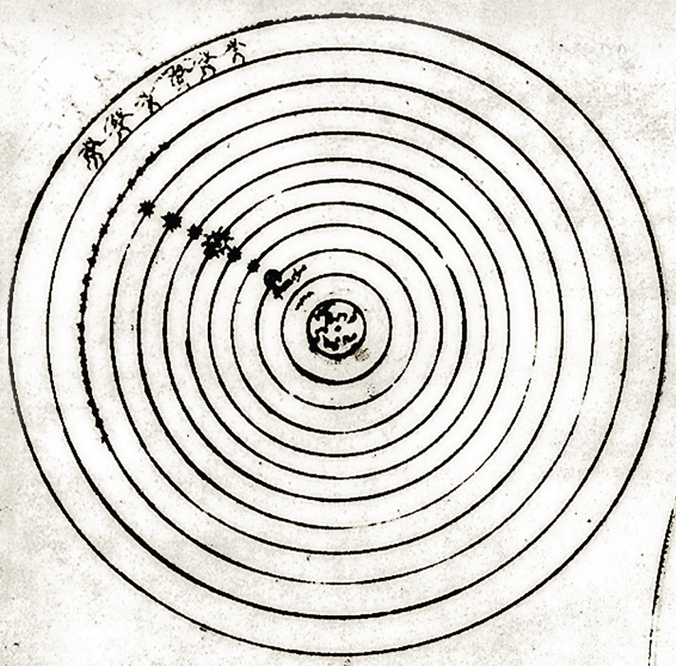
A Cultural History of Physics

Károly Simonyi





CRC Press
Taylor & Francis Group
AN A K PETERS BOOK

A Cultural History of Physics



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Károly Simonyi

A Cultural History of Physics

Translated by David Kramer



CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
AN A K PETERS BOOK

Originally published in Hungarian as *A fizika kultûrtörténete*, Fourth Edition, Akadémiai Kiadó, Budapest, 1998, and published in German as *Kulturgeschichte der Physik*, Third Edition, Verlag Harri Deutsch, Frankfurt am Main, 2001. First Hungarian edition 1978.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2012 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

International Standard Book Number: 9781032690988 (Hardback)
DOI: 10.1201/9781032697697

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

The Open Access version of this book, available at www.taylorfrancis.com, has been made available under a Creative Commons Attribution-NonCommercial-NoDerivatives (CC-BY-NC-ND) 4.0 International license. Funded by Charles Simonyi estate.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data
A catalog record for this title has been requested

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Foreword by Charles Simonyi xi
Preface xiii

Introduction: The History of Physics and Its Relevance to Our Lives Today
0.1 The History of Physics and Its Relevance to Our Lives Today 1
0.2 Assessment and Division into Epochs 3
0.2.1 A Historical Timeline Based on the Intensity of Scientific Activity 3
0.2.2 Scientific Knowledge from the Viewpoint of the Physicist of Today 4
0.2.3 Division into Epochs Based on Theoretical Synthesis 7
0.2.4 The Role of Modeling 8
0.3 Elements of the Philosophy of Science 10
0.3.1 Illusory Simplicity 10
0.3.2 Reason and Experience 12
0.3.3 Pitfalls of the Inductive Method 14
0.4 The Dynamism of History 15
0.4.1 Forces for Progress 15
0.4.2 Limits, Possibilities, and Dangers 19
0.4.3 Uncertainty in the Precision 21
0.4.4 Physics in a New Role 23
0.4.5 Characterization of Epochs in Physics 24

Chapter 1: The Classical Heritage
1.1 What the Greeks Inherited 31
1.1.1 The Beginnings of Science 31
1.1.2 Egypt and Mesopotamia 32
1.2 The Harmonious, Beautiful Order 45
1.2.1 Overview: Temporal, Spatial, and Causal Connections 45
1.2.2 Mysticism and Mathematics: Pythagoras 50
1.2.3 Idea and Reality 54
1.2.4 Plato on Insight and Ideas 57
1.3 Matter and Motion: The Aristotelian Synthesis 60
1.3.1 Atoms and Elements 60
1.3.2 Motion under Terrestrial Conditions: Peripatetic Dynamics 65
1.3.3 Celestial Motion 72
1.3.4 The Aristotelian Worldview 74
1.3.5 A Selection from Aristotle's Metaphysics 76
1.4 The Greatest Achievements of the Ancient Sciences 78
1.4.1 Archimedes 79
1.4.2 The Ptolemaic System for Describing Celestial Motion 89
1.4.3 Astronomy and Geography 90
1.4.4 Geometry 95
1.4.5 Scientific Instruments and Technology 98

1.5 The Twilight of Hellenism **99**
1.5.1 Pessimistic Philosophers **99**
1.5.2 Augustine on the Absurdity of Astrology **104**
1.5.3 Augustine on Time **105**

Chapter 2: The Stewards of the Heritage

2.1 The Thousand-Year Balance Sheet **113**
2.1.1 Why Did Progress Stall? **113**
2.1.2 Europe Takes Shape **115**
2.1.3 The Technological Revolution **121**
2.1.4 Monasteries and Universities **123**
2.2 The Salvage of Ancient Knowledge **129**
2.2.1 The Direct Path **129**
2.2.2 Byzantium **131**
2.2.3 The Arab Transmission **132**
2.2.4 Return to the Source **133**
2.3 The Indian and Arab Worlds **136**
2.3.1 The Decimal System **136**
2.3.2 Algebra and Algorithm **137**
2.3.3 Some Outstanding Contributions of Arab Science **138**
2.4 The West Awakens **140**
2.4.1 Fibonacci: The Artist of Computation **140**
2.4.2 Jordanus Nemorarius: Structural Engineer **142**
2.4.3 Descriptive Kinematics: Nicole Oresme and Merton College **143**
2.4.4 Peripatetic Dynamics Reformed **145**
2.4.5 Buridan's Theory of Impetus **146**
2.4.6 Physics in Astronomy **147**
2.4.7 Results **148**
2.4.8 Nicole Oresme on the Motion of Earth **149**
2.5 Medieval Natural Philosophy **151**
2.5.1 Faith, Authority, and Science **151**
2.5.2 Faith and Experience **154**
2.6 The Renaissance and Physics **157**
2.6.1 Art, Philology, and Science **157**
2.6.2 Progress in Mechanics **159**
2.6.3 The Science of Artists **161**
2.6.4 Leonardo da Vinci **162**
2.6.5 The Professional Astronomers Take the Stage **164**
2.6.6 The Role of the Printing Press **165**

Chapter 3: Demolition and the Construction of a New Foundation

3.1 The World in 1600 **171**
3.2 Numerology and Reality **176**
3.2.1 Back to Plato in a New Spirit **176**
3.2.2 The Retrograde Revolutionary: Copernicus **177**
3.2.3 A Compromise: Tycho Brahe **185**
3.2.4 Celestial Harmony: Kepler **189**

3.3 Galileo and Those Who Stood in His Shadow **194**
3.3.1 The Unity of the Celestial and Terrestrial Spheres **194**
3.3.2 Inclined Planes, Pendulums, and Projectile Motion **201**
3.3.3 Galileo's Greatness **208**
3.3.4 In the Background: Stevin and Beeckman **210**
3.3.5 The Possibility of Connection **212**
3.4 The New Philosophy: Doubt Becomes Method **213**
3.4.1 Francis Bacon and the Inductive Method **213**
3.4.2 A Method for Discovering Certain Truth: Descartes **216**
3.4.3 The Cartesian Laws of Motion **219**
3.4.4 The First Cosmogony **220**
3.4.5 On the Periphery of Western Culture **223**
3.5 Light, Vacuum, and Matter through the Middle of the Seventeenth Century **226**
3.5.1 The Snell–Descartes Law **226**
3.5.2 Fermat's Principle **230**
3.5.3 Vacuum and Air Pressure **233**
3.5.4 Uncertain Steps on the Path to Modern Chemistry **237**
3.6 After Descartes and before Newton: Huygens **240**
3.6.1 Huygens's Axioms on Dynamics **240**
3.6.2 The Mathematical Pendulum **245**
3.6.3 The Cycloidal Pendulum **246**
3.6.4 The Physical Pendulum **249**
3.6.5 The Collision Laws as Consequences of the Equivalence of Inertial Systems **252**
3.6.6 Circular Motion **254**
3.7 Newton and the Principia: The Newtonian Worldview **255**
3.7.1 The Tasks Awaiting the Advent of Newton **255**
3.7.2 A Force Is Not Required to Maintain a State of Motion but to Change It **256**
3.7.3 The Law of Universal Gravitation **260**
3.7.4 Selections from the *Principia* **264**
3.7.5 Newton as Philosopher **270**

Chapter 4: The Completion of Classical Physics

4.1 Starting Capital for the Eighteenth Century **281**
4.1.1 Prior Results **281**
4.1.2 Waves or Particles? **281**
4.1.3 Analytic Geometry **289**
4.1.4 Differential and Integral Calculus: The Battle of the Titans **291**
4.1.5 For and against Descartes **297**
4.1.6 Voltaire and the Philosophers **299**
4.2 Worthy Successors: d'Alembert, Euler, and Lagrange **301**
4.2.1 Possible Directions for Progress **301**
4.2.2 Results in Statics **304**
4.2.3 Newtonian Mechanics in Euler's Hands **305**
4.2.4 The First Variational Principle in Mechanics: Maupertuis **309**
4.2.5 The First "Positivist": d'Alembert **311**
4.2.6 Modern Ideas **314**
4.2.7 Mechanics as Poetry **316**

4.3 The Century of Light **319**
4.3.1 The Enlightenment **319**
4.3.2 The Great Encyclopedia **321**
4.3.3 d'Alembert: Preface to the Encyclopedia **322**
4.3.4 Belief in the Solid Foundation of Physics: Kant **326**
4.4 From Effluvium to the Electromagnetic Field **329**
4.4.1 Peter of Maricourt and Gilbert **329**
4.4.2 The Chronology of Progress **330**
4.4.3 Qualitative Electrostatics **332**
4.4.4 Quantitative Electrostatics **338**
4.4.5 Flow of Electric Charge **340**
4.4.6 The Magnetic Field of Electric Currents: Cross-Fertilization from Natural Philosophy **345**
4.4.7 The Interaction of Currents: An Extension of Newton's Ideas **346**
4.4.8 Faraday: The Greatest of the Experimentalists **349**
4.4.9 Maxwell: The Fundamental Laws of Electromagnetic Fields **354**
4.4.10 The Electromagnetic Theory of Light **359**
4.4.11 Lorentz's Theory of the Electron **363**
4.5 Heat and Energy **365**
4.5.1 The Thermometer **365**
4.5.2 Progressive in Its Day: The Caloricum Theory of Joseph Black **366**
4.5.3 Rumford: But Heat Is Still a Form of Motion! **368**
4.5.4 Fourier's Theory of Heat Conduction **369**
4.5.5 Caloricum and the State Equation **372**
4.5.6 The Carnot Cycle **373**
4.5.7 The Kinetic Theory of Heat: First Steps **374**
4.5.8 The Law of Conservation of Energy **376**
4.5.9 The Kinetic Theory of Gases **377**
4.5.10 The Second Law of Thermodynamics **379**
4.5.11 Entropy and Probability **381**
4.6 The Structure of Matter and Electricity: The Classical Atom **387**
4.6.1 Chemistry Hinting at the Atomic Structure of Matter **387**
4.6.2 The Electron: J. J. Thomson **388**
4.6.3 Chemistry to the Rescue Again: The Periodic Table **393**
4.6.4 First Ideas about the Structure of the Atom **395**
4.6.5 The Line Spectrum and the Reappearance of the Integers **398**
4.6.6 A Farewell to the Nineteenth Century **400**

Chapter 5: The Physics of the Twentieth Century

5.1 "Clouds on the Horizon of Nineteenth-Century Physics" **405**
5.1.1 A Conclusion or a New Start? **405**
5.1.2 Mach and Ostwald **407**
5.2 The Theory of Relativity **409**
5.2.1 Antecedents: Failed Attempts at Measuring Absolute Velocity **409**
5.2.2 Attempts at Adaptation **412**
5.2.3 The Protagonists: Lorentz, Einstein, and Poincaré **416**
5.2.4 The Measurement of Distance and Time **421**

5.2.5 The Equivalence of Energy and Mass **425**
5.2.6 Matter and the Geometry of Space **429**
5.2.7 Einstein on Space, Ether, and the Field Problem of Physics **433**
5.3 Quantum Theory **437**
5.3.1 Blackbody Radiation in Classical Physics **437**
5.3.2 Planck: Entropy Points the Way to the Solution **441**
5.3.3 The Appearance of the Energy Quantum **444**
5.3.4 Einstein: Light Is Also Quantized **448**
5.3.5 Bohr: The "Classical" Quantum Theory of the Atom **448**
5.3.6 The Statistical Derivation of the Radiation Formula as Prelude to Quantum Electronics **451**
5.3.7 Heisenberg's Matrix Mechanics **452**
5.3.8 Einstein and Heisenberg **457**
5.3.9 Schrödinger's Wave Mechanics **458**
5.3.10 Heisenberg: The Copenhagen Interpretation of Quantum Theory **465**
5.3.11 Operators. Quantum Electrodynamics **470**
5.3.12 The Causality Problem **477**
5.3.13 John von Neumann on Causality and Hidden Parameters **481**
5.3.14 Quantum Mechanics as a Tool and as a Philosophy **483**
5.3.15 What Remains of Classical Physics? **486**
5.4 Nuclear Structure, Nuclear Energy **489**
5.4.1 A Backward Glance at the First Three Decades **489**
5.4.2 Stations of the Study of the Atomic Nucleus **496**
5.4.3 Becquerel: Why Do Uranium Salts Fluoresce? **498**
5.4.4 The Protagonists of the Heroic Age: The Curies and Rutherford **501**
5.4.5 The Rutherford–Bohr Model Begins to Take Shape **505**
5.4.6 The First Artificial Nuclear Transformation **507**
5.4.7 Quantum Mechanics Can Be Applied to Nuclear Phenomena **507**
5.4.8 Predicted by Rutherford, Found by Chadwick: The Neutron **508**
5.4.9 Nuclear Structure and Nuclear Models **509**
5.4.10 Nuclear Fission: Experimental Evidence, Theoretical Doubt **514**
5.4.11 The Chain Reaction: The Large-Scale Liberation of Nuclear Energy **518**
5.4.12 Energy through Nuclear Fusion: The Fuel of the Stars in the Hands of Mankind **522**
5.4.13 The Responsibility of Physicists **523**
5.5 Law and Symmetry **524**
5.5.1 The Historian's Role in the Description of Modern Physics **524**
5.5.2 The Elementary Particles, in Order of Appearance **525**
5.5.3 A Few Words about Cosmic Rays **529**
5.5.4 Particle Accelerators and Detectors **530**
5.5.5 Fundamental Interactions **532**
5.5.6 The Conservation Laws **536**
5.5.7 Symmetry, Invariance, Conservation **537**
5.5.8 Mirror Symmetry? **540**
5.5.9 "A Bit of Asymmetry Improves the Aesthetics" **543**
5.5.10 Back to the Apeiron? **546**
5.5.11 The Quark Theory Is Completed **548**

5.6 Mankind and the Universe **549**
5.6.1 New Information Channels **549**
5.6.2 Energy Production in the Stars **550**
5.6.3 Birth, Life, Death on a Cosmic Scale **554**
5.6.4 The Formation of the Universe **556**
5.7 Summary and Preview **561**
5.7.1 Physics, Philosophy, and Society at the Turn of the Millennium **561**
5.7.2 The Standard Model and Beyond **565**
5.7.3 Groups and Symmetries **567**
5.7.4 The Grand Unification **569**
5.7.5 The Great Laboratory **571**
5.7.6 Questions and Doubts Multiply **575**
5.7.7 "Between Nothing and Infinity" **578**

Epilogue: Looking Ahead in Physics by Edward Witten **581**
Bibliography **585**
Name Index **601**
Subject Index **617**

Foreword by Charles Simonyi

An exceptional book such as this could have been created only under exceptional circumstances. My father was a working physicist and a beloved university professor who taught a whole generation of Hungarian electrical engineers. His textbooks on the foundations of electrical engineering have been translated into many languages. Yet, in the politically charged atmosphere of the 1960s in Hungary, his quasi-apolitical personal conduct, based on the age-old virtues of hard work, good character, and charity, was interpreted as political defiance that could not be countenanced by the state. Hence, he progressively lost his directorship at the Physics Research Institute, his post as department head, and finally his teaching position altogether. I was still a minor when I left the country—and my parents—in search of a better life. It was understood by all that my doing so—a political act in a totalitarian era—would make my father's situation even more difficult.

Besides being a scientist, my father was a great humanist, not only in terms of his concern for his fellow man but also in the sense of a scholar of the humanities: he was extremely well read in the classics as well as in contemporary literature and history. The break in his career at midlife did not drive him to despair; his humanism instead commanded him to work on the subject he had perhaps always wanted to work on: the history of the interplay of science and the humanities. His first notes became a lecture series, first given off campus, in the evenings at the invitation of student organizations. Much later, when I was able to return to Hungary, I was privileged to listen to one of these lectures, still filled to more than capacity with students and young intellectuals, hearing my father convey the excitement and wonder of scientific development—how difficult it was to make progress in science, not simply because of ignorance but because the arguments were complex and the evidence was often ambiguous, and how the scientists gained courage or were otherwise influenced by the humanities. The success of these lectures gave rise to the present book that he continued to revise and extend almost until his death in 2001.

All history books that treat the modern period face a problem: when should the discussion close? My father prided himself on keeping the book up-to-date as it progressed through five Hungarian editions and three German editions. Now, nearly a decade after his death, we edited the story down to what was firmly settled by the year 2000, and asked the noted and brilliant physicist Ed Witten to write an epilogue bringing us up to date with the current scientific outlook, as opposed to the already stale speculations made in the recent past.

The English edition of the book was a dream for my father that he was unfortunately unable to realize due to the costs and the difficulties of supervising the translation. I was very fortunate in having found an experienced and courageous publisher, A K Peters (now part of CRC Press, a Taylor and Francis Group), who was willing and able to undertake the task. The translation was based on the third German edition, but being cognizant of the dangers

involved in a second-generation translation of a translation, we carefully compared the results with the original Hungarian text and, wherever necessary, the more direct and conversational tenor of the original was restored.

The goal of the English edition is to be a "world book"—not just for the US and for other English-speaking countries but for all nations. Just as in the Middle Ages when Latin was the language of international scholarship, now we have a true world language of great expressive power, beauty, and flexibility, namely English, and it is our earnest hope that this translation will be enjoyed by everyone interested in the subject regardless of their native language.

Special thanks are due to Alice and Klaus Peters for the direction of this multifaceted project, from the typographic design to the supervision of the translation, editing, and production. The base translation was done by David Kramer. The text was reviewed by Robert Schiller and Alex Farba DeLeon. Charlotte Henderson did the final copy editing, including that of the mathematical formulas. Others involved in the project were Camber Agrelius, Sarah Chow, Julie Nicolazzo, and Sandra Rush.

This republishing of my father's main work would not have been possible without the support of the family: my mother, Zsuzsa, my brother, Tamas, and my wife, Lisa. Special help from Ildikó Csurgay with the illustrations is gratefully acknowledged.

Finally, I am tremendously grateful for Prof. Edward Witten of the Institute for Advanced Study for contributing the epilogue.

Preface

Today, the history of science is a discipline in its own right, with its own subject matter and methodology, its own journals, and its own university chairs. And, of course, it has its own professional practitioners, a group that the author of this book does not belong to. His profession is teaching and research in physics and he has simply taken delight in the history of his subject, a delight that he wishes to share with others. The reader may therefore take those parts of this book that deal with physics and technology to be authentic—to the extent that any book can be regarded as such—while the interpretation of the historical and philosophical background bears some of the stamp of the subjective and, to a certain, perhaps permissible degree, that of dilettantism.

This book has been written for a broad audience. The author hopes that the nonspecialist reader will be able to follow the presentation—to be sure not without a certain measure of intellectual effort—and at the same time that the professional physicist will also find it of value. Such a twofold goal should not be attained at the cost of compromise: the level of discourse cannot just be set somewhere between that of the educated member of the general public and the professional physicist. Rather, it was the author's intention to set apart, wherever possible—if necessary by typography—the more easily assimilated portions from those requiring specialized knowledge. These latter segments appear in the present book in a smaller typeface, and they may be skipped by someone reading the main text, without loss of continuity. Yet, these technical passages can be also useful for the general reader, for the formulas and illustrations there—even from a cursory examination—should help to fend off false impressions. For example, one feels Greek literature and art to be of importance not only for their time, but for all time, since they have something of value to say to us even today. On the other hand, with regard to the greats of ancient science, we might consider it to be self-evident that they were largely prisoners of their time, and that today the knowledge of a schoolchild may well exceed that of a learned man of antiquity, Archimedes, for instance. Perhaps we would say the same about the ancient artists if we were unable to marvel at the sculptures of Praxiteles and Myron of Eleutherae, in the originals or copies, or if we could not read Homer at home, or see the plays of Euripides at the theater. If we would immerse ourselves as thoroughly in the ideas of Archimedes—to stay with the above example—we would see that to reconstruct them requires, even of the scientifically educated, a significant intellectual effort and that doing so can give one great intellectual delight. The reader may therefore look upon the technical passages as the analogues of the indispensable illustrations or quotations in works on the history of art or literature.

This book is, therefore, a work for the public understanding of science, and it may also serve as a textbook for college students. It has been the author's intention that to these two goals a third should be added, and he is fully aware of the danger that in attempting too many tasks he might succeed satisfactorily in none of them. This third goal is to be a primer in the history of physics because it contains almost as much in the way of quotations as it does of main text. In order to separate the quotations as little as possible from the text and to interrupt it no more than necessary, the quotations—printed on shaded background—are in most cases presented

in side notes to the text that they accompany, or on occasion are inserted directly into the main flow of the book.

Biographical information that could not be organically integrated into the main text, as well as additional facts that require no special commentary, can be found in the extended figure captions. Thus there is a fourth use to which this book can be put, namely, as a sort of encyclopedia.

The color plates should serve—or so the author has intended—apart from their decorative and informational functions, to provide in their coordinated entirety a skeleton for the book or—more generally—for the cultural history of physics.

The author of a book such as this one must—if only from the scope of the project—rely on a host of other books. Some of the books listed in the bibliography have served as inspirational sources, others offer the reader introductory material, and still others provide a more wide-ranging view. The author has tried to indicate the origins of his ideas and to give proper credit for the figures and quotations, referring where possible to original sources. The figures have been taken—again, wherever possible—from first editions, primarily those to be found in Hungarian libraries.

Finally, the author would like to thank all those who participated in the creation of this book. First of all, his thanks are due to his assistant Ildikó Csurgay, who participated in preparing the manuscript for publication and in solving a number of technical and stylistic problems.

The author's thanks go also to the following Hungarian libraries for their help in locating ancient works and for permission to make photocopies to be used in the present book: the Library of the Technical University, Budapest; the Budapest University Library; the National Széchényi Library; the main library of the Benedictine Abbey in Pannonhalma; the Székesfehérvár diocesan library; the Memorial Library of the University for Heavy Industry, Miskolc; and the Library of the Hungarian Academy of Sciences, Budapest.

The author would also like to thank the following museums and institutions, which made illustrative material in their collections available without charge: CERN, Geneva; the Zwettl collegiate church; the Museum of the City of Paris; the Museum of Versailles; the Naples Museum, the Herzog August Library, Wolfenbüttel; and the Berlin State Museums.

The author thanks the Urania publishing company, in Leipzig, the publisher Akadémiai Kiadó, in Budapest, Harry Deutsch publishers, and in particular Bernd Müller, for a pleasant and sympathetic collaboration. Also to be mentioned are the valuable comments of Dr. Martin Franke as well as those of Dr. Otto Haiman, which were of great use in the final formulation of the text.

The last part of Chapter 5 was read critically by Professor András Patkós, and we have adopted many of his suggestions, some of them directly into the text. I thus owe him a special measure of thanks.

In addition to the collaborators mentioned above, I thank my son Tamás, who reviewed the Arabic texts and helped with editorial work. I thank as well my son Charles for his generous support, both material and emotional.

The author has striven to recognize all those to whom he owes thanks, especially for comments and suggestions, and to do so as precisely as possible. He is aware that he has been unable to realize these intentions completely. Finally, the author thanks his wife, who offered tireless assistance in preparing the manuscript, in bibliographic work, and in discussions of stylistic and pedagogical questions.

Károly Simonyi, 2000

INTRODUCTION

0.1 The History of Physics and Its Relevance to Our Lives Today

In today's industrialized societies, it has become possible for an ever-increasing number of individuals to pursue a life free of want. For this achievement we may thank the ever larger number of specialists working in very narrow fields of endeavor. Individuals yearn for a general overview of the cultural values created by the whole of humanity; or if not, we would like to awaken such desire in them. But is it possible to arouse in specialists—the "cultural barbarians"—an enthusiasm for art or literature? And conversely, can those versed in the humanities be convinced that discoveries in the various branches of science constitute an integral part of universal human culture? Or to put it in more general terms, employing a notion made popular in the twentieth century by C. P. Snow (1905–1980), is it possible to bridge the gap between the "two cultures," that is, between the humanities and the natural sciences? (See Figure 0.1 and Quotation 0.1.) Are today's citizens capable of making such a synthesis, and is it even possible or useful for a society to set such a goal? After all, the capacity of the individual to absorb knowledge is very limited; moreover, is it not the mark of truly great specialists that their work within their professional fields represents a calling, a life's work, a complete source of satisfaction and self-fulfillment?

In this context, what does the history of physics offer us? For physicists, the triumphs in the history of their science could stand as points of reference, as criteria for measuring the value of significant accomplishments in other cultural domains, while those with a more humanistic education or inclination could find in the history of the natural sciences, and in particular that of physics, those elements—research methods, principles of establishing the validity of results, and of course the results themselves—that in the course of history have become significant milestones of universal human culture, indeed often serving as a cultural driving force. In any case, one thing should be stated plainly: *Human culture is a single, unified whole, and it is only for us, the consumers of culture, that the problem arises how its significant elements are to be selected, appropriated, and transmitted* (see Quotations 0.2–0.4). Yet we must also note that paradoxically, the greatest creative personalities, both artists and scientists, of necessity operate as laws unto themselves, which often means that they are completely one-sided in their views.

There is much to be found in the history of physics that can make instruction at all levels, from elementary school to university, interesting, indeed exciting: There are amusing anecdotes, to be sure, but also tales of tragic conflicts, entertaining accounts of historical events, and the birth pangs of the clarification of concepts and methods—which has a philosophical aspect as well. All of this is highly suited to awaken interest in the student and provide valuable educational experiences. Moreover, the history of physics abounds in examples of lofty ideals and guidelines for ethical behavior.

On the other hand, we might ask what, in contrast to general expectation, is *not* to be expected from instruction in the history of physics? There is much talk in the contemporary pedagogical literature about education for independent, "critical," thinking. In scientific education attempts are often made to achieve this goal not by presenting students with the laws of nature as established fact or the opinion of



▲ **Figure 0.1** A scientist with the ecstasy of a saint or of an artist. This image of a Greek scholar in a medieval cathedral stands as a symbol of the unity of human culture (statue of Ptolemy in the Ulm cathedral, sculpted in 1470 by Jörg Syrlin).

Quotation 0.1

I believe the intellectual life of the whole of western society is increasingly being split into two polar groups. ...

Literary intellectuals at one pole—at the other scientists, and as the most representative, the physical scientists. Between the two a gulf of mutual incomprehension—sometimes (particularly among the young) hostility and dislike, but most of all lack of understanding. They have a curious distorted image of each other. ...

The non-scientists have a rooted impression that the scientists are shallowly optimistic, unaware of man's condition. On the other hand, scientists believe that the literary intellectuals are totally lacking in foresight, peculiarly unconcerned with their brother men, in a deep sense anti-intellectual, anxious to restrict both art and thought to the existential moment. ...

[There was a somewhat well-known scientist] who, when asked what books he read, replied firmly and confidently: "Books? I prefer to use my books as tools." It was very hard not to let the mind wander—what sort of tool would a book make? Perhaps a hammer? A primitive digging instrument? ...

continued on next page

DOI: 10.1201/9781032697697-1
This chapter has been made available under a CC-BY-NC-ND license.

Quotation 0.1, ***continued***

It isn't that they lack the interests. It is much more that the whole literature of the traditional culture doesn't seem to them relevant to those interests. They are, of course, dead wrong. As a result, their imaginative understanding is less than it could be. They are self-impoverished. But what about the other side? They are impoverished too—perhaps more seriously, because they are the vainer about it. ...

As with the tone-deaf, they don't know what they miss. They give a pitying chuckle at the news of scientists who have never read a major work of English literature. They dismiss them as ignorant specialists. Yet their own ignorance and their own specialization is just as startling. A good many times I have been present at gatherings of people who, by the standards of the traditional culture, are thought highly educated and who have with considerable gusto been expressing their incredulity at the illiteracy of scientists. Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold: it was also negative. Yet I was asking something which is about the scientific equivalent of: Have you read a work of Shakespeare's? I now believe that if I had asked an even simpler question—such as, What do you mean by mass, or acceleration, which is the scientific equivalent of saying, Can you read?—not more than one in ten of the highly educated would have felt that I was speaking the same language. So the great edifice of modern physics goes up, and the majority of the cleverest people in the western world have about as much insight into it as their Neolithic ancestors would have had. ...

In our society (that is, advanced western society) we have lost even the pretence of a common culture. Persons educated with the greatest intensity we know can no longer communicate with each other on the plane of their major intellectual concern. This is serious for our creative, intellectual, and, above all, our normal life. It is leading us to interpret the past wrongly, to misjudge the present, and to deny our hopes of the future. It is making it difficult or impossible for us to take good action.

There is, of course, no complete solution ... But we can do something. The chief means open to us is education—education mainly in primary and secondary schools, but also in colleges and universities. There is no excuse for letting another generation be as vastly ignorant, or as devoid of understanding and sympathy, as we are ourselves.

—C. P. Snow, *The Two Cultures and the Scientific Revolution,* 1959 [pp. 4–16]

experts, but by providing them with the experimental setup used by the discoverer of the law, so that with the help of the experiments, the students can *rediscover the laws for themselves.* Yet in reality, such attempts are frequently good only for giving the student a completely false impression of the degree of difficulty in discovering new laws, with the result that they are unable to appreciate adequately the contributions of the great thinkers. Only a historical approach can show the decisive steps in a new discovery, which often required not only genius, but in many cases also an unusual degree of human courage.

As an example, we might mention here the inclined plane, a typical means by which students rediscover the laws of motion under constant acceleration, allowing them to reproduce an intellectual leap of Galileo (1564–1642). But if we give the students an inclined plane and some smooth spheres to roll down it, *we have removed the only significant new element of discovery and originality.* All that remains is mechanical work, which from the point of view of an experimental psychologist would not be much different from studying the behavior of a laboratory rat traversing a maze.

Galileo's great achievement was to have found an experimental approach that does not use materials found in nature. In this case, *the description of the experimental setup is the crucial step,* which by abstraction from reality simplifies the problem to such an extent that it can be investigated with the tools of mathematics. To be sure, experimental procedures like that of the inclined plane have their purpose, but they fulfill a quite different pedagogical goal. Thus reproducing some of the measuring apparatus used by Galileo, for instance, brings the student through his or her own experience nearer to both the natural phenomenon under investigation and the historical background of the scientific discovery. It is even capable of eliciting ideas indirectly: Galileo measured time by measuring the mass of the water that flowed out a vessel during the time to be measured; similarly, in today's technology, very small intervals of time can be measured by the tiny amounts of electrical charge that flow out of a capacitor.

Care must be taken not to present an over idealized image of the great personalities who changed the course of physics: We should not forget that their opponents were nearly their equals in intellectual outlook and were even their superiors on occasion in matters of ethics and morality. If we insist that these opponents were simply expounding nonsense or were somehow less moral, we are in effect devaluing the importance of the scientific discovery.

One could even imagine the traditional introductory course in physics being replaced by one on the history of physics—in both the basic courses for would-be physicists and engineers and those for students in the humanities.

The history of physics can also serve as a motivation for practicing physicists. In this regard we turn to two famous scholars of the twentieth century to demonstrate that the science, the philosophy, and the physics of the ancient Greeks remains a living force even to this day. The first is Bertrand Russell (1872–1970), who considered Plato (ca. 427–ca. 347 BCE) to be the predecessor of the modern work of clarifying the concept of a number that was undertaken by Gottlob Frege (1848–1925), and Alfred North Whitehead (1861–1947), and himself. The second is Werner Heisenberg (1901–1975), who attributed to Plato the fundamental idea behind his unified field theory (see Quotation 1.20).

The history of physics can also be of use to active scientific researchers in other, more direct, ways: Namely, the theoretical underpinnings of some new theory are typically debated in great detail when it first appears. Opponents of the theory

will bring forth all possible arguments against it for which reassuring answers have to be supplied. The next generation will in its turn accept the theory as given and the doubts will have been forgotten. But when further scientific progress again requires an investigation of the foundations of new theories, those old disputes can provide many useful ideas.

And, we must not forget that revolutionary ideas arise from the revolt of creative genius against received wisdom; see Quotations 0.5 and 0.6.

Questions that arise in the history of physics can challenge the human intellect. Even today, new documents continue to be unearthed that cast a whole new light not only on the history of physics—or more generally the history of science—but on the history of human culture and indeed on world history itself.

For example, at the beginning of the twentieth century, a letter was discovered written by Archimedes (ca. 287–ca. 212 bce) on a palimpsest—an erased and reused parchment on which the earlier writing is still discernible—in which he describes his research methods and gives advice that could of value to scientists of today. Likewise, in the recent past, researchers succeeded in explaining the purpose of a device retrieved from an ancient shipwreck. Described with some exaggeration as an ancient computer, this device nevertheless revolutionizes our understanding of the period, for it casts the metal-working technology of the Hellenistic era in an entirely new light (see Section 1.4).

At present, we are witnessing a reevaluation of an entire historical epoch brought about through a study of the history of physics. An examination of a number of manuscripts from the so-called "dark" Middle Ages, often thought of as an era of cultural stagnation, has led to the conclusion that the era when modern science might have germinated has to be dated three centuries earlier than what was previously believed. All this will be discussed in detail in Section 2.4.

Finally, an understanding of the history of physics helps us to evaluate properly the achievements of our own era. We speak readily of a growth in scientific knowledge at an unprecedented rate, of a "quickening of time," and we worry about how students and teachers are to deal with this growing mass of information. It is the considerable *technological progress* of our time that underlies this attitude. But our perception of the fundamental relationships that reflect the inner structure of our physical world follows a different developmental tempo. The opinion—one shared by the author of this book—has been expressed that the results achieved in this area in recent years cannot be compared with the revolutionary insights of the first decades of the twentieth century or the last part of the seventeenth century, which burst forth at a truly dazzling rate.

0.2 Assessment and Division into Epochs

0.2.1 *A Historical Timeline Based on the Intensity of Scientific Activity*

Historical events do not conveniently align themselves into decades, but at every decade one can step back and summarize what happened up until that point. An even, metronome-like division of the timeline also has the added advantage that it strikingly shows changes between periods crowded with events, and seemingly empty periods when preparation, evaluation, and refining was taking place. George Sarton (1884–1956), one of the most important historians of science of our time, has chosen this method of description in one of his books.

Quotation 0.2

Science, after all, is a branch of literature; and working on science is a human activity like building a cathedral. No doubt there is too much specialization and too much professionalism in contemporary science, which makes it inhuman; but this unfortunately is true of contemporary history or psychology also, almost as much as of the natural sciences. ...

Laboring the difference between science and the humanities has long been a fashion, and has become a bore.

—Karl Popper, *Objektive Erkenntnis*, 1973 [p. 185]

Quotation 0.3

I think there have always been two cultures and always should be. And even if we thought it undesirable, there would still continue to be two.

—Melvin J. Lasky [Gingerich 1978, p. 484]

Quotation 0.4

... with the collection of the summary of all of our knowledge concerning the sciences, the arts, and commerce into an Encyclopedia we wanted to make felt the mutual assistance they lend each other. [author's translation]

—d'Alembert, Program for the *Encyclopédie*, 1751

Quotation 0.5

Great conquerors, we read, have been both animated, and also, in great measure, formed by reading the exploits of former conquerors. Why may not the same effect be expected from the history of philosophy to philosophers? May not even more be expected in this case? The wars of many of those conquerors, who received this advantage from history, had no proper connection with former wars: they were only analogous to them. Whereas the whole business of philosophy, diversified as it is, is but one; it being one and the same great scheme, that all philosophers, of all ages and nations, have been conducting, from the beginning of the world; so that the work being the same, the labors of one are not only analogous to those of another, but in an immediate manner subservient to them; and one philosopher succeeds another in the same field, and pursuing the same conquests, in the same country. In this case, an intimate knowledge of what has been done before us cannot but greatly facilitate our future progress, if it be not absolutely necessary to it.

continued on next page

Quotation 0.5, ***continued***

These histories are evidently much more necessary in an advanced state of science, than in the infancy of it. At present philosophical discoveries are so many, and the accounts of them are so dispersed, that it is not in the power of any man to come at the knowledge of all that has been done, as a foundation for his own inquiries. And this circumference appears to me to have very much retarded the progress of discoveries.

Not that I think philosophical discoveries are now at a stand. On the other hand, as quick advances seem to have been made of late years, as in any equal period of time past whatever. Nay, it appears to me, that the progress is really accelerated.

— Joseph Priestley, *The History and Present State of Electricity*, 1767

Quotation 0.6

I want to discuss now the art of guessing nature's laws. It is an art. How is it done? One way you might suggest is to look at history to see how the other guys did it. So we look at history.

[Here follows a brief survey of the methods of Newton, Maxwell, Heisenberg, and Schrödinger.]

I am sure that history does not repeat itself in physics, as you can tell from looking at the examples I have given. The reason is this. Any schemes—such as "Think of symmetry laws," or "put the information in mathematical form," or "guess equations"—are known to everybody now, and they are tried all the time. When you are stuck, the answer cannot be one of these, because you will have tried these right away. There must be another way next time. Each time we get into this log-jam of too much trouble, too many problems, it is because the methods that we are using are just like the ones we have used before. The next scheme, the new discovery, is going to be made in a completely different way. So history does not help us much.

—Richard Feynman, *The Character of Physical Law*, 1967

We thus have at our disposal an informal procedure for historical, or more precisely, chronological, division into epochs: We shall investigate the intensity of scientific achievement as a function of time. This curve, depicted in Figure 0.2, shows some surprising features. Right away, we note that in the course of the past two and a half millennia, scientific achievement had at least two maximally intensive periods extending over a few centuries. The maximum that stretches from 500 to 200 bce. falls in the era known in the history of humanity as the "Greek miracle." In the figure, the curve running down the right-hand side represents the intensity of scientific achievement, while that on the left represents the intensity of work in other areas (art and literature). We can see that these two realms have developed more or less in synch with each other, though to be sure, one can also find some imbalances. Thus, for example, Roman antiquity can boast cultural achievements in some fields that are equal to those of the Greeks. We are thinking here particularly of Roman jurisprudence, which has had a long-term influence—continuing into the present—over how a society regulates its behavior, or of the outstanding works of Roman literature, for example those of Virgil (70–19 bce) and Horace (65–8 bce). But during this period, the Romans produced no original work in physics or mathematics, unless we consider perhaps the further development of Greek atomism by Lucretius (ca. 96–ca. 55 bce) as the single exception. The situation is similar in the Renaissance, which preceded the great explosion in the natural sciences of the seventeenth century. It is clear from Figure 0.2 that here as well, the sciences take a back seat to the unique level of activity in the arts.

The period of almost two thousand years between the flowering of ancient Greece and the seventeenth century can be seen as an era of transition and rediscovery, to which original discoveries were contributed only sporadically, as for example those in the Arab world and Byzantium as well as those of the late Scholastic period.

It thus seems suitable to divide the history of science into the following divisions: the epoch of ancient Greek knowledge, the period of transition, and the era of modern science.

Perhaps the most exciting part of the history of physics was the period during which the rediscovered legacy of antiquity was discarded, because an all-encompassing and closed system was not suitable for use as a basis for new development. This was also the era in which new foundations were beginning to be established. It rightfully deserves a chapter of its own. The physics of the modern period will be divided into classical physics and the physics of the twentieth century. Though these differ in their details, they nonetheless demand of the physicist the same attitude toward the methods used in the acquisition of new knowledge.

0.2.2 *Scientific Knowledge from the Viewpoint of the Physicist of Today*

Any chronological listing, now matter how objective it may appear, is a conscious or unconscious value judgment. While the evaluation of historical facts is always with respect to a certain ideological and aesthetic point of view, the situation with regard to the history of physics might seem to be somewhat simpler: We start with the physics of today, and we judge the physics of past epochs in reference to it. Such a procedure would indeed be sufficient if physics were nothing more than a collection of facts. However, physics does not merely observe and describe, that is, it does not only answer the question, "What is there to be found in nature?" but also interprets and seeks connections between the facts. In other words, it asks "Why is it so?" and then seeks the answer. Physicists formulate concepts for the description and for the interpretation of phenomena; they create methods to be used for the demonstration

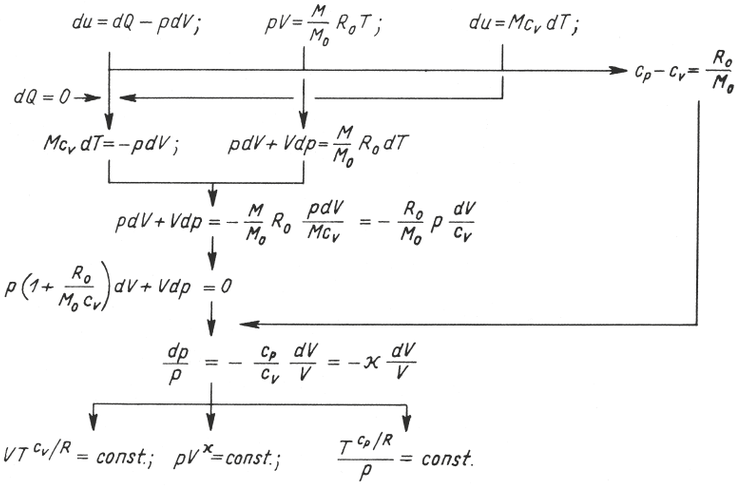
of the truth of achieved results and also for the discovery of new results. The attitude toward the method, toward the truth of the knowledge, the attitude that spans all branches of science, is a philosophical or even ideological posture, and because of this it may play an even more significant role in the history of science than particular concrete results. It is immediately apparent that in such case it would be pointless to base our appraisal on the intensity of scientific activity alone. Since in the following evaluation we will be placing great weight on the methods of scientific inquiry, let us begin by assembling the basic principles that must be observed, consciously or not, by today's experimental or theoretical physicists.

As an example we shall consider a problem that is not terribly complicated, but not trivial either. Our aim here is to investigate the kinds of questions that can be answered by physics as well as the methods and concepts employed by physicists.

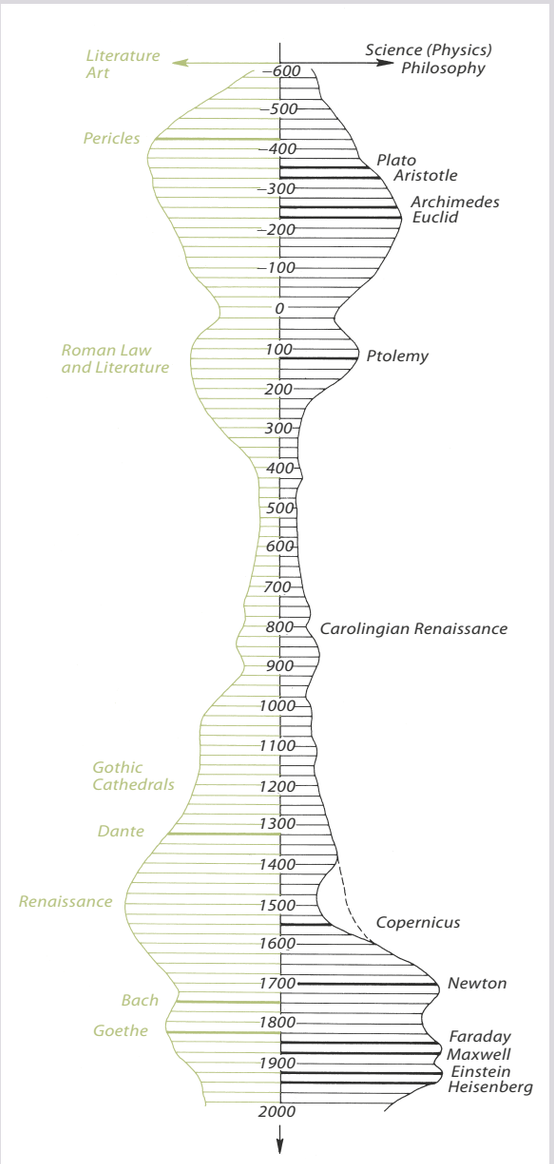
The problem that we shall consider is the following: We are given a sealed cylinder that is closed at one end by a frictionless piston. By pushing on the piston, we rapidly compress the gas within the cylinder. We observe that the temperature of the gas rises (Figure 0.3). The question to be asked now is, How can this process be explained, and what are the quantitative relationships among the physical quantities involved in the process? Take, for example, the question of what the pressure and temperature of the gas will be for a given ratio of initial to final volume. This problem—the behavior of a gas under an adiabatic change of state—is well known.

As we can see, first a concrete physical situation or action is presented; that is, the conditions to be investigated or results to be explained are described. A qualitative explanation of the phenomenon that we are considering here is, as we know, rather easy. If the gas is compressed by the piston, then an external force—in this case our muscles doing the pushing—is doing work. This work increases the internal energy of the gas if we assume that the compression takes place so rapidly that the gas cannot transfer any heat to the environment, that is, that the compression is *adiabatic*. The increase in the internal energy of the gas is expressed by an increase in temperature.

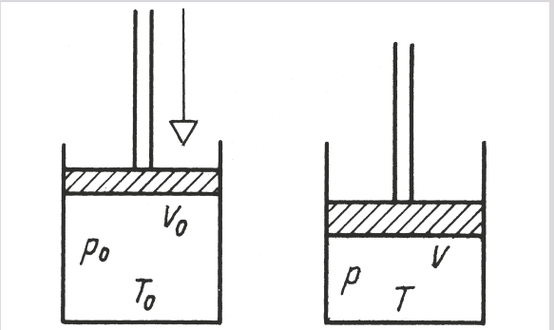
For a quantitative interpretation, we require the principle of conservation of energy, the general state equation of the gas, as well as the relationship between internal energy and temperature. From these equations we derive in the following diagram the equation for the adiabatic change of state —under the assumption that the state equation is adiabatic, that is, that $\Delta Q = 0$.



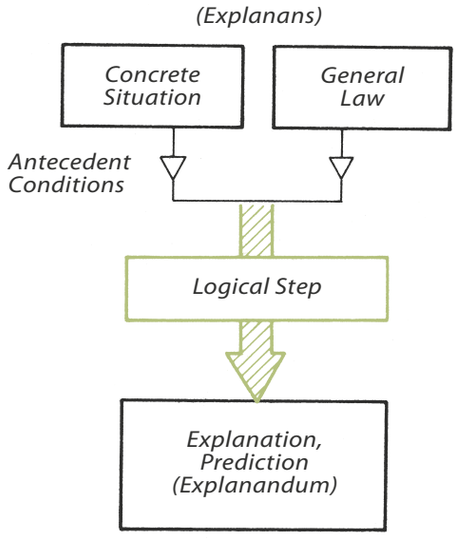
The three initial equations stand at the top of the diagram. The notation is as follows: Q is the quantity of heat absorbed by the gas, V the volume of the gas, μ the internal energy, R_0 the universal gas constant, M the mass of the gas, M_0 the mass of a kilomole of the gas, T the absolute temperature, c_p, c_v the molar specific heats respectively under constant pressure and constant volume, p the pressure, $\kappa = c_p / c_v$, and $R = R_0 / M_0$ the universal gas constant divided by the mass unit.



▲ **Figure 0.2** The intensity of intellectual activity from the pre-Socratic times to the present.

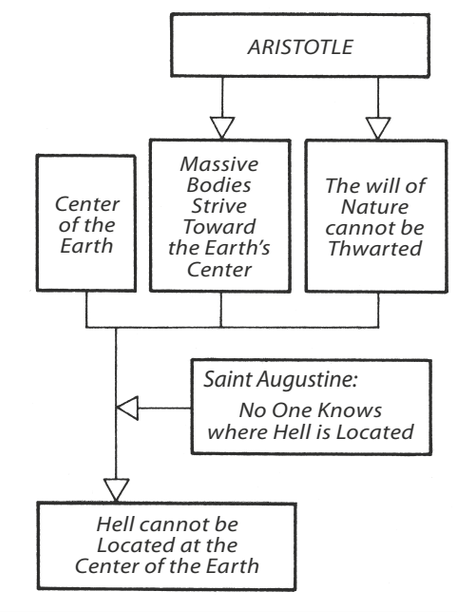


▲ **Figure 0.3** A model illustrating the method of physical explanation of a phenomenon: Why does the temperature of a gas increase when it is rapidly compressed?



▲ **Figure 0.4** The logic of explaining a physical phenomenon.

> [Science is an] activity characterized by three features:
>
> **1.** It is a search for understanding … some aspect of reality.
>
> **2.** … achieved by means … laws applicable to the widest possible variety of phenomena.
>
> **3.** [which] can be tested experimentally.
>
> —M. Goldstein and I. F. Goldstein [1978, p. 6]



▲ **Figure 0.5** Aquinas's answer to the question as to whether hell is located at the center of the Earth. Note that the two "authorities" are integral to the argument. (*Declaratio 36 quaestionum*).

We see that a number of physical laws have to be invoked in order to explain an everyday occurrence, and several complex logical steps must be executed.

From the final formula, one can determine not only the qualitative result that the temperature rises with increasing pressure, but the quantitative increase in temperature as well. This means not only that we have interpreted our observation, but also that we know in advance what the result of a given physical situation will be, since a quantitative explanation enables us to predict the course of future events.

The individual steps in explicating a physical phenomenon can be listed as follows (see Figure 0.4):

1. We begin with a concrete situation and describe it. In the philosophy of science, the initial assumptions are called *antecedents*.
2. We look for physical laws that may be involved in the given problem. Through these laws and the antecedents, one then determines—again in the language of the philosophy of science—the *explanans* (generalized antecedent).
3. We proceed from the results obtained in the first two steps and derive using the methods of scientific deduction the phenomenon to be explained, which in the philosophical jargon is called the *explanandum*. We use only the rules of logic and the equivalent rules of mathematical derivation.

In order for the explanation of a physical phenomenon to be considered satisfactory, the concepts and magnitudes that appear must satisfy certain conditions. For example, a problem can be viewed as a physical problem only if both the antecedent assumptions and the inferences made from them contain only assertions that can be checked by observation. We must insist on this requirement in order to separate questions about our physical environment from the many other possible types of questions, such as those of logic and mathematics, or even metaphysics and theology. Thus, for example, the train of thought depicted in Figure 0.5 can be seen neither as a physical problem nor as an explanation of a physical problem.

Another important requirement is that the antecedents involve a relationship to physical laws. Consider the following statements:

- All the clocks in cupboard *C* are ticking.
- Clock *A* is in cupboard *C*.
- Therefore, clock *A* is ticking.

This sequence of statements satisfies all the conditions outlined above, since it describes a concrete physical situation that can be empirically investigated, and the logical conclusion is also correct. Nonetheless, no one would argue that we have provided an explanation of why clock *A* is ticking.

It is precisely to avoid such obstacles that a general law—a law of nature—must appear in the antecedents. The statement that all the clocks in cupboard C are ticking is certainly not such a general law.

Naturally, it is also a crucial condition that the *explanandum* be derivable solely by the application of logic. In this way, not only do we satisfy the desire of obtaining a logically watertight explanation of a physical phenomenon, we also make possible predictions of future phenomena, and that is what makes technological progress possible.

We have not yet mentioned the importance of the fact that for both the description of the concrete physical situation—in our example the description of the physical properties of the cylinder and piston (such as their geometric dimensions, the initial state of the enclosed gas, etc.)—and the formulation of the physical laws, we used measurable physical quantities. So before quantitative laws can emerge, it is necessary first that the concepts of the quantities that play a role are formulated, and at the next level of development that they become measurable.

It is possible to base an evaluation of progress in the history of physics and the associated classification into historical periods on the formation of physical concepts, and on the turning of these concepts into measurable quantities. It was by means of such an argument that researchers arrived at a reevaluation of the late Scholastic period, for at that time, the question of the intensity of *qualities* was raised, and the process toward them becoming quantitative began.

We have not yet spoken about how we arrive at the laws underlying the starting point of our investigations. Apparently, a universal law should emerge from the various kinds of empirical evidence by applying the inductive method (see Figures 0.6 and 0.7). Although the inductive method remains useful, it has by its very nature a limited range of application. However, with care, the result of an inductive process can be applied in the form of a hypothesis to phenomena that are not identical to those on which the inductive derivation was based, that is, to phenomena that, for example, are related to other values of the variables or even to other ranges of values. In the experimental confirmation of a hypothesis, this can eventually lead to a law that can be applied either in general or within a specified domain of validity.

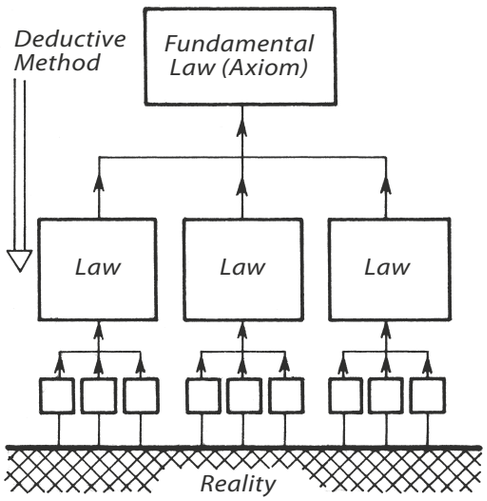
The model sketched in Figures 0.6 and 0.7 needs to be extended: We now regard the laws that cover various domains of reality as laws that directly express the experience. Then they are combined into a general law (Figure 0.8). If we attribute the status of axiom to this general law, then it is possible to discuss phenomena in the given domains deductively—that is, by deriving the particular from the general.

Figure 0.9 shows a simplified diagram of how research into physical laws is carried out and at the same time how the truth of assertions with physical contents can be established. What is significant here is that we must begin with physical reality and must come back to this reality. An assertion that can be properly integrated into this model is held to be true.

However, we should not forget that this model has something to say only about the communication and verification of a discovery, and nothing about how an idea related to physical reality first comes into being.

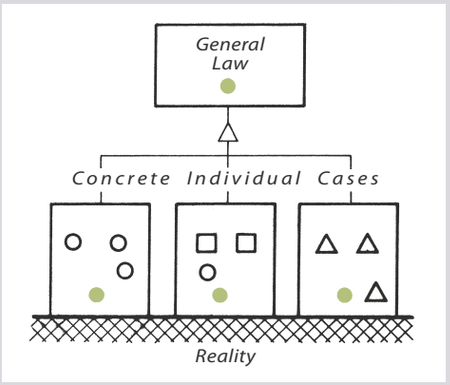
0.2.3 *Division into Epochs Based on Theoretical Synthesis*

The history of physics can also be divided into periods on the basis of discoveries that offered a unified explanation for seemingly disparate phenomena from previously unconnected branches of physics and thereby led to a merging of these branches.



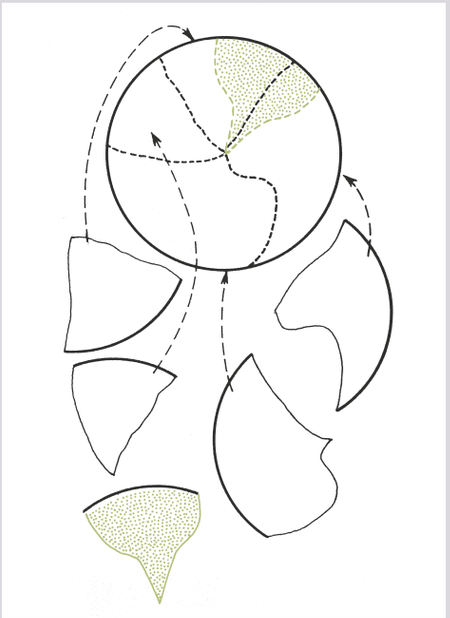
◀ **Figure 0.8:** The hierarchy of laws. Higher-level laws encompass larger classes of phenomena. At the beginning of the seventeenth century, Francis Bacon summarized the method:

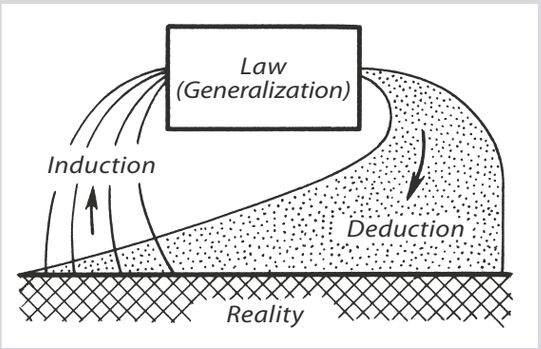
> [Proceed] from works and experiments to extract causes and axioms, and again from those causes and axioms new works and experiments.



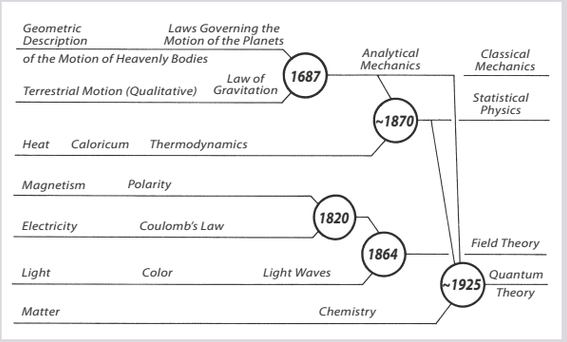
▲ **Figure 0.6** With the inductive method, a general law emerges from a large number of individual cases.

▼ **Figure 0.7** A perhaps more realistic representation of the process by which a general law is derived from concrete observations. The process is similar to the construction of a jigsaw puzzle. The colored region indicates how a new concrete case may be selected for study on the basis of a general law.





▲ **Figure 0.9** Simplified schematic representation of the scientific method. A law encompasses a greater range of phenomena than the individual observations on which it is based. The limitation of the domain of validity is also indicated in the figure.



▲ **Figure 0.10** Junction points in the history of physics. The dates indicate when the relationships between disparate groups of phenomena were recognized:

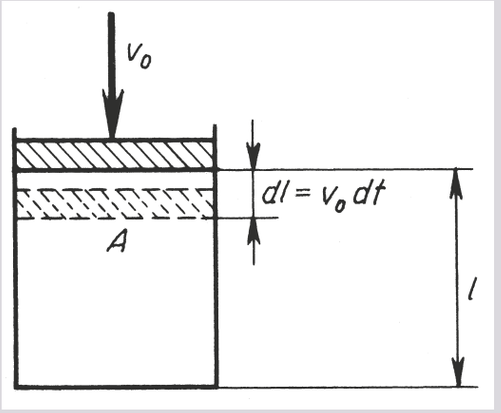
1687: Publication of Newton's *Principia*

1820: Ørsted's discovery of the magnetic properties of electric current

1864: Maxwell's electrodynamics

1870: Development of statistical mechanics

1925: Birth of quantum mechanics (after [Hund 1972, p. 16]).

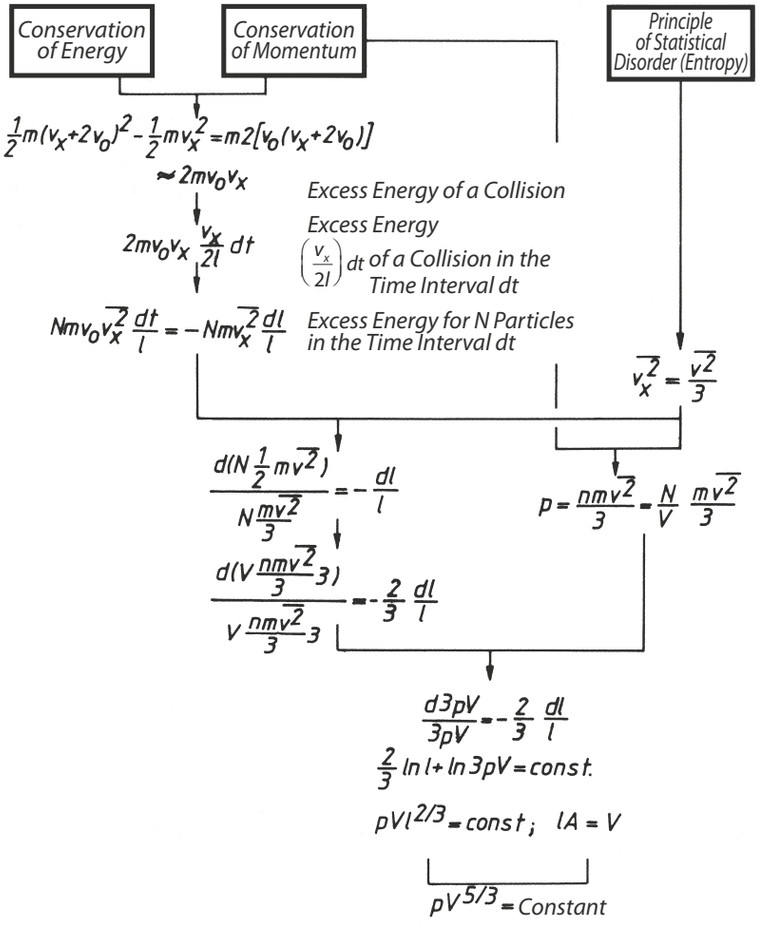


▲ **Figure 0.11** Molecular interpretation of the adiabatic state equation.

Figure 0.10 displays a logical and plausible chronological division of the history of physics into periods due to Friedrich Hund (1896–1997); the points at which branches of physics were united are indicated by the dates in circles.

Let us return now to the concrete physical example that we considered above and investigate in detail the theoretical and aesthetic consequences of the unification of two branches of physics and the practical results to which such unification has led. We can imagine, for example, that the gas particles enclosed in the cylinder are small spheres that on impact with the walls of the cylinder exert a pressure. If the piston enclosing one end of the cylinder is moving, then the speed of the particles on impact with the piston will be different from what it was earlier. The result will be a change (in this particular case an increase) in kinetic energy, and an increase in kinetic energy means an increase in the temperature of the gas (Figure 0.11).

In diagram form, a quantitative solution to the problem would appear as follows:



As we can see, among the fundamental laws from which we started, there is now a new one, which is also of a different nature, namely a law based on statistical probability. The result of this derivation contains more than the previous result, for we have now found a numeric value for the quantity κ. This value is furthermore in good agreement with the value found by experiment.

The derivation above is more satisfying than the previous one in a more general sense as well, since we have the feeling that by reducing macroscopic phenomena to processes in the world of atoms that also obey the laws of nature derived from macroscopic physics, we achieve a deeper insight into nature's workshop.

0.2.4 *The Role of Modeling*

If we look at the periods when particular areas of physics become quantitative, we see differences of nearly two thousand years. We may ask why is it that, in certain areas, our contemporary level of physical understanding was already reached in ancient

Greece—here we are thinking primarily of the principle of the lever, hydrostatics, and the kinematic description of the motions of the heavenly bodies—while the description of other phenomena, such as the motion of bodies here on earth, just to take a simple and important example, was not achieved until the seventeenth century.

It is clear that a significant impetus for the study of a given phenomenon is provided by the role it plays in daily life. As a first approximation, we might even say that for the development of a given area of knowledge, practical applications play the decisive role. But this is only very roughly the case. Specifically, this cannot be said for the Greeks, the founders of science: they were careful to make sure that their results had no practical use, since they wished to indulge only in the kind of science "befitting free men."

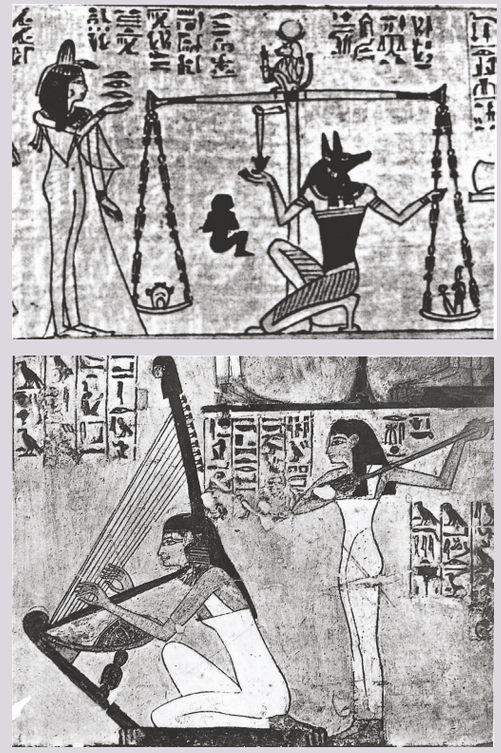
If we consider once more the example that we have been studying in detail, in both of the forms in which it has been presented, we may conclude that idealization and the abstraction from reality that results are of decisive importance. The experimental setup is already artificial, for in nature one will certainly not find cylinders closed at one end by a piston. But even in the context of such an artificial setup, we have to make further idealizations: we assume that the piston is perfectly sealed and that it moves without friction. During compression there is no heat transfer because it happens so *rapidly*. The gas is ideal, which is to say that its internal energy depends only on the temperature.

After so much abstraction, one cannot help asking if some essential element of the problem has been abstracted completely out of the picture, so that a comparison of the theoretical conclusion with experimental observation, that is, with the results of measurement, becomes illusory. For example, let us think of the treatment of adiabatic compression from the atomic viewpoint, in which we have assumed that the gas is compressed at an infinitely *slow* rate, and with this assumption have come to a conclusion that agrees with the macroscopic description involving rapid compression. Here we must now ask whether the two idealizations—the infinitely slow and the infinitely rapid compression—can be brought into any kind of agreement.

Although it is not terribly important for our further discussion, for the sake of completeness, we should mention that in using such expressions as "very rapidly" and "very slowly" in physics, one must always specify *in relation to what*. Clearly, in the two descriptions of compression, we are dealing with two different time constants. The first is the time constant for the transfer of heat from the gas to the environment, and the second is the time constant for the relaxation process of the gas itself. The time during which the piston is in motion should be small in relation to the first time factor, but large in relation to the second. A thorough investigation of the process shows that the two assumptions are compatible with each other to a sufficient degree.

Phenomena in nature—whether or not they are connected to human activities—generally appear in combinations. Just consider that the gas laws described above—in a more complex form and combined with the laws of motion—underlie atmospheric phenomena or the respiration process, the latter also involving chemical reactions.

For a science to become quantitative, the decisive factor, besides the practical importance of the phenomena in question, or the intellectual interest triggered by religious, mystical, or any other motives, is the possibility of abstraction. The sciences became sciences in the order of how close the situations found in nature or in practice were to the abstract situations that would make scientific treatment possible. It seems that in the history of science the most difficult step is that of abstraction, that is, the simplification of a phenomenon in a manner that does not affect its fundamental character while at the same time allows for quantitative investigation. In a similar manner, novelists as well frequently place their characters in abstract or paradoxical situations in order to emphasize the truth of what they are trying to say.



▲ **Figure 0.12** The balance and the harp as used in daily life are almost identical to the "abstract" balance and harp that can be treated mathematically. It is no coincidence that it was precisely for these instruments that the first quantitative natural laws were formulated. (*The Book of the Dead,* twenty-first dynasty, Metropolitan Museum of Art, New York; Wall painting from the tomb of Rekhmire, eighteenth dynasty.)

Quotation 0.7

It is a most remarkable demand that is often made, but never fulfilled, even by those who make it: One should present one's experience without any theoretical apparatus, and let the reader, the student, form his own conclusions. For merely looking at a subject cannot advance our understanding. Every look passes over into an observation, every observation into a meditation, every meditation into a connection, and so one may say that we theorize with every attentive glance at the world. However, to do so with consciousness, with self-knowledge, with freedom, and to make use of a daring word, much skill is needed if the abstraction that we fear is to be harmless and the result of our experience that we hope for is to come alive and be useful.

—Goethe, *On the Theory of Colors*, Preface

Conclusive knowledge rests in every individual as in a seedcase. Through the observation of things we come to wondering, from wondering to concluding, and from concluding, we come to true knowledge.

—Albertus Magnus [Liertz 1932, p. 15]

Quotation 0.8

Every age has scoffed at its predecessor, accusing it of having generalized too boldly and too naively. Descartes used to commiserate with the Ionians. Descartes in his turn makes us smile, and no doubt some day our children will laugh at us. Is there no way of getting at once to the gist of the matter, and thereby escaping the raillery which we foresee? Cannot we be content with experiment alone? No, that is impossible; that would be a complete misunderstanding of the true character of science. The man of science must work with method. Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house. Most important of all, the man of science must exhibit foresight. Carlyle has written somewhere something after this fashion. "Nothing but facts are of importance. John Lackland passed by here. Here is something that is admirable. Here is a reality for which I would give all the theories in the world." Carlyle was a compatriot of Bacon, and, like him, he wished to proclaim his worship of the God of Things as they are.

But Bacon would not have said that. That is the language of the historian. The physicist would most likely have said: "John Lackland passed by here. It is all the same to me, for he will not pass this way again."

—Henri Poincaré, *Science and Hypothesis*, 1905

Quotation 0.9

[Heisenberg quoting Bohr:] "Science is the observation of phenomena and the communication of the results to others, who must check them. Only when we have agreed on what has happened objectively, or on what happens regularly, do we have a basis for understanding. And this whole process of observation and communication proceeds by means of the concepts of classical physics.... It is one of the basic presuppositions of science that we speak of measurements in a language that has basically the same structure as the one in which we speak of everyday experience. We have learned that this language is an inadequate means of communication and orientation, but it is nevertheless the presupposition of all science."

[Later, while washing the dishes after a meal taken at a hikers' hostel, Bohr continues:] "Our washing up is just like our language," Niels said. "We have dirty water and dirty dishcloths, and yet we manage to get the plates and glasses clean. In language, too, we have to work with unclear concepts and a form of logic whose scope is restricted in an unknown way, and yet we use it to bring some clarity into our understanding of nature.

—Werner Heisenberg, *Physics and Beyond: Encounters and Conversations*, 1971

Consider the balance scale. It has been used by mankind since ancient times. The gods of the underworld weighed the value of human deeds, and the kings of this world weighed gold and rare spices (Figure 0.12). The balance is therefore a practical tool, and at the same time if we look at the "abstract balance"—that is, one consisting of a rigid two-armed lever supported at a point at which the lever can move freely without friction—we see that the abstract balance and the real balance used in practice are identical to a high degree of approximation.

But now let us consider the laws of motion for objects set in motion under the power of people or draft animals. If we wish to remove all secondary phenomena, in order to make an abstract treatment possible, then we have to seek out conditions that are very different from those presented at the outset; one might say they do not even resemble the original practical problem. For example, Aristotle (384–322 BCE) derived the law of motion from the everyday observation that two horses can pull a wagon faster than one horse alone. To derive a proper—scientific—law of motion, Galileo had to work with smooth inclined planes and smooth spheres that are not found anywhere in nature, and strictly speaking, he really should have investigated the motion of the spheres in a vacuum.

A division into historical epochs better adapted to the internal structure of science than those previously suggested might begin with the points at which in a particular branch of physics a level of abstraction has been reached that makes the formulation of general laws possible.

Credit must be given to the Greeks—despite their exaggeration in this respect—for the recognition that laws cannot be formulated without abstraction and idealization. Their exaggeration consisted of viewing the abstracted concepts as well as the laws derived from them as the true nature of things; between the model and reality, they gave priority to the model. This point of view, while today outmoded in its fundamental conception, has played a decisive role in the history of science by contributing to the emergence of the fundamental characteristics of modern natural science, namely the tight coupling between physics and mathematics.

0.3 Elements of the Philosophy of Science

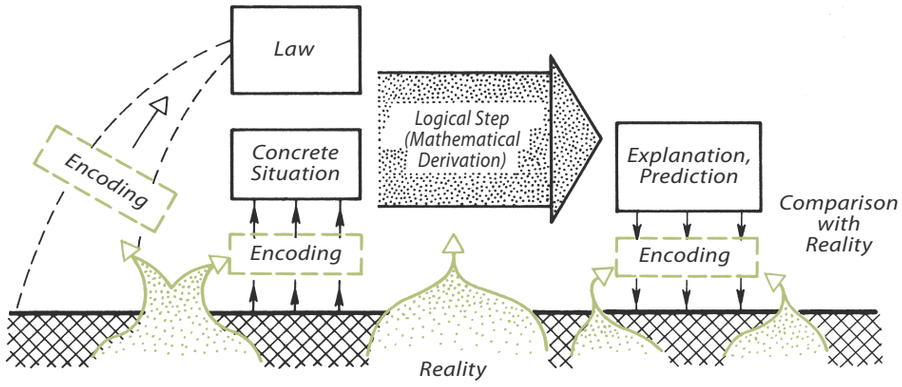
0.3.1 *Illusory Simplicity*

From reality by way of abstraction to natural law, and from law back again to reality—it is over this closed path that science walks. The correctness of a theory, and indeed the correctness of the whole methodology, is thus ensured by this twofold connection with reality.

As we shall see, this insight was long in coming, and it established itself only after significant intellectual struggle. No matter how obvious we consider this method to be today, historically it was not so at all. But even now, if we look more closely at the individual steps of the method, we come up against a jumble of questions for which we are able to supply only a more-or-less satisfactory answer.

In Figure 0.13, the simple diagram of Figure 0.9 is represented in greater detail. The problems that arise are detailed here:

1. For scientific investigation, reality is a formless raw material. Every measurement is already an intervention. By this we do not refer to the quantum-mechanical laws of the microworld, but simply note that the numerical values that our measuring instruments return are in units that are based on an also already established model that is itself based on a simplified structure. This

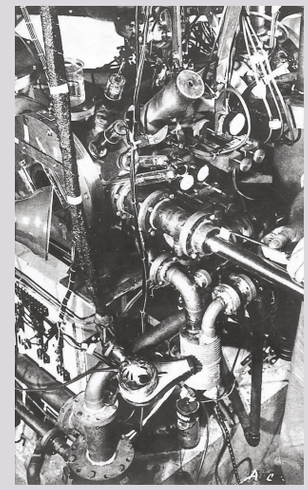


▲ **Figure 0.13** The path of science "from reality to reality" in greater detail.

means that reality appears to us to be already encoded into an "experimental language" even when we wish to describe a concrete situation in its "unadulterated reality" (Figure 0.14, Quotations 0.7 and 0.8). In other words, every experimental attempt already assumes a theory.

2. To interpret a phenomenon, we have assumed the existence of general physical laws. Of course, we have arrived at these laws based on perceived reality, using the inductive method. We thus extrapolate from a finite set of experiences collected in the past and from experiences to be observed in the future, the number of which is in principle infinite. What is to guarantee the validity of this extrapolation?
3. Beginning with a concrete situation and general laws, we arrive, via logical procedures—mathematics and geometry—at new hypotheses, which we wish to check against reality. But what guarantees do we have that the laws of mathematics and geometry apply to reality?
4. The results thus derived must of course again be checked against reality, which as we have already noted is just raw material to be observed and analyzed. Hypotheses derived from theory can be confronted only with hypotheses that in turn have been confirmed or refuted by yet other hypotheses.
5. In connection with all this, there arise the notions of subject and object, which are linked to the separation between the observer and the observed phenomenon. This issue does not arise because physicists have doubts about the objective existence of an external world independent of the observer, but rather because they view the experimental apparatus as an integral component of the external world. Furthermore, it also cannot be denied that the physicists themselves and their system of concepts are also a part of reality.

   Physicists believe in the objective nature of physics as a science, in that within physics, in relation to natural phenomena, assertions can be made that are "intersubjective," that is, that can be understood by anyone of sound mind possessing the requisite education. Moreover, these assertions are assumed to embody the possibility for persons of said sound mind and requisite education to experimentally check, reverify, or reproduce their contents. It is here that social practices for judging objectivity and for establishing the criteria for truth make their appearance (Quotation 0.9). The social practice is a long process: it verifies in the present the assertions of the past. The real problem arises when we want to make assertions for the future with some expectation of certainty.
6. Still another point has to be considered: In order to construct a diagram like that of Figure 0.13, we have to assume that the phenomena of reality can be subdivided into recognizable partial phenomena; that is, the identi-



◀ **Figure 0.14** Confrontation with reality: We use this apparatus to ask of nature directly what the magnetic moment of a particular particle is. Contact with "unadulterated reality" is possible only if a well-developed system of concepts and laws has been established. (*Brookhaven National Laboratory*, from *L'ère atomique*.)

The measurement method serves as the operational definition of those concepts that form the elements of the language of science, and as such, it does not methodologically belong to the theory. But, in addition, the measurement itself is also a physical process, and in this regard it ***is*** a part of the full theory. In this manner the very same process appears both as a premise and as a result in the theory and with this, the language of physics reaps a cyclic, self-consistent structure that is unheard of in other formal languages.

—Mittelstaedt, *Die Sprache der Physik* [***The Language of Physics***], 1972 [p. 85]

Quotation 0.10

The ancients established the axiom that all our ideas come from our senses; and this great truth is, today, no longer a subject of controversy. ...

However, all the sciences do not draw on the same experimental foundation. Pure mathematics requires less than all the others; next come the physic-mathematical sciences; then the physical sciences. ...

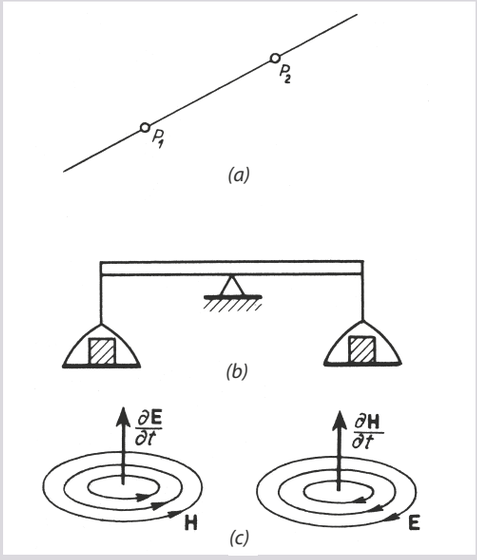
Certainly it would be satisfactory to be able to indicate exactly the point at which each science ceased to be experimental and became entirely rational [read: in order to develop rationally, starting from principles obtained from experiment]; that is, to be able to reduce to the smallest number the truths that it is necessary to infer from experiment and which, once established, suffice to embrace all the ramifications of the science, being combined by reason alone. But this seems to be very difficult. In the desire to penetrate more deeply by reason alone, it is tempting to give obscure definitions, vague and inaccurate demonstrations. It is less inconvenient to take more information from experiment than

continued on next page

Quotation 0.10, ***continued***

would strictly be necessary. The development may seem less elegant, but it will be more complete and more secure. ...

—Lazare Carnot, *Principes généraux de l'équilibre et du mouvement* [Dugas 1988, pp. 323–324]



▲ **Figure 0.15** Axioms and their truth content:

(a) Euclid's first postulate: Given any two points, there exists a straight line joining them. This theorem is easy to accept, but behind the acceptance stands experience.

(b) Archimedes' axiom on static equilibrium: Equilibrium exists when two equal weights are placed on identical arms of a lever. We accept this also, because there is no good reason to think that the lever should tilt in one direction rather than the other.

(c) The first two Maxwell equations: Here it is clearly seen that nature has the last word, for these two equations hold whether we accept them or not.

cal boundary conditions around the subdivisions can be realized again and again. In other words, the joining ofa separated phenomenon at its boundaries with the rest of the world is made possible with the help of characteristic quantities that are determined for the given moment in time.

Classical epistemology as well as modern philosophy of science has grappled with these problems. Naturally, it cannot be our goal here to examine each of these questions in extensive detail. However, we must touch on some of these problems, for in what follows we shall have occasion to refer back to them.

0.3.2 *Reason and Experience*

We have taken it as self-evident that science always has to start with reality and must follow an empirical path to arrive at the fundamental laws embodying the greatest level of abstraction: to the axioms (Quotation 0.10). In the history of science such a schema first became possible in the field of geometry, namely Euclidean geometry, which has exactly this kind of structure (Quotation 0.11). It has been held up for two thousand years as the model for an exact science, so that *more geometrico*, the "geometric method," came to be seen as the ideal for all sciences. As we shall see later, the philosopher Baruch Spinoza (1632–1677) wished to treat even ethics by the "geometric method." But what is the relationship between geometry and reality? Clearly the basic concepts of geometry—more precisely, those of Euclidean geometry—such as line, plane, and triangle, have been abstracted from corresponding forms in our everyday lives. It was the practical value of geometry together with its relatively easy abstraction that assured geometry its standing as a model science. But geometry ungratefully severed the umbilical cord to reality: according to Euclid (ca. 365–ca. 300 BCE) and later geometers, the axioms were true because they could be immediately understood and were self-evident, requiring no further proof.

We should mention here that the basic equations, or axioms, of the axiomatized subfields of physics such as mechanics and electrodynamics—for the former, the Newtonian laws, and for the latter, Maxwell's equations—are true not because they can be immediately understood, but rather because the inferences drawn from them agree with reality (Figure 0.15).

Eventually it turned out—based on the work of, among others, János Bolyai (1802–1860)—that the Euclidean axioms are not that self-evident after all. That is, their particular formulation is not a logical necessity, and indeed, it is possible to create a logically consistent geometry by leaving out one of Euclid's axioms. Later, this question was extended to the physical realm: does reality indeed follow Euclidean geometry? Now the logical role of the geometric axioms necessarily shifted. They could no longer function as inviolable truths; rather, they describe relationships among previously undefined concepts. Since the initial axioms now take on the character of a definition, the question of their truth content no longer arises. The theorems of geometry are no longer seen as making claims about reality. Instead, their truth is of the form, "if–then." Of course, a whole host of interesting questions can still be asked, for example whether the logical structure of the statements implied by the axioms is *consistent*, which is to say, free of internal contradiction, and whether every reasonable question that can be asked within the axiomatic structure can be decided, which is to say, whether the axiom system is itself complete. The most exciting question for the natural sciences is, of course, the question, if geometry is nothing more than a logical construct, then what assurance do we have that it can be applied to reality?

If we choose our logical axioms completely arbitrarily and call any structure thus produced a science—and there is much to be said for such a point of view—then one could say the same about the theory of any number of games, where the rules of the game assume the role of axioms. Even the opinions of authoritative persons could be the basis for such a schema (Figure 0.16).

We sketch here as an example an idea of John Mckinsey (1908–1953) and Patrick Suppes (1922–) on how Newtonian mechanics might be formally, axiomatically, abstracted from reality. These authors asked what such an axiomatized structure would have to look like in order for one of the models, or "realization possibilities," to precisely correspond with the laws of classical mechanics. This logical structure is defined by the following axioms:

1. P is a nonempty finite set.
2. T is a nonempty finite set.
3. If p is an element of the set P, and t an element of the set T, then let $s(p, t)$ be a three-dimensional vector such that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} s(p,t)$$

exists.

4. If p is an element of the set P, then there exists a positive real number $m(p)$.
5. If p is an element of the set P and t an element of the set T, then the quantities

$$f(p,t,1), \quad f(p,t,2), \quad \ldots, \quad f(p,t,i)$$

are three-dimensional vectors such that the series $\sum_i^\infty f(p,t,i)$ is absolutely convergent.

6. If p is an element of the set P and t an element of the set T, then one has the relation

$$m(p)\frac{\mathrm{d}^2}{\mathrm{d}t^2} s(p,t) = \sum_{i=1}^{\infty} f(p,t,i).$$

The abstract structure represented above is brought to life if we associate with the quantities defined in the model the following set of particular quantities taken from the domain of physical reality:

- P: a set of objects considered as point masses;
- T: an interval of time;
- t: a time point in the interval T;
- m: mass;
- s: distance;
- f: force.

We would like to make two observations with regard to these theoretical logical structures.

First, we should like to answer the question, how can one convince oneself that a given structure has application to objective reality, and if it does, what are the limits of that applicability? We might think here of Euclidean geometry. An answer might take a form wherein we can identify elements in reality that correspond to the elements in the abstract structure, and if we can experimentally verify that the relationships between the elements of reality are the same as those between the elements of the structure, then it is clear that all laws present in the logical structure will also be found in reality, or at least to the extent that the elements of the logical structure and the elements of reality correspond to each other. In the concrete example given above from theoretical mechanics, this means that we must restrict our attention to point masses and to speeds that are small compared to the speed of light.

We would now like to illustrate the situation with the comparison depicted in Figure 0.17. The drawing of accurate maps of mine tunnels situated deep beneath the surface of the Earth is a task of great practical importance. The connection to the external world is established only by a short baseline that is projected down the mineshaft. The entire underground map is built off of this baseline. A requirement

Quotation 0.11

Let us now cast an eye over the development of the theoretical system, paying special attention to the relations between the content of the theory and the totality of empirical fact. We are concerned with the eternal antithesis between the two inseparable components of our knowledge, the empirical and the rational, in our department.

We reverence ancient Greece as the cradle of western science. Here for the first time the world witnessed the miracle of a logical system which proceeded from step to step with such precision that every single one of its propositions was absolutely indubitable—I refer to Euclid's geometry. This admirable triumph of reasoning gave the human intellect the necessary confidence in itself for its subsequent achievements. If Euclid failed to kindle your youthful enthusiasm, then you were not born to be a scientific thinker.

But before mankind could be ripe for a science which takes in the whole of reality, a second fundamental truth was needed, which only became common property among philosophers with the advent of Kepler and Galileo. Pure logical thinking cannot yield us any knowledge of the empirical world; all knowledge of reality starts from experience and ends in it. Propositions arrived at by purely logical means are completely empty as regards reality. Because Galileo saw this, and particularly because he drummed it into the scientific world, he is the father of modern physics—indeed, of modern science altogether.

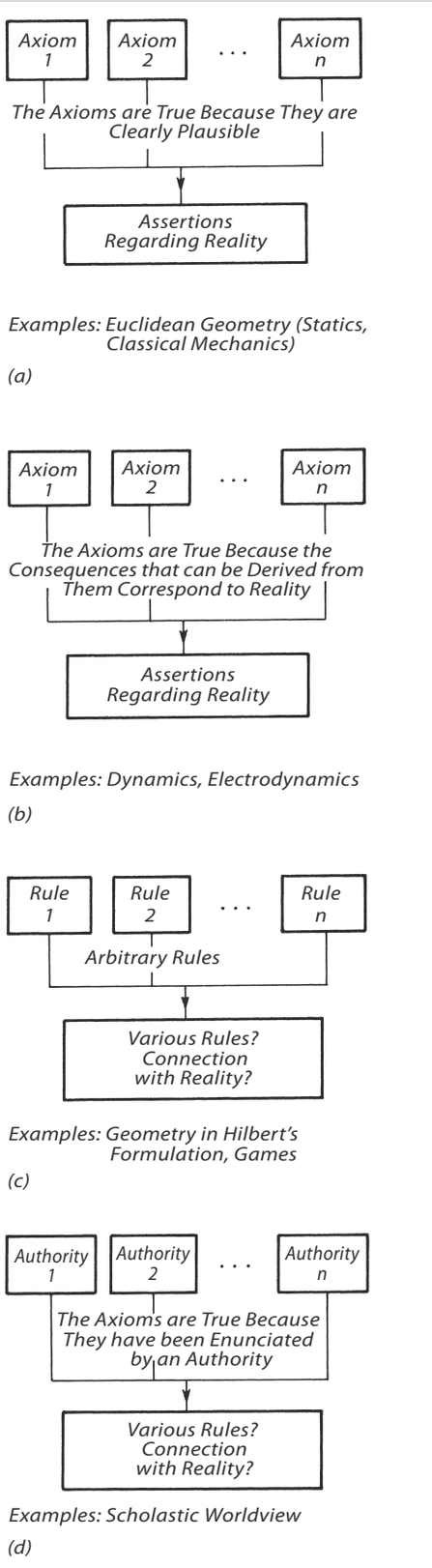
If, then, experience is the alpha and the omega of all our knowledge of reality, what is the function of pure reason in science?

A complete system of theoretical physics is made up of concepts, fundamental laws which are supposed to be valid for those concepts and conclusions which must correspond with our separate experiences; in any theoretical treatise their logical deduction occupies almost the whole book.

This is exactly what happens in Euclid's geometry, except that there the fundamental laws are called axioms and there is no question of the conclusions having to correspond to any sort of experience. If, however, one regards Euclidean geometry as the science of the possible mutual relations of practically rigid bodies in space, that is to say, treats it as a physical science, without abstracting from its original empirical content, the logical homogeneity of geometry and theoretical physics becomes complete.

—Albert Einstein, *The Herbert Spencer Lecture*, 1933

▼ **Figure 0.16** Structures that can be derived deductively.



for such a map is, of course, that if we start boring at a given point on the surface, we should get to a well-defined place within the tunnels. In our metaphor, either of the two coordinate systems can be selected to play the role of the logical structure. So let's arbitrarily choose the coordinate system on the Earth's surface to be the one. The mutually corresponding elements are the baseline integrated into the coordinate network on the surface and its image on the floor of the mineshaft. Instead of the statement that for every proposition within the logical structure we can find a corresponding proposition in reality, in the example we can ask whether junctions on the surface map lie exactly above the corresponding junctions on the map surveyed in the tunnels. Of course, as a consequence of errors in measurement, or because some subtle condition is not reproduced in the underground map, the correspondence between the points can become increasingly fuzzy as one moves away from the baselines. For example, the map on the surface might follow the Earth's curvature, while the map in the tunnels might, incorrectly, fail to take this into account.

We now discuss another observation concerning the application of abstract structures. The fact that certain branches of physics such as mechanics, electrodynamics, and quantum mechanics have been axiomatized independently of one another gives rise to the impression that all natural phenomena can be forced into independent logical structures. This is indeed the case to some extent, but the moment we identify these logical structures with physical content, connections are revealed that were always present but not in that form. Consider, for example, the laws of mechanics. In item 6 of the definition above given by Mckinsey and Suppes, the right-hand side of the equation can be identified with the forces acting on the point mass, namely in concrete cases the gravitational force, a macroscopic electrodynamic force, or the phenomenological expression of microscopic forces, the elastic force. As another example, let us consider that in concrete applications physics and technology require numerical conclusions—that is, are spelled out in actual numbers. But such numerical statements already contain the units of measurement of physical quantities, and in the definition of a single physical quantity, such as in the definition of the unit of time (the second), practically all of physics is already involved. Thus we have the following:

> The unit of time is the second. The second is 9,192,631,770 times the period time of the radiation that corresponds to a transition between to the $F = 4$, MF $= 0$ and $F = 3$, $MF = 0$ hypefine levels of the $^2S½$ base state of the atoms of cesium-133 isotope.

We see that for the precise measurement of the most fundamental concept of the second, we need to be familiar with both nuclear and quantum physics, as well as with all of classical physics. From this, it also follows that in the final analysis, any one particular theorem of physics, by itself, can never be proven either by logical or by experimental means. At any given time in history, the whole system, the system of concepts, the method, and the concrete physical theories must all be taken into account in order for us to accept any single theory as true. This is how the integrity and the wholeness of reality are revealed to physicists.

0.3.3 *Pitfalls of the Inductive Method*

The road to physical laws seems obvious to us: from the knowledge acquired from experience and experimental results, we use induction, that is, we focus on the common traits to find the general, to find the regularities. Here we are not concerned with the question of whether physicists indeed follow this path: that will be discussed later.

We shall see that the important thing is not how the truth was found, but rather the method by which the truth of a statement—a physical law—is determined.

It is easy to see, however, that in the case of laws derived inductively, one can never have complete confidence about their validity for all time. This is because a law—precisely because of its generality—refers to infinitely many individual cases, clearly not all of which can be tested by experiment. This is conventionally expressed as follows: A natural law is not verifiable but it is falsifiable. A single experimental result contradicting a law suffices to refute it. But such statements originated from a time when the task of physics was seen as finding *eternal truths.* Today, on the other hand, we live in the era of approximations. An experiment that "refutes" something does not necessarily lead to the downfall of a theory, but instead helps in defining its boundaries and limitations (Quotation 0.12).

In connection with the issue of the general versus the particular, we would like to mention two paradoxes. Paradoxes often seem forced and thus can be simply a waste of time. However, we should not forget that the study of paradoxes such as "A certain Cretan asserts that every Cretan always lies"—where one is to determine whether the Cretan in question has lied or told the truth—has contributed to the clarification of the foundations of mathematics.

The paradox that arises when a general theorem is derived from assertions about individual cases is frequently called a *raven paradox*, for the following reason.

One is given the task of proving experimentally the general theorem, "all ravens are black." Clearly, the observation of a black raven is to be considered an individual case, representing a partial proof of the general theorem. Many such individual observations then yield a general result. However, the assertion "every raven is black" is logically equivalent to the assertion that "every object that is not black is not a raven." A verification of this assertion is likewise logically equivalent to a verification of the original assertion that "every raven is black." Thus, for example, if we observe that a page of the book we are reading is white, then we have found evidence for the assertion that "every object that is not black is not a raven," since a piece of paper is not black and neither is it a raven. Since this assertion is equivalent to the original assertion, the experiential fact that a piece of paper is white increases our confidence in the general assertion that all ravens are black, and that is surely absurd.

A further paradox along these lines is called the Goodman paradox, after Nelson Goodman (1906–1998), who discovered this paradox in his investigation into the connections between conclusions reached by induction and semantics.

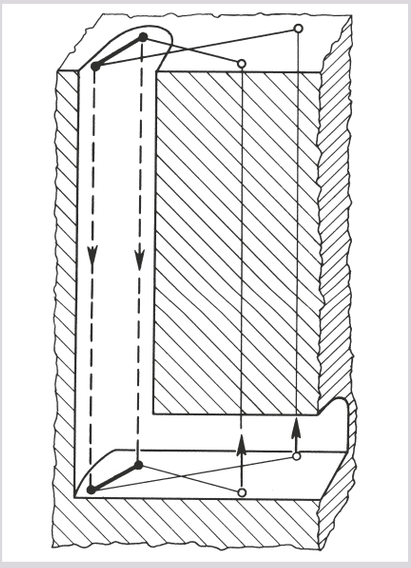
We begin with the assertion, "every emerald is green." We can establish that every emerald found to date has indeed been green, and thus emeralds found in the future should be green as well. We now define a concept, to which we assign its own name. To this end, we form from the words "green" and "blue" the new term "grue." We define an object to be grue if it was green when examined before some time t. All other objects whose color was not investigated before time t will be assigned the color grue if they turn out to be blue when examined after time t. (Although the notion of grueness seems to be rather forced, life can produce much more forced notions than this one.)

Let us now investigate the truth of the assertion that every emerald is grue. Up till the present moment t we have experienced every emerald as green, and by definition, if an object has previously been observed to be green, then it is grue. From this it follows that the assertion "every emerald is grue" has been proved by the inductive method and may therefore be taken as true into the future. In other words, this statement must be confirmed by the results of observations carried out in the future. By definition, we say that for times later than t, an object is grue if it is observed to have the color blue. This means, then, that from the fact that emeralds in the past have been green, we must conclude that emeralds investigated in the future will be blue, which of course is absurd.

0.4 The Dynamism of History

0.4.1 *Forces for Progress*

As emphasized before, we are discussing in this introduction the physics of today. Figure 0.13 is an illustration of the practice of a particular subfield of physics, as physics is learned and taught today. Experimental physicists lay emphasis on the themes that stand in close relation to reality, while theoretical physicists prefer the deductive approach. The detail at which material is presented at various levels of our education mostly differ in the degree of detail in which the tools of physics



▲ **Figure 0.17** Metaphor for the illustration of the connection between the elements of a logical structure and physical phenomena: if one can find an unambiguous two-way correspondence between the basic elements of the logical structure and the fundamental physical phenomena, then the structure can correctly yield all the phenomena in some specific branch of physics. The abstract structure is represented here by the map on the Earth's surface, and the physical phenomena by the map within a tunnel of the mine.

Quotation 0.12

By the traditional philosophical problem of induction I mean some formulation like the following (which I will call "Tr"):

Tr. What is the justification for the belief that the future will be (largely) like the past? Or, perhaps, What is the justification for inductive inferences?

Formulations like these are wrongly put, for several reasons. ...

But every action presupposes a set of expectations; that is, of theories about the world. Which theory shall the man of action choose? Is there such a thing as a rational choice?

This leads us to the pragmatic problems of induction:

Pr_1. Upon which theory should we rely for practical action, from a rational point of view?

Pr_2. Which theory should we prefer for practical action, from a rational point of view?

My answer to Pr_1 is: From a rational point of view, we should not "rely" on any theory, for no theory has been shown to be true, or can be shown to be true.

My answer to Pr_2 is: But we should prefer as a basis for action the best-tested theory.

In other words, there is no "absolute reliance"; but since we have to choose, it will be "rational" to choose the best-tested theory.

—Karl Popper, *Objektive Erkenntnis,* 1973 [pp. 2, 22–22]

Quotation 0.13

Scientific method is what working scientists do, not what other people or even they themselves may say about it. ...

I think that the objectives of all scientists have this in common—that they are all trying to get the correct answer to the particular problem in hand. This may be expressed in more pretentious language as the pursuit of truth. Now if the answer to the problem is correct there must be some way of knowing and proving that it is correct—the very meaning of truth implies the possibility of checking or verification. ...

But to the working scientist himself all this [the usual methods and criteria of truth and their confirmation—the inductive method, the rejection of all authority, the avoidance of all possible sources of error] appears obvious and trite. What appears to him as the essence of the situation is that he is not consciously following any prescribed course of action, but feels complete freedom to utilize any method or device whatever which in the particular situation before him seems likely to yield the correct answer. In his attack on his specific problem he suffers no inhibitions of precedent or authority, but is completely free to adopt any course that his ingenuity is capable of suggesting to him. No one standing on the outside can predict what the individual scientist will do or what method he will follow. In short, science is what scientists do, and there are as many scientific methods as there are individual scientists.

—P. W. Bridgman, *Reflections of a Physicist*, 1950 [pp. 81–83]

Quotation 0.14

It is said that Gauss, when asked by a friend about the progress of some urgent research, replied, "All the formulas and results are finished; I only need to find the way by which I will arrive at them." I don't believe that Gauss said this: he was not so frank; but he must certainly have often thought it.

—Ludwig Boltzmann, *Vorlesungen über Maxwells Theorie der Elektricität und des Lichtes*, 1891 [p. iii]

are deployed; we are thinking here of tools of experimentation and those of mathematics. Lower-level textbooks typically stress on both the inductive and the deductive sides of the subject the parts most closely connected to reality, while books written for a lay audience tend to emphasize the general, the important principles, with a generous dose of simplifications that are necessary in these cases.

We then might say: physical theory—or perhaps, more concretely, an explanatory schema depicted in Figure 0.13—as presented in textbooks provides us with a fossilized, static picture of a given subfield of physics; it gives methods for the discovery and verification of true statements and delimits what questions are meaningful—that is to say answerable—within this structure.

A sentence earlier asserted that all statements that fit into this schema are correct. For example, let the schema represent the structure of classical electrodynamics. The reader should attempt to integrate the following questions and their answers into this schema.

Why do you believe—if indeed you do—the following statements to be correct?

- Between a current of magnitude I flowing through a resistor of resistance R and the measurable voltage V at the two terminals of the resistor one has the relation $V = IR$.
- A change in magnetic induction in a closed circuit induces a voltage.
- In a region of space devoid of conductive matter, the curl of the magnetic field is equal to the time rate of change of the dielectric displacement.
- A unit of current, that is, 1 ampere, will flow in a conductor when a current of also 1 ampere flowing in a parallel conductor exerts a force of exactly 2×10^{-7} newtons on a segment of 1 meter in length of the former conductor.

If you doubt the correctness of any of these assertions, consider what sort of proof you would require to accept their validity.

A brief digression from our subject gives us the opportunity to think about the differences between assertions of the natural sciences and ethical or aesthetic value judgments. Let us consider, for example, why we might come to the following conclusion: "The most beautiful part of J. S. Bach's *St. Matthew Passion* is the B minor aria." And let us then consider how someone might be able to convince us of this if we were not initially of that opinion.

At this point it seems that the intellectual tools with which our physicists and engineers begin their careers consist of ossified knowledge. Where within this schema is there a place for creative activity? It should not be forgotten, however, that most of our physicists and engineers rely in their daily work on received knowledge of this kind. New questions that arise in such cases appear on the left side in Figure 0.13: The familiar structures must be applied to new and newer concrete situations. Remaining with electrodynamics as the example, consider this: Maxwell's equations must continually be solved for new geometric configurations or for materials with new physical properties whenever new technologies are created for practical use (antennas with new directional characteristics, switches, cavity resonators, etc.).

Based on our recognition of the problems of the present, we may now formulate our program for the history of physics—or more precisely, the problem of writing a history of physics—in greater detail. We have essentially to answer the question, how can the physics as presented in textbooks and thus our common property become a living, evolving organism, or in other words, why and when should an old physics book be thrown away and an entirely new one be written in its place?

The historical process can be and typically is approached from a variety of viewpoints. According to one conception, the accent should be placed on the achievements of the great historical figures, the great thinkers; we understand the development of physics when we understand how the ideas of those thinkers developed. From this point of view, the history of physics is essentially restricted to the investigation of the psychology of individual creativity.

It would go beyond the aims of this book, and also beyond the powers of the author, if he wished to carry out such investigations in detail. Nevertheless, it is most instructive to capture a number of historical moments. Thus, for example, we should consider the statement of Percy Bridgman (1882–1960) (Quotation 0.13) that the search for truth proceeds along as many paths as there are investigators.

The author of this book finds himself under the spell of the confession of a great poet that happens also to demonstrate the parallels between the creative processes of artists and scientists. According to this poet, the moment of inspiration does not mean a state of intellectual or emotional stress, but rather it is a completely empty but tense *waiting*, not unlike a hunter stalking his prey. He thinks about nothing in particular; he simply waits expectantly. And when the thought comes to him, its appearance is indeed like that of the wild animal breaking out of the forest: It was already there before in its entirety and had only to reveal itself.

That poet's formulation accords excellently with the anecdote about Carl Friedrich Gauss (1777–1855) in Quotation 0.14: The appearance of solutions is an essentially irrational event, and the role of logic is relevant only in their communication. Naturally we do not mean to say here that the discovery, the creative act, is entirely a result of irrational factors: To the creation of the waiting state, of the empty tense state of readiness, also belongs the laborious preparation, the traversal through the well-worn paths of logic.

From this point of view, the history of physics is the history of the great physicists, and the chapters of a book on the subject could be named for the great physicists. The important questions for such a history are who discovered this or that law, and the most interesting issues are the debates over priority. Physics here is essentially a cumulative process: One giant stands on the shoulders of another, and thus sees further and further.

Nowadays, people are more and more of the opinion—primarily on account of the work of Thomas Kuhn (1922–1996)—that the questions just raised are not the crucial ones. We may attempt to identify Figure 0.13 with Kuhn's notion of *paradigm* (Quotation 0.15) and may regard the development of physics as a process of paradigm changes. The focus then shifts from the psychology of creativity to the laws of how the structures of physics evolve. From these structures one can then only with difficulty tease out the individual achievements and individual laws. The result of all this is Kuhn's theory, according to which the history of physics is the chronicle of a series of revolutionary transformations, in which one paradigm is replaced by another. Typical examples of such *paradigm shifts* are those that we associate with the names Copernicus (1473–1543), Newton (1643–1727), and Einstein (1879–1955) (Quotation 0.16). According to this theory, the history of physics is not cumulative; indeed, subtraction as well as addition plays an important role. Or as the adherents of the theory graphically put it: the creative physicist is not only a master builder, but also a demolition specialist.

There are opinions—such as given by Stephen Toulmin (b. 1922) and the author himself—to the effect that in the development of physics, addition and subtraction indeed both play roles, not in the framework of revolutionary transformation, but rather in something along the lines of the zoological processes of natural selection: In the course of history, new theories appear in excessive numbers. From among these, those that are fit are selected by history, and the remainder are consigned to the ash heap after they have fulfilled their social functions—such as to

Quotation 0.15

According to Kuhn (1962) each period in the development of a scientific discipline is governed by a "paradigm," or to use his more recent terminology (1974), a "disciplinary matrix." The paradigm includes not only the conceptual framework of accepted theories but also a set of approved worked-out problem solutions ("exemplars") used in training students, and a general viewpoint or set of criteria for determining what kinds of problems and solutions are scientifically acceptable.

—Stephen G. Brush, *The Temperature of History: Phases of Science and Culture in the Nineteenth Century*, 1978 [p. 5]

Quotation 0.16

In recent years, however, a few historians of science have been finding it more and more difficult to fulfill the functions that the concept of development by accumulation assigns to them. As chroniclers of an incremental process, they discover that additional research makes it harder, not easier to answer questions like: When was oxygen discovered? Who first conceived of energy conservation? Increasingly, a few of them suspect that these are simply the wrong sorts of questions to ask. Perhaps science does not develop by the accumulation of individual discoveries and inventions. Simultaneously, these same historians confront growing difficulties in distinguishing the "scientific" component of past observation and belief from what their predecessors had readily labeled "error" and "superstition." The more carefully they study, say, Aristotelian dynamics, phlogistic chemistry, or caloric thermodynamics, the more certain they feel that those once current views of nature were, as a whole, neither less scientific nor more the product of human idiosyncrasy than those current today. If these out-of-date beliefs are to be called myths, then myths can be produced by the same sorts of methods and held for the same sorts of reasons that now lead to scientific knowledge. If, on the other hand, they are to be called science, then science has included bodies of belief quite incompatible with the ones we hold today. Given these alternatives, the historian must choose the latter. Out-of-date theories are not in principle unscientific because they have been discarded. That choice, however, makes it difficult to see scientific development as a process of accretion. The same historical research that displays the difficulties in isolating individual inventions and discoveries gives ground for profound doubts about the cumulative process through which these individual contributions to science were thought to have been compounded.

continued on next page

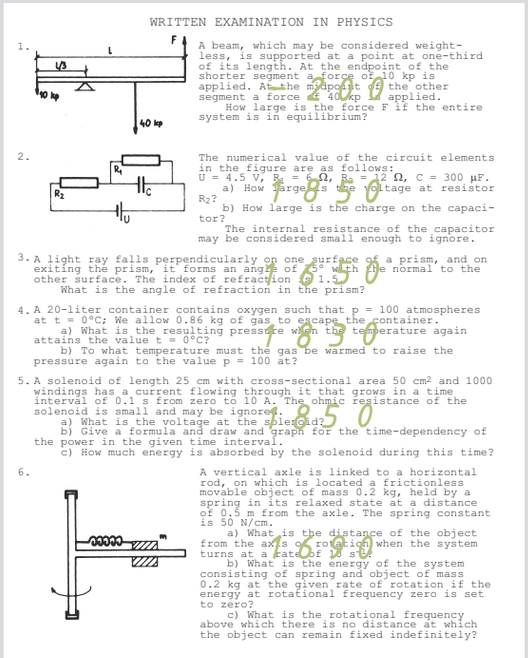
Quotation 0.16, ***continued***

The result of all these doubts and difficulties is a historiographic revolution in the study of science, though one that is still in its early stages. ...

The most obvious examples of scientific revolutions are those famous episodes in scientific development that have often been labeled revolutions before associated with the names of Copernicus, Newton, Lavoisier, and Einstein. More clearly than most other episodes in the history of at least the physical sciences, these display what all scientific revolutions are about. Each of them necessitated the community's rejection of one time-honored scientific theory in favor of another quite incompatible with it. Each produced a consequent shift in the problems available for scientific scrutiny and in the standards by which the profession determined what should count as an admissible problem or as a legitimate problem-solution. And each transformed the scientific imagination in ways that we shall ultimately need to describe as a transformation of the world in which scientific work was done. Such changes, together with the controversies that almost always accompany them, are the defining characteristics of scientific revolutions.

—Thomas Kuhn, *The Structure of Scientific Revolutions*, 1967 [pp. 2–3, 6]

▼ **Figure 0.18** In many countries, to take up the study of physics or engineering one has to be able to solve problems like those shown in the figure. For each problem, the year is given in which the physical laws necessary for a solution were known. In 1700, a very well educated young man would have been able to achieve a grade of 50% on this test and could consequently have qualified to study in many of today's institutions.



have shown that a supposedly productive direction leads to a dead end, or to have provided their author with suitable reputation and a tenured professorship.

Figure 0.18 presents a striking demonstration of the extent to which the scientific knowledge of distant ages still makes up, practically without changes, the subject matter of today's high-school physics.

The idea of evolution suggests another connection. Just as the human embryo recapitulates in nine months the developmental history of perhaps billions of years, so does the development of a child's ideas about the physical world recapitulate—appropriately sped up—the history of physics. Adherents of this theory may find ideas in this direction in the work of Jean Piaget (1896–1980).

If we consider these two, maybe a bit exaggerated, points of view of the history of physics and compare them with the problems of writing history in general, we may perhaps observe a healthy tendency that instead of kings, generals, and statesmen or great battles and peace treaties, the everyday lives of the great masses of people and the changes in their everyday lives get to play the decisive role in the unfolding of historical events.

Let us now leave behind our investigations into the internal genesis of the great thoughts of the great physicists and the theories of paradigm shifts and proceed one step further, to investigate the influences to which the great physicists are subject as social beings, whether they are favorable or unfavorable and which help or hinder the development of new ideas. We are thinking here primarily of influences of a philosophical or artistic nature.

From a host of examples and counterexamples we would like to consider just a few. At the beginning of the nineteenth century, the natural philosophy of the Romantic era proclaimed the unity of all that exists in the world and thereby supported the integration of various types of phenomena (such as the discovery of the magnetic effect of electrical current by Hans Christian Ørsted (1777–1851)) and assisted in the search for unifying principles among disparate phenomena (such as the formulation of the principle of conservation of energy). In the course of our later more detailed treatment, we shall also, see how the same principle in fact delayed Ørsted's discovery and led brilliant minds like Michael Faraday (1791–1867) astray. The second example—and simultaneously a counterexample—is that of Einstein, on whom the philosophy of Ernst Mach (1838–1916) had a great effect (although in his later years, Einstein considered Mach to be outdated) and who was guided by aesthetic considerations in his formulation of the theory of gravitation. However, those same considerations failed him when it came to quantum mechanics (Quotation 0.17).

Finally, it must be noted that physics is a social phenomenon, and thus its development is closely bound up with the *entirety* of such phenomena. Therefore, a complete understanding of physics requires knowledge of its relationship—or to put it more precisely, the historical development of such relationships—to methods of production, technology, art, and philosophy.

It is at this point that Marxism has something to say in its ambitious program. According to its teachings, everything related to production—that is, the entire economic structure—forms the foundation of society in a given historical epoch, and in the end, this foundation is the determining factor for the societal superstructure, namely the legal, political, and ideological institutions and relationships among, for example, the state, the law, morality, religion, philosophy, and art. Accordingly, the

motive force behind every fundamental change in society is the development of methods of production, in particular the perfection of production technology.

However, while emphasizing the determinative, primary character of this foundation, historical materialism is aware of the opposite force of the superstructure on the foundation; that is, it recognizes the active role of the superstructure in societal progress. A very large obstacle in realizing this program is that of placing physics and more generally the natural sciences in the above simplified model: They belong neither to the foundation nor to the *superstructure.*

The history of mankind is a unique and unrepeatable process. All of human activity is inextricably linked to natural phenomena such as geography and weather. In this interplay of influences, physics stands in close relationship with technology on one side, to mathematics and philosophy on another, and, a little bit further, to religious ideology. Physics is inseparable from the *foundations* of every social order through its connection with technology and methods of production. Yet philosophy and religious ideology bind it to the superstructure. If we consider as well that in the course of history the individuals who have taken part in the development of physics have each taken their places in the given political systems by virtue of being members of particular social classes or institutions, we may then begin to appreciate with how many strands physics is linked to other social processes that determine the history of mankind. It is Marxism's contribution to have emphasized economic factors, which have frequently been disregarded.

In this book we shall not ally ourselves inflexibly with any schema. We hope with our eclectic approach to be better able to emphasize the aspects of the subject that we consider important.

Each of the above-mentioned modes of representation and analysis attempts, naturally, to force the history of physics into some pattern or other, into a framework governed by certain natural laws—one could say almost into a procrustean bed. Yet to the concept of natural law belong the requirements of reproducibility and repeatability. When unique processes are under investigation—and the histories of physics, humanity, and the universe are such processes—then the notion of natural law must be given an interpretation different from the usual one. In such cases, natural law is itself the process, or, perhaps better, a description of the process that is as accurate as possible. The possibility of prediction—the principal criterion for the correctness of a scientific assertion—is in principle absent in this case.

If instead of conformity to natural law we speak more modestly of analogies, these could certainly be of help to us in organizing our knowledge, and therefore their pedagogical utility is beyond doubt. One might also dare to make predictions, provided that they are no more trusted than weather forecasts.

0.4.2 *Limits, Possibilities, and Dangers*

The question of the limits of research in physics can be framed from a number of different points of view. First of all, they can be viewed *temporally*: Does progress in physics have a limit, or to put it another way, could such a time arrive when nothing of significance is left to be investigated? Is the course of research in physics similar to that of geographical discovery, in which an object of finite extent—the Earth's surface—is to be discovered, a process that requires a finite amount of time, and where, moreover, one can see today that this task has been more or less completed?

Quotation 0.17

H. L. Anderson: There is no logical way for Einstein, for example, to develop his theory of relativity. He had to take up the idea just from a sort of feeling that there was a unifying principle involved. His belief was that a simple statement would describe more accurately the way in which the universe was organized. ...

R. C. Henry: Concerning your remark on Einstein, I think it is interesting that the same feeling Einstein had, that nature had to be a certain way, the feeling that led him to relativity, caused Einstein to reject the probabilistic quantum theory. Everything since then has tended to indicate that the quantum theory and probabilistic view was quite right, and Einstein was wrong.

So it seems to me that it is not that there really is this sort of "god" underlying nature such that if you follow "his" path, then you will succeed in finding the underlying laws of the universe. If a person has a certain kind of spirit or way of thinking and the spirit happens to fit the particular spirit of the time, he is going to be successful insofar as the philosophy of the time is capable of success. It doesn't mean his idea or approach is right in any larger context.

—Owen Gingerich, *The Nature of Scientific Discovery*, 1975 [pp. 495, 496]

Quotation 0.18

What of the future of this adventure? What will happen ultimately? We are going along guessing the laws; how many laws are we going to have to guess? I do not know. Some of my colleagues say that this fundamental aspect of our science will go on; but I think there will certainly not be perpetual novelty, say for a thousand years. This thing cannot keep on going so that we are always going to discover more and more new laws. If we do, it will become boring that there are so many levels underneath the other. It seems to me that what can happen in the future is either that all the laws become known—that is, if you had enough laws you could compute consequences and they would always agree with experiment, which would be the end of the line—or it may happen that the experiments get harder and harder to make, more and more expensive, so you get 99.9 per cent of the phenomena, but there is always some phenomenon which has just been discovered, which is very hard to measure, and which disagrees; and as soon as you have the explanation of that one there is always another one, and it gets slower and slower and more and more uninteresting. That is another way it may end. But I think it has no end in one way or another.

continued on next page

Quotation 0.18, ***continued***

We are very lucky to live in an age in which we are still making discoveries. It is like the discovery of America—you only discover it once. The age in which we live is the age in which we are discovering the fundamental laws of nature, and that day will never come again. It is very exciting, it is marvelous, but this excitement will have to go. Of course in the future there will be other interests. There will be the interest of the connection of one level of phenomena to another—phenomena in biology and so on, or, if you are talking about exploration, exploring other planets, but there will not still be the same things that we are doing now.

Another thing that will happen is that ultimately, if it turns out that all is known, or it gets very dull, the vigorous philosophy and the careful attention to all these things that I have been talking about will gradually disappear. The philosophers who are always on the outside making stupid remarks will be able to close in, because we cannot push them away by saying, "If you were right we would be able to guess all the rest of the laws," because when the laws are all there they will have an explanation for them. For instance, there are always explanations about why the world is three-dimensional. Well, there is only one world, and it is hard to tell if that explanation is right or not, so that if everything were known there would be some explanation about why those were the right laws. But that explanation would be in a frame that we cannot criticize by arguing that that type of reasoning will not permit us to go further. There will be a degeneration of ideas, just like the degeneration that great explorers feel is occurring when tourists begin moving in on a territory.

In this age people are experiencing a delight, the tremendous delight that you get when you guess how nature will work in a new situation never seen before. From experiments and information in a certain range you can guess what is going to happen in a region where no one has explored before.

—Richard Feynman, *The Character of Physical Law,* 1967 [pp. 172–173]

Quotation 0.19

As regards the elementary particles, which today are at the center of interest among physicists, the question arises in many places whether once the problems related to this area have been solved, we will have arrived at the end of physics.

continued on next page

In addition, one may ask about limits on the *applicability* of the *methodology* of physics: To what degree—if at all—do the successful methods of physics prove useful in other domains of human knowledge, such as in biology or sociology?

Finally, one may ask whether our ethical beliefs might force us to restrict research in certain areas.

In the course of the history of physics it has seemed more than once as though a terminal point had been reached, or was at least at hand. Think of the Aristotelian worldview, within whose framework, it was believed, an answer to every fundamental question was to be found. As we shall see, at the turn of the nineteenth and twentieth centuries, with the extension of classical physics, many people thought that the most important part of research in physics had been completed. Even in the more recent past, such luminaries as Einstein and Heisenberg believed the vision of an all-encompassing unified field theory or a general equation of everything to be a realistic one, whereby the development of physics would likewise have reached its conclusion. However, history has proven that this was but wishful thinking, and today, perhaps no one believes in such a closure.

Yet it seems completely plausible that sooner or later, an answer will be found to every question formulated by science—and there appear to be finitely many of them. We know, however, that every new answer gives birth to a new question, the nature of which we have perhaps no notion whatsoever at present. And finally, it should not be forgotten that physical research implies an interaction between man and nature. The experimenter might create conditions that never before existed anywhere. The investigation of such phenomena can thus lead to the development of entirely new areas of research. In any case, we would like to point out that the great physicists have spoken carefully on this topic (Quotations 0.18–0.20).

From the spectacular successes of the methods of physics, it appears reasonable that an attempt should be made to find applications of these methods to other areas of human endeavor. Just as in earlier times, geometry was considered the model for all exact sciences, in the last two centuries, physics has taken over this role. Indeed, this point of view has frequently been exaggerated to the point where the assertion has been advanced that within any field of research, there is only as much of value as that which can be achieved and verified by application of the scientific method. The success of the method in certain domains is doubtless of great significance, in biology as well as in sociology, to give two examples. To what unhealthy consequences this can lead, however, can best be seen perhaps by the so-called prayer test described in detail in *The Temperature of History* by Stephen Brush. In the latter part of the nineteenth century, some physicists suggested that it might be worth investigating the effect of prayer on the health of sick persons using the methods of experimental physics. (It may be of interest that a hundred years later, practically in our own day, similar research has been carried out.) It should come as no surprise to hear that practitioners of other sciences have launched passionate protest against the hegemony of the physicists' methods (Quotations 0.21–0.23).

The question of ethical limits immediately brings to mind the danger to all mankind posed by the existence of atomic, chemical, and biological weapons. It is now a fact that despite all moral reflection by individuals, science—through open or tacit toleration by society—has long gone beyond such bounds. Genetic engineering today offers the biological sciences the opportunity to bend society's ethical standards. The past teaches us that scientists will always continue to investigate the unknown, wherever a challenge exists, despite any and all prohibitions. It should

also not be forgotten that the ethics of the ages change with time. To quote Saint Thomas Aquinas (1225–1274), all knowledge, even black magic, has its uses; whether for good or ill depends on the uses to which such knowledge is applied.

0.4.3 *Uncertainty in the Precision*

Let us examine closely from where physicists derive their confidence that the methods, and hence the discoveries, of their science are reliable and correct. The answer is quite simple: They have no such confidence! In contrast to general opinion, physicists are not nearly as certain that their assertions are correct as one might assume given the results achieved by applied physics. The birth and development of Western science, and physics in particular, owe much to the idea that the world is somehow ordered, rational, and therefore can be grasped by human reason. Nature reveals itself not as arbitrary and capricious—and therefore unpredictable—acts of gods, but by laws of nature exhibiting a regularity that can be investigated.

This belief in an ordered world is first shaken in the course of a physicist's or engineer's education after they come to believe that with the knowledge of lofty general theories they are in possession of everything needed to solve the theoretical or practical problems that may come their way in the course of their careers. But then it turns out that in their daily work they require above all weighty tomes containing collections of tables and graphs of empirical data, with apparently only a very loose relationship to the lofty general theories.

Despite all this, it has never yet happened in the history of physics that a general belief in the rationality and order of the world has been seriously shaken. Indeed, apparent deviations from such a viewpoint—such as the statistical interpretation of the fundamental laws—evoked strong opposition from some of the best minds of the twentieth century. On the other hand, we note doubts as to the rationality of physical law expressed by Bohr (1885–1962), as reported by Heisenberg (Quotation 0.24). One can arrive at similar doubts if one follows the paradigm model of Kuhn, according to which, as we have noted above, the development of physics is driven by tearing down an old paradigm and constructing a new one in its place.

But what happens if the demolition experts finish their work before the architects and builders are done with theirs? Such could be said to be the case when experimental physics proceeds too rapidly, not giving the theorists time to analyze all the experimental data and work out a theory that encompasses them. In such a situation, no universally recognized overall theory is available to reflect the order and rationality of nature, and indeed, in principle, such a state of affairs could well become permanent.

Physics belongs among the exact sciences. In fact, with respect to exactitude, it holds pride of place. "Logical and precise," is how physics is described, and by those in the humanities this is often stated with a note of reproach in their voices. Physicists are said to be arrogant in the certainty of their knowledge and to be inflexible in their opinions. The author hopes, however, that the foregoing has convinced the reader that this is only an appearance. The physicist knows only too well that the starting point of all logical thinking is some sort of hypothesis, the truth of which is always susceptible to questioning. Thus any conclusions derived from a hypothesis, even with the most precise logic, contain within themselves the original uncertainty of the initial hypotheses. With some exaggeration perhaps one might say that logical thought is the well-worn path that leads with utmost certainty and without a trace of ambiguity from one quagmire to another.

Quotation 0.19, ***continued***

The newest realm of our experience, particle physics, can no longer be dealt with within the framework of the earlier theories that have been worked out and completed—quantum mechanics and the theory of relativity—even though these theories represent very general idealizations. Namely, quantum mechanics postulates—like the old Newtonian mechanics—the existence of unchanging, constant mass points; there was no talk in these theories of a significant transformation of rest energy into kinetic energy. Conversely, the theory of relativity disregarded every unusual aspect of nature that was connected with Planck's constant; thus it postulated the objectivization of phenomena in the sense of classical physics. For particle physics, one had to search for an even more general idealization that would include both quantum theory and the theory of relativity as limiting cases and explain the complex spectrum of elementary particles similarly to how in its time, quantum mechanics explained the complex optical spectrum of the iron atom. Doubtless, this idealization will be expressed in mathematical form. ... However, one may ask whether with this idealization, physics will have come to an end. After all, every physical object consists of particles; one could assert that knowledge of the laws that describe the behavior of particles comes to the same thing as knowledge of the behavior of all physical objects, and in this sense, one could speak of the end of physics.

However, such a conclusion is false, for it deals with a certain point in a way that is less than satisfactory. Even a closed theory of particles—whether or not it is called a "universal theory"—is still to be interpreted as an idealization. It provides an account of an extended domain of phenomena, but there may be additional phenomena that cannot be explained within the framework of this idealization.

—Werner Heisenberg, *Is Physics Coming to an End?* 1976

Quotation 0.20

Interviewer: A few years ago, Stephen Hawking said that he thought the end of theoretical physics might be in sight. I think he was referring to the recent successes in attempting to unify all of physics into a single descriptive scheme. This seems a very provocative statement. How do you feel about this, having spent a lifetime in attempting to unify certain aspects of physics?

Feynman: I've had a lifetime of that, and I've had a lifetime of people who believe that the answer is right around the corner. But again and again it's been a failure. Eddington, who thought that with

continued on next page

Quotation 0.20, ***continued***

the theory of electrons and quantum mechanics everything was going to be simple and then guessed everything, because it was going to be simple, but guessed wrong. Einstein, who thought that he had a unified theory just around the corner, but didn't know anything about nuclei and was unable of course to guess it. And today, there are a large number of things that are not understood. That isn't fully appreciated, and people think they're very close to the answer, but I don't think so.

—P. C. W. Davies and Julian Brown, *Superstrings: A Theory of Everything?* 1988 [pp. 192–193]

Quotation 0.21

On Friday noon, July the twentieth, 1714, the finest bridge in all Peru broke and precipitated five travelers into the gulf below. This bridge was on the high-road between Lima and Cuzco and hundreds of persons passed over it every day. ...

St. Louis of France himself protected it, by his name and by the little mud church on the further side. The bridge seemed to be among the things that last forever; it was unthinkable that it should break. ...

Everyone was very deeply impressed, but only one person did anything about it, and that was Brother Juniper. By a series of coincidences so extraordinary that one almost suspects the presence of some Intention, this little red-haired Franciscan from Northern Italy happened to be in Peru converting the Indians and had happened to witness the accident. ...

It seemed to Brother Juniper that it was high time for theology to take its place among the exact sciences and he had long intended putting it there. What he had lacked hitherto was a laboratory. Oh, there had never been any lack of specimens; any number of his charges had met calamity—spiders had stung them; their lungs had been touched; their houses had burned down and things had happened to their children from which one averts the mind. But these occasions of human woe had never been quite fit for scientific examination. They had lacked what our good savants were later to call proper control. The accident had been dependent on human error, for example, or had contained elements of probability. But this collapse of the bridge of San Luis Rey was a sheer Act of God. It afforded a perfect laboratory. Here at last one could surprise His intentions in a pure state. ...

This was not the first time that Brother Juniper had tried to resort to such methods. Often on the long trips he had to make (scurrying from parish to parish, his robe tucked up about his knees, for haste) he would fall to dreaming of experiments that justify the ways of God to man. For instance, a complete record of the Prayers for Rain and their results.

—Thornton Wilder, *The Bridge of San Luis Rey,* 1927 [pp. 15–21]

However, physicists may at least rest assured that although they may not be in possession of irrefutable knowledge, the models that they have created for use in their work provide a range of application that extends as far as the validity of the theory. Thus, for example, the laws of classical mechanics can be applied with confidence to elastic spheres of finite macroscopic size.

Unfortunately, this does not resolve all of our problems. An assertion can be made with confidence only if it is clear at the outset that the phenomenon under investigation can in fact be described with the given model. Let us take as an example the following simple situation: An experimental physicist is presented with two uniform spherical objects with diameters of order of magnitude 10 centimeters and is asked how the objects will behave if they are made to collide with each other. The experimental physicist establishes that both spheres are composed of the same homogeneous metal and without hesitation predicts that the collision will obey the laws of classical physics. The physicist can make more-precise predictions about the outcome of the experiment—that is, the changes in the spheres' velocities—upon determining the precise initial data. If the experiment is carried out, it turns out that not only the experimental physicist, but the entire laboratory as well, is turned into plasma at a temperature of several million degrees. It so happens that the two spheres were made of the metal uranium-235, and their size placed each of them just below critical mass, so that on collision, critical mass was exceeded. What happened is easily explained after the fact: The event lay outside the laws of classical physics. But the reliability of physics is assured precisely by its ability to make predictions.

The physicist will be made still more uncertain by the following fact: Let us assume that a set of phenomena can be precisely delineated as the range of validity of a theory. We would then be in possession of a logical system for describing the given phenomena whose statements, conclusions, etc., match uniquely the corresponding phenomena in the real world. However, in regard to such logical systems, Kurt Gödel (1906–1978) formulated in 1931 a theorem that represents a milestone in the history of mathematics: Within every sufficiently powerful consistent logical system, there exist assertions the truth or falsity of which cannot be determined within that system. Translated into the language of physics, this means that even a description of phenomena within the bounds of a given model cannot guarantee an unambiguous answer to every question.

All this speculation, it would seem, is rather artificial and lies far outside the set of problems relating to reality and daily life, where one can see everywhere that the certainty of science's pronouncements—the accuracy of its predictions—is a fact. Admittedly, the questions we have raised are ideological rather than practical questions for a physicist. However, it could happen in everyday life that physicists would be unable to prove their claims "beyond a shadow of a doubt." These words were not chosen carelessly. Let us imagine, for example, that an inventor wishes to patent an invention that purports to function as a perpetual motion machine. The patent examiner expects that an expert consultant will offer proofs of the impossibility of such an apparatus. In the final analysis, any such proof will rest on the law of conservation of energy. As for the validity of that law, physics texts always appeal to the fact that all of our experience supports the law, and to this day, no one has ever succeeded in creating and operating a perpetual motion machine. The reader may have noticed immediately how weak this argument sounds—especially to the inventor.

In the foregoing, we have attempted to convince the reader that physicists are not self-satisfied aggressive beings, blinded by their belief of the incontestability of their results, and forming their worldview according to their preconceptions —or at least to assure the reader that they have no right to be so.

Most people are aware of the danger of overestimating the value of logic, but they can also see the danger lying in the opposite direction, namely of shifting the emphasis from reason to emotion (Quotation 0.25).

Many people lead their lives according to other ideals. Perhaps humanity will succeed in establishing a culture that considers a variety of approaches to be compatible with one another, one in which the attitudes of physicists also have a place (Quotations 0.26 and 0.27).

0.4.4 *Physics in a New Role*

To a physicist, physics, with all its doubts and uncertainties and in spite of a potentially ignoble role in the future history of mankind, is nevertheless the most marvelous creation of the human imagination, a creation capable of filling an active life with meaning. Physics is by itself incapable of pointing the way to answering the great questions of human existence. Physics is ethically neutral, although the physicist is not. Yet the physicist is convinced that although no ethical categories are applicable to theories about the physical universe, aesthetic ones certainly are.

It is still the case today that the utility of the sciences takes precedence over all other arguments in their favor. Wherever one looks—on the front pages of the daily newspapers, in the proceedings of learned societies, or in long-range planning documents—everywhere one reads or hears about the implementation of scientific discoveries. Throughout the world, even scientific institutions involved exclusively in basic research justify their existence with claims that the results of their work will sooner or later find practical application. All this is as it should be, and considering the current state of the world economy, the usefulness of a scientific result is indeed of paramount importance. Nevertheless, especially in more highly developed societies, physics—and the other sciences as well—should gradually be given a new role—something similar to the role of the arts. In addition to their usefulness, we need to bring to the forefront the aesthetic qualities of the sciences and thus recognize the beauty of scientific results (Quotations 0.28 and 0.29). Saying it another way, we should extend the idea of usefulness: Science has been viewed as useful because it has contributed to mankind's material wellbeing by satisfying immediate material needs. However, beyond that, it should be recognized as being of crucial social significance in satisfying cultural and spiritual needs. The thirst for true knowledge about the world, independent of the practical utility of the knowledge, is a real social phenomenon; the joy of understanding, of learning is for many people the same as the joy and pleasure they derive from works of art. If we ask ourselves whether the study of semiconductors is of greater or lesser utility than theoretical speculation about the curvature of the universe, the answer will assuredly be in favor of semiconductors. We can offer as justification that with the use of semiconductors, cheap radio receivers can be built. And what do we need radios for? We need them, among other things, to listen to broadcasts about all those interesting theories regarding the nature of the cosmos, and in particular, about the curvature of the universe.

We do not need to further analyze the joy that one experiences from new knowledge. Nonetheless, it is undeniable that in order to derive pleasure from the beauty of science, one has to study in preparation; in effect one has to develop an "eye" for it. But the same holds for the aesthetic enjoyment of a work of art. Figure 0.19 shows the fundamental equations of general relativity, a modernist sculpture, and a poem of our era. Most physicists see in Einstein's equation derivable laws that, building on the Newtonian worldview, provide a vision of the cosmos, joining geometry and physics in an intimate bond, not only a concrete scientific and logical system, but also a work of art whose significance lies in its aesthetic value. Admittedly, it takes more effort to see the aesthetics behind the symbols than in the sculpture or poem. But the appreciation of abstract works of art also requires of the layman who is open to art a particular kind of education and preparation as well as a certain effort. It thus becomes a question of individual preference what sort of intellectual investment brings about the greatest satisfaction.

$$R_{ik} - \frac{1}{2} R g_{ik} = -\varkappa T_{ik}$$

(a)



(b)

Who, if I cried out, would hear me among the Angelic Orders?
And even if one were to suddenly
take me to its heart, I would vanish into its
stronger existence. For beauty is nothing but
the beginning of terror, that we are still able to bear,
and we revere it so, because it calmly disdains
to destroy us. Every Angel is terror.

…Early successes, Creation's favourite ones,
mountain-chains, ridges reddened by dawns
of all origin – pollen of flowering godhead,
junctions of light, corridors, stairs, thrones,
spaces of being, shields of bliss, tempests
of storm-filled, delighted feeling and, suddenly, solitary
mirrors: gathering their own out-streamed beauty
back into their faces again…

—Rainer Maria Rilke, *Duino Elegies*

(c)

▲ **Figure 0.19** To appreciate the beauty of **(a)** the general theory of relativity, **(b)** a sculpture, or **(c)** a poem, one requires, in each case, a certain willingness to learn and a considerable investment of intellectual effort. Einstein's equation, which brings together the ideas of mass and the geometry of space, yields astounding new knowledge about our entire universe.

> Einstein introduced the idea that something which is beautiful is very likely to be valuable in describing fundamental physics. This is really a more fundamental idea than any previous idea. I think we owe it to Einstein more than to anyone else that one needs to have beauty in mathematical equations which describe fundamental physical theories
>
> —Dirac [Kragh 1980, p. 285]

Waiting for Godot, a statue by the Hungarian sculptor Miklós Borsos, inspired by the play of Samuel Beckett. *Duino Elegies* by Rainer Maria Rilke, translated by A. S. Kline (© 2001 All Rights Reserved).

Quotation 0.22

I receive mathematics as the most sublime and useful science, so long as they are applied in their proper place; but I cannot commend the misuse of them in matters which do not belong to their sphere, and in which, noble science as they are, they seem to be mere nonsense. As if, forsooth! things only exist when they can be mathematically demonstrated. It would be foolish for a man not to believe in his mistress's love because she could not prove it to him mathematically. She can mathematically prove her dowry, but not her love.

—Johann Peter Eckerman, in Conversation with Goethe, December 20, 1826 [Moorhead 1930, p. 139]

Quotation 0.23

"Big Science" in no way implies a science concerned with the most important things on our planet, nor is it the science of the human psyche and intellect: it is exclusively that science that promises money, energy, or power, even if it is only the power destroying the really great and beautiful. The primacy of physics among the sciences is not of man's making. In the system of the sciences, physics forms the basis. Every successful analysis, on every, even the highest integrational plane of natural systems, is a step "downward" to physics. ...

We contend that physics, too, rests on a foundation, and that this foundation is a biological science, namely the science of the functioning of the human mind, in other words, epistemology. ...

Well-known sayings, such as that all research is science insofar as it involves mathematics, or that science consists in "measuring what is measurable, and making measurable what is not measurable," are epistemologically the greatest nonsense that ever came from the lips of those who should know better.

Although these pseudo-wise dicta are demonstrably false, their influence still dominates the picture of science. It is fashionable to make use of methods as physicslike as possible, irrespective of whether or not they promise success in the investigation of the particular object.

—Konrad Lorenz. *Civilized Man's Eight Deadly Sins,* 1973 [pp. 92–94]

Quotation 0.24

Heisenberg: I remember that I once had a discussion with Niels Bohr where he doubted whether we would ever find such a mathematical scheme. He felt nature might be so irrational that we could never get any kind of good mathematical description.

—Gingerich, 1975 [pp. 568–569]

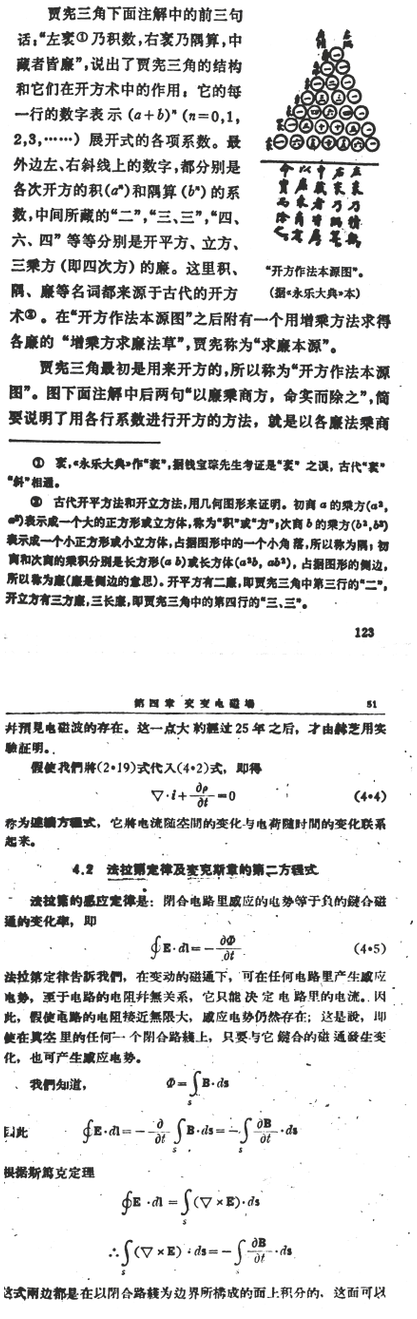
0.4.5 *Characterization of Epochs in Physics*

To summarize our foregoing discussion, we may distinguish the following significant milestones in the history of physics: As we have mentioned, one of the most important steps forward occurred when out of curiosity, mankind began to pose questions of nature and expect answers in return. Both this step and another—the uniting of mathematics and physics—were carried out by the Greeks about 600 BCE. From that point on, for a long time, the philosophical assumption was that the human mind can grasp the basic laws of nature by pure reason alone and, by utilizing these laws, can explain the illusory world of phenomena. We must add immediately that such a conception is not alien, even to many—above all theoretical—physicists of the twentieth century and that it has proven itself fruitful throughout the entire history of science.

After the classical and Hellenistic periods, significant progress was made in the Middle Ages through the development from the qualitative to the quantitative. The beginning of the modern era in the history of science coincides with the period in which mankind began to doubt whether the fundamental laws of nature are capable of being grasped by reason alone. A crucially important factor that made possible the development of European science is, however, the belief that in spite of the limits of the human mind, a rational—that is, graspable by human reason—world order exists, one that can be described by mathematical concepts. From the seventeenth century on, we find the interplay between experiment and theory that characterizes our epoch. However, we shall see later how difficult it was even for Galileo to free himself from the shackles of the Aristotelian mode of thought.

The beginning of the seventeenth century, that is, the year 1600, did not mark a turning point, although the pyre where Giordano Bruno (1548–1600) was burned at the stake in this year is a suitable beacon marking the beginning of fundamental changes. At that time, there existed a closed worldview, which encompassed the entire classical heritage in a version that seemed compatible with Christian dogma. This homogeneous worldview, unique in human history, united science, faith, and philosophy, in which everything—God, man, and matter—had its assigned place. No detail of this picture could be altered, for even phenomena in the physical world of seeming insignificance fit seamlessly into a general principle that itself was bound up with faith and therefore with the omnipotence of the Church. But it was precisely such small, concrete, and gradually accumulating facts that eventually revealed fissures and internal contradictions and tensions in the system, and led ultimately, a half century later, to the collapse of the imposing structure erected by Aristotle and Aquinas (1225–1274).

Around 1650, intellectually open-minded individuals were examining questions similar to those under investigation today and using modern methods to do so. But only Descartes (1596–1650), ahead of his time, attempted to realize a synthesis of physics and philosophy free of the auspices of religion. At the end of the seventeenth century, all conditions for the birth of the new, Newtonian, worldview were in place. From that point on, the history of physics coincides with the historical representation of physics as it is taught today. A significant portion of physics currently taught in universities, such as the laws of Kepler, Descartes, Snellius, Newton, Boyle, and Mariotte were already known at this time. Even the theoretical underpinnings of the Apollo moon-landing program are prescribed by the Newtonian laws of gravitation and motion.



贾宪三角下面注解中的前三句话，"左袤①乃积数，右袤乃隅算，中藏者皆廉"，说出了贾宪三角的结构和它们在开方术中的作用：它的每一行的数字表示 $(a+b)^n$ $(n=0,1,2,3,\cdots\cdots)$ 展开式的各项系数。最外边左、右斜线上的数字，都分别是各次开方的积 (a^n) 和隅算 (b^n) 的系数，中间所藏的"二"，"三、三"，"四、六、四"等等分别是开平方、立方、三乘方（即四次方）的廉。这里积、隅、廉等名词都来源于古代的开方术②。在"开方作法本源图"之后附有一个用增乘方法求得各廉的"增乘方求廉法草"，贾宪称为"求廉本源"。

"开方作法本源图"。（据《永乐大典》本）

贾宪三角最初是用来开方的，所以称为"开方作法本源图"。图下面注解中后两句"以廉乘商方，命实而除之"，简要说明了用各行系数进行开方的方法，就是以各廉法乘商

① 袤，《永乐大典》作"衮"，据钱宝琮先生考证是"袤"之误，古代"袤""斜"相通。

② 古代开平方法和开立方法，用几何图形来证明。初商 a 的乘方 (a^2, a^3) 表示成一个大的正方形或立方体，称为"积"或"方"；次商 b 的乘方 (b^2, b^3) 表示成一个小正方形或小立方体，占据图形中的一个小角落，所以称为隅；初商和次商的乘积分别是长方形 $(a\,b)$ 或长方体 (a^2b, ab^2)，占据图形的侧边，所以称为廉（廉是侧边的意思）。开平方有二廉，即贾宪三角中第三行的"二"，开立方有三方廉，三长廉，即贾宪三角中的第四行的"三、三"。

123

第四章 交变电磁场 51

并预见电磁波的存在。这一点大约经过25年之后，才由赫兹用实验证明。

假使我们将(2·19)式代入(4·2)式，即得

$$\nabla\cdot i+\frac{\partial\rho}{\partial t}=0 \qquad (4\cdot4)$$

称为继续方程式，它将电流随空间的变化与电荷随时间的变化联系起来。

4.2 法拉第定律及麦克斯韦的第二方程式

法拉第的感应定律是：闭合电路里感应的电势等于负的链合磁通的变化率，即

$$\oint \mathbf{E}\cdot d\mathbf{l}=-\frac{\partial\Phi}{\partial t} \qquad (4\cdot5)$$

法拉第定律告诉我们，在变动的磁通下，可在任何电路里产生感应电势，至于电路的电阻并無关系，它只能决定电路里的电流。因此，假使电路的电阻接近無限大，感应电势仍然存在；这是說，即使在真空里的任何一个闭合路綫上，只要与它链合的磁通發生变化，也可产生感应电势。

我們知道，$\Phi=\int_s \mathbf{B}\cdot d\mathbf{s}$

因此 $\oint \mathbf{E}\cdot d\mathbf{l}=-\frac{\partial}{\partial t}\int_s \mathbf{B}\cdot d\mathbf{s}=-\int_s \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{s}$

根据斯篤克定理

$$\oint \mathbf{E}\cdot d\mathbf{l}=\int_s (\nabla\times\mathbf{E})\cdot d\mathbf{s}$$

$$\therefore \int_s (\nabla\times\mathbf{E})\cdot d\mathbf{s}=-\int_s \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{s}$$

这式兩边都是在以閉合路綫为边界所構成的面上积分的，这面可以

▲ **Figure 0.20** The Chinese are justifiably proud of their ancient culture. In a 1978 book on the history of ancient Chinese science and technology, we see a drawing of Pascal's triangle that appeared long before Pascal. However, in another textbook on electrodynamics, we find the names of Maxwell and Faraday and their equations. The underlined Chinese characters for the respective names can be transliterated as follows: fǎ lā dì (Faraday) and mài kè sī wěi (Maxwell).

Of greater importance is that at that time, the methodological principles of scientific investigation, marked by a productive interplay between theory and experiment, had been established, with experiment having the veto power.

The intellectual revolution in seventeenth-century Europe is without parallel in human history. One could make similar claims for the ancient Greek, Indian, and Chinese cultures (Figure 0.20) or the achievements of the European and non-European cultures at the time of the Renaissance, although such comparisons leave a great deal of latitude for national bias on account of the lack of objective standards. There is room for a great diversity of opinion regarding the *Venus de Milo*, Michelangelo's *David*, Aristotle, and Aquinas. The culture of the natural sciences in the seventeenth century, on the other hand, has become the common property of all of mankind, independent of Europe, even though it is a product of European culture. An affirmation of this culture is not a question of taste, determined by one or another cultural group, but rather represents for every group and every nation the necessary—though alas not also sufficient—conditions for survival.

The greatest contribution of the following century is the dissemination of a new way of thinking and the results that such thinking achieved. Physics, and with it science in general, became fashionable, above all in the salons of the day. The *Encyclopédie* of Denis Diderot (1713–1784) was a unique attempt to assign to mankind—no longer occupying the center of the universe—a new place, one that accorded with the new discoveries of science. Through the end of the century, the implications of Newtonian mechanics were worked out in detail, and the worldview of physics was simplified. Pierre Laplace (1749–1827) could still seriously believe that a mind of superhuman capacity could in principle determine all that was taking place in the present and would take place in the future given the equations of state and motion of the universe at its inception.

Quotation 0.25

Heisenberg: Well, it may be that they consider this over-emphasis of rationalism in science as a danger. They feel that this over-rationalizing everything really makes it impossible to get to any kind of stability and, therefore, they tend towards irrational. Now everybody, of course, knows how dangerous political movements can be when they become irrational. I am thinking of my country. But of course you can look at other countries nowadays and it's almost the same thing. This entering of irrational motives in political life is a great danger, but at the same time it may be necessary in order to get the new stability, because rationalism alone, so far as I can see, is not sufficient to establish a sound basis of a social community. This is simply because, when you go to the rational arguments, they always lead from assumptions to results, but you can never prove whether the assumption is right.

—Gingerich, 1975 [p. 562]

Quotation 0.26

... Adherence to science, logic and rationalism is in reality only a more or less dogmatic adherence to a certain Western form of life with its origin in ancient Greek philosophy. [...] [Hodgson] argues that science cannot solve all kinds of problems. It cannot tell us what is good or beautiful, and thus cannot solve esthetic and ethic problems, which perhaps are at least as important as the problem of giving scientific explanations and predictions of phenomena.

—Gunnar Andersson, quoting Feyerabend [1979, pp. 5–6]

Quotation 0.27

... And for just this reason, I maintain that we shall not escape from this crisis of today, indeed, we shall not avoid the catastrophe if we do not undertake a comprehensive cultural dialogue among all nations, all races, all continents.

There is no peace, no culture, no genuine civilization without the exchange of ideas. Just consider the great civilizations of history: Indeed, they were all mixed civilizations. Mixed in a biological sense and mixed in a cultural sense.

continued on next page

Quotation 0.27, ***continued***

The limitations of your Western culture are clearly visible. This culture alone cannot make the world happy. No matter how much good this culture has done, it is clear that the one-sidedness with which it sets technological mastery over nature as its goal endangers our entire planet. This culture inspires a blind development without any truly human goal, a development that consists in infinite growth of production and consumption.

For us, culture means something quite different.

Since Aristotle, you have emphasized the role and importance of argumentational reasoning. For you, God means above all "reason." For us, he is above all the "force," the great "life force," the source of all life and all life force into which these same forces return. You place the accent on *thought*. We place it on *feeling*. With you, *ratio* dominates, while with us it is *intuition*.

A truly universal civilization must strive toward an equilibrium between these two tendencies.

—Léopold Senghor, Interview, 1980

Quotation 0.28

And since goodness is distinct from beauty (for it is always in actions that goodness is present, whereas beauty is also in immovable things), they are in error who assert that the mathematical sciences tell us nothing about beauty or goodness; for they describe and manifest these qualities in the highest degree, since it does not follow, because they manifest the effects and principles of beauty and goodness without naming them, that they do not treat of these qualities. The main species of beauty are orderly arrangement, proportion, and definiteness; and these are especially manifested by the mathematical sciences. And inasmuch as it is evident that these (I mean, e.g., orderly arrangement and definiteness) are causes of many things, obviously they must also to some extent treat of the cause in this sense, i.e., the cause in the sense of the Beautiful.

—Aristotle, *Metaphysics*, XIII.iii

Quotation 0.29

The truly wise ask what the thing is in itself and in relation to other things, and do not trouble themselves about the use of it,—in other words, about the way in which it may be applied to the necessities of existence and what is already known. This will soon be discovered by minds of a very different order—minds that feel the joy of living, and are keen, adroit, and practical.

—Goethe, *Maxims and Reflections* [p. 142]

In the first half of the nineteenth century mankind was first confronted with a real physical phenomena that did not fit into a world of gears and levers, or even into the world of gravitational attraction: the electromagnetic field. The enormous progress made in the second half of the century led to the completion, by the century's end, of the entire edifice of classical physics. As equal partners, mechanics and electrodynamics laid claim to having integrated all areas of physics. This included both optics and thermodynamics, which could be understood with the help of statistical mechanics. Besides Newton's name, the name of a new authority appears: James Clerk Maxwell (1831–1879).

By the end of the nineteenth century, there were only a few "recalcitrant" phenomena that were waiting to be reconciled with the classical framework. These included the speed of light in moving media, blackbody radiation, and radioactivity—phenomena that then gave impetus to the development in the twentieth century of the theories of relativity and quantum mechanics as well as all of atomic physics. These new theories altered, or at least put into a different light, the fundamental conceptions of classical physics—how we view space and time and even causality.

Yet this new revolution in physics differs fundamentally from that of the seventeenth century. Here there is no rubble to be taken away, nothing has been demolished, all things are just put in their proper place. The trust in Newtonian

	-600	+529	1543	1687	1900
Epoch	**Legacy from Antiquity**	**Protectors of the Legacy**	**Ending and New Beginning**	**The Development of Classical Physics**	**Physics of the Twentieth Century**
Relationship of science to religion	independent	"the hand-maiden of theology"	independent		
Relationship between reason and experience	THE HUMAN MIND IS CAPABLE OF UNDERSTANDING THE WORLD				
	including fundamental principles	with the help of faith	with the help of experience		
Important results	geometry, statics, descriptive astronomy	quality, quantity, impetus theory of projectile motion	unification of celestial and terrestrial laws, clarification of methodology	classical mechanics, electrodynamics	theory of relativity, quantum mechanics
Central personalities	ARISTOTLE	THOMAS AQUINAS	GALILEO, DESCARTES	NEWTON, MAXWELL	PLANCK, EINSTEIN, HEISENBERG
Character of natural laws	ETERNAL TRUTHS			APPROXIMATIONS	
	deterministic			statistical	

Table 0.1 Characteristics of the epochs, greatly simplified.

Significance of the (arbitrarily chosen) epoch boundaries:

- -600 Beginning of the teachings of the pre-Socratic philosophers.
- +529 The emperor Justinian closes the last philosophical schools in Athens.
- 1543 Publication of Copernicus's book *De revolutionibus*....
- 1687 Publication of Newton's *Principia*.
- 1900 Birth of Max Planck's quantum theory.

mechanics has become, if such is possible, greater precisely because of the greater understanding of the limits of its validity. It has lost its status of "absolute truth," but such a title was not accorded to the newer theories, either. From this point on, apart from problems that are still open, physics provides us with more and more precise laws that have limited validity but will continue to survive precisely because of their limited validity.

In its details, science is becoming ever more complex, yet in its fundamentals it is becoming simpler. The antique and medieval final cause (*causa finalis*), the striving toward perfect harmony, were replaced by causality; later this faded into a functional connection in space-time. Today, it is probabilistic processes operating under the restrictions set by conservation laws that are seen as fundamental.

In Table 0.1 we have collected the principal characteristics of the individual epochs.

Plate I shows the stage on which the history of physics has played. Mere words could not represent more convincingly how human events are interwoven with the unity of the cosmos.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

CHAPTER 1

THE CLASSICAL HERITAGE

To the illustrious athletes who had won the Olympic, Pythian, Isthmian, and Nemean Games, the forebears of the Greeks awarded such great honor that not only are they given palms, garlands and praises as they stand before the assembled public, but also, when they return home victorious to their cities, they are carried in triumph in four-horse chariots through their city walls to their homes, and at public expense they enjoy the rest of their lives on a pension. Now when I observe this, I am amazed that the same honors—or honors greater still—are not bestowed on writers, who provide every nation with endless utility for everlasting ages. For this would have been a much more worthy institution to have set up, because athletes make their own bodies stronger by exercising, whereas writers strengthen not only their own wits, but indeed everyone's, by preparing books for learning and the sharpening of the minds.

What good does it do humanity that Milo of Croton was undefeated, or the others who were champions of this kind, other than that, so long as they were alive, they held distinction among their own fellow citizens? The valuable precepts of Pythagoras, on the other hand, of Democritus, Plato, Aristotle, and the other sages, cultivated by daily industry, not only produce ever fresh and flourishing fruit for their own fellow citizens, but indeed for all the nations. And those who from an early age enjoy an abundance of learning develop the best judgment, and in their cities they have established civilized customs, equal justice, and those laws without which no community can exist safely.

Since so many private and public gifts have been prepared for humanity by the wisdom of writers, I conclude that more than palms and garlands should be awarded them—indeed triumphs should be declared for them and to them it ought to be decided to dedicate thrones among the gods.

—Vitruvius, *The Ten Books on Architecture*

DOI: 10.1201/9781032697697-2
This chapter has been made available under a CC-BY-NC-ND license.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

The Classical Heritage

1.1 What the Greeks Inherited

At the time the Greeks were beginning to lay the foundations of the natural sciences in the modern sense—meaning that they tried to understand nature on the basis of universal laws, with the exclusion of supernatural powers and requiring provability and, through that, certainty—they had at their disposal a body of knowledge that had been gathered by mankind over the course of millennia, more or less systematized, and in some details already astoundingly advanced. Aeschylus provides—though in a sense more metaphorical than literal—an inventory of the Greeks' starting intellectual capital, with a reference to the divine origin of knowledge just to underscore its value (Quotation 1.1).

The Greeks were well aware of the true source of their knowledge, even though from today's perspective we can see that they have accorded a disproportionate importance to Egypt at the expense of Mesopotamia. For example, according to Herodotus, geometry and astronomy were founded by the Egyptians. The Greek philosophers, when they were given to boasting— as they had very high self esteem and at times were drunk with pride over their own achievements—compared themselves to the Egyptian priests or to the "rope-stretchers" (as surveyors were called).

In this section, we will investigate in more detail the inheritance that the Greeks had at their disposal and on which they were able to build.

1.1.1 *The Beginnings of Science*

Knowledge is classified through the use of abstract concepts: this definition is not enough to pinpoint the beginnings of science (Quotation 1.2). Language itself already presupposes general concepts, and in this in turn is already found the two significant elements of science, namely abstraction and classification. The abstract concept of *bison* already goes beyond the concept of one particular concrete bison. The division of a kill or the organization of work unconsciously gives rise in men to groupings of different things and then the relating of these groups too each other. In modern terms we would say that sets were formed, relationships were established among the elements of these sets, and then the correspondences between the sets were developed. The decisive step toward science was when different sets (persons, animals, plants) were mapped onto one set, for example, a series of notches; this is how man arrived at the idea of numbers (Figure 1.1).

A similarly decisive step was the abstraction of the concept of line from the contours of objects, from the tracks in sand or the string of a bow. Geometry appeared, both geometry as an end in itself and as the geometry of decoration, which at some later point inspired a very modern branch of geometry, namely the study of symmetry laws that are of such great importance today in physics (Figure 1.2).

If one examines the drawings or statuettes made by early man, one is confronted with the question of the level of scientific achievement that must be postulated for such accomplishments. For example, a two-dimensional representation of three-dimensional objects requires a very high level of abstraction, and here the abstraction is joined with very sharp powers of observation, which is also a prerequisite for scientific investigation (Figure 1.3).

Quotation 1.1

... I made them ...
... intelligent and possessed of minds.
...
At first they looked about them, but looked in vain;
hearing they did not hear, but like mere shapes
of dreams they led their long lives randomly
... They did not know how to build
houses of brick against the sun, nor carpentry,
but they lived underground like scurrying ants
in the dark and sunless recesses of caves.
They had no sure sign, either, of winter
nor of flower-fragrant spring, nor of fruitful
summer, but they carried on entirely
without rational thought, until I showed them
the stars' risings and settings, difficult to discern.
And more: number, that mental feat par excellence,
I discovered and gave them, and the combining of
letters,
memory's helper, hardworking mother of the arts.

—Aeschylus, *Prometheus Bound*, fourth century BCE [p. 44]

Quotation 1.2

The deeper we delve and the farther we press and grope into the underworld of the past, the more totally unfathomable become those first foundations of humankind, of its history and civilization, for again and again they retreat farther into the bottomless depths, no matter to what extravagant lengths we may unreel our temporal plumb line.

—Thomas Mann, *Joseph and His Brothers* [p. 3]



◀ **Figure 1.1** The grouping of notches proves that the concept of number must already have existed (one group of 25 notches, another group of 30 on a wolf's shinbone). (Dolní Véstonice, Czech Republic. Kolman: *Istorija matematiki v drevnosti*.)

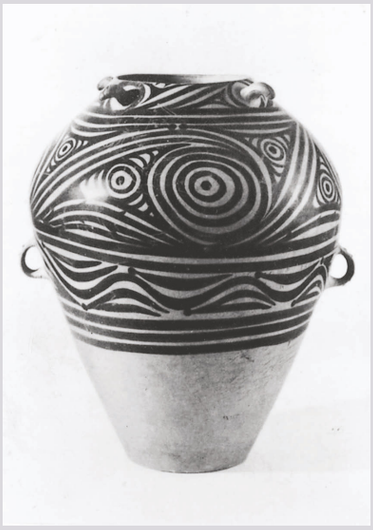


Figure 1.2
A sense for the symmetry of geometric figures developed early. The figure shows a vessel from the Neolithic Yangshao culture, about 2000 BCE. (*Products of Ancient Chinese Science and Technology*, Beijing 1978.)



Figure 1.3
Beyond its artistic merits, this figure shows two elements of science: abstraction and realistic observation. Part of a bone engraving from Lortet, Pyrenees (after BREUIL).

Quotation 1.3

... Of every tree of the garden thou mayest freely eat: but of the tree of knowledge of good and evil, thou shalt not eat of it: for in the day that thou eatest thereof thou shalt surely die. ...

And when the woman saw that the tree *was* good for food, and that it *was* pleasant to the eyes, and a tree to be desired to make *one* wise, she took of the fruit thereof, and did eat, and gave also unto her husband with her; and he did eat. ...

And the Lord God said, Behold, the man is become as one of us, to know good and evil: and now, lest he put but forth his hand, and take also of the tree of life, and eat, and live for ever: Therefore the Lord God sent him forth from the Garden of Eden, to till the ground from whence he was taken.

—Genesis 2:15, 3:6, 3:22

In the bodies of the drawn animals, we frequently see an equally abstract drawn arrow. This indicates that early man already understood the parallel between the general image in his mind and reality, not only in the direction that we expect today—abstracting from reality to arrive at the general—but in the opposite direction also: the imagined stories that are expressed in an abstract form have an effect on reality itself. The killing of the abstract bison with an abstract arrow should result, or at least assist, in the death of the real bison by a real arrow. Primitive man still saw the world as a unified whole and took the mystical direct connection between abstract representation and reality for granted. If we wish, we might say that this is the starting point of the long debate over the relationship between the general concept and the concrete object, or between idea and reality: from PARMENIDES, through the medieval debates between the nominalists and realists, all the way to HEGEL.

Apart from speculation over early man's notions of existence and consciousness, there has also been research with the tools of modern science that might point directly to the existence of scientific knowledge in primitive society. But at times such investigations appear quite fanciful. We see in Figure 1.4, for example, an attempt to interpret a series of notches on a bone, long viewed as mere decoration, as the representation of the periodic changes in the phases of the Moon. The demonstration is not particularly convincing, yet the method might bring any number of surprises in the future.

From the oldest myths that have come down to us in written form, we may infer that since earliest times, mankind has held knowledge to be a great treasure whose origin could only be divine. The appreciation of knowledge is not only proved by the belief in its divine origin but also by the fact that, according to the myths, the gods jealously guarded it. For mankind to share in this knowledge, either a god or a person had to rebel against the divine order. Punishment was certain, whether that of PROMETHEUS or of the first man and woman in Eden after the Fall (Quotation 1.3). It is perhaps the gods of Mesopotamia who showed the greatest understanding: They understood that the gods needed man, to honor them, worship them, and please them with sacrifices. And so the gods taught mankind all that was needed to render them proper homage (Quotation 1.4).

Somewhere around the fifth millennium BCE, major civilizations began to develop in a number of river valleys. Along the Nile, the Tigris and Euphrates, the Indus, and the Yellow River, there arose, almost simultaneously, civilizations with a highly developed urban life, local administration, and social stratification (Figures 1.5 and 1.6). Among these, only the Egyptian and Mesopotamian civilizations had a direct influence on Greek culture and, through it, on that of Europe. Therefore, in what follows, we shall focus our attention on those two cultures, while emphasizing here that similar accomplishments were achieved along the Indus and the Yellow River.

1.1.2 *Egypt and Mesopotamia*

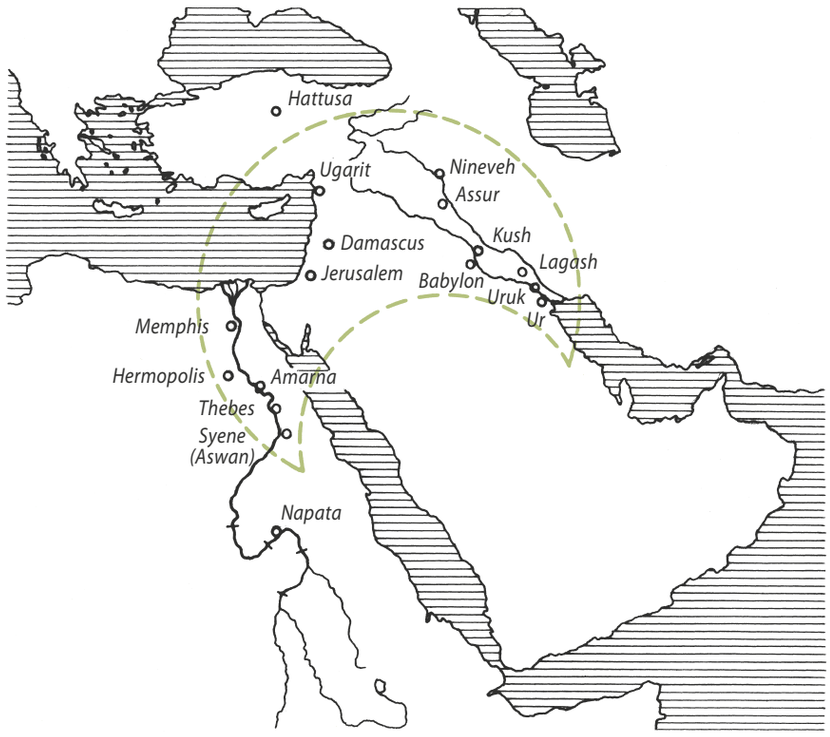
The histories of cultures that arose along riverbanks show remarkable similarities. The rivers offered to those living along their banks identical opportunities for fishing and the irrigation of fields and identical difficulties as well. At particular times of the year, it was necessary to defend against devastating floods; at others, against drought. All this required a certain level of technological development as well as organization for the division of labor, and hence a high degree of social organization.

Because these problems were recurrent, repeating themselves from one year to the next and from generation to generation, both the technologies and organizational structures that developed were prone to ossification. Of course, this also made for a certain degree of strength and stability. This is what makes it possible to understand the survival of Egypt for nearly four thousand years and also explains why the level of Egyptian culture plateaued relatively early.

Aside from their similarities, the cultures of Egypt and Mesopotamia exhibit significant differences, in certain respects resulting from their differing geographical circumstances. Egypt comprises a closed unit. Its borders are mostly unpopulated deserts. From there came neither enriching influences nor external threats. It is open only to the south, where there is but one, militarily easily defended gateway to the civilized world: the Sinai Peninsula. It is therefore easy to appreciate that, except for the invasion by the Hyksos, Egypt never came under foreign domination.



▲ **Figure 1.5:** Fluvial valley cultures, about 2000 BCE.



▲ **Figure 1.6** Egypt and Mesopotamia (the Fertile Crescent).

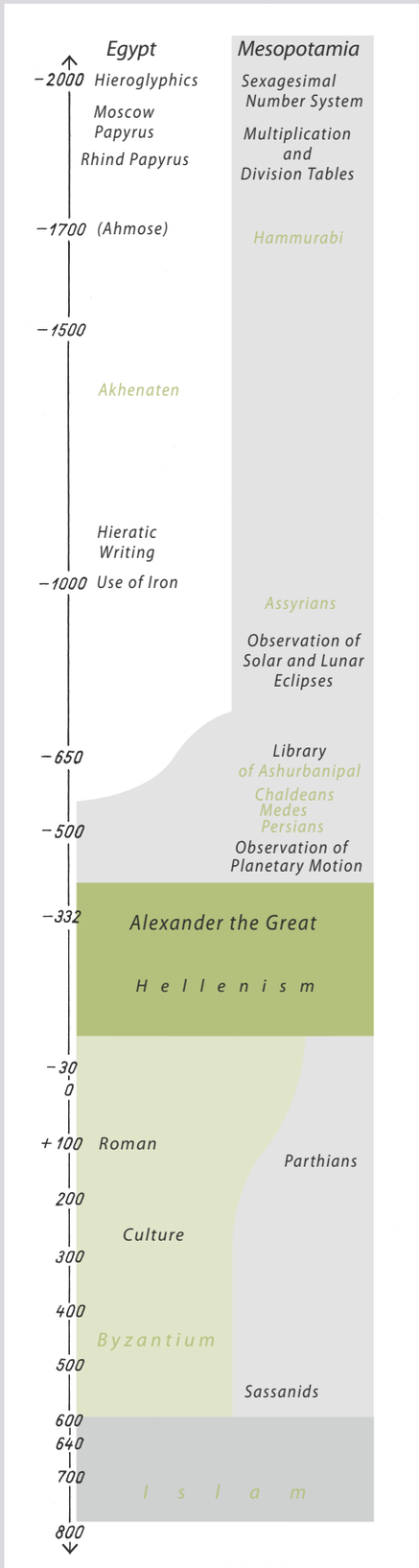
Quotation 1.4

... let us create mankind.
The service of the gods be their portion,
for all times,
to maintain the boundary ditch,
to place the hoe and the basket
into their hands
for the dwelling of the great gods,
which is fit to be an exalted sanctuary,
to mark off field from field
for all times
to maintain the boundary ditch,
to give the trench (its) right course,
to maintain the boundary stones,
...
to make the field of the Anunnaki produce plentifully,
to increase the abundance in the land,
to celebrate the festivals of the gods,
to pour out cold water [on the altar]
in the great house of the gods, which is fit to be an exalted sanctuary.

—Babylonian cuneiform tablet [Heidel, 1963, pp. 69–70]



▲ **Figure 1.4** The first astronomical observation (?): Alexander Marshack interprets the signs on this nearly 30,000-year-old bone as representations by Cro-Magnon man of the phases of the Moon [1972].



▲ **Figure 1.7** Important events, from our point of view, in the history of Egypt and Mesopotamia.

The situation in Mesopotamia was quite different. In the land around the valley of the Tigris and the Euphrates, a number of energetic nomadic tribes lived around the city-states that had achieved a higher level of cultural development. The tribes tried to trade with the cities peacefully, or they plundered or even conquered them. Therefore, the history of this "land between the rivers" is more eventful and colorful. Time and again, new conquerors snatched the legacy of their predecessors and would enrich it with their own distinctive imprint. Thus we can speak here also of the continuity of a unified culture.

The raw materials available in the two areas were also different. In Egypt, stone was used for building, and papyrus for writing. In contrast, Mesopotamia is poor in wood and naturally occurring stone, and therefore baked clay bricks were used for building and their method for writing also used clay in the form of carved tablets. The Egyptian monuments have withstood the ravages of time better than those erected at the same time in Mesopotamia, which can be readily understood from the difference in building materials. Conversely, far more examples of writing have survived on clay tablets than on papyrus scrolls. They fill entire libraries of several thousand volumes with reports of the daily life, the science, and the entire culture of the Mesopotamian realm, in contrast with the relative dearth of papyrus scrolls. To be sure, the surviving inscriptions on Egyptian tombstones and temples provide valuable information on that civilization's culture, but their purposes—memorials for glorious military victories, or the listing of the good deeds of the deceased for judgment in the netherworld—inherently excluded any systematic discussion of scientific problems.

We must mention one additional circumstance that determined the course of life in Egypt and affected Egypt's scientific advancement. The Nile overflowed its banks with the regularity of a calendar, year after year depositing nutrient-rich silt from inner Africa on the fields. The layering of silt rendered the boundary markers of estates unrecognizable, so that the fields had to be surveyed anew every year. It is for this reason that the surveyors, known to the Greeks as rope-stretchers, acquired such significance. It was also of great importance to be able to predict the start of the flooding season. Thus, in Egypt, astronomical observations served not only to fix the time of religious festivals but also to aid in economic planning.

Perhaps because the two cultures were in contact, and in any case because identical needs under similar conditions give rise to similar solutions, both areas became well acquainted with copper in the fourth millennium BCE, with bronze a millennium later and iron a millennium after that. At the same time, the plow had become a tool for agriculture; for the transport of goods, in addition to beasts of burden, the wheeled cart was employed. Sailing ships were built, and craftsmen, such as potters and metalworkers, had begun to form their own social stratum. The organization of production furthered the emergence of slavery as a social system in the fourth millennium BCE. Continually expanding cities arose centered on crafts and trade, with temples for the priestly class and palaces for the aristocracy.

Figure 1.7 shows the most important historical and cultural events in Egypt and Mesopotamia lined up side by side. The timeline includes information of interest from the point of view of the history of science: for example, the fact that writing was already in use well before 2000 BCE. We can see when the papyri were created to which we owe our knowledge of ancient Egyptian mathematics and geometry. As for Mesopotamian science, the information came from, among other sources, the library of ASHURBANIPAL in Nineveh. Since the center of Hellenic culture emerged in Alexandria, in the land of the ancient Egyptian civilization, it is worth-

while taking a closer look at the last epoch in the history of ancient Egypt, specifically at the political transformations that led up to the conquest of the area by the Arabic world in about the year 640 CE.

After a continuous history of some nearly 4000 years, Egypt was conquered in the sixth century BCE, first by the Chaldeans, relatives of the Assyrians, and then by the Persians. After the collapse of the Persian Empire in 332, Egypt became part of the empire of ALEXANDER THE GREAT. Upon ALEXANDER's death, PTOLEMY, one of his generals, "inherited" the empire, becoming the first ruler of Egypt's Ptolemaic dynasty. Under that dynasty, the city of Alexandria, founded by ALEXANDER in the third century BCE, flourished. In 30 BCE, Egypt became a Roman province. After the division of the Roman Empire, it belonged to the Eastern Roman Empire until its conquest by the Arabs.

We mentioned above that our knowledge of Egyptian science rests on surviving papyri; in particular, two larger ones and several fragments. One of these is the Rhind papyrus, named for the Scottish archeologist HENRY RHIND, who discovered it in 1858 [Peet, 1923], and the other is the Moscow papyrus [Struve, 1930]. The Rhind papyrus is also known as the AHMES papyrus, as the author identifies himself and at the same time promises that he will reveal to the reader all existing secret knowledge. The Rhind papyrus was written around 1700 BCE, though the knowledge contained in it is quite probably much older.

We now present some of the characteristics of ancient Egyptian mathematics and some of the outstanding results that the Egyptians obtained [van der Waerden, 1954; Neugebauer, 1931].

The Egyptians used a decimal system in which special symbols correspond to ones, tens, hundreds, and so on (Figures 1.8 and 1.9). The number 7, for example, was represented with seven "one" symbols, placed next to one another, above and below, or in any other regular or irregular arrangement. Analogously, the number 60, for example, could be written using six "ten" symbols: inverted U symbols arranged alongside or below one another (Figure 1.10).

A very important role was played in ancient Egyptian mathematics by the reciprocal of a whole number (Figures 1.11 and 1.12). Special tables were used for representing arbitrary fractions as sums of such reciprocals. For example, the following decompositions were known:

$$\frac{2}{5}=\frac{1}{3}+\frac{1}{15},$$

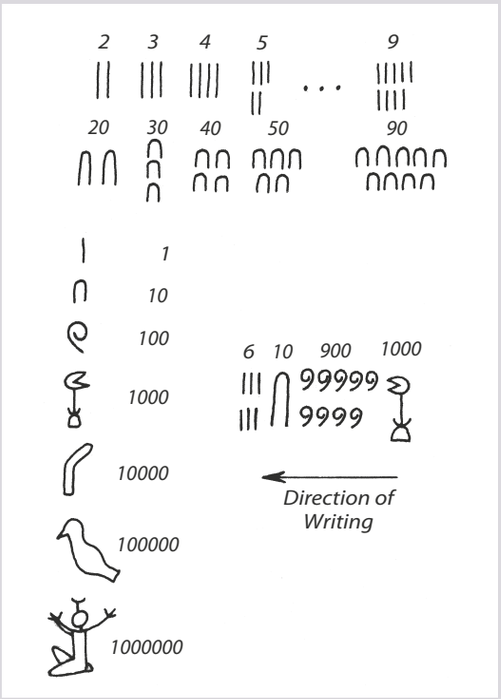
$$\frac{2}{43}=\frac{1}{42}+\frac{1}{86}+\frac{1}{129}+\frac{1}{301}.$$

A characteristic problem presented in the RHIND papyrus and its solution are as follows: A number of objects increased by one-seventh of that number yields 19 objects. What is the number? Today, such a problem is easily solved by using the techniques of basic high school algebra:

$$x+\frac{x}{7}=19,$$

from which we get

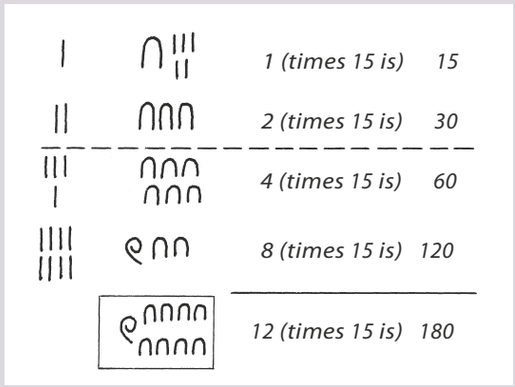
$$x=\frac{7\cdot 19}{8}.$$



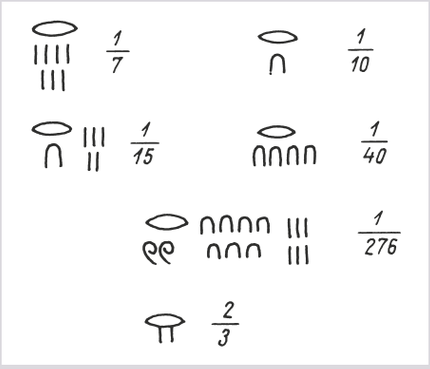
▲ **Figure 1.8** Hieroglyphic representation of numbers. The large units of a number were written first, then the small units. Because the writing was from right to left, on papyri we find the large units on the right. In books published today, the customary left-to-right notation is often used, but the symbols for the numbers are always written in the direction opposite to that of the writing.

▼ **Figure 1.9** Numbers from the yearbook of THUTMOSE III [Taton, 1957]. (Louvre.)

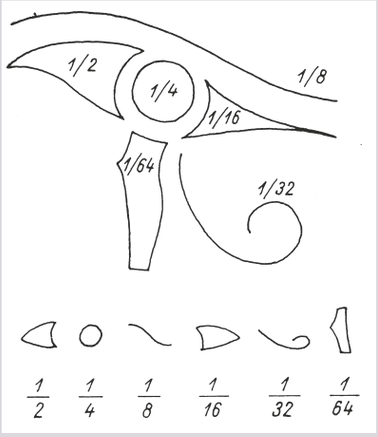




▲ **Figure 1.10** The figure shows how the Egyptians carried out multiplication. To solve the problem 12 × 15, the latter number is doubled step by step. That is, it is multiplied in sequence by 1, 2, 4, 8, … = 1, 2, 2^2, 2^3, …, from which the steps for 4 and 8 (adding up to 12) are selected for summation. The Egyptians thus operated with a sort of binary system, such as is used in today's computers. Division was reduced to this form of multiplication: the denominator was multiplied sequentially by powers of 2 (that is, sequentially doubled), and then it was checked how the numerator could be represented as a sum of the numbers thereby obtained. For example, let us attempt the problem 45 ÷ 5. Sequentially doubling the denominator gives the sequence $1 \times 5 = 5$, $2 \times 5 = 10$, $4 \times 5 = 20$, $8 \times 5 = 40$. We see that $1 \times 5 + 8 \times 5 = 45$, and so the result is $1 + 8 = 9$.



▲ **Figure 1.11** The representation of fractions: By writing the symbol for "part" above a number *n*, we get the fraction of the form 1/*n*. Among the fractions not of this form, only 2/3 had its own symbol.



◀ **Figure 1.12** An old notation for fractions that was used to express quantities of grain. The symbols represent parts of the eye of the falcon god Horus, which was torn into pieces by the evil god Seth.

Egyptian mathematicians solved such problems in the following way: In place of the unknown we take an arbitrary number as a trial solution, carrying out the operations on this number as prescribed in the problem. For example, let us assume that the solution—that is, the number of objects—is 7. If we take the seventh part of this number, we obtain 1. From this, it follows that the number of objects increased by one-seventh is 8, since $7 + 1 = 8$. However, from the statement of the problem, the result should be 19, and we see that the trial solution is incorrect. Ahmes now instructs his students, correctly, as follows: We must now allow the number taken by us as a provisional solution to grow in relation to the amount by which the required value 19 differs from 8. The Egyptian scribe thereby arrives at the same solution that we obtained with algebraic methods.

It is worth noting that this procedure is used even today, above all in electrical engineering, for the calculation of linear networks. For example, suppose the current in a particular branch of a complex network is sought for a given source voltage. To solve the problem, it is frequently easiest to assume that the current is known and from the assumed value of the current to calculate the other currents and potentials, including the source voltage. We obtain the correct value of the desired current by changing all the values—or in fact the entire diagram if we are using a graphical procedure—in proportion to the relationship between the calculated and given source voltages.

The most significant result of Egyptian mathematics is to be found in the Moscow papyrus. It deals with the calculation of the volume of the frustum of a pyramid with square upper and lower faces. Of course, in the papyrus concrete numerical values are used, but from the calculation one can clearly see that the formula being used is

$$V = \frac{h}{3}\left(a^2 + ab + b^2\right),$$

which is entirely correct! This formula is today part and parcel of high school mathematics, although its derivation is too complex to be presented there. There is no indication as to how the Egyptians arrived at such a high level of geometric knowledge. It is certain that they could not have acquired this formula from the Babylonians, whose mathematical sophistication was otherwise at a higher level, because Mesopotamians were using an incorrect formula for this volume, even though the correct formula would surely have been useful in the construction of their dams.

Figure 1.13 shows parts of this papyrus together with a present-day interpretation.

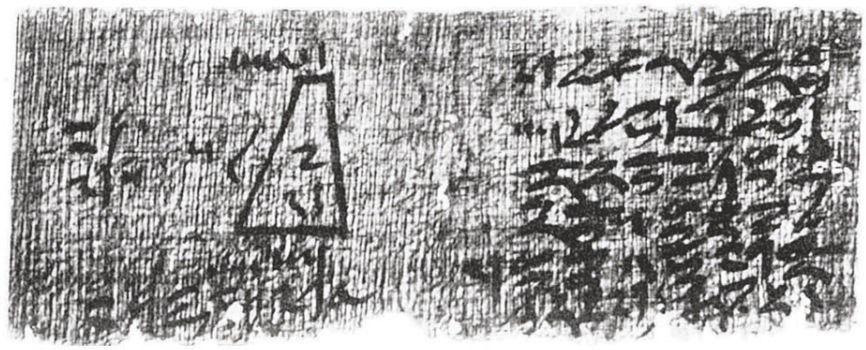
The highly precise north–south orientation of the great pyramids also represents an outstanding achievement in relating astronomy with geometry. In the case of the pyramid of Khafra, the deviation is a mere tenth of one degree!

From the mathematical problems that they posed, we can determine that the Egyptians used for the ratio of the circumference of a circle to its diameter, that is, for the number p, the value

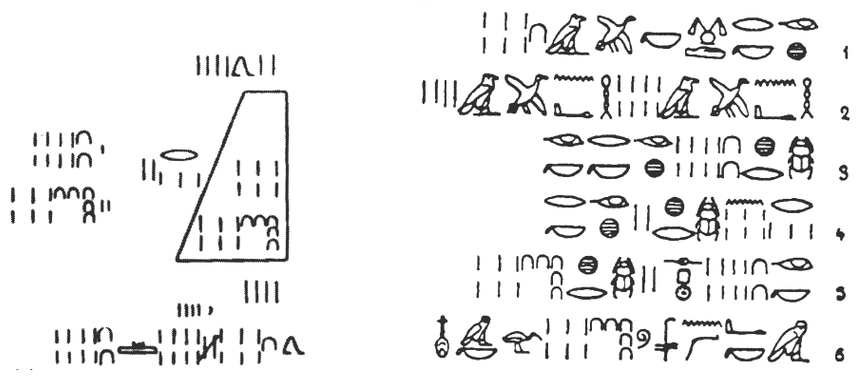
$$\pi = 4\left(\frac{8}{9}\right)^2 = 3.1605$$

(the correct value being $p = 3.14159\ldots$).

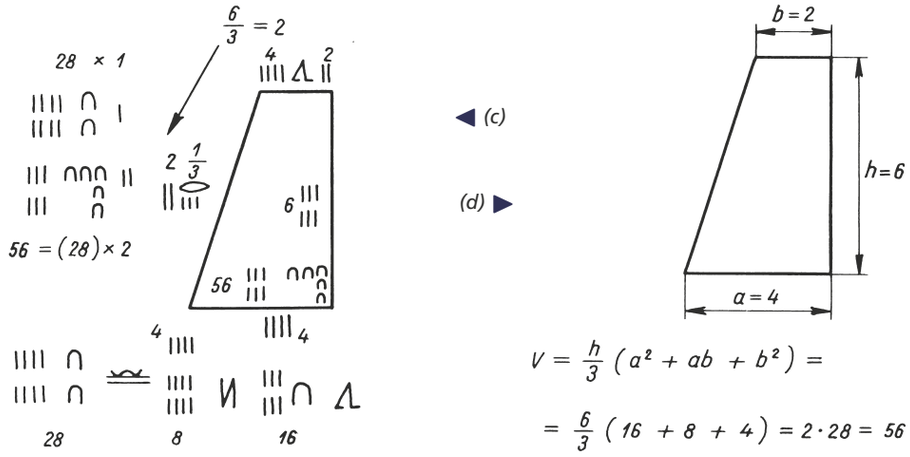
In many respects, Mesopotamian mathematics attained a higher level of sophistication than that of the Egyptians. For example, the Babylonians knew of a procedure for finding Pythagorean triples, that is, triples of integers satisfying the relationship $a^2 + b^2 = c^2$ and therefore representing the possible legs and the hypotenuse of a right triangle. The Egyptians, in contrast, knew only the triple (3, 4, 5) (Figure 1.14). In modern notation, the Babylonians derived integer Pythagorean



▲ (a)



▲ (b)



▲ **Figure 1.13** Determining the volume of the frustum of a pyramid (after [van der Waerden, 1954] with additions). **(a)** The original papyrus. **(b)** Transcription into hieroglyphic symbols by Yuri Yakovlevich Perepelkin [Peet, 1923]. **(c)** Interpretation of the symbols. **(d)** Solution by today's method.

triples from the relations

$$x = p^2 - q^2, \quad y = 2pq, \quad z = p^2 + q^2 \qquad (1)$$

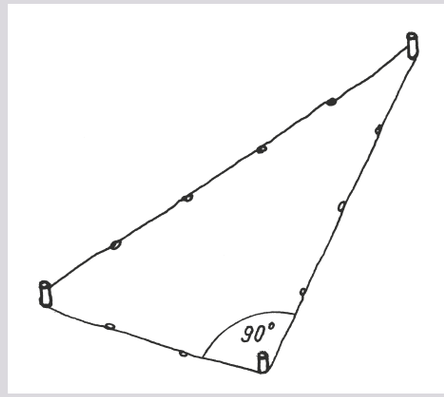
(p and q integers, $p > q$). For example, if we set $p = 2$ and $q = 1$, then we obtain $x = 3$, $y = 4$, $z = 5$. With $p = 2$ and $q = 2$, in turn, we obtain the numbers (5, 12, 13) respectively, which are connected by the Pythagorean relation

$$5^2 + 12^2 = 13^2.$$

One can prove that all numbers x, y, z found with the help of equation (1) satisfy the relation

$$x^2 + y^2 = z^2$$

by simple substitution, after which the left and right sides appear identical. The



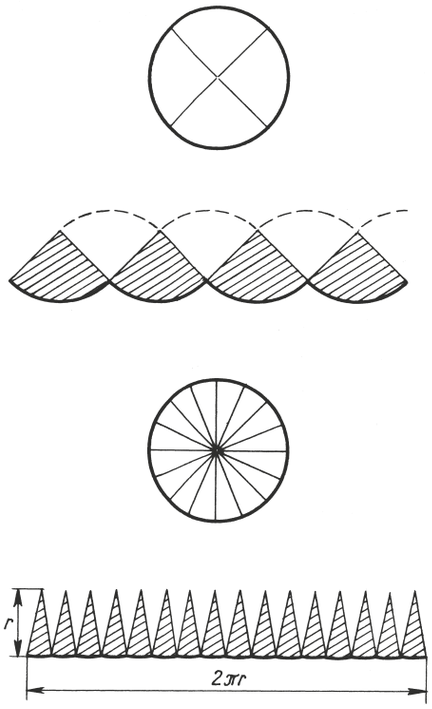
▲ **Figure 1.14** According to Cantor's opinion at the end of the nineteenth century, the Egyptian rope-stretchers might have constructed right triangles in the following way: A rope of length 5 + 4 + 3 units was used to span a triangle with side lengths 3, 4, and 5. The angle opposite the longest side would have been a right angle. It is unclear from primary sources whether such a technique was in fact used.

Quotation 1.5

And he made a molten sea, ten cubits from the one brim to the other: *it was* round all about, and its height *was* five cubits: and a line of thirty cubits did compass it round about. ...

It stood upon twelve oxen, three looking toward the north, and three looking toward the west, and three looking toward the south, and three looking toward the east: and the sea *was set* above upon them, and all their hinder parts *were* inward.

—1 Kings 7:23–26



▲ **Figure 1.15** Perhaps this is how the formula for the area of a circle was derived in antiquity.

Babylonians knew the formula for calculating the square of a sum,

$$(a+b)^2 = a^2 + 2ab + b^2,$$

and the relation

$$(a+b)(a-b) = a^2 - b^2$$

as well. They were able to solve the quadratic equation in one unknown,

$$x^2 + ax = b,$$

but they also worked, among others, with systems of equations in two unknowns of the form

$$x + y = a, \quad xy = b.$$

Interestingly, for the ratio of the circumference of a circle to its diameter, the rather crude value $\pi = 3$ was used. According to a document discovered in 1950, the value $\pi = 3 + 1/8 = 3.125$ was also considered. However, in geometric calculations, the circumference of a circle was calculated as six times the radius, and the area as three times the square of the radius, which implies the value $p = 3$.

It is also interesting to note that the Jews, who had come into contact with the Mesopotamian culture, also used $\pi = 3$ in the Old Testament (Quotation 1.5).

To calculate the area of the circle, they apparently multiplied the circumference by the radius and took one half of the result. To us today, this result seems obvious. The Mesopotamians arrived at it perhaps in the following way: Decompose the circle into sectors with smaller and smaller central angle and arrange them in a row, as shown in Figure 1.15. It is then easy to see that the sectors fill up half the area of a rectangle whose length is the circumference and whose height is the radius of the circle. We note here that in the diaries of Leonardo da Vinci, who was self-taught in mathematics, as in most everything else, there is a drawing illustrating a similar line of argument.

As we mentioned earlier, for the volumes of the frustum of a pyramid and the frustum a cone, the Babylonians used the incorrect relationships

$$V = \frac{h}{2}\left(a^2 + b^2\right)$$

and

$$V = \frac{h}{2}\left(3R^2 + 3r^2\right).$$

It is clear where these formulas came from. They must have begun with the very plausible, but for the given circumstances incorrect, assumption that the volume of the frustum would be the average of the volumes of two columns of the same height as the frustum: one having a cross section the same as the base of the frustum and one as the top of the frustum.

The most significant contributions of the Babylonians to mathematics that continue to be used today are the sexagesimal (base-60) number system and the positional notation for numbers [Neugebauer, 1935].

As seen in Figure 1.16, like the Egyptians, the Babylonians represented the integers 1 through 9 by the corresponding number of notches that were imprinted into soft clay tablets as wedge-shaped indentations. A special symbol was used for the number 10. The numbers 20, 30, 40, and 50 were represented by the requisite number of symbols for 10 arranged in a row. For 60, however, the symbol for 1 was again used. Thus the initial symbol in a number is to be multiplied by 60, just as today the first digit in a two-digit number is to be multiplied by 10, the result then being added to the second digit. If we consider a three-digit number in our decimal system, such as 346, then the notation means that 3 is to be multiplied by 100, that is, by the second power of the base 10 of our number system, and 4 is to be multiplied by the first power of 10, that is, by 10 itself, and finally, these two results are to be added together along with the rightmost digit, 6:

$$346 = 3 \cdot 10^2 + 4 \cdot 10 + 6.$$

In a sexagesimal system, this number is expressed using only two digits, since we have

$$346 = 5 \cdot 60 + 46,$$

and thus the Babylonians wrote this number as



The introduction of positional notation had one serious difficulty. Namely, if we want to write the number 60, then its symbol is the same as the symbol for 1. That we would have to multiply the 1 by 60 is usually signaled by the number having two digits. But what if we want to write just 60? Simply stated, the Babylonians lacked the symbol that was later introduced to signify an empty space, then became generally used, and finally was recognized as a full-fledged member of the family of numbers: the *zero.* In Babylonian texts, one frequently has to infer the place value of a given digit from its context.

The image reproduced in Figure 1.17 may be considered to be the highest achievement of Babylonian mathematics. We recognize a square, on whose diagonals are drawn the symbols



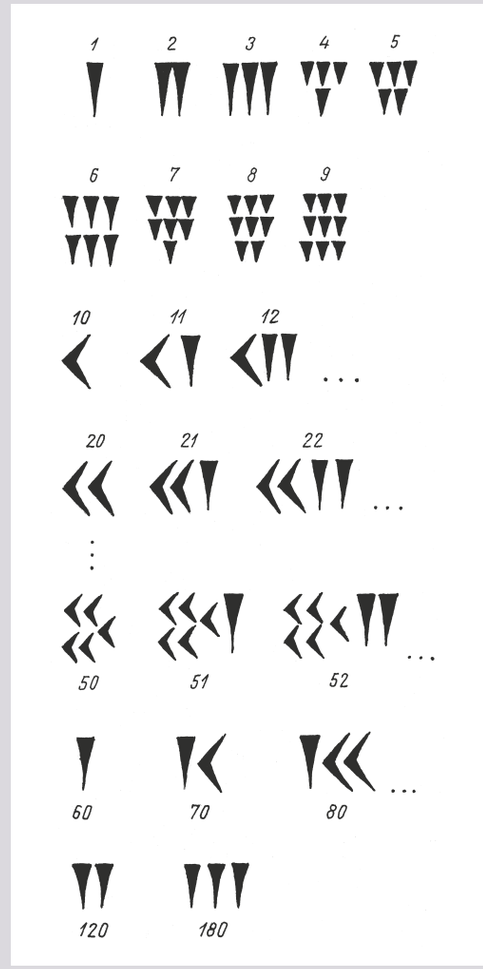
which in contemporary notation correspond to the sequence of digits

1, 24, 51, 10.

If we interpret these digits in the sexagesimal system and assign appropriate place values according to their positions, then we obtain

$$1 \cdot 60^0 + 24 \cdot 60^{-1} + 51 \cdot 60^{-2} + 10 \cdot 60^{-3} = 1.4142,$$

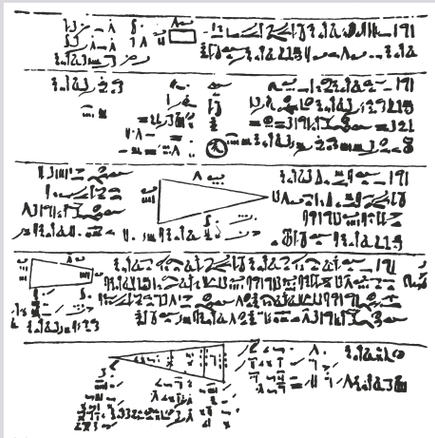
that is, the length of the diagonal of a square with side length 1, or the value of the square root of 2, with accuracy to the fourth decimal place. Here the question naturally arises how the Babylonians could determine the root of a number that was not a perfect square. They probably used an approximation procedure, which,



▲ **Figure 1.16** The Babylonian symbols for numbers. In the sexagesimal system the symbol for 1 could have also indicated 60 or $60^2 = 3600$, and so on, or even the fractions $60^{-1} = 1/60$ or $60^{-2} = 1/3600$, and so on.



▲ **Figure 1.17** An outstanding achievement of Babylonian mathematics. The length of the diagonal of a square of side length 1 (that is $\sqrt{2}$) is given to a precision equivalent to four decimal places.



(a)



(b)

▲ **Figure 1.18**

(a) Facsimile of the Rhind papyrus, produced by an Egyptian firm. The manufacturer claims that it was made with the ancient methods from the papyrus plant.

(b) A contemporary reconstruction of a Babylonian multiplication table. A wooden stylus with a tip worked into a V-shape was used to carve cuneiform symbols into wet clay. The table shows symbols representing the following:

7 a ra	1	7
a ra	2	14
a ra	3	21
a ra	19	133 (= 2 × 60 + 13)
a ra	20	140 (= 2 × 60 + 20)
a ra	30	210 (= 3 × 60 + 30)
a ra	40	280 (= 4 × 60 + 40)

Thus, this is a table for multiplication by the constant factor 7. The symbol "a ra" is equivalent to our multiplication sign ×.

for simplicity, might be represented in today's notation as follows: If we denote the value to be calculated by a, then the equation

$$a \cdot a = 2$$

must be solved. This can be transformed into

$$a = \frac{2}{a},$$

and substituting this into an obvious identity, we get

$$a = \frac{a+a}{2} = \frac{1}{2}\left(a + \frac{2}{a}\right). \qquad (2)$$

Let's assume an approximation for a, say, $a = 1.5$. This value is larger than the value sought, since its square, 2.25, is larger than the prescribed number 2. However, the second summand in the parentheses of equation (2), that is, the number $2/a$, is smaller than the target number. If we now compute the quantity

$$\frac{1}{2}\left(a + \frac{2}{a}\right),$$

then the average of the two numbers, one too large and one too small, will be closer to the true value than the original number. This first approximation step, or iteration, already leads from the quite coarse approximation 1.5 to the value 1.415, not far from the correct 1.414…. If we repeat the above calculation with this new value, we obtain an even better approximation.

Figure 1.18 shows some amateur and professional reproductions of traditional mathematical texts. Figure 1.19 shows the circuitous route by which positional representation of numbers found its way to Europe. The path leads from Babylon to Alexandria, where PTOLEMY used the sexagesimal system and indicated a zero by an empty place. From Alexandria, the system found its way to India. There, the notational system was transformed to one of base 10, and zero was introduced and interpreted. Finally, the path of the notational system led via the Arabs to Europe, where we find its description in full for the first time in a work of FIBONACCI published in the year 1202.

In both Egypt and Mesopotamia, astronomical calculations were used for the preparation of calendars, which made possible the determination of religious festivals and the times for initiating certain agricultural practices. We have already mentioned the flooding of the banks of the Nile, which occurred immediately after the appearance of the star Sothis (Sirius, or the Dog Star) in the sky. The Egyptians divided the year into 12 months of 30 days each, plus five additional days. This way, however, their year was shorter than the real year whose length is very close to 365 and one-fourth days. This means that their calendar shifted by one-fourth of a day every year. The Egyptian Empire lasted long enough for such a shift to become noticeable; indeed, the seasons shifted through the entire calendrical year. Only after each period of $4 \times 365 = 1460$ years did they return to the same place. The Egyptians were aware of this cycle and called it the Sothis cycle, after the star. At the suggestion of the Egyptian astronomer SOSIGENES, JULIUS CAESAR improved the calendar by declaring a leap year every four years. The resulting Julian calendar was used in Europe until the sixteenth century (indeed, in England until 1752, and in Russia until 1918).

The development of astronomy in Mesopotamia progressed significantly further, which is due to the facts that the continuity of the ancient culture was not disrupted by either the Persian or the Greek conquests—in fact, many results were obtained during the latter period—and that astronomical knowledge was of great importance to astrology, the practice of which was associated with Chaldea in the Bible; it was also practiced by the Greeks, and in the Roman Empire as well. The Chaldeans determined the positions of the Sun and planets with great accuracy. To them can also be attributed the division of the ecliptic—the apparent path of the Sun through the heavens—into the twelve signs of the zodiac and the naming of those signs. The Chaldeans knew that 235 lunar months was equivalent to almost exactly 19 (tropical) years, a fact discovered again in ancient Greece and named the Metonic cycle for its Greek discoverer Meton. From the measurements of the Chaldeans, it is apparent that they had determined the length of the year to an accuracy of 4.5 minutes.

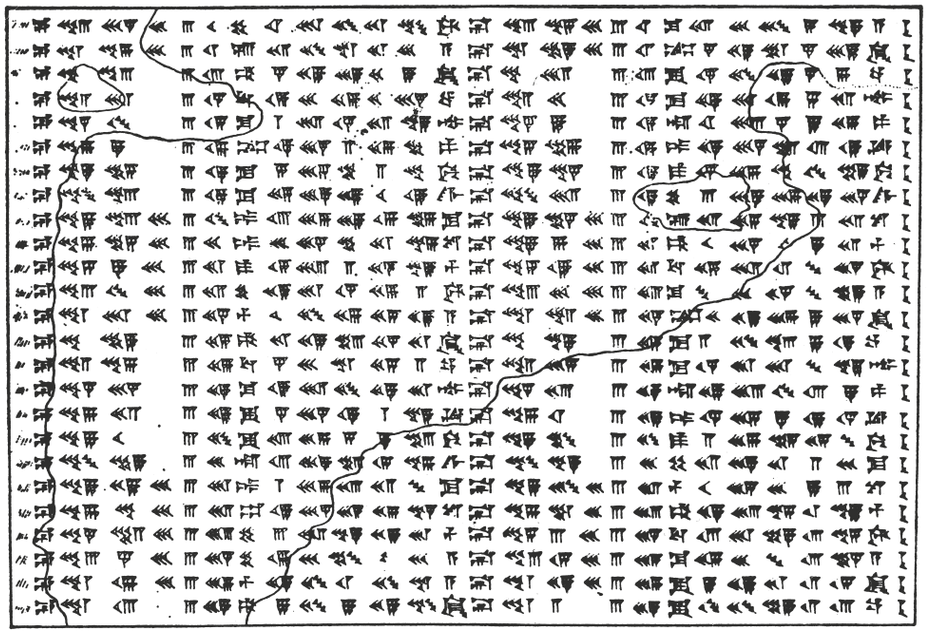
While we are speaking of the greatest achievements of the Mesopotamian peoples, we should mention the Chaldean astronomical tables. The Chaldeans, a Semitic people, were the founders of the Neo-Babylonian Empire. But even later, following the conquests of the Medes, Persians, Greeks, and Parthians, they continued to hold leading positions in intellectual occupations, above all in astronomy, which was essential to priestly tasks. The most valuable part of their observations that have come down to us stems from the Hellenistic and even later periods, yet without showing any sign of Greek influence. Indeed, it can be determined that the Chaldeans—in contrast to the Greeks, with their preference for geometry—had a decided preference for arithmetic methods in their search for and description of natural laws.

It is not surprising that these tables involve primarily the major heavenly bodies, that is, the Sun, the Moon, and the "wanderers" (the planets). The purpose of these tables was to show regularities and periodicities—on the basis of which it was possible to predict events or to determine when in the past an event occurred—and which were crucial in determining dates of religious festivals and in astrology, which required times of conjunctions, oppositions, and solar and lunar eclipses.

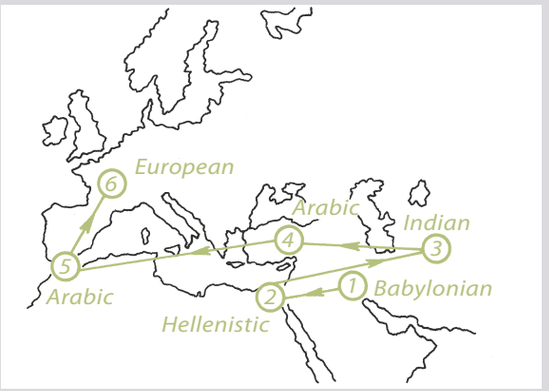
It is above all the lunar tables that fill us with awe at the meticulous recording of minute detail as well as awe at those who deciphered them.

We shall examine the Chaldeans' methods by looking at a relatively simple table, that dealing with the planet Jupiter. Figure 1.20 shows a copy of the original table, made by J. N. Strassmaier. Like all the other tables, this one is completely devoid of any explanatory remarks or legend.

The table gives the changes in the speed with which Jupiter crosses the various signs of the zodiac. The main observation of the table is that the successive positional data form arithmetic progressions whose



▲ **Figure 1.20** Table of successive second directional changes of the planet Jupiter (Marduk). Copy of a clay tablet, after Strassmaier.

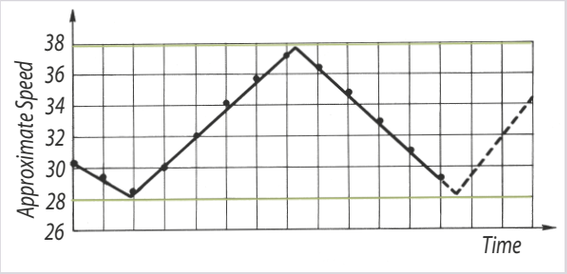


▲ **Figure 1.19** The likely path by which the system of positional notation for numbers traveled from Babylonia to Europe.

▼ **Figure 1.21** Interpretation according to Kugler of a portion of the Jupiter data.

A		B		C		D		E	F
3	10	XII	11	31	29	21	49	♋	ush
3	11	XIII	22	29	41	21	30	♌	ush
3	13	II	4	28	38	20	8	♍	ush
·	·	·		30	26	20	34	♎	ush
·	·	·		32	14	22	48	♏	ush
3	16	V	17	34	2	26	50	♐	ush
3	17	VII	5	35	50	2	40	♒	ush
3	18	VII	25	37	38	10	18	♓	ush
3	19	IX	13	36	38	16	56	♈	ush
3	20	IX	30	34	50	21	46	♉	ush
3	21	XI	15	33	2	24	48	♊	ush
3	22	XII	28	31	14	26	2	♋	ush
3	24	I	10	29	26	25	28	♌	ush

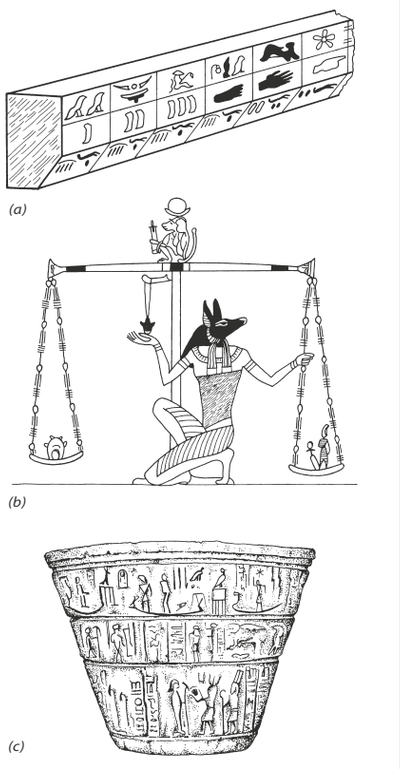
▼ **Figure 1.22** Modern graphical representation of the Jupiter data.



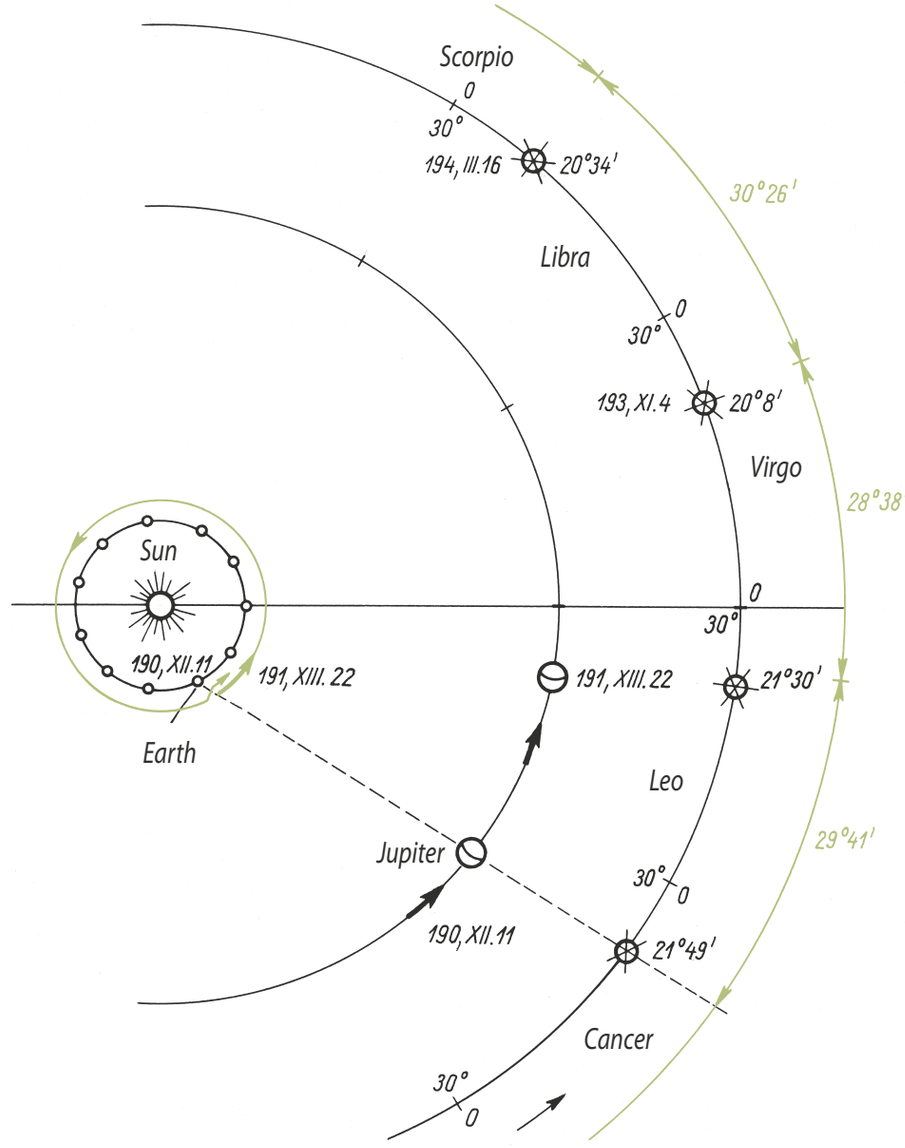
♈	Aries	March 21–April 20
♉	Taurus	April 21–May 20
♊	Gemini	May 21–June 21
♋	Cancer	June 22–July 22
♌	Leo	July 23–August 23
♍	Virgo	August 24–September 23
♎	Libra	September 24–October 23
♏	Scorpio	October 24–November 22
♐	Sagittarius	November 23–December 21
♑	Capricorn	December 22–January 20
♒	Aquarius	January 21–February 19
♓	Pisces	February 20–March 20

Sun	Mercury	Venus	Earth	Moon	Mars	Jupiter	Saturn
☉	☿	♀	⊕	☽	♂	♃	♄

▲ **Figure 1.24** The signs of the zodiac and of the most important celestial bodies. Their names are associated with Greek and Latin mythology.



▲ **Figure 1.25** The Egyptians were in possession of measuring devices for determining the three basic quantities, length, mass, and time. **(a)** Royal measuring rod, ca. 1500 BCE: Its total length is about 52 cm (20 in.). **(b)** Illustration of a balance in a Book of the Dead from the twenty-first dynasty. **(c)** A water clock (clepsydra): the markings on the inner wall of the vessel showed the passage of time as water dripped from a small hole in the base.



▲ **Figure 1.23** Illustration of the physical meaning of the Jupiter data based on current knowledge; it could almost serve as an explanation of KEPLER's second law (the law of areas).

differences at certain points abruptly change sign. If we represent the tabulated data in modern graphical form, then the graph will consist of alternating segments of rising and falling lines of equal steepness. The significance of the table's content can be observed in Figure 1.21, which is an enlargement of a portion of Figure 1.20 together with the decipherment given by the Jesuit priest KUGLER. Based on what we have learned about Babylonian numerical notation, the reader should be able to decipher the numbers that appear in the table; thus the transcriptions of the columns labeled A, C, and D should require no further explanation.

Let us consider the columns one by one. Column A gives the year numbers—of course written in the sexagesimal system—calculated according to a chronology in which the year 1 was that of the accession to the throne of SELEUCUS in 311 BCE. Thus the pair of numbers 3 and 10 indicates year 3 × 60 + 10 = 190 of the Seleucid dynasty. The next numbers then represent the years 191, 193, 194, etc. The right-hand part of column B contains—as the reader may easily decipher—numbers corresponding to the various days of the month, and the left-hand part represents the month. Thus, for example, in the first row can be found the representation of the month Adaru, but in its place we shall use, as is the custom, Roman numerals. What is striking is the appearance of a 13th month. The extra month arises because the compilers of the table calculated with months of 30 days each, thus 12 months per year, for a total of 360 days. To avoid calendrical drift, from time to time (on average every six years) a 13th month of 30 days was intercalated.

What is the significance of a particular day given by columns A and B—for example, the 11th day of the 12th month of the year 190? What occurred on that day? The answer is to be found in column F: the

notation there ("ush") indicates that at issue is the second change in direction of the planet Jupiter. We know that the planets as observed from Earth traverse a complicated path with respect to the background of "fixed" stars (cf. Figure 1.56). They appear to move forward, reverse direction, go in the retrograde direction, again reverse direction, and then move forward. A second such reversal occurs at the given point in time. If we look now at the data for 22.XIII.191, we again find the symbol "ush," and for the following dates as well. The data for columns A and B are therefore those for the second reversals of Jupiter. Columns E and D now provide information about where Jupiter was to be observed at these times. The precise position is given by specifying the relevant zodiacal sign together with the location within the sign which is divided into thirty degrees. Thus, for example, the location of the second Jovian reversal on 11.XII.190 is to be found in Cancer at longitude 21° 49′. The dates and positions for the remaining reversals are given similarly, where, of course, the numbering of the degrees begins anew upon entry into the next zodiacal sign. Finally, column C tells how far apart (measured along the ecliptic) two successive second reversals of Jupiter are to be observed. Thus the data of column C are simply the result of subtracting two successive data from column D, with the restart of the degree scale at the new sign taken into account. The first angle recorded in column C must have been calculated with the help of a previous position that is not in the table. The next angle, namely 29° 41′, is calculated from the data in column D as follows: From the positional angle 21° 49′ in Cancer there is an angle of $30° - 29° 41′ = 8° 11′$ up to the zero mark in Leo. Together with angle of 21° 30′ in the Lion, we obtain altogether $8° 11′ + 21° 30′ = 29° 41′$.

In Figure 1.22, time appears on the horizontal axis, and the dots on the graph represent the data from the table inserted at the appropriate points. As can be seen, these points are equidistant each from the next, namely, 402 days apart. That is not surprising, for, after all, we are dealing here with the so-called synodic period of Jupiter. (The synodic period of a planet is the time it takes for the planet to reappear at the same point in the sky, relative to the Sun, as seen from Earth.) In Figure 1.22, we have also recorded on the vertical axis the corresponding positional values from column C. It is now possible to check quantitatively the main observation mentioned near the beginning of this digression: Successive values differ by exactly 1° 48′, in both forward and retrograde motion.

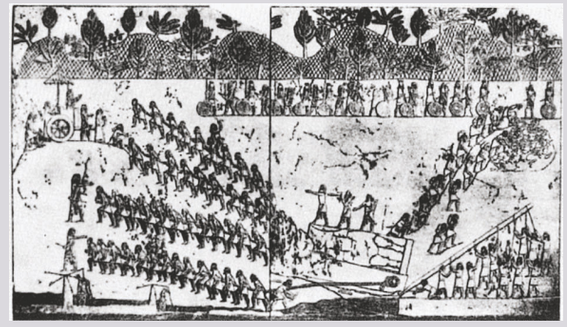
Figure 1.23 shows a sketch of what (according to our current knowledge) the Chaldean priests described in their tables. The sketch shows the path of Jupiter together with the positions of this planet and Earth for selected time points taken from the table. Finally, Figure 1.24 shows the signs of the zodiac and the symbols (regularized in a later era) of the most important celestial bodies.

Alas, we have no information at all about how the Chaldean priests went about preparing these tables. Direct observation would not have given them such linear results because reality is not that simple. Probably some theoretical consideration led them to choose the zigzag functions to represent periodic phenomena.

The greatest achievements of the Egyptians and Mesopotamians in the fields of mathematics and astronomy are immediately visible in the written records that have come down to us. The extent of their knowledge of physics—and mechanics, in particular—can only be inferred (Figures 1.25 and 1.26), principally from their achievements in structural engineering. If one considers that it was necessary to raise stone blocks weighing roughly 50 tons each to heights of several dozens of meters and to place them at precisely specified locations, and to move boulders weighing several hundreds of tons and to transport them using nothing but skids, it would seem unimaginable that among the overseers directing such work no practical sense had been developed for concepts that today we call friction, torque, and vector sums of forces. Even more astounding, however, is the organization of these construction projects. A look at Figure 1.26 should make this clear. No wonder that Lewis Mumford, in his book *The Myth of the Machine*, sees in such organized work collectives the prototypes of modern complex machines. In fact, they are machines, in the strict sense of the word, for the function of the individual human being has been reduced to that of a component; the systems of transmission of forces, the gears, hinges, drive mechanisms, are here replaced by human bone and muscle. The regulators in this enormous machine are the whips of the overseers. The term "megamachine" is appropriate, considering that tasks of comparable magnitude and complexity could be accomplished today only by using the machinery of the twentieth century.

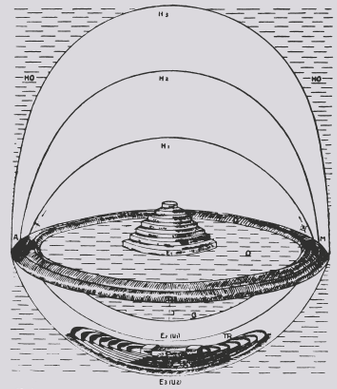
The Egyptians' and Mesopotamians' ideas about the cosmos do not go beyond the mythical (Figure 1.27). The heavens, the stars, the Moon, and the planets are

▼ **Figure 1.26** The Mesopotamian "megamachine," as Mumford called the organized and coordinated use of human muscle power. Friction, vector addition of forces, stability—the theoretical explanation of such concepts would remain a task for the distant future, but there must already have been a "feel" for them in order to carry out such tasks.





(a)



(b)

◀ ▼ **Figure 1.27**
(a) The heavenly queen Nut spans the celestial dome and holds aloft the stars (Saqqara, thirtieth dynasty. Metropolitan Museum, New York).
(b) A more objective image: the Mesopotamian universe (after Meissner: *Babylonien und Assyien*, 1925, p. 109).

deities that ply the heavens in boats; they are devoured and given birth to, one by the other; for example, the sky goddess, Nut, every night devours the Sun god, Ra, and every morning gives birth to him.

But the gods did more than this: they inspired the great king of the Neo-Babylonian Empire, Hammurabi (d. ca. 1750 BCE), to enact laws introducing a common legal standard to all of his domains. Hammurabi thus stands at the head of the long line of great lawgivers in the history of civilization (Figure 1.28).

When we are discussing the sources of knowledge available to the ancient Greeks, we also have to include the Phoenicians. Being a nation of seafarers, colonizers, and merchants, they made an enduring contribution to the simplification and transmission of the kind of knowledge important for everyday life—in particular, that of the alphabetic characters used in writing. Until quite recently, it was generally believed that their phonetic writing was a later development of a simplified form of the official Egyptian script, the demotic form (Figure 1.29(a)). The Greeks then only needed to add the vowels, which produced the alphabet that became the precursor of all European writing systems. Today, this view is questioned by many scholars (Figure 1.29(b)). Figure 1.29(c) shows the classical Greek alphabet, which today is used throughout the world in geometry, in mathematics, and in the physics of elementary particles.

Although the Egyptians and Babylonians possessed a great deal of mathematical knowledge, rather than formulating general rules of calculation, they always demonstrated particular procedures with numerical examples. Their pedagogy therefore always had the character of a set of instructions, and there is no trace of mathematical proof or any acknowledgment of the necessity for such proof. Nor did these peoples attempt to interpret their astronomical observations in the context of a physical model.

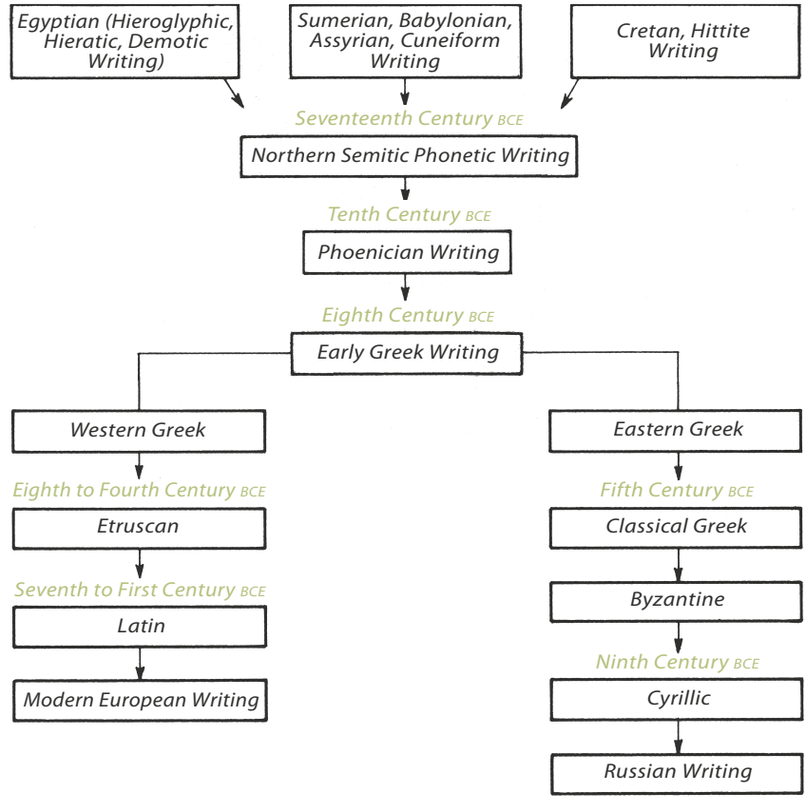


▲ **Figure 1.28**
The most important collection of laws from the ancient Near East is preserved on a 225-cm-high diorite stele. It is the *codex Hammurabi* (ca. 1700 BCE). The bas relief shows King Hammurabi before the sun god Shamash, who is handing him the book of the law.

> The righteous laws, which Hammurabi, the wise king, established and (by which) he gave the land stable support and pure government. ... I am the guardian governor. ... In my bosom I carried the people of the land of Sumer and Akkad ... in my wisdom I restrained them; that the strong might not oppose the weak, and that they should give justice to the orphan and the widow ... Let any oppressed man, who has a cause, come before my image as king of righteousness! Let him give heed to my weighty words! And may my monument enlighten him as to his cause and may he understand his case! May he set his heart at ease! (and he will exclaim): "Hammurabi indeed is a ruler who is like a real father to his people ... he has established prosperity for the people for all time and given a pure government to the land."... In the days that are yet to come, for all future time, may the king who is in the land observe the words of righteousness which I have written upon my monument.
>
> —Code of Hammurabi [Harper 1904, pp. 99–101]



▲ **Figure 1.29 (a)** The letter A might have developed this way.



▲ **Figure 1.29 (b)** The letters of the European alphabets attained their present form through a complex sequence of interactions (after the *Great Soviet Encyclopedia*).

The question, therefore, now arises as to whether they were perhaps not in possession of some deeper secret that has not come down to us. As a response to this question, a noted historian of science has pointed out that the level of development of a science could be inferred today even from engineering handbooks. Similarly, the ancient instructions written for the use of the scribes are evidence to support the assumption that neither the Egyptian nor the Chaldean priests possessed any kind of secret knowledge.

1.2 The Harmonious, Beautiful Order

1.2.1 *Overview: Temporal, Spatial, and Causal Connections*

In the history of science, a thorough treatment of the contributions of the Greeks typically begins with the sixth century BCE. By that time, Greece already had a long history: the events narrated by HOMER and the Mycenaean flowering were half a millennium in the past. The Greeks had already colonized the Mediterranean shoreline and the Aegean islands. Flourishing city-states were to be found not only in the mother country, but in Asia Minor, southern Italy, and Sicily; and there were settlements on the shores of the Black Sea and near what is now Marseille in the western Mediterranean. In addition to science and art, ancient Greece could thank these mutually competitive cities for another contribution, namely, democracy (though to be sure, the democracy of slaveholders). As seafarers, traders, and conquerors, the Greeks came into contact with the highly developed cultures to the south and east, that is, with those of Egypt and Mesopotamia. Both of these, but above all Egypt, had for several centuries already been hiring Greek mercenaries, who brought knowledge from these areas back home with them. The Greek merchants also obtained information, but mostly of the kind useful in trade and commerce. There were also some people at that time who had the means to undertake travel for the sole purpose of education and diversion, people we would today call tourists.

In the fifth century BCE, the Greeks basically maintained contact with only one great empire, from which they could acquire scientific knowledge and civilized customs, but which was also powerful enough to threaten their very existence: this was Persia. In the west, in Italy—a relatively uncivilized peasant republic annexed its neighbors obstinately and inflexibly one after another, absorbing the cultures of the peoples they conquered, notably the Etruscans—the Persians were able to raise the level of their own culture. There was as yet no indication that this country would one day become a serious rival in power, and finally the heir to both the Persian and Hellenic Empires. On the Mediterranean coast and on the distant Iberian Peninsula, Phoenician colonies were in flower, the most notable of which was Carthage. They also had no inkling of the mortal danger that Rome would pose for them.

In the interior of the European continent, from the Atlantic Ocean to the steppes of southern Russia, dwelt a wide variety of peoples (Celts, Scythians), some on a relatively high cultural plane. However, they had not reached the level of social organization that would have allowed them to influence the historical development in the Mediterranean region. Nonetheless, the Greeks carried on a lively trade with them.

It is easy to see that the Greeks could have had no detailed knowledge of the other great cultures—in particular, of the cultural and scientific achievements of the Indians and Chinese. Yet, one may observe at about the same time in these different cultures, around the fifth century BCE, a more intense level of questioning concerning the relation of man to the heavenly powers and to nature, as well as that of the individual to society. Such questions had engaged humanity from the most ancient times, and they have retained their currency to the present day.

Α	α	ἄλφα	Alpha
Β	β	βῆτα	Beta
Γ	γ	γάμμα	Gamma
Δ	δ	δέλτα	Delta
Ε	ε	ἒ ψιλόν	Epsilon
Ζ	ζ	ζῆτα	Zeta
Η	η	ἦτα	Eta
Θ	θ, ϑ	θῆτα	Theta
Ι	ι	ἰῶτα	Iota
Κ	κ	κάππα	Kappa
Λ	λ	λάμβδα	Lambda
Μ	μ	μῦ	Mu
Ν	ν	νῦ	Nu
Ξ	ξ	ξεῖ	Xi
Ο	ο	ὂ μικρόν	Omicron
Π	π	πεῖ	Pi
Ρ	ρ	ῥῶ	Rho
Σ	σ,ς	σίγμα	Sigma
Τ	τ	ταῦ	Tau
Υ	υ	ὗ ψιλόν	Upsilon
Φ	φ	φεῖ	Phi
Χ	χ	χεῖ	Chi
Ψ	ψ	ψεῖ	Psi
Ω	ω	ὦ μέγα	Omega

▲ **Figure 1.29 (c)** Greek letters used frequently in this book.

ا	*Alif*	*ā*	ض	*Ḍād*	*ḍ*
ب	*Bā'*	*b*	ط	*Ṭā'*	*ṭ*
ت	*Tā'*	*t*	ظ	*Zā'*	*ẓ*
ث	*Ṯā'*	*ṯ*	ع	*'Ain*	*ʿ*
ج	*Ǧīm*	*ǧ*	غ	*Ġaim*	*ġ*
ح	*Ḥā'*	*ḥ*	ف	*Fā'*	*f*
خ	*Ḫā*	*ḫ*	ق	*Qāf*	*q*
د	*Dāl*	*d*	ك	*Kāf*	*k*
ذ	*Ḏāl*	*ḏ*	ل	*Lām*	*l*
ر	*Rā'*	*r*	م	*Mīm*	*m*
ز	*Zāy*	*z*	ن	*Nūn*	*n*
س	*Sīn*	*s*	ه	*Hā'*	*h*
ش	*Šīn*	*š*	و	*Wāw*	*w, ū*
ص	*Ṣād*	*ṣ*	ى	*Yā'*	*y, ī*

▲ **Figure 1.29 (d)** Arabic script.

Quotation 1.6

The word of the Lord came unto me again, saying,

What mean ye, that ye use this proverb concerning the land of Israel, saying, The fathers have eaten sour grapes, and the children's teeth are set on edge?

As I live, saith the Lord God, ye shall not have *occasion* any more to use this proverb in Israel.

Behold, all souls are mine: the soul that sinneth, it shall die.

But if a man be just, and do that which is lawful and right ...

And hath not oppressed any, *but* ... hath given his bread to the hungry, and hath covered the naked with a garment; ...

Hath walked in my statutes, and hath kept my judgments, to deal truly; he *is* just, he shall surely live, saith the Lord God.

If he beget a son *that is* a robber, a shedder of blood, and ...

Hath oppressed the poor and needy, hath spoiled by violence ... shall he then live? He shall not live; he hath done all these abominations; he shall surely die; his blood shall be upon him.

Now, lo, *if* he beget a son, that seeth all his father's sins which he hath done, and considereth, and doeth not such like ...

Neither hath oppressed any, ... hath given his bread to the hungry, and hath covered the naked with a garment ... he shall not die for the iniquity of his father, he shall surely live.

As for his father ... he shall die in his iniquity.

Yet say ye, Why? Doth not the son bear the iniquity of the father? When the son hath done that which is lawful and right, *and* hath kept all my statutes, and hath done them, he shall surely live.

The soul that sinneth, it shall die. The son shall not bear the iniquity of the father, neither shall the father bear the iniquity of the son: the righteousness of the righteous shall be upon him, and the wickedness of the wicked shall be upon him.

—Ezekiel 18.

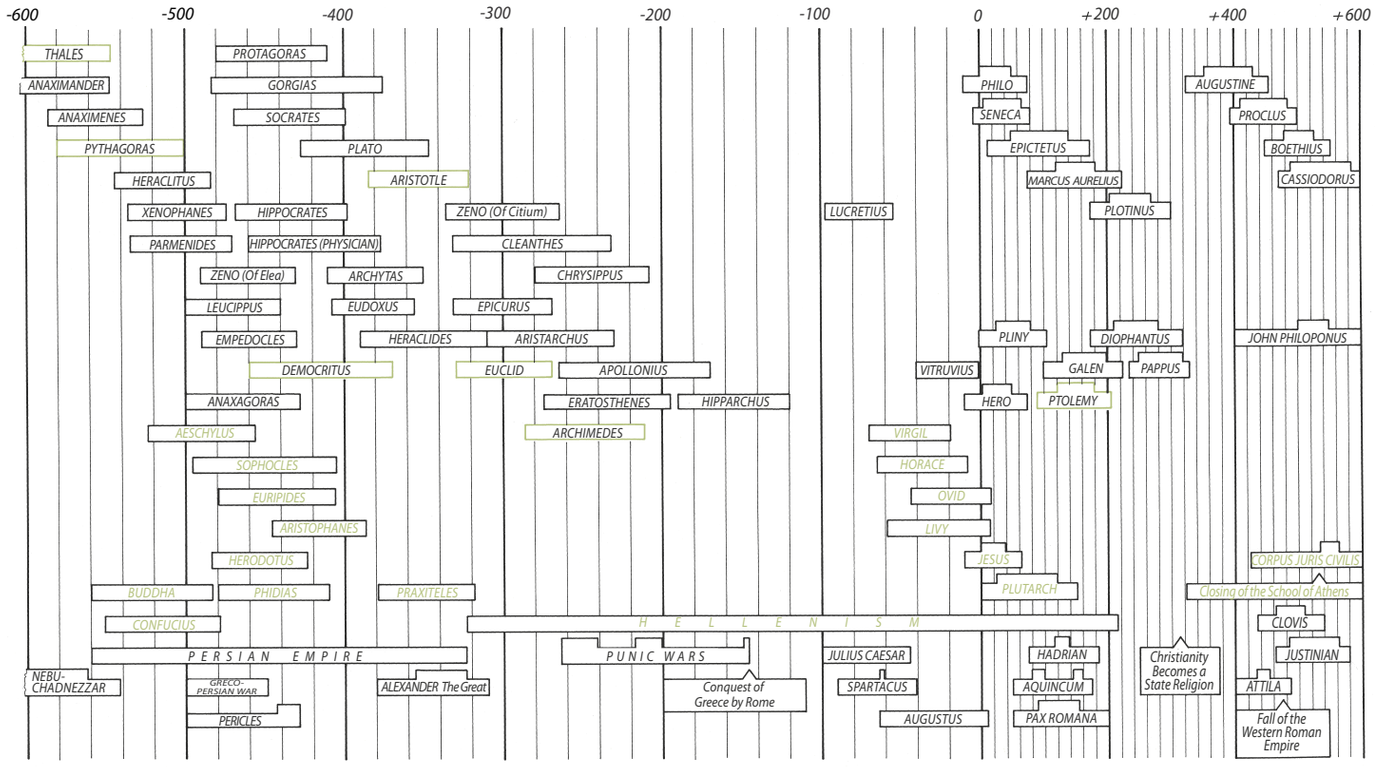
Thus, at the same time that Thales (ca. 625–ca. 547 BCE) considered the relations between man and nature while excluding heavenly powers, in Judea, the prophet Ezekiel was emphasizing the direct responsibility of the individual. Today, his words reach millions through the Old Testament (Quotation 1.6). In the same period, Buddha taught that individual happiness could be achieved through the complete renunciation of everything worldly. His teachings are today followed by hundreds of millions. And finally, Confucius investigated the place of the individual in society as well as the norms of moral behavior. His teachings—whether praised or condemned—even today influence the lives of millions (Quotations 1.7 and 1.8). These teachings have a certain relevance to the history of science. Thus, for example, some see in the connection between sin and repentance the germ of the idea of causality linking cause and effect.

The Greeks certainly had to fight to defend their freedoms against the expansionist Persian Empire pressing against them from the east. Their successful struggle that united all the Greeks—one has only to think of the now legendary victory at Marathon or the naval battle at Salamis—gave them a belief in their own physical, moral, and spiritual strength, which was a necessary precondition, among other things, for the ushering in of the Golden Age of Pericles in Athens (Table 1.1, Plate II).

Pericles lived to experience the outbreak of the civil war after the Golden Age (the Peloponnesian War, a battle for hegemony between Sparta and Athens) and thereby the beginning of the Greeks' decline. Yet the flowering of Greek philosophy and science falls precisely in the next century, the fourth century BCE, which is the century of Plato and Aristotle. The last third of that century was under the reign of Philip, king of Macedonia, who gradually brought all of Greece under his dominion. His son, Alexander the Great, who was educated by Aristotle, set out in the year 334 BCE to subjugate the Persian Empire and to conquer all the world of political significance then known to the Greeks. Thanks to Alexander's sterling ability as a military leader, and his novel martial tactics, but also to the internal weaknesses of the Persian Empire, his campaign was a total success. Alexander pressed as far as India to the east and Bactria to the north. In the south, he subdued the entire Egyptian Empire. The cultural-historical significance of these military conquests is immeasurable. Greek culture spread to all the conquered lands, even as far as India where we can see Greek influence in Indian art. Conversely, Greek culture was enriched by its contacts with the Mesopotamian and Egyptian cultures. The result of all these cultural processes is Hellenism, whose most brilliant contributions from the point of view of science took place not on Greek territory, but in the city Alexandria, founded by Alexander in 332 BCE in Egypt.

The flowering of Hellenism, in the third and second centuries BCE, was followed by a slow decline, interrupted by some significant accomplishments in the first and second centuries CE. The quiet decay of Hellenism—first within the unified Roman Empire, then in the Byzantine Empire—was given its final blow by the Arab conquest.

Besides the chronology of Table 1.1, it is also worthwhile to look at the geographical connections depicted in Figure 1.30. We see that the development of the sciences had their start at the beginning of the fifth century BCE with the Ionian natural philosophers who lived on the coast of Asia Minor. That is not surprising because this is where the cultural influences from the east and south were strongest. Beginning in Asia Minor, the center of gravity moved to the Pythagorean school as well as the school of Elea, in what is now southern Italy, lying geographically to the west of the motherland. Thereafter, philosophy experienced its heyday in Athens, with the great trio of Socrates, Plato, and Aristotle.



▲ **Table 1.1** Chronology of the great names from the Greco-Roman and Hellenistic worlds.

The names of the main players in this book are shown in colored frames, and the names of other significant persons from the realms of art, literature, and religion are shown in colored type. The table covers the six centuries before and six centuries after the beginning of the Common Era, with the latter appearing in a time scale compressed by half, which emphasizes the stagnation of intellectual life in that period.

We would like to give brief information about some of the names in the table that are mentioned in the main text only in passing or not at all.

Xenophanes moved from Colophon, in Asia Minor, to the southern Italian town of Elea, where he founded the Eleatic school of philosophy; he may thus be considered to be the teacher of Parmenides. He mocked the mysticism of Pythagoras and the anthropomorphism of the Homeric gods: "If oxen and horses or lions … could draw … as men do, horses would draw the forms of the gods like horses, and oxen like oxen, and lions like lions."

For Xenophanes there was but one god, a god dissimilar from man in both appearance and manner of thinking. A namesake of Zeno of Elea (who viewed all change as impossible) is Zeno of Citium, the founder of Stoicism. His school branched out to include Cleanthes and Chrysippus. More or less closely associated with the Stoic school are Epicurus, Lucretius, Seneca, the slave Epictetus, and the emperor Marcus Aurelius.

Anaxagoras (500–428 bce), a friend of Pericles, was of the opinion that the Sun is a burning rock, perhaps larger than the entire Peloponnesus. He understood correctly the cause of solar and lunar eclipses.

The physician Hippocrates is more important than his mathematician namesake. The Hippocratic oath goes back to him. His work was continued by Galen, whose teachings on medicine influenced medical practice in Europe until the sixteenth and seventeenth centuries.

Diophantus, Pappus, and Proclus were important mathematicians. Their names are recalled in Diophantine equations and several propositions that go under the name *Pappus's theorem*. Of great importance to the history of science are Proclus's commentaries on the works of Plato and Euclid.

The scholar Philo, who worked in Alexandria, was of Jewish background. He attempted to harmonize the teachings of Plato and the Stoics with the Old Testament. Boethius, Cassiodorus, and Philoponus will play a role in a later chapter of this book.

Vitruvius Pollio lived shortly before the beginning of the Common Era. He was an outstanding architect and the author of the ten-volume *De architectura*. In its Italian translation, it was a very important work even in the sixteenth century. In addition to Vitruvius, two other noteworthy practitioners of Roman civil engineering were Sextus Julius Frontinus (40–113 ce, *De aquaeductibus urbis Romae*) and the Damascene Apollodorus (first century ce, a famous builder of bridges).

The Roman Empire can lay claim to no significant natural scientists. However, we should not forget the network of roads, covering some 50,000 miles with streets about 20 feet wide and including numerous bridges and tunnels, many of which are still in use. A unified legal code was created that applied to 50 million people. As early as 451 bce, at a time when Pericles was ruling in Greece, the particular jurisprudential capabilities of the Romans were demonstrated in the Law of the Twelve Tables (*Lex Duodecim Tabularum*, or simply *Duodecim Tabulae*). Among the legal scholars, we mention only Gaius (ca. 140–180) and his textbook *Institutiones* as well as Ulpian (170–228). From the latter come the definitions of right conduct: *honeste vivere, alterum non laedere, suum cuique tribuere* (live honorably, injure no one, give to each that to which he is entitled). Tribonian (ca. 500–547) was a minister under Justinian, and as director of the commission on legal codification led the preparation of the *Corpus Iuris Civilis*, which has influenced the European legal system even to this day.

Interesting synchronisms:

407 bce: Plato joins Socrates; production of Euripides' plays *Orestes* and *Iphigenia in Aulis*; construction of the Erechtheum.

387 bce: Plato's *Symposium*, founding of the Academy, whose motto was "Let no one ignorant of geometry enter here." The Gallic chieftain Brennus uttered the famous *vae victis* (woe to the vanquished) to the defeated Romans.

201 bce: Battle to the death between Hannibal and the Romans; end of the Second Punic War; Apollonius studies the conic sections; Eratosthenes measures the Earth's diameter at Alexandria.

170 ce: Marcus Aurelius battles the Marcomanni; in Pannonia, he writes his *Mediations*; Ptolemy works in Alexandria on his *Almagest* and *Geographia*.

Quotation 1.7

1. Now this, O monks, is the noble truth of pain: birth is painful, sickness is painful, old age is painful, sorrow, lamentation, dejection and despair are painful ...
2. Now this, O monks, is the noble truth of the cause of pain: that craving, which leads to rebirth, combined with pleasure and lust, finding pleasure here and there, namely, the craving for passion, the craving for existence, the craving for non-existence.
3. Now this, O monks, is the noble truth of the cessation of pain: the cessation, without a reminder, of that craving; abandonment, forsaking, release, non-attachment.
4. Now this, O monks, is the noble truth of the way that leads to the cessation of pain: this is the noble Eightfold Way: namely, right views, right intention, right speech, right action, right living, right effort, right mindfulness, right concentration.

—Buddha [Durant, 1954, p. 430]

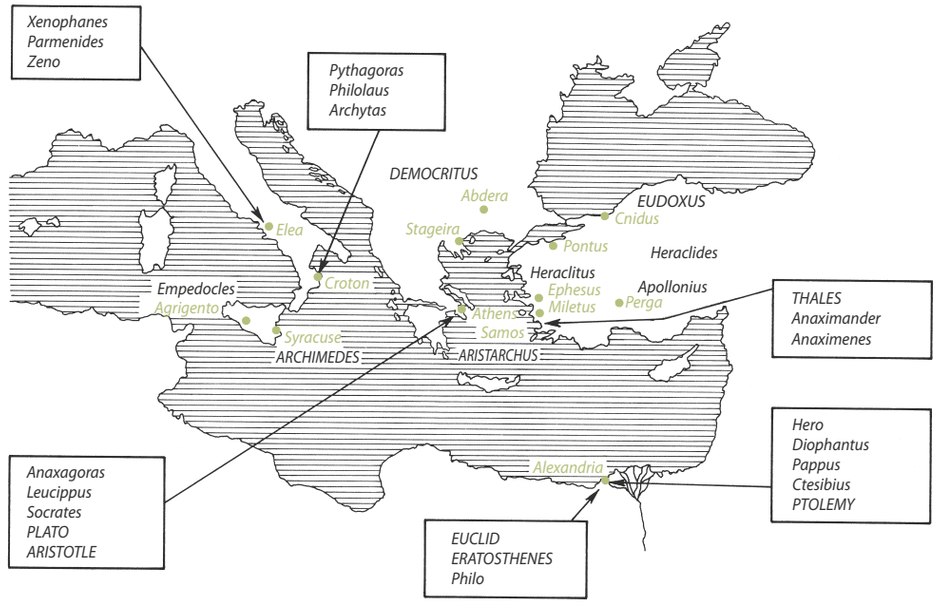
Quotation 1.8

The Master said, Guide them with government The Master said, Guide them with government orders, regulate them with penalties, and the people will seek to evade the law and be without shame. Guide them with virtue, regulate them with ritual, and they will have a sense of shame and become upright.

—Confucius, *Analects*, book 2, verse 3 [cited in Chai, 1973, p. 95]

When the great *Tao* prevailed, the world was common to all. Men of talent and virtue were selected; mutual confidence was emphasized and brotherhood was cultivated. Therefore, men did not regard as parents only their own parents, nor did they treat as sons only their own sons. A competent provision was secured for the aged till their death; employment [was provided] for the able-bodied; the means of growing up [was provided] for the young. Kindness and compassion were shown to widows, orphans, childless men, and cripples, so that they were all well cared for. Men had their respective occupations, and women their homes. They hated to see the wealth lying about in waste, and yet they did not hoard it for their own use.

—Anonymous, *Li Yün* ("Evolution of Rites") [cited in Chai, 1973, pp. 95–96]



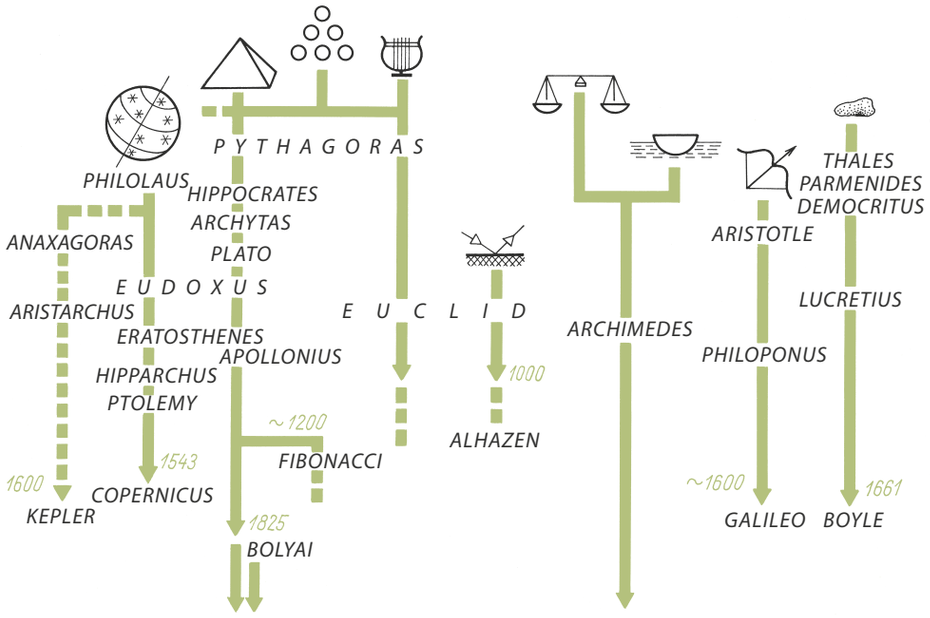
▲ **Figure 1.30** Greek philosophy and the natural sciences were born in the colonies of Italy and Asia Minor. In the age of Pericles and afterward, Athens was the center of cultural life; later, the sciences flourished in Alexandria.

With the exception of Aristotle, all the scholars whose ideas were taken up by Arab science, and later exerted an enormous influence on European science, lived and taught in Alexandria, among the most important of whom were Euclid, Eratosthenes, and Ptolemy.

During the flowering of Alexandria Romans made their appearance on the stage of the history of science. It began with a tragic event. In their struggle for domination of the Mediterranean, the Romans had to eliminate their one great obstacle to world conquest: Carthage. A bone of contention between the two great powers of Rome and Carthage was Sicily, home to the greatest Greek mathematician and physicist of the time, Archimedes. One of the greatest thinkers ever, he was in close contact with scholars at Alexandria and likely had worked there as well. During the capture of Syracuse, he was killed by Roman legionnaires.

Depicted in Figure 1.31 are the collections of themes that constituted Greek science, showing the practitioners and, where possible, the temporal relationships; we have restricted our attention here to mathematics and physics. Also noted are the times when the achievements of ancient Greek science were finally assessed at their true value, that is, when they were rejected, had their validity questioned, or were finally superseded. The eternal significance of Greek science can be seen by the fact that even today, some of their conclusions are still considered correct and continue to be taught in high school and even in universities. Here we may mention the fields of geometry, hydrostatics, and the statics of rigid bodies. Furthermore, some of their discoveries in astronomy, kinematics, and the structure of matter remain valid two thousand years later—both in the general and in the scientific cultures—as crowning achievements of human understanding in these areas.

If we pose the question of why the findings of geometry and statics have remained in force all these years, we might recall some of the ideas presented in the introduction to this book, according to which the most difficult step in the beginning of any scientific enterprise is the creation of the abstract model. In statics and



▲ **Figure 1.31** Themes in Greek natural science. The arrows indicate the length of time for which an idea was accepted as scientific truth.

geometry, the abstract model making possible a mathematical treatment is given, so to speak, by nature. In contrast, in the description of motion under terrestrial conditions and particularly in the question of the fundamental components of matter, much more abstraction is required. It is again no coincidence that the Greeks were able to give, with the Ptolemaic system, a precise and quantitative description of the motions of the heavenly bodies, for here again abstraction is easy because outside influences are insignificant. We might say that the Ptolemaic system as a *description* of the motions of the heavenly bodies can be viewed even today as correct, to be sure, with restricted validity—though at the same time the "restricted validity" qualification could be applied to all modern theories as well. The Ptolemaic system was displaced in the sixteenth and seventeenth centuries only because it offered no physical explanation of the observed phenomena and so was not suitable for the further development of dynamics.

From the viewpoint of the history of science, it is surprising with what boldness—equal at least to their courage in battle—the Greeks started with the most difficult questions at the outset: what is the fundamental nature of things, or—a bit simplified—what are the fundamental matter or matters that combine to form the diversity of the world? Regarding this question, they could give, of course, only qualitative pictures, and the qualitative picture left as a legacy by Aristotle could not be the source of continued development; only its complete abandonment could open the way to new ideas about the composition of matter. In contrast, the ancient atomic theory, even though also only qualitative, had the possibility of its transformation into a quantitative theory hidden within it.

In what follows, all the lines of development sketched here will be presented, even if we cannot go into every detail. In general, we shall content ourselves—as we did in the section on Egyptian and Babylonian science—to present the most important results. All the while, however, we should not neglect to indicate how certain ideas were formed and perfected, because only in this way can we truly discover the process by which great ideas develop.

Quotation 1.9

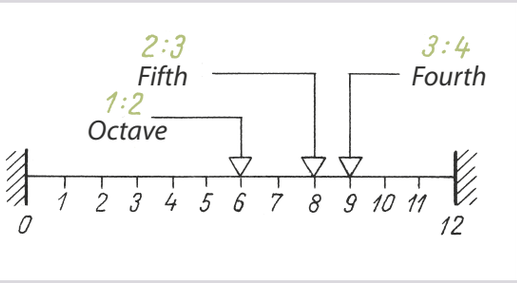
Pythagoras is one of the most interesting and puzzling men in history. Not only are the traditions concerning him an almost inextricable mixture of truth and falsehood, but even in their barest and least disputable form they present us with a very curious psychology. He may be described, briefly, as a combination of Einstein and Mrs. Eddy. He founded a religion, of which the main tenets were the transmigration of souls and the sinfulness of eating beans. His religion was embodied in a religious order, which, here and there, acquired control of the State and established a rule of the saints. But the unregenerate hankered after beans, and sooner or later rebelled.

Some of the rules of the Pythagorean order were:
To abstain from beans.
Not to pick up what has fallen.
Not to touch a white cock.
Not to break bread ...
Not to stir the fire with iron ...
Not to walk on highways.
Not to let swallows share one's roof.
When the pot is taken off the fire, not to leave the mark of it in the ashes, but to stir them together.

—Bertrand Russell, *A History of Western Philosophy*, 1945 [pp. 31–32]



◀ **Figure 1.32** Pythagoras and music (miniature, about 1210; Bayrische Staatsbibliothek).



▲ **Figure 1.33** The first concrete quantitatively formulated law of nature. Harmony among sounds is achieved when the lengths of the strings are in proportions of small integers. It is likely that the Pythagoreans performed experiments to determine this.

▼ **Figure 1.34** Orpheus can tame wild animals because they, too, resonate to the harmonies of the world. But harmony is a number, and so the essence of the world is the number. (Roman floor mosaic, first half of the second century CE [Honolka, 1968].)

> The spectral series that are controlled by integral numbers are basically the metaphorical generalizations of the harmonic chords of the ancient lyre, from which the Pythagoreans 2500 years ago derived the "integer" harmonies of nature; our quanta are indeed reminiscent of the role that the integer numbers must have played for the Pythagoreans: for they express not the characteristics of physical events, but their very essence. [author's translation]
>
> —Arnold Sommerfeld, *Atomic Structure and Spectral Lines*, 1925



1.2.2 *Mysticism and Mathematics: Pythagoras*

We know very little about the life of Pythagoras. There are some who question whether he was actually a living person, or—similar to other founders of religions or sects—that his person, life, deeds, and the sayings attributed to him were merely the inventions of disciples (Figure 1.32).According to recent research, however, we have to accept that Pythagoras really lived. He was born around 580 BCE on the island of Samos, and then apparently fled, from the tyrant Polycrates, to Croton, in southern Italy, where surrounded with his disciples he formed a religious sect.

As sketchy and unreliable as the details about Pythagoras's life are, so is the range of evaluation of his works. Many authors are of the opinion that he merits the title of "founder of modern science," whereas others see in him more the poet, the mystic, and the prophet. His role in society is condemned by some as reactionary ("idealistic philosopher, served the aristocracy") and praised by others as progressive ("plebian origins, represented the interests of the craftsmen").

It is difficult to separate the work of the master from that of his pupils, so when we speak in the following of Pythagoras, we should rather credit the Pythagoreans.

Pythagoras's work—his achievements in mathematics and his scientific views about music and the cosmos—should be appraised highly, but only if we disregard the way these ideas were obtained and the mystical-religious mold into which they were forced.

The novel ideas of the Pythagoreans were born in the midst of such a welter of religious mysticism and superstitions that they seem a continuation of the cultic taboos of tribal times. Viewed as such, their attitudes and work are quite far from what would today be called scientific. We can see this in some of the Pythagoreans' compulsory rules for living, given in Quotation 1.9. Yet perhaps it is natural that the beginning of scientific discovery should arise in such a manner; indeed, it cannot be expected that right at the beginning of the history of science, both the results and the correct principles behind the results would appear together in their pure forms. The introductory chapter of this book stressed that a science should lead to true results and must possess a method for putting truth to the test. The means by which a truth is discovered is irrelevant for science. For example, if a mathematician dreams of a theorem, and can prove it by using generally accepted mathematical methods, then from the point of view of science, this result is just as valid as one derived systematically. This is not to say that there are no scientific methods for the seeking of truth. As we shall see (Section 1.4), Archimedes himself gives advice for how to "intuit" new results in a scientific way.

It is of particular interest to the history of science and instructive for today's scientific researchers to follow the path of associations, which may be confused and sometimes fantastic, but which leads to a scientific theory that we can objectively test and show to be true today.

The entire Pythagorean worldview can be summed up in the following statement: Number is nature and the essence of all things.

By number, the Pythagoreans, of course, meant the positive integers. In physics, this general principle can be witnessed in relation to musical harmony: the Pythagoreans found a simple relationship between the lengths of stretched strings that produce harmonic chords: for the octave the ratio is 1:2, whereas for the fifth, it is 2:3, and for the fourth, 3:4. Thus, we see that the harmonic intervals are intimately related to the first few positive integers (Figures 1.33 and 1.34).

Conversely, it may be that this musical harmony is precisely what led the Pythagoreans to their general thesis. Numbers have a role in the structure of the cosmos; this is an idea that is closely linked to the Orphic cult that spread in southern Italy. Orpheus was able to influence both the living and the dead with his music because the whole world was believed to be constructed according to certain harmonic relationships and could therefore be made to resonate with particular harmonies. The Pythagoreans called the world *cosmos*, meaning "beautiful order" (Quotation 1.10). The harmonic–harmonious beautiful order—like the lengths of the stretched strings—can be expressed in whole numbers. Thus, in the final analysis, the cosmos, as the apotheosis of harmonious order, stands in an intimate relationship to the integers.

It is often stated that the discovery of the relationships between the lengths of the strings of a musical instrument was the first law of nature—in the modern sense of the word—in the history of physics because it expresses a physical fact in mathematical form.

Accordingly, investigations of the cosmos also mean investigations of numbers, and with this a mysticism of numbers came into being, from which we derive many interesting results, if only we separate them from their mystical features.

The Pythagoreans had particular regard for the number 10, since it could be represented as the sum of the first four positive integers (10 = 1 + 2 + 3 + 4), which themselves had a particular significance. They arranged these numbers as dots in the form of an equilateral triangle to make the sacred symbol of the Pythagoreans, the tetractys (Figure 1.35), on which new adherents to the sect had to swear an oath (Quotation 1.10).

To emphasize the fundamental quality of the number 10, the Pythagoreans distinguished exactly ten different qualitative pairs of opposites, and they also assigned a special role to the number 10 in their astronomy (the cosmos of Philolaus; see Quotation 1.11).

The Pythagoreans assigned particular properties to certain numbers and also associated certain numbers with certain qualities. This association is of considerable interest because, in a certain sense, it leads to number-theoretic problems of current interest. Thus, for example, a number was considered perfect if it is equal to the sum of its proper divisors. Examples of perfect numbers are 6, since

$$6 = 1 + 2 + 3,$$

and 28, since

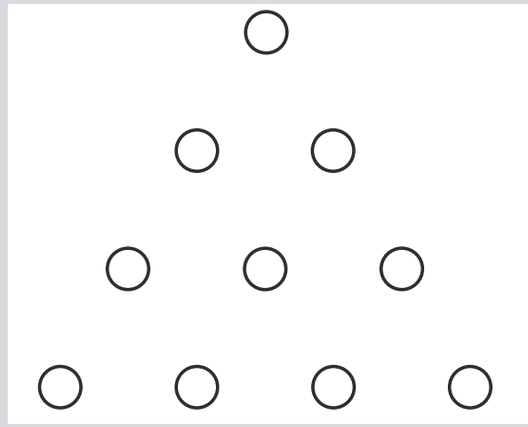
$$28 = 1 + 2 + 4 + 7 + 14.$$

The following theorem, which is to be found in Euclid, holds for perfect numbers: If a prime number p can be written in the form

$$p = 1 + 2 + \cdots + 2^n = 2^{n+1} - 1,$$

then the number $2^n p = 2^n (2^{n+1} - 1)$ is a perfect number. As of 2010, the largest known perfect number was $2^{43,112,608} (2^{43,112,609} - 1)$.

The Pythagoreans also associated geometric figures with the integers, creating with their *figurate numbers* the field of *arithmogeometry*, or the geometry of numbers, a synthesis of arithmetic (number theory) and geometry. For example, they distinguished triangular and square numbers (Figure 1.36) and derived interesting relationships from such representations. Thus they discovered—just to mention one result—that the sum of two successive triangular numbers is a square number.



▲ **Figure 1.35** The "holy tetractys." The number ten has mystical significance because it is the sum of the first four natural numbers.

Quotation 1.10

And philosophers tell us, Callicles, that communion and friendship and orderliness and temperance and justice bind together heaven and earth and gods and men, and that this universe is therefore called Cosmos or order, not disorder or misrule, my friend.

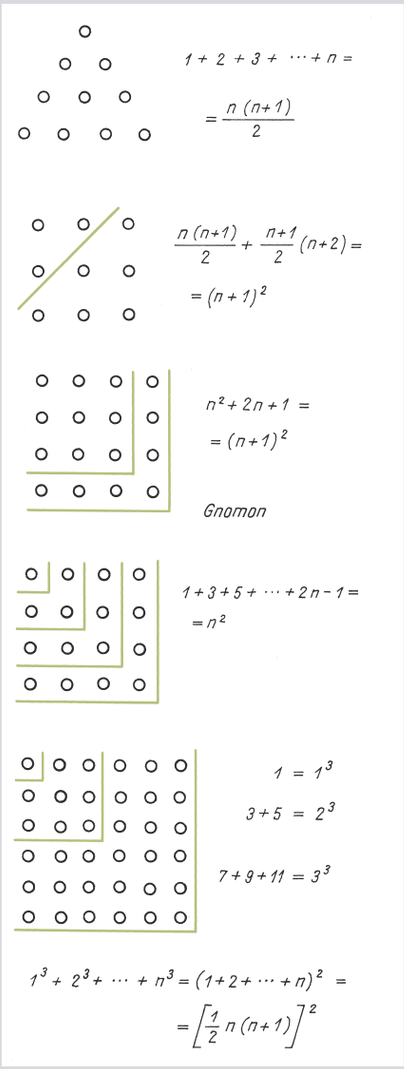
—Plato, *Gorgias*

Quotation 1.11

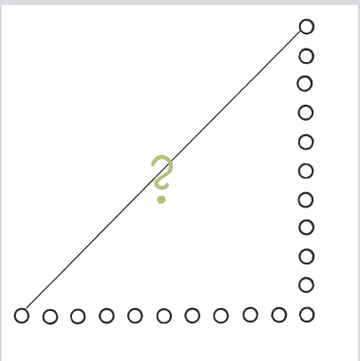
The so-called Pythagoreans, who were engaged in the study of mathematical objects, were the first to advance this study, and having been brought up in it, they regarded the principles of mathematical objects as the principles of all things. ... And whatever facts in numbers and harmonies could be shown to be consistent with the attributes, the parts, and the whole arrangement of the heaven, these they collected and fitted into a system; and if there was a gap somewhere, they readily made additions in order to make their whole system connected. I mean, for example, that since ten is considered to be complete and to include every nature in numbers, they said that the bodies which travel in the heavens are also ten; and since the visible bodies are nine, they added the so-called "Counter-Earth" as the tenth body. ...

Other members of this same school say there are ten principles, which they arrange in two columns of opposites: limit and unlimited, odd and even, one and plurality, right and left, male and female, resting and moving, straight and curved, light and darkness, good and bad, square and oblong.

—Aristotle, *Metaphysics,* Book A, Chapter 5 [Apostle, pp. 20–21]



▲ **Figure 1.36** Interesting and significant relationships can be derived from the figurative arrangement of numbers.



▲ **Figure 1.37** What can be said about the length of the hypotenuse of a right triangle when the two legs have integer lengths? The Pythagoreans answered this question correctly: the hypotenuse can be represented neither by an integer nor by the quotient of two integers. But this answer meant the bankruptcy of their entire number mysticism.

Using these relationships, various summation formulas for series were derived or, rather—from the viewpoint of current pedagogy, illustrated.

There are opinions that the Pythagoreans considered the points in their geometric diagrams as real, that is, as existing in the phenomenal world, and indeed, that they can be seen as constituting the elementary building blocks of matter. Thus in Pythagorean numerological mysticism one may detect the seeds of modern atomic theory.

The most significant finding of the Pythagoreans in the area of number theory is certainly the discovery of irrational numbers. However, this discovery came as a severe shock to the sect, for it heralded a complete collapse of their numerological mysticism. It was a shock from which later Greek science never fully recovered. One legend has it that the Pythagoreans threw the discoverer of irrational numbers into the sea, since his discovery had offended their belief in the beautiful, harmonious, cosmic order.

The discovery of irrational numbers arose from the question of the length of the diagonal of a square with side length 1 (Figure 1.37). The Pythagoreans were aware of the relationship between the lengths of the legs of a right triangle and the hypotenuse; indeed, this relationship is known today as the Pythagorean theorem. Therefore, they knew that the problem of determining the diagonal of a square could be solved by way of the relationship

$$a^2 = 1^2 + 1^2 = 2,$$

where the number a is the unknown whose value is sought. Clearly—thus argued the Pythagoreans—this number a must be a fraction, the quotient of two integers, lying somewhere between 1 and 2. That is, it should be representable as

$$a = \frac{m}{n},$$

where at least one of the integers m, n is odd, for if both of them were even, then the two numbers could be continually divided by 2 until at least one of them was odd. Therefore, the relation $a^2 = 2$ can be written in the form

$$\frac{m^2}{n^2} = 2, \quad m^2 = 2n^2.$$

It is clear that m must be even, since only the square of an even number can be even; thus by the assumption given above, n must be odd. If we write $m = 2p$, to represent m as an even number, then the relationship between m and n becomes

$$(2p)^2 = 2n^2,$$

or, after division by 2,

$$2p^2 = n^2.$$

But from this relationship we are forced to conclude that in contradiction to our original assumption, n is an even integer as well. From this contradiction we conclude that the number a cannot in fact be represented as the quotient of two integers. In other words, the diagonal of the square cannot be represented as a sequence of points as seen in Figure 1.37. The Greeks soon recognized that if they tried to assign an integer value to the diagonal instead, then the length of a leg of a right triangle, or equivalently, the side of a square, becomes irrational (ἄλογος). As a result, right from the very beginning, geometry and arithmetic went their separate ways in Greek science. Because they were unable to analyze the irrational

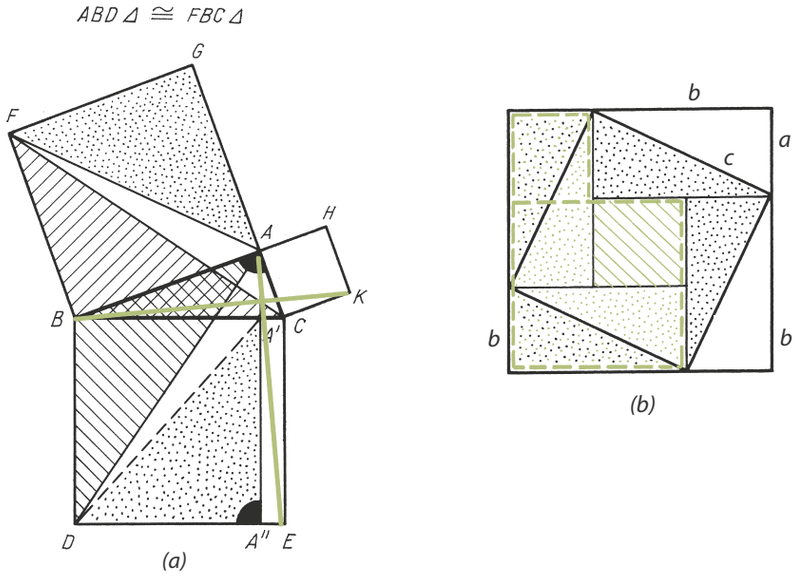
numbers in a way that conformed to their own strict logical requirements, ancient scholars either assumed that the numbers in a given problem could be represented as fractions or employed geometric proofs (Figure 1.38).

The great significance that the Pythagoreans ascribed to various averages, or means, also was a result of their numerology. It was they who introduced the notions of arithmetic, geometric, and harmonic means, which they used to define the perfect proportions and the musical proportions (Table 1.2). These proportions were later to play a role in art; they also appear in the contemporary aesthetics of our daily lives, for example, in the proportions used in books and writing paper.

The most significant of the Pythagoreans' discoveries in the area of geometry is the Pythagorean theorem, which states that in a right triangle, the sum of the areas of the squares with side lengths equal to the lengths of each of the legs is equal to the area of the square whose side length is the hypotenuse. Only a few theorems in all of mathematics have attained the renown of the Pythagorean theorem. Over the past 2500 years, students have turned it into verse, sometimes in jest and sometimes as a mnemonic formulation. And there are legends as to its origin. According to one of these, Pythagoras expressed his joy and gratitude for the discovery by sacrificing one hundred oxen. (It is usually added that for this reason, "dumb oxes" are always afraid of new scientific discoveries.)

However, the legend is almost certainly without historical basis because the Pythagoreans were vegetarians, and moreover, they respected all living beings and believed in reincarnation.

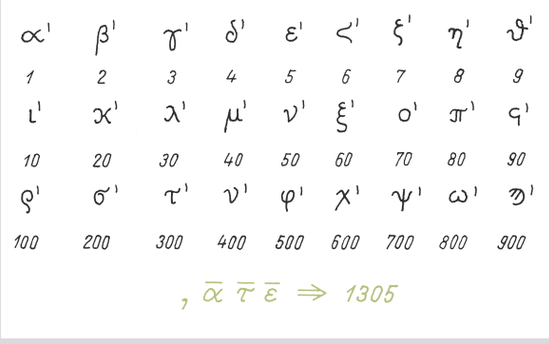
It is unknown whether Pythagoras actually proved the theorem that bears his name. Some writers have expressed the opinion that no proof in the modern sense of the word could have been given. The first proof that has come down to us stems from Euclid, who wrote it three hundred years after the time of Pythagoras; it is given in Figure 1.39.



▲ **Figure 1.39** **(a)** The Pythagorean theorem in the form given by Euclid. (Proclus attributes this proof to Euclid himself.) The triangles *BFC* and *BAD* are congruent (namely, *BA* = *BF*, *BC* = *BC*, and ∠*DBA* = ∠*CBF*). The area of triangle *BFC* is equal to that of triangles *BFA* ≥ *FAG*. The area of triangle *BFC* is equal to that of the triangles *BAD* = *DA′A″* since they have equal bases and altitudes. Therefore, the area of the square *BAGF* is equal to that of the parallelogram *B A′A″D*. Similarly, one can prove that the remaining portion of the square *BCED*, namely, the parallelogram *A′CEA″*, has the same area as the square *AHKC*. **(b)** The Pythagorean theorem as proved by the Chinese.

▼ **Table 1.2** The means and proportions introduced by the Pythagoreans.

Arithmetic mean:	$A=\frac{p+q}{2}$
Geometric mean:	$g=\sqrt{pq}_{\,2}$
Harmonic mean:	$H=\frac{2pq}{p+q}$
Mean proportional:	$A:G=G:H \quad G=\sqrt{AH}$
Musical proportion:	$p:A=H:q$

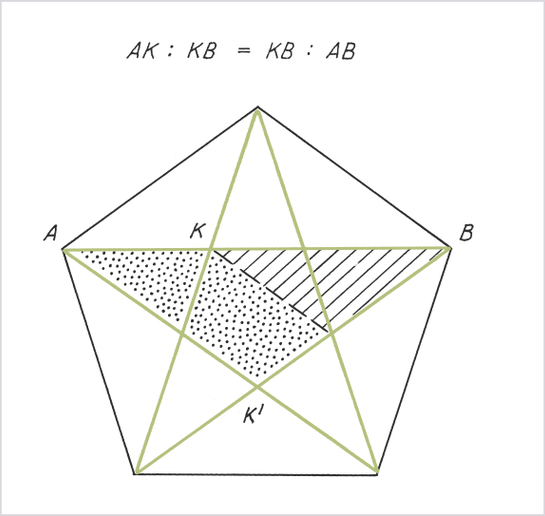


▲ **Figure 1.38** The Greeks' preference for geometry at the expense of arithmetic is illustrated, for example, by their convoluted number system. In addition to the alphabet, three obsolete letters were used to represent numbers: the digamma (or stigma) for 6; koppa for 90, which corresponds to the Latin Q; and sampi for 900. Numbers without their own letter were written as follows: ϡ = 11, ιβ′ = 12, ϡ = 13, etc. Thousands were written thus: ′α = 1000, ′β = 2000, ′γ = 3000, …, ′ι = 10,000, ′ια = 11,000, etc.

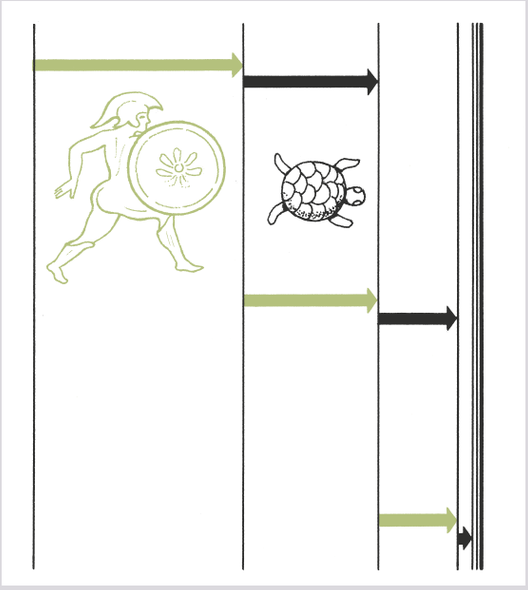
Before the Ionian system, presented here, the Attic system was in use from the fifth to the first century BCE.



The symbol for the number 5 is a variant of π (from the Greek *pente* = five). The symbols Δ, H, X, M come from *deka*, *hekaton*, *khilioi*, and *myrioi*.



▲ **Figure 1.40** The pentagram, the Pythagoreans' sign of recognition. The point *K* divides the segment *AB* in the golden section: The ratio of the smaller of the two segments thus obtained to the larger is equal to the ratio of the larger segment to the entire segment; that is, $AK : KB = KB : (AK + KB)$. In view of what will appear later, we note that for $KB/AK = x > 1$, this relation leads to the equation $x^2 - x - 1 = 0$, whose larger root is equal to $x_1 = (1 + \sqrt{5})/2 = 1.618\ldots$, known as the *golden ratio* or *golden section*.



▲ **Figure 1.41** The most famous of Zeno's paradoxes: Achilles, fleet of foot, can never overtake the tortoise (after [Russell, 1972]).

The five-pointed star (pentagram; see Figure 1.40) played a special role in Pythagorean geometry because its sides divide one another in accordance with the golden ratio. The pentagram served the Pythagoreans as a sign of recognition.

The Pythagoreans knew of four of the five regular (Platonic) solids—the tetrahedron, cube, octahedron, and dodecahedron. Sometimes the discovery of the fifth, the icosahedron, is ascribed to them as well, although it is not mentioned explicitly until Plato, two hundred years later.

The Pythagoreans' great accomplishment in the realm of astronomy is their assertion that Earth is a sphere. They arrived at this conclusion not from any physical observations, but rather on the basis of ideas of symmetry and perfection, which does not in any way detract from the significance of the claim (Plate I). This way, it is no longer necessary to imagine supports for the Earth in the form of elephants, turtles, or Atlas. To the contrary, an Earth moving freely in space must be unfettered, and in fact, the Pythagoreans considered such a possibility. According to their worldview, Earth revolves about a central fire, although one not identical to the Sun. In order to bring the number of heavenly bodies up to the sacred number ten, they had to invent an additional body, the counter-Earth, or antichthon, which is neither visible nor discernible by its effects. Notwithstanding all of its fantastic features, this image of the universe became the starting point for all later cosmologies—that is, the relationships among the Sun, Earth, and planets—propounded by later Greek scholars; see Section 1.4.

The mysticism and cultic taboos have been forgotten, but their belief in a "harmonious, beautiful order" has remained with us. Even two thousand years later, Copernicus proudly referred to Pythagoras as his teacher.

The belief in a relationship between reality and mathematics was well supported by the Pythagoreans' discoveries. We may thank Pythagoras and his school, above all, for the theorem that bears his name, for the proof for the existence of irrational numbers, for the image of Earth as a sphere in motion through the heavens, and for the relationship between musical pitches and the lengths of strings, just to mention the most important ones.

1.2.3 *Idea and Reality*

Two characteristic properties of Greek thought are of great importance for us. What stands out first of all is its enormous range. There are likely few, if any, significant movements in the history of philosophy whose roots are not to be found in the thought of some ancient Greek philosopher. Secondly, it is characteristic of the Greek thinkers that they actually used, and used to the full, all the possibilities opened up to them by the methods of formal logic that they had discovered and developed. With stubborn determination, they followed every line of thought through to the bitter end, even if this led to absurd conclusions. On the contrary, they were delighted to astonish themselves and others with paradoxical deductions. Having arrived by logical deduction at some conclusion, they boldly denied even the simplest facts, such as the existence of motion. The fact that motion can actually be observed was not acceptable as a counterargument. They knew that logically drawn conclusions were not to be refuted with sensory impressions but only by other logically drawn conclusions, which is something often overlooked today.

One Greek philosopher denied the existence of all motion, while another maintained that motion was all that existed. Hence the motto of Heraclitus was that everything flows, and nothing stands still (*panta rhei,* πάντα χωρεϊ καί

οὐδέν μένει). This means that reality can be understood as a result of competition between opposites and the arrival at equilibrium. But Heraclitus also indicated the traps created by reliance on the senses: "Evil witnesses are eyes and ears of men, if they have souls that do not understand their language." It is no coincidence that Heraclitus saw fire, the symbol of motion, as the foundation (*arché,* ἀρχή) of all things (Quotation 1.12).

As a result of his thinking about the questions of being, existence, and the passage of time, Parmenides (fl. ca. 500 BCE) came to the conclusion that everything that exists is one and indivisible. There is no change, and the phenomenal world is an illusion. With Parmenides, we arrive for the first time at an idea that can be traced throughout the entire history of philosophy: to think of something and for the object of the thought to exist are the same.

For us, the primary significance of the philosophy of Parmenides is that it leads directly to the atomism of Democritus.

Antisensationalism, whose most famous advocate was Zeno, is a protest against the truth of the knowledge acquired through our sensory organs. In our earlier discussion of the negation of motion, it was, in fact, Zeno that we had in mind. His aporias (ἀπορία), that is, his apparently unsolvable logical conundrums and paradoxes, are well known. One of these is about a race between Achilles and the tortoise: Achilles, fleet of foot, is unable to catch up with the tortoise (Figure 1.41), because in the time it takes for Achilles to reach the point where the tortoise had been when Achilles started, the tortoise has already moved a small distance farther. Subsequently, when Achilles covers that second distance, again the tortoise has moved along a bit, though to be sure through a smaller distance. As we know, the source of the paradox is that Zeno has divided the period of time that it would take Achilles to catch up with the tortoise into infinitely many intervals and concludes from the existence of infinitely many intervals of time that the tortoise will never be overtaken. However, we are well aware of the fact that the infinitely many intervals in the scenario have a finite sum and therefore represent a finite amount of time.

In particular, we are dealing with the infinite geometric series

$$t = \frac{l}{v_\mathrm{A}} + \frac{v_\mathrm{T}(l/v_\mathrm{A})}{v_\mathrm{A}} + \frac{v_\mathrm{T}(l/v_\mathrm{A})}{v_\mathrm{A}}\frac{v_\mathrm{T}}{v_\mathrm{A}} + \cdots = \frac{l}{v_\mathrm{A}}\left[1 + \frac{v_\mathrm{T}}{v_\mathrm{A}} + \left(\frac{v_\mathrm{T}}{v_\mathrm{A}}\right)^2 + \cdots\right],$$

where t is the total amount of time it takes for Achilles to overtake the tortoise, l is the original distance from Achilles to the tortoise, v_A is Achilles' velocity, and v_T is the tortoise's velocity. The sum of the series is given by

$$t = \frac{l}{v_\mathrm{A}}\left(1 - \frac{v_\mathrm{T}}{v_\mathrm{A}}\right)^{-1} = \frac{l}{v_\mathrm{A} - v_\mathrm{T}}.$$

Thus, Achilles clearly overtakes the tortoise, as it must be in accordance with common sense.

Protagoras is an outstanding representative of relativism. "Man is the measure of all things (πάντων χρημάτων μέτρον μέτρν άνθρωπος), of things which are, that they are, and of things which are not, that they are not ... I can know nothing of the gods, whether they are or whether they are not. ..."

Gorgias drew the ultimate conclusion:

Quotation 1.12

This world, which is the same for all, no one of gods or men has made; but it was ever, is now and ever shall be an ever-living. Fire, with measures kindling and measures going out. (Clement of Alexandria)

We step and do not step into the same rivers; we are and are not. (Heraclitus)

Men do not know how what is at variance agrees with itself. It is an attunement of opposite tension, like that of the bow and the lyre. (Hippolytus)

War is the father of all and the king of all; and some he has made gods and some men, some bond and some free. (Hippolytus)

All things are exchanged for Fire, and Fire for all things, even as wares for gold, and gold for wares. (Plutarch)

The lord whose is the oracle at Delphoi neither utters nor hides his meaning, but shows it by a sign. (Plutarch)

Dogs bark at every one they do not know. (Plutarch)

Eyes and ears are bad witnesses to men, if they have souls that understand not their language. (Sextus Empiricus)

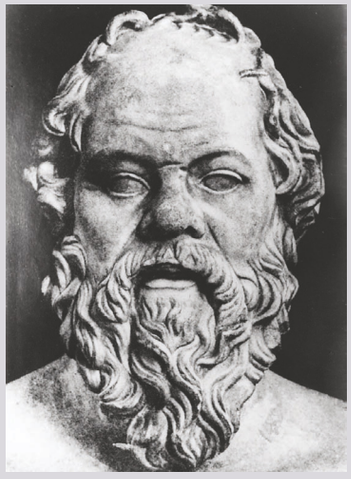
The Ephesians would do well to hang themselves, every grown man of them, and leave the city to beardless lads; for they have cast out Hermodoros, the best man among them, saying: "We will have none who is best among us; if there be any such, let him be so elsewhere and among others." (Diogenes Laërtius)

—Heraclitus, *Fragments*

Quotation 1.13

Ah, creature who yearn for grand wisdom from us, how blessed you will become among the Athenians and all Greeks,
if you're retentive and a cogitator,
if endurance abides in your soul,
if you don't tire out either standing or walking,
if you're not too annoyed by the cold or too keen on having breakfast,
if you stay away from wine and gymnasiums and all other follies,
and if, as befits a clever man, you consider absolute excellence to be victory in action, in counsel, and in tongue warfare.

—Aristophanes, *Clouds* [p. 69]



◀ Figure 1.42 Socrates (469–399 bce) was born ten years after the naval battle of Salamis. His father was a sculptor; his mother, a midwife. He studied under Anaxagoras and later under Archelaus. In 399, he was accused of disrespect toward the gods of his city and of having corrupted the Athenian youth. He was found guilty by a vote of 280 to 220. The vote on his sentence was more heavily against him because of his proud self-defense and he was condemned to death by poison.

In Plato's *Phaedo*, we can read of Socrates' relationship to the natural sciences:

> When I was a young man I was wonderfully keen on that wisdom which they call natural science, for I thought it splendid to know the causes of everything, why it comes to be, why it perishes and why it exists. I was often changing my mind in the investigation, in the first instance, of questions such as these: Are living creatures nurtured when heat and cold produce a kind of putrefaction, as some say? Do we think with our blood, or air, or fire, or none of these, and does the brain provide our senses of hearing and sight and smell, from which come memory and opinion, and from memory and opinion which has become stable, comes knowledge? Then again, as I investigated how these things perish and what happens to things in the sky and on the earth, finally I became convinced that I have no natural aptitude at all for that kind of investigation, and of this I will give you sufficient proof. This investigation made me quite blind even to those things which I and others thought that I clearly knew before, so that I unlearned what I thought I knew before.

i. Nothing exists.
ii. Even if existence exists, it cannot be known.
iii. Even if it could be known, it cannot be communicated.

In the Athenian democracy, "logic became a commodity," one might say, for in daily life, logic and rhetoric could bring one advantages (Quotation 1.13). It was the Sophists who devoted themselves to the task of teaching—for a fee—the art of argumentation. This, of course, eventually led to a misuse of logic, and moreover, such misuse has been observed in every age. Today as well, logical argumentation is used to prove the truth or falsity of an inference. However, we often forget the simple fact that a logical inference is derived from a set of prior assumptions, the premises. It is characteristic of logical operations that they leave the truth content of an assertion unchanged (invariant) or else transform it in a particular way. Therefore, even with the most rigorous logic, the truth content of inferences is no greater than that of the premises, which is to say, of the prior assumptions.

From antiquity, we have a number of examples of sophistic argumentation that could easily serve as material for a standup comic.

Aristophanes ridiculed the Sophists in his play *Clouds*, although he unjustly chose Socrates (Figure 1.42), a man who had himself fought against the Sophists, as the object of his mockery. It is maintained that with his play, Aristophanes contributed to the condemnation and the ultimate death of Socrates. But of greater interest are certain lines in the play, intended ironically but through which Socrates is brought closer to the physicists of today than in the best of Plato's dialogues (Quotation 1.14). The quotation is especially interesting because it involves a process of measurement. Even though it is a measurement of distance using the "foot of a flea"—whereby a clever technological procedure is specified for taking the measurement of a flea's foot—everything here suggests that Socrates clearly dealt with problems other than those of ethics. It is likely that this ironic reference relates to ideas about the measurement of distances, perhaps in astronomy.

Of even greater interest is our next quotation (Quotation 1.15). The subject here is lightning, and even in the distortions of Aristophanes, we find a core of truth: the fact that lightning is accompanied by compression of the air. All of this suggests that Socrates may have been concerned with a great variety of natural phenomena, and at a rather deep level. At his trial and conviction, if indeed Aristophanes' play had anything to do with it, it is precisely the section about lightning that could have played a role. Here, lightning is not the weapon with which Zeus punishes perjurers, for this lightning strikes at random into cork oaks, even though "an oak tree doesn't perjure itself." We shall meet this sentence again in a poem of Lucretius (Section 1.5).

That the real Socrates also argued somewhat "sophistically" can be seen from Quotation 1.16. Here, Plato has Socrates saying that it is philosophers who should rule the state. Using the Socratic method of teacher as midwife, he allows his own opinion to be revealed bit by bit from the mouths of his listeners. The reservation voiced by Adeimantus is brought up in discussions even today.

This is how rhetoric as the art of argumentation, and logic as the science of valid inference, developed. Contradictions appeared in everyday language, and the need to formulate clearer concepts increased.

A separate discussion of Plato (Figure 1.43), the most significant of the idealistic thinkers in the history of science, is required. An evaluation of his work exceeds the powers of a working physicist. Regarding his epistemology as well, most can only state a few platitudes: True existence (τὸ ὄντως ὄν) is to be found in the world of ideas; it can be understood only by the mind. Knowledge is therefore memory

(ἀνάμνησις), and our sensory perceptions help us, as imperfect shadow images, in our understanding of the primary ideas hidden behind them. Thus concealed behind the imperfect, poorly represented triangle is the idea of the triangle, and it is to this ideal, which cannot be realized in full by any actual triangle, to which the theorem that the sum of the internal angles is 180 degrees refers. The manifestations of the world of our senses should be described as well as possible ("phenomena must be rescued"); one must not assign too great a significance to the deviations between the theoretical construction and reality (Quotation 1.17). But we would do better to let Plato explain his ideas in his own words.

1.2.4 *Plato on Insight and Ideas*

Socrates: The soul by its nature is immortal; having been born again many times there is nothing it has not already seen here and in the underworld. It is therefore no wonder that it can remember virtue and the other ideas that it has seen. Because the whole world was created from the same essence, and the soul has seen the whole, nothing can prevent it start from a single recollection, that people call "learning", and from it figure out everything else with determined and strenuous work. This is why all enquiry and learning is nothing other than recollection. We must not listen to the sophistry that says that enquiry is impossible because it would make us lazy and it flatters the idle as my argument make us active and inquisitive. This is my belief and now I will gladly enquire with you into the nature of virtue.

Meno: Yes, Socrates, but what does it mean that we do not learn, and that what we call learning is recollection? Can you teach me how this is so?

Socrates: As I just told you, Meno, you are crafty. You now ask me if I can teach you, when I am saying that there is no teaching but recollection, in order to make me contradict myself. [translated by Bela Hamvas]

—Plato, *Meno*

And now, I said, let me show in a figure how far our nature is enlightened or unenlightened: Behold human beings living in a underground den, which has a mouth open towards the light and reaching all along the den; here they have been from their childhood, and have their legs and necks chained so that they cannot move, and can only see before them, being prevented by the chains from turning round their heads. Above and behind them a fire is blazing at a distance, and between the fire and the prisoners there is a raised way; and you will see, if you look, a low wall built along the way, like the screen which marionette players have in front of them, over which they show the puppets.

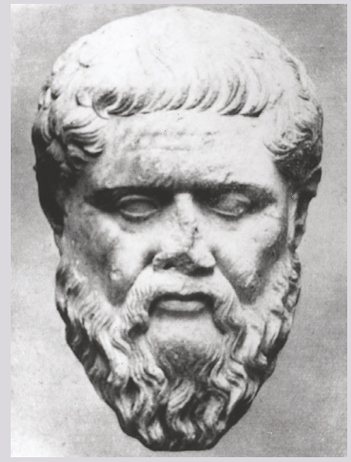
I see.

And do you see, I said, men passing along the wall carrying all sorts of vessels, and statues and figures of animals made of wood and stone and various materials, which appear over the wall? Some of them are talking, others silent.

You have shown me a strange image, and they are strange prisoners.

Like ourselves, I replied; and they see only their own shadows, or the shadows of one another, which the fire throws on the opposite wall of the cave?

True, he said; how could they see anything but the shadows if they were never allowed to move their heads?



◀ Figure 1.43
Plato (427–347 bce), student of Socrates and teacher of Aristotle. At the age of 20, he went to Athens to study under Socrates. After his teacher's death, he traveled throughout the known world. In southern Italy, he became familiar with the teachings of the Pythagoreans, and he likely visited Egypt as well. On his return, he founded a school in the grove of the mythological hero Akademos (whence the name "Academy"). In 368 bce, he attempted to realize his political ideas in Syracuse, at the court of his pupil Dionysius II. In this he was unsuccessful. Plato wrote his works in the form of dialogues. In almost all of these, the principal interlocutor is Socrates (an exception is Plato's last work, *The Laws*). *Republic*, written in 374 bce, gives perhaps the broadest overview of Plato's philosophy. His thoughts on natural philosophy are to be found in *Timaeus*.

After Plato's death, his nephew Speusippus became the director of the Academy. The astronomer Heraclides Ponticus also studied there.

Quotation 1.14

Pupil: I'll tell you, but you've got to consider these matters holy secrets. Just now Socrates asked Chaerephon how many of its own feet a flea can jump, because one had bitten Chaerephon's eyebrow and jumped off onto Socrates' head.

Strepsiades: And how did he measure it off?

Pupil: Very cleverly. He melted some wax, then picked up the flea and dipped both its feet in the wax, and then when the wax cooled the flea had Persian slippers stuck to it. He took these off and went about measuring the distance.

Strepsiades: Lord Zeus, what subtlety of mind!

—Aristophanes, *Clouds*

Quotation 1.15

Strepsiades: ... But now explain this: where does the lightning bolt come from, blazing with fire, that incinerates us on contact and badly burns the survivors? It's quite obvious that Zeus hurls it against perjurers.

Socrates: How's that, you moron redolent of the Cronia, you moon-calf! If he really strikes perjurers, then why hasn't he burned up Simon or Cleonymus or Theorus, since they're paramount perjurers? On the other hand, he strikes his own temple, and Sunium headland of Athens, and the great oaks. What's his point? An oak tree certainly doesn't perjure itself!

Strepsiades: I don't know; but you seem to have a good argument. Very well, what is the thunderbolt, then?

Socrates: When a dry wind rises skyward and gets locked up in these Clouds, it blows them up from within like a bladder, and then by natural compulsion it bursts them and is borne out in a whoosh by dint of compression, burning itself up with the friction and velocity.

Strepsiades: By Zeus, exactly the same thing happened to me one time at the Diasia, when I was cooking a haggis for my relatives and forgot to make a slit. So it bloated up, then suddenly it exploded, spattering gore in my eyes and burning my face.

—Aristophanes, *Clouds*

Quotation 1.16

Adeimantus: No one would be able to contradict the things you've said, Socrates, but on each occasion that you say them, your hearers are affected in some such way as this. They think that, because they're inexperienced in asking and answering questions, they're led astray a little bit by the argument at every question and that, when these little bits are added together at the end of the discussion, great is their fall, as the opposite of what they said at the outset comes to light. Just as inexperienced checkers players are trapped by the experts in the end and can't make a move, so they too are trapped in the end and have nothing to say in this different kind of checkers, which is played not with disks but with words. Yet the truth isn't affected by this outcome. I say this with a view to the present case, for someone might well say now that he's unable to oppose you as you ask each of

continued on next page

And of the objects which are being carried in like manner they would only see the shadows?

Yes, he said.

And if they were able to converse with one another, would they not suppose that they were naming what was actually before them?

Very true.

And suppose further that the prison had an echo which came from the other side, would they not be sure to fancy when one of the passers-by spoke that the voice which they heard came from the passing shadow?

No question, he replied.

To them, I said, the truth would be literally nothing but the shadows of the images.

That is certain.

And now look again, and see what will naturally follow it; the prisoners are released and disabused of their error. At first, when any of them is liberated and compelled suddenly to stand up and turn his neck round and walk and look towards the light, he will suffer sharp pains; the glare will distress him, and he will be unable to see the realities of which in his former state he had seen the shadows; and then conceive someone saying to him, that what he saw before was an illusion, but that now, when he is approaching nearer to being and his eye is turned towards more real existence, he has a clearer vision; what will be his reply? And you may further imagine that his instructor is pointing to the objects as they pass and requiring him to name them; will he not be perplexed? Will he not fancy that the shadows which he formerly saw are truer than the objects which are now shown to him?

Far truer.

And if he is compelled to look straight at the light, will he not have a pain in his eyes which will make him turn away to take and take in the objects of vision which he can see, and which he will conceive to be in reality clearer than the things which are now being shown to him?

True.

And suppose once more, that he is reluctantly dragged up a steep and rugged ascent, and held fast until he's forced into the presence of the sun himself, is he not likely to be pained and irritated? When he approaches the light his eyes will be dazzled, and he will not be able to see anything at all of what are now called realities.

Not all in a moment, he said.

He will require to grow accustomed to the sight of the upper world. And first he will see the shadows best, next the reflections of men and other objects in the water, and then the objects themselves; then he will gaze upon the light of the moon and the stars and the spangled heaven; and he will see the sky and the stars by night better than the sun or the light of the sun by day?

Certainly.

Last of he will be able to see the sun, and not mere reflections of him in the water, but he will see him in his own proper place, and not in another; and he will contemplate him as he is.

Certainly.

He will then proceed to argue that this is he who gives the season and the years, and is the guardian of all that is in the visible world, and in a certain way the cause of all things which he and his fellows have been accustomed to behold?

Clearly, he said, he would first see the sun and then reason about him.

And when he remembered his old habitation, and the wisdom of the den and his fellow-prisoners, do you not suppose that he would felicitate himself on the change, and pity them?

Certainly, he would.

And if they were in the habit of conferring honors among themselves on those who were quickest to observe the passing shadows and to remark which of them went before, and which followed after, and which were together; and who were therefore best able to draw conclusions as to the future, do you think that he would care for such honors and glories, or envy the possessors of them? Would he not say with Homer, "Better to be the poor servant of a poor master, and to endure anything, rather than think as they do and live after their manner?"

Yes, he said, I think that he would rather suffer anything than entertain these false notions and live in this miserable manner.

Imagine once more, I said, such an one coming suddenly out of the sun to be replaced in his old situation; would he not be certain to have his eyes full of darkness?

To be sure, he said.

And if there were a contest, and he had to compete in measuring the shadows with the prisoners who had never moved out of the den, while his sight was still weak, and before his eyes had become steady (and the time which would be needed to acquire this new habit of sight might be very considerable) would he not be ridiculous? Men would say of him that up he went and down he came without his eyes; and that it was better not even to think of ascending; and if any one tried to loose another and lead him up to the light, let them only catch the offender, and they would put him to death.

No question, he said.

This entire allegory, I said, you may now append, dear Glaucon, to the previous argument; the prison-house is the world of sight, the light of the fire is the sun, and you will not misapprehend me if you interpret the journey upwards to be the ascent of the soul into the intellectual world according to my poor belief, which, at your desire, I have expressed whether rightly or wrongly God knows. But, whether true or false, my opinion is that in the world of knowledge the idea of good appears last of all, and is seen only with an effort; and, when seen, is also inferred to be the universal author of all things beautiful and right, parent of light and of the lord of light in this visible world, and the immediate source of reason and truth in the intellectual; and that this is the power upon which he who would act rationally, either in public or private life must have his eye fixed.

I agree, he said, as far as I am able to understand you.

Moreover, I said, you must not wonder that those who attain to this beatific vision are unwilling to descend to human affairs; for their souls are ever

Quotation 1.16, ***continued***

your questions, yet he sees that of all those who take up philosophy—not those who merely dabble in it while still young in order to complete their upbringing and then drop it, but those who continue in it for a longer time—the greatest number become cranks, not to say completely vicious, while those who seem completely decent are rendered useless to the city because of the studies you recommend.

Socrates: Do you think that what these people say is false?

Adeimantus: I don't know, but I'd be glad to hear what you think.

Socrates: You'd hear that they seem to me to speak the truth.

—Plato, *Republic,* Book VI, 487b-d [p. 111]

Quotation 1.17

Let us now face the question, suppose a discrepancy had appeared, well confirmed and substantiated, between the theory and the observations. How should one react to it? How would Einstein himself have reacted to it? Should one then consider the theory to be basically wrong? I would say that the answer to the last question is emphatically No. The Einstein theory of gravitation has a character of excellence of its own. Anyone who appreciates the fundamental harmony connecting the way Nature runs and general mathematical principles must feel that a theory with the beauty and elegance of Einstein's theory has to be substantially correct. If a discrepancy should appear in some application of the theory, it must be caused by some secondary feature relating to this application which has not been adequately taken into account, and not by a failure of the general principles of the theory.

One has a great confidence in the theory arising from its great beauty, quite independent of its detailed successes. It must have been such confidence in the essential beauty of the mathematical description of Nature which inspired Einstein in his quest for a theory of gravitation.

—P. A. M. Dirac, "The Excellence of Einstein's Theory of Gravitation," 1978

Quotation 1.18

Nothing can change into something absolutely different.

The universe is infinite because it has not been produced by a creator. The causes of what now exists had no beginning ...

Among my contemporaries, I am the one who has traveled most about the Earth, where I have investigated the most widely and seen the most lands and parts of the heavens and heard the most learned men. And none has surpassed me in putting together lines with proofs, not even the so-called rope-knotters (land surveyors) the Egyptians. With these I have been after all five(?) years together on foreign soil. ...

I would rather discover a single (geometric) proof than gain the Persian throne. ...

Sweet exists by convention, bitter by convention, color by convention. Atoms and void alone exists in reality ...

In reality, however, we know nothing; for truth lies deep... We, however, understand in truth nothing certain, rather only what changes, according to the perception of our bodies and the atoms flowing within them or opposing them. ...

Culture is better than wealth ... No power and no treasure can compensate the expansion of our knowledge.

—Democritus, *Fragments*

hastening into the upper world where they desire to dwell; which desire of theirs is very natural, if our allegory may be trusted.

Yes, very natural.

And is there anything surprising in one who passes from divine contemplations to the evil state of man, misbehaving himself in a ridiculous manner; if, while his eyes are blinking and before he has become accustomed to the surrounding darkness, he is compelled to fight in courts of law, or in other places, about the images or the shadows of images of justice, and is endeavouring to meet the conceptions of those who have never yet seen absolute justice?

Anything but surprising, he replied.

Anyone who has common sense will remember that the bewilderments of the eyes are of two kinds, and arise from two causes, either from coming out of the light or from going into the light, which is true of the mind's eye, quite as much as of the bodily eye; and he who remembers this when he sees any one whose vision is perplexed and weak, will not be too ready to laugh; he will first ask whether that soul of man has come out of the brighter light, and is unable to see because unaccustomed to the dark, or having turned from darkness to the day is dazzled by excess of light. And he will count the one happy in his condition and state of being, and he will pity the other; or, if he have a mind to laugh at the soul which comes from below into the light, there will be more reason in this than in the laugh which greets him who returns from above out of the light into the den.

—Plato, *The Republic*, Book VII

1.3 Matter and Motion: The Aristotelian Synthesis

1.3.1 *Atoms and Elements*

From the very beginning, Greek scholars attempted to provide answers to the most difficult questions, such as, What is the world made of? What is the fundamental principle that holds the world together, the fundamental principle of all existence (ἀρχή τών ὄντων)? The earliest attempts at answering these questions—for example, that of Thales (ca. 625–ca. 547 BCE) claiming that water is the material basis of all existence, meaning that all that exists is created from water and eventually becomes water again—give us the impression that science still had a long way to go before being able even to pose such questions properly, let alone to try to answer them. The questions and answers of the Ionian natural philosophers, among whom Thales is numbered, nevertheless contain a decisive element of scientific rigor in the assertion that nature should be understood and interpreted on its own terms, without the intervention of supernatural powers or mythological figures.

Anaximander (ca. 610–ca. 546 BCE), an Ionian natural philosopher who followed Thales, no longer begins with a substance as the foundation of matter, but calls such a foundation *apeiron* (τὸ ἄπειρον), something boundless and formless, from which the concrete things of our world are created. Anaximenes (ca. 585–ca. 525 BCE) returns to a single substance, the air, as the fundamental constituent of matter. It is understandable that Heraclitus (ca. 535–475 BCE), whose motto was "everything flows," found the fundamental principle (*archê*) in fire, while Parmenides, who rejected all change, speaks of the world and existence as a homogeneous sphere.

We first meet the Aristotelian four elements with Empedocles (ca. 490–430 BCE), according to whom the diversity of our world emerges from the composition in

various proportions of earth, air, water, and fire, compositions which are also subject to decomposition.

We should not forget that the branch of physics whose subject is the structure of matter was very closely connected to philosophy. With all these theories, the object was to answer *philosophical* questions about existence, constancy, and change. Viewed from today's standpoint, or rather from the standpoint of the nineteenth century, Democritus (ca. 460–ca. 370 BCE) was, with his atomic theory, best able to describe constancy in change, that is, the possibility of change in a context of permanence. There are historians of science who view Democritus's atomism as a premature and immature intellectual adventure and who support the critique presented by Aristotle—on account of which Democritus's theory remained unnoticed for two thousand years. Others, however, see in it one of the most brilliant germs of an idea in the histories of both physics and philosophy, out of which arose our modern scientific worldview based on atomic theory.

The philosophical basis of atomism can best be understood if we look at its logical kinship to the theories that were already in existence at the time. The Eleatics (Parmenides and his school) with their paradoxes gave arguments that sounded convincing that all change is logically absurd. They thus arrived at a model of the world in the form of a homogeneous sphere, which is in turn absurd to our senses. Democritus accepted—in philosophical terminology—the nonexistent as existing, or, expressed in physical terms, that the emptiness (τό κένον) or the vacuum exists. With this, the homogeneous world of Parmenides becomes decomposable, since parts of the whole now had a place to be in space. The world model of Democritus consists only of atoms, whose substance is completely homogeneous, and the space between them. The diversity of the world is the result of the various connections between the atoms and of their motions. The atoms themselves differ in their forms, sizes, positions, and velocities (Quotations 1.18 and 1.19).

The atomism of Democritus is strikingly similar to the picture presented by the kinetic theory of gases. In relation to philosophy, several conclusions arise. One is expressed by the following fragment from his writings: nothing arises from nothing and nothing can be destroyed to become nothing. Accepting the existence of space gives credence to the idea that the world is infinite. In fact, Democritus states that formally and unambiguously, for the first time in the history of science, and with a surprising justification: The world is infinite because it was not created by an otherworldly power. Democritus also confronted epistemological questions. In his view, valid insights are transmitted to us from the external world because fine atomic layers (εϊδωλα, *eidola*) that represent the images of things become detached from those things and then find their way to our sensory organs. Thus, our sensory impressions give us real knowledge, but that knowledge can be partly deceiving and in part understandable only with the help of conventions.

Democritus also considered the relationship between the body and the soul. He assumed that the soul also consists of atoms that are particularly smooth and reside between the atoms of the body. After the decomposition of the body, the atoms of the soul scatter—the same way as the atoms of the body—and in this manner the soul ceases to exist, the soul decomposes, just as the body decomposes.

Democritus's atomism was propagated, explained, and further developed by Lucretius. In his didactic poem *De rerum natura* (On the Nature of Things), he describes, with some poetic license, vividly and in great detail, how with our senses we can perceive the teeming swarms of atoms as macroscopic bodies that are perhaps at rest or are moving slowly (see Section 1.5). It is a historical irony that

Quotation 1.19

Leucippus, however, and his disciple Democritus hold that the elements are the Full and the Void – calling the one "what is" and the other "what is not." Of these they identify the full or solid with "what is," and the void or rare with "what is not" (hence they hold that what is not is no less real than what is, because Void is as real as Body); and they say that these are the material causes of things. And just as those who make the underlying substance a unity generate all other things by means of its modifications, assuming rarity and density as first principles of these modifications, so these thinkers hold that the "differences" are the causes of everything else. These differences, they say, are three: *shape, arrangement,* and *position;* because they hold that what is differs only in contour, inter-contact, and inclination. (Of these contour means shape, inter-contact arrangement, and inclination position.) Thus, e.g., A differs from N in shape, AN from NA in arrangement, and Z from N in position. As for motion, whence and how it arises in things, they casually ignored this point, very much as the other thinkers did. Such, then, as I say, seems to be the extent of the inquiries which the earlier thinkers made into these two kinds of cause.

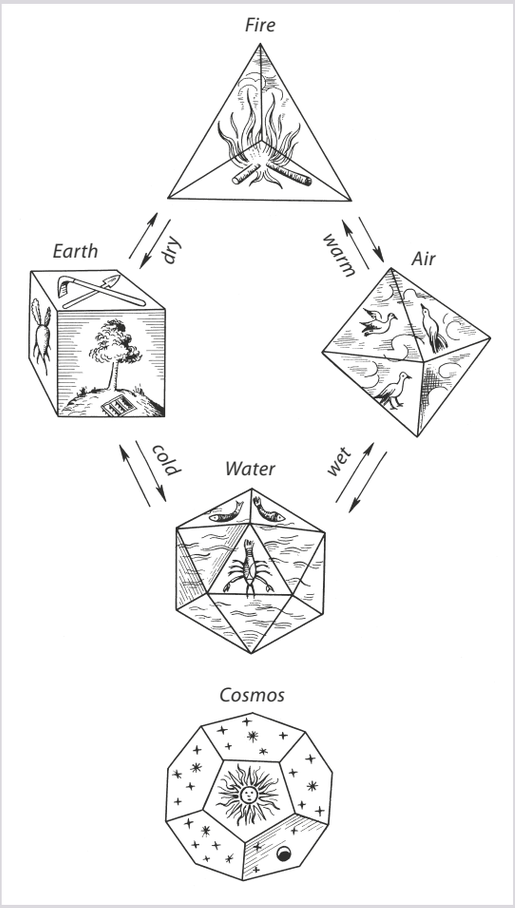
—Aristotle, *Metaphysics,* 985b [Tredennick]

Quotation 1.20

"In the beginning was symmetry" is certainly a better expression than Democritus' "In the beginning was the particle." Elementary particles embody symmetries; they are their simplest representations, and yet they are merely their consequence. Accident came later on in the development of the universe, but it, too, fits neatly into the original forms; it satisfies the statistical laws of quantum theory....

Our elementary particles are comparable to the regular bodies of Plato's *Timaeus*. They are the original models, the ideas of matter. Nucleic acid is the idea of the living being. These primitive models determine all subsequent developments. They are representative of the central order. And though accident does play an important part in the subsequent emergence and development of a profusion of structures, it may well be that accident, too, is somehow related to the central order.

—Werner Heisenberg, *Physics and Beyond: Encounters and Conversations* [pp. 240–241]



▲ **Figure 1.44** Association of the four Platonic solids with the four elements of Parmenides. The fifth Platonic solid represents the idea of the cosmos (after Kepler).

New groups were appended to this group of four over the course of time, with repercussions detectable to this day. Hippocrates founded, and Galen refined, the idea of "humoral pathology": Health is determined by the proper proportions (*eukrasia*) of the four humors. Sickness is caused by a disturbance in these proportions (*dyskrasia*). The four humors are blood, phlegm, yellow bile, and black bile. The differences in human temperaments can be traced to an excess of the humors: sanguine (*sanguis* = blood), choleric (*chole* = bile, anger, hate), melancholy (*melancholia* = black bile), phlegmatic (*phlegma*; fire, inflammation, phlegm causing inflammation).

We thus arrive at the following groups of four:

octahedron	air	warm–wet
tetrahedron	fire	warm–dry
hexahedron	earth	cold–dry
icosahedron	water	cold–wet

spring	blood	sanguine
summer	yellow bile	choleric
autumn	black bile	melancholic
winter	phlegm	phlegmatic

continued on next page

the Western world completely ignored the atomic theory until as late as Gassendi (1592–1655)—even though Lucretius's poem, written in Latin, did not need to be transmitted from the Islamic world—in part because of the poem's atheistic point of view, and in large measure because the Aristotelian tradition—even while assigning a certain value to it—rejected the atomic theory.

Plato approaches the problem of constancy and change from quite a different point of view. He projects *all* true knowledge, including truths relating to the structure of matter, into the realm of ideas. The four elements of Empedocles that appear to us through our senses—earth, water, air, and fire—assume abstract forms in the world of ideas that are as perfect as possible. Thus the four elements are the shadow images of their underlying ideals, that is, four regular solids, and these images are what our senses recognize. The tetrahedron is the ideal behind fire, while the hexahedron (cube) belongs to earth, the octahedron to air, and the icosahedron to water. The fifth regular solid, the dodecahedron, assumes the role of the ideal underlying the structure of the cosmos (Figure 1.44). This association of regular solids with the four elements postulated by Aristotle and Empedocles held sway for over two thousand years. Figure 1.44, for example, comes from a book by Kepler. Plato further subdivided the elements. He realized that their faces can be assembled from various right triangles: for the tetrahedron, octahedron, and icosahedron, one needs, for each face, six triangles with an angle of 30°, and for the cube, one needs four triangles with two 45° angles (Figures 1.45 and 1.46).

In the abstract world of ideas, relationships can be found among the various elements. Thus eight equilateral triangles can form an atom of air, but also two atoms of fire, which might be expressed by a formula as $A = F_2$. From the twenty equilateral triangles of an icosahedron, representing the idea of water, we could make two air atoms and one fire atom, and thereby obtain the formula $W = A_2F$. As absurd as these ideas may seem, they contain, like all of Plato's ideas, a kernel of truth, or at least one leading to truth. In this case, we may see the kernel of truth in the attempt to find an abstract model of reality that makes a description or characterization using numerical proportions possible. The path leading from here to the formula H_2O for water will be very long indeed, but one can at least say that the path began with Plato. The idea that the structural elements of matter could be imagined in terms of regular solids or in terms of the shapes of their faces is not at all absurd in contemporary particle physics. One of the most important, and indeed macroscopically detectable, properties of fundamental particles is their symmetry. Since regular solids can also be characterized by their symmetry properties, they can serve here as very practical examples. Heisenberg himself emphasized the close relationship between the fundamental ideas of his field theory and the ideas of Plato (Quotations 1.20 and 1.21).

Aristotle maintained the four elements of Empedocles. With each of the elements earth, water, air, and fire, he associated two characteristics from the pairings dry–wet and cold–warm. Thus, as can be seen in Figure 1.44, earth is dry and cold; fire, dry and warm; air, wet and warm; and water, wet and cold. According to Aristotle, the elements can be transformed into one another under certain conditions, and such a transformation is easiest between two elements that possess a common characteristic. In the figure, these simple or natural transformations are indicated by arrows.

In the real world, these four elements exist in various combinations, or mixtures (*mixtio*, μῖξις). Aristotle left much room for the opinions of later commenta-

tors regarding the properties of elements in mixtures, that is, with regard to the question of whether the elements retain their original properties in mixtures or form different substances altogether. Aristotelian chemistry assumed that matter is infinitely divisible and that every part after a division will have the same structure as the whole had prior to the division. The limits of divisibility, the *minima naturalia*, appear at the time of the Late Scholastics, but they fail to bring the Aristotelian theory any closer to an atomic theory, even to the slightest degree.

1.3.1.1 Plato and the "Elementary Particles"

First of all, everyone knows, I'm sure, that fire, earth, water and air are bodies. Now everything that has bodily form also has depth. Depth, moreover, is of necessity comprehended within surface, and any surface bounded by straight lines is composed of triangles. Every triangle, moreover, derives from two triangles, each of which has one right angle and two acute angles. Of these two triangles, one [the isosceles right-angled triangle] has at each of the other two vertices an equal part of a right angle, determined by its division by equal sides; while the other [the scalene triangle] has unequal parts of a right angle at its other two vertices, determined by the division of the right angle by unequal sides. This, then, we presume to be the originating principle of fire and of the other bodies, as we pursue our likely account in terms of Necessity. Principles yet more ultimate than these are known only to the god, and to any man he may hold dear.

We should now say which are the most excellent four bodies that can come to be. They are quite unlike each other, though some of them are capable of breaking up and turning into others and vice-versa. If our account is on the mark, we shall have the truth about how earth and fire and their proportionate intermediates [water and air] came to be. For we shall never concede to anyone that there are any visible bodies more excellent than these, each conforming to a single kind. So we must wholeheartedly proceed to fit together the four kinds of bodies of surpassing excellence, and to declare that we have come to grasp their natures well enough.

Of the two [right-angled] triangles, the isosceles has but one nature, while the scalene has infinitely many. Now we have to select the most excellent one from among the infinitely many, if we are to get a proper start. So if anyone can say that he has picked out another one that is more excellent for the construction of these bodies, his victory will be that of a friend, not an enemy. Of the many [scalene right-angled] triangles, then, we posit as the one most excellent, surpassing the others, that one from [a pair of] which the equilateral triangle is constructed as a third figure. Why this is so is too long a story to tell now. But if anyone puts this claim to the test and discovers that it isn't so, his be the prize, with our congratulations. So much, then, for the selection of the two triangles out of which the bodies of fire and the other bodies are constructed—the [right-angled] isosceles, and [right-angled] scalene whose longer side squared is always triple its shorter side squared [i.e., the half-equilateral].

At this point, we need to formulate more precisely something that was not stated clearly earlier. For then it appeared that all four kinds of bodies could turn into one another by successive stages. But the appearance is wrong. While there are indeed four kinds of bodies that come to be from the [right-angled] triangles we have selected, three of them come from triangles that have un-

Figure 1.44 *continued*

Below, we quote from the *Labyrinthus medicorum* (1538) of Paracelsus to show the first uncertain steps from humoral pathology to iatrochemistry (that is, to the idea that life processes and methods of cure have a chemical basis).

And such the physician should understand of the cause that he not indicate the qualtates and humores: but the elements as mothers and their procreationes as species, not [as] humores. Not that one should say: Cuius humoris? Melancholici: yet melancholia is nothing other than a mad, senseless phantastica illness, not a pillar of the four [elements]. Also do not say: Cuius Complexionis? Cholericae, but Calidae Sectae. Now, Cholera is not one of the four pillars, but a Morbus of all types of expulsion. Thus also: Cuius Qualitatis? Sanguineae: thus Sanguis is not one of the four pillars, but the Corpus Venarum, like the wine in a barrel. Thus also not: Cuius Naturae? Phlegmaticae: now, is not Phlegma snot from the nose, what has it to do with the stomach? But Cuius Elementi? Aquae, Terrae, Ignis, Aeris: now the answer arises: from what element does sickness arise? From fire, not Cholera: from earth, not Melancholia: from water, not Phlegmate: from air, not Sanguine. So it would appear. And do not say: that is Melancholicum, because neither heaven nor earth knows about melancholy. Do not say: that is Cholera, Phlegma, Sanguis, etc. So it is that nature knows nothing of its process and order. So as the physician becomes acquainted with the element, so will he find in the Generatis all illnesses of which mankind suffers.

We can also mention the following noteworthy parallel. H. J. Eysenck (1916–1997) finds the precursors of his model of personality in Hippocrates. Based on his statistical investigations, Eysenck maintained that individuals can be grouped in the categories stable/unstable and introvert/extrovert. Thus he arrives at the following four groups, where we have added the corresponding Hippocratic types:

stable–extrovert	→	sanguine
unstable–extrovert	→	choleric
unstable–introvert	→	melancholic
stable–introvert	→	phlegmatic

Quotation 1.21

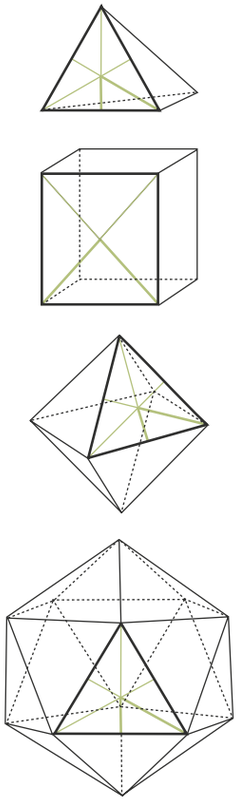
This attempt to assign geometrical figures to the simple bodies is on all counts irrational. In the first place, the whole of space will not be filled up. Among surfaces it is agreed that there are three figures which fill the place that contains them—the triangle, the square, and the hexagon: among solids only two, the pyramid and the cube. But they need more than these, since they hold that the elements are more. Secondly, the shape of all the simple bodies is observed to be determined by the place in which they are contained, particularly in the case of water and air. The shape of the element therefore cannot survive, or it would not be everywhere in contact with that which contains the whole mass. But if its shape is modified, it will no longer be

continued on next page

Quotation 1.21, ***continued***

water, since its shape was the determining factor. Clearly therefore the shapes of the elements are not defined. Indeed it seems as if nature itself here shows us the truth of a conclusion to which more abstract reasoning also points. Here, as in everything else, the underlying matter must be devoid of form and shape, for so, as is said in the *Timaeus*, the "receiver of all" will be best able to submit to modification. It is like this that we must conceive of the elements, as the matter of their compounds, and this is why they can change into each other, and lose their qualitative differences.

—Aristotle, *On the Heavens* [Guthrie, p. 319]



▲ **Figure 1.45** The four Platonic solids associated with the four elements together with the "elementary particles" that constitute them. Plato considered the possibility of imagining the regular solids, that is, the elements, as built up from two types of triangles, that is, from still simpler elements. The triangles of the first kind are isosceles right triangles, whereas those of the second type are right triangles whose hypotenuse is twice the length of the shorter leg, that is, triangles obtained by halving an equilateral triangle. The regular solids correspond in a way to our chemical elements; the triangles, to the elementary particles.

continued on next page

equal sides, whereas the fourth alone is fashioned out of isosceles triangles. Thus not all of them have the capacity of breaking up and turning into one another, with a large number of small bodies turning into a small number of large ones and vice-versa. There are three that can do this. For all three are made up of a single type of triangle, so that when once the larger bodies are broken up, the same triangles can go to make up a large number of small bodies, assuming shapes appropriate to them. And likewise, when numerous small bodies are fragmented into their triangles, these triangles may well combine to make up some single massive body belonging to another kind.

So much, then, for our account of how these bodies turn into one another. Let us next discuss the form that each of them has come to have, and the various numbers that have combined to make them up. Leading the way will be the primary form [the tetrahedron], the tiniest structure, whose elementary triangle is the one whose hypotenuse is twice the length of its shorter side. Now when a pair of such triangles are juxtaposed along the diagonal [i.e., their hypotenuses] and this is done three times, and their diagonals and short sides converge upon a single point as center, the result is a single equilateral triangle, composed of six such triangles. When four of these equilateral triangles are combined, a single solid angle is produced at the junction of these plane angles. This, it turns out, is the angle which comes right after the most obtuse of the plane angles. And once four such solid angles have been completed, we get the primary solid form, which is one that divides the entire circumference [sc. of the sphere in which it is inscribed] into equal and similar parts.

The second solid form [the octahedron] is constructed out of the same triangles which, however, are now arranged in eight equilateral triangles and produce a single solid angle out of four plane angles. And when six such solid angles have been produced, the second body has reached its completion.

Now the third body [the icosahedron] is made up of a combination of one hundred and twenty of the elementary triangles, and of twelve solid angles, each enclosed by five plane equilateral triangles. This body turns out to have twenty equilateral triangular faces. And let us take our leave of this one of the elementary triangles, the one that has begotten the above three kinds of bodies and turn to the other one, the isosceles [right-angled] triangle, which has begotten the fourth [the cube]. Arranged in sets of four whose right angles come together at the center, the isosceles triangle produced a single equilateral quadrangle [i.e., a square]. And when six of these quadrangles were combined together, they produced eight solid angles, each of which was constituted by three plane right angles. The shape of the resulting body so constructed is a cube, and it has six quadralinear equilateral faces....

Let us now assign to fire, earth, water and air the structures which have just been given their formations in our speech. To earth let us give the cube, because of the four kinds of bodies earth is the most immobile and the most pliable—which is what the solid whose faces are the most secure must of necessity turn out to be, more so than the others. Now of the [right-angled] triangles we originally postulated, the face belonging to those that have equal sides has a greater natural stability than that belonging to triangles that have unequal sides, and the surface that is composed of the two triangles, the equilateral quadrangle [the square], holds its position with greater stability than does the equilateral triangle, both in their parts and as wholes. Hence, if we assign this solid figure to earth, we are preserving our "likely account." And of the solid figures that are left, we shall next assign the least

mobile of them to water, to fire the most mobile, and to air the one in between. This means that the tiniest body belongs to fire, the largest to water, and the intermediate one to air—and also that the body with the sharpest edges belongs to fire, the next sharpest to air, and the third sharpest to water. Now in all these cases the body that has the fewest faces is of necessity the most mobile, in that it, more than any other, has edges that are the sharpest and best fit for cutting in every direction. It is also the lightest, in that it is made up of the least number of identical parts. The second body ranks second in having these same properties, and the third ranks third. So let us follow our account, which is not only likely but also correct, and take the solid form of the pyramid that we saw constructed as the element or the seed of fire. And let us say that the second form in order of generation is that of air, and the third that of water.

Now we must think of all these bodies as being so small that due to their small size none of them, whatever their kind, is visible to us individually. When, however, a large number of them are clustered together, we do see them in bulk. And in particular, as to the proportions among their numbers, their motions and their other properties, we must think that when the god had brought them to complete and exact perfection (to the degree that Necessity was willing to comply obediently), he arranged them together proportionately.

Given all we have said so far about the kinds of elemental bodies, the following account [of their transformations] is the most likely: When earth encounters fire and is broken up by fire's sharpness, it will drift about—whether the breaking up occurred within fire itself, or within a mass of air or water—until its parts meet again somewhere, refit themselves together and become earth again. The reason is that the parts of earth will never pass into another form. But when water is broken up into parts by fire or even by air, it could happen that the parts recombine to form one corpuscle of fire and two of air. And the fragments of air could produce, from any single particle that is broken up, two fire corpuscles. And conversely, whenever a small amount of fire is enveloped by a large quantity of air or water or perhaps earth and is agitated inside them as they move, and in spite of its resistance is beaten and shattered to bits, then any two fire corpuscles may combine to constitute a single form of air. And when air is overpowered and broken down, then two and one half entire forms of air will be consolidated into a single, entire form of water.

Let us recapitulate and formulate our account of these transformations as follows: Whenever one of the other kinds is caught inside fire and gets cut up by the sharpness of fire's angles and edges, then if it is reconstituted as fire, it will stop getting cut. The reason is that a thing of any kind that is alike and uniform is incapable of effecting any change in, or being affected by, anything that is similar to it.

—Plato, *Timaeus* [pp. 1256–1259]

1.3.2 *Motion under Terrestrial Conditions: Peripatetic Dynamics*

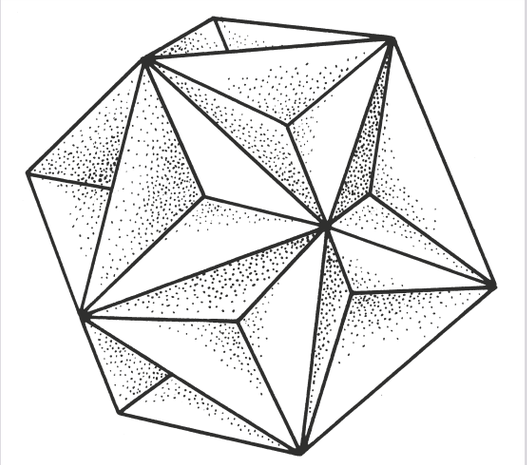
Dynamics and the theory of the structure of matter are the two branches of science in which, behind everyday manifestations, the fundamental nature of the phenomena that can be reduced to a mathematical model is most deeply hidden. The puzzles here cannot be solved on the basis of immediate experience, not even the sharpest observation and cleverest thinking. Instead, experiments have to be

Figure 1.45 *continued*

As mentioned earlier, the Pythagoreans knew that there were *at least* five regular solids; the proof that there are *exactly* five is due to Plato's friend Theaetetus, who appears as an interlocutor in the *Dialogues*. In his *Elements*, Euclid gives a proof that relies on a previously proved theorem, that the sum of the dihedral angles at the vertex of a solid must be less than 360°. If we now wish to obtain a regular solid whose faces are equilateral triangles, we may have three, four, or at most five such triangles at each vertex. Thus arise the tetrahedron, octahedron, and icosahedron. With squares as faces, there must be three—neither more nor less—and we obtain the cube. Finally, we may have three regular pentagons at a vertex, giving us the dodecahedron.

Today, one may obtain the proof from a 1758 theorem of Euler: The numbers v of vertices, e of edges, and f of faces in a convex solid are related by the formula

$$v - e + f = 2.$$



▲ **Figure 1.46** Euclid mentions only convex regular solids. In 1809, Poinsot constructed four regular nonconvex solids. Then in 1811, Cauchy proved that there are exactly four such solids. The figure shows one of these (from the *Great Soviet Encyclopedia*).

Archimedes considered semiregular solids, that is, solids composed of regular polygons, but not all identical. Plate XIII shows one of these *Archimedean solids*, the rhombicuboctahedron, composed of eight triangles and eighteen squares (on the left in Pacioli's painting).

In our time, the Archimedean solid composed of twelve regular pentagons and twenty regular hexagons—the truncated icosahedron—has acquired a particular significance (see also Figure 2.66). Sixty carbon atoms arranged at the vertices of this solid yield a stable carbon molecule, called buckminsterfullerene or (C_{60}-I_h)[5,6]fullerene. It is named for the architect Buckminster Fuller. Its discoverers were awarded the 1996 Nobel Prize in chemistry.

Buckminsterfullerene is also called the "soccer-ball molecule," since the stitching pattern on many of the world's soccer balls is that of this Archimedean solid.

Quotation 1.22

The Pythagoreans ... did not accept, in things divine and eternal, such disorder as [the heavenly bodies] moving sometimes more quickly, sometimes more slowly, and sometimes standing still. ... One would not accept such anomaly of movement in the goings of an orderly and well-mannered man. The business of life is often the cause of slowness or of swiftness for men. But in the case of the incorruptible nature of the stars, it is not possible to adduce any cause of swiftness or slowness. For this reason, they put forward the question: how would the phenomena be accounted for by means of uniform and circular motions?

—Geminos of Rhodes, *Introduction to the Phenomena* [pp. 117–118]

Quotation 1.23

It is instructive to see how the ideas in Quotation 1.22 are later echoed in the writings of the great Jewish philosopher Maimonides (1135–1204):

That the sphere is endowed with a soul is clear upon reflection. However, he who hears this may deem this a matter that is difficult to grasp or may regard it as impossible because of his imagining that when we say, "endowed with a soul," the soul referred to is like the soul of a man, or an ass and a bull. Now this is not the meaning of the dictum. This meaning is rather that the local motion of the sphere is a proof of there indubitably being in it a principle in virtue of which it is moved. And this principle is undoubtedly and incontestably a soul. This may be explained as follows. It is absurd that the circular motion of the sphere should be similar to the rectilinear motion of the stone downwards or to the motion of the fire upwards, so that the principle of that motion would be a nature and not a soul. For what is moved in natural motion is only moved by the principle subsisting in it, when the object to be moved is not in its place, and it is moved in order that it may seek to come to its place. However, when the object in question reaches its place, it comes to rest. The sphere, on the other hand, is moved in its own place in a circular motion. ... In consequence this circular motion can only come about in virtue of a certain mental representation, which determines the sphere's moving in that particular way. Now there is no mental representation without intellect. In consequence the sphere must be endowed with an intellect. ... [T]he soul, in virtue of which there is the

continued on next page

set up in ways that differ greatly from what is found directly in nature, and new concepts have to be created with a bold departure from immediate experience. It is therefore understandable that in ancient Greece, little progress was made in these fields. Even Aristotle—despite his skills as a very acute observer to whom biology owes some observations that remain valid to this day—made no attempt to carry out any experiments to clarify the fundamental laws of motion. Furthermore, he lacked the soaring imagination of his teacher Plato, as well as Plato's abstract mathematical point of view. Aristotle's dynamics is therefore only a summation of everyday observations that can be viewed as science only because Aristotle with his outstanding capacity for systematization was able to fit his dynamics into an all-encompassing worldview. However, it was precisely the effect of the universality of the Aristotelian worldview that compelled posterity to accept even the erroneous details. We shall return to this issue later.

Aristotelian dynamics, which for over two thousand years was the final word in this area, could be superseded only by casting off the *entire* Aristotelian worldview. In general, one may say that the Peripatetic theory of motion, while on the surface it correctly described the relevant phenomena and systematically organized them, it nevertheless lacked depth, and in the end hindered rather than advanced the development of dynamics as a science.

Here, we would like to demonstrate that Aristotle was led to the establishment of his fundamental laws of dynamics and to his classifications of motion by perfectly clear-headed considerations, so that some of his rules remain valid today as limited cases. Conversely, we would like to identify the decisive step that Aristotle was not able to take, nor were many generations of physicists after him able to do so (Plate III).

Let us imagine that we have forgotten everything we learned in school about Newton's laws of motion and seek to discover, from our everyday observations, without the help of experiment, some system that will account for the various forms of motion and at the same time establish a connection between some characteristic of motion, such as velocity, for example, and its effective cause. We think about these matters as we stroll along the shore of a lake, where we observe the various motions caused by man and by nature. In the shallow water, many small fish are swimming, gulls fly about in the air, and other people out for a walk stroll past us. All of these motions have something in common: the cause of independent motion is to be found in the fact that these are living beings.

As we play with a handful of pebbles, we drop one accidentally. Of course, it falls to the ground. From the stack of a passing ship, a stream of smoke is belching skyward, which also strikes us as natural. A child is pulling a small wagon on a string, while a toy car races in front of another child without any visible connection between car and child. We do not consider these two motions to be naturally caused; in each case we look for a cause of the motion, for the impetus. Evidently, in both cases the motion is caused by some action, in one case by the child, that is a living being, and in the other, by some mechanism inside the toy.

It is evening; the Sun is slowly sinking below the horizon, and the stars are becoming visible. The promenade along the shore is becoming empty, but we still hear the wind blowing and see the play of the waves. Compared to movement on Earth, we see something strikingly different in the sky: movements there with their measured stateliness and everlasting regularity are in sharp contrast with our experiences on Earth. It is clear, or in any case it is our immediate impression, that the laws of motion in the heavens and those on Earth *must be different*.

We have thus arrived at the Aristotelian classification of motions:

1. Eternal motions: motions of the celestial spheres (*motus a se*).
2. Terrestrial motions:
   a. Motions of living organisms (*motus a se*).
   b. Natural motions or restoration of a disturbed order: heavy bodies fall downward; light bodies move upward (*motus secundum naturam* or *motus naturalis*).
   c. Violent motion due to a force forcing (*motus violentus*).

The celestial bodies move of their own accord, so they must have a soul, and they move—in contrast to human beings in their daily lives—in perfect circles, so they must therefore be of a godly nature of a higher order than that of mankind. The perfect circular motion, or motion composed of several perfect circles, is fitting only to such beings (Quotations 1.22 and 1.23).

For the motion of terrestrial living beings, we cannot give such laws

Here we have arrived at the first characteristic assertion of Peripatetic dynamics: Fundamentally different laws hold for the celestial and terrestrial (sublunar) spheres. As we shall see, the distinction between these two domains will be important in all respects.

The second characteristic of Peripatetic dynamics follows from the first: Since a definite order exists in the cosmos, heavy bodies have their place below, light bodies above, and heavenly bodies must be in the firmament. Thus, the nature of a body determines the types of motion that it is able to undergo.

In our everyday lives, when something moves we ask: why does it move? This is also the question posed by Aristotelian dynamics, and the third thesis provides an answer: Every motion is due to an effective cause, or as we would say today, is due to a force (*omne quod movetur ab alio movetur*). Therefore, motion is a process and not a state, which means that when the cause ceases, the motion ceases, too.

Peripatetic dynamics also assumes that the effective cause or force can be applied only through immediate contact; that is, for every motion, we need to look for some connected driving force (*motor conjunctus*). In Aristotelian dynamics, the motive cause or force is related to the velocity of the body. Of course, both velocity and force are to be considered here only qualitatively. The proportionality between the two was built into the theory only by later Aristotelian commentators (Quotation 1.24). From our point of view, it follows immediately that a body in motion must overcome resistance and that the velocity is inversely proportional to the force of resistance. We are going well beyond the ideas of both Aristotle and his commentators in the following when we use today's customary notations and concepts in stating the fundamental law of Peripatetic dynamics.

The velocity of a body is determined by the motive force and the resistance. Velocity and motive force are closely linked, since a large velocity is associated with a large force, and a small velocity with a small force. In today's notation, we may write

$$\text{velocity} \approx \frac{\text{effective cause}}{\text{resistance}} \Rightarrow v \approx \frac{F}{R}.$$

We stress once again that, although this law states the fundamental law of dynamics falsely, it nevertheless corresponds to everyday observations. It expresses the simple fact that, for example, a chariot travels more rapidly if there are more horses pulling it, or that a block of stone of a certain weight can be pulled faster by a larger number of slaves than by a smaller number.

Quotation 1.23, ***continued***

motion, and the intellect, by which the object is represented to oneself, are not both of them together sufficient to account for the coming-about of such a motion until desire for the notion represented is conjoined with them. Furthermore, it follows necessarily from this that the sphere has a desire for that which it represents to itself and which is the beloved object: namely, the deity, may He be exalted. Aristotle says that it is in this manner that the deity causes the sphere to move, I mean to say through the fact that the sphere desires to come to be like that which it apprehends, which is the notion represented—a notion that is most exceedingly simple, in which there is no change and no coming-about of a new state, and from which good always overflows. This is impossible for the sphere qua a body unless its activity be a circular motion and nothing else. For this is the final perfection of what is possible for a body to have as its perpetual activity. ...

As for the assertion that the spheres are living and rational, I mean to say endowed with apprehension, it is true and certain also from the point of view of the Law; they are not dead bodies similar to fire and earth—as is thought by the ignorant—but they are—as the philosophers say—living beings who obey their Lord and praise Him and extol Him greatly. Thus Scripture says: *The heavens tell of the glory of God,* and so on.

—Maimonides, *The Guide of the Perplexed*, Chapters 4 and 5 [pp. 255, 256, 259]

Quotation 1.24

We observe that the same weight or body travels faster for two *reasons*, either because there is a difference in the medium through which it travels, as through water or earth or air, or because, other things being the same, the traveling body has an excess of density [or weight] or of lightness. The medium through which the body travels is a cause by the fact that it obstructs the body, most of all if it [the medium] is travelling in the opposite direction, but even if it is *resting*; and it does so more if it is not easily divisible, and such is a more viscous medium.

continued on next page

Quotation 1.24, ***continued***

The body *A*, then, will travel through medium *B* in time *C* and through medium *D* (which is less viscous) in time *E*, these [*C* and *E*] being proportional to the obstructing medium, if the lengths of *B* and *D* are equal. For let *B* be water and *D* air. Then the extent to which air is less viscous or less corporeal than water is proportional to the extent to which *A* travels faster through *D* than through *B*. Let the two speeds have the same ratio as that by which air differs from water. Then if air is half as viscous as water, *A* will travel through *B* in twice the time as it will through *D*, and *C* will be twice as long as *E*. And always, the more incorporeal or less obstructive or more easily divisible is the medium, the faster will the body travel through it.

But there is no ratio in which the [weight of the] void is exceeded by a body, just as there is no ratio of zero to a number.

—Aristotle, *Physics*, IV.8 [Apostle, pp. 73–74]

In Table 1.3, we compare Peripatetic dynamics with Newtonian dynamics, whose correctness we do not doubt, at least not in the macroscopic world. In Newtonian dynamics, motion is a state that can be maintained without an effective cause or force. Today, we formulate this law of motion as follows:

$$\frac{\text{change in velocity}}{\text{time}} = \frac{\text{force}}{\text{mass}} \Leftrightarrow \frac{dv}{dt} = \frac{F}{m}.$$

An external force is required only to *alter* the state of motion. The degree of abstraction that was required to reach this conclusion can be seen in the natural tendency of even scientifically educated persons to ask, regarding a moving object, "Why does it move?" rather than, "Why did its state of motion change?" Furthermore, if a person were to claim that some macroscopic terrestrial system could remain forever in motion without the influence of any outside force, we would not only doubt his scientific knowledge, but his sanity as well.

It is revealing to look for points of agreement between Peripatetic and Newtonian dynamics, in other words, to ask under what approximating assumptions would we be led from Newtonian dynamics to the superficial laws of motion of everyday experience, that is, to Peripatetic dynamics.

Let us assume that a constant force is applied to an object, and in addition, a resistive force operates proportional to the velocity of the object. As an example, we might consider an object in freefall in air or in some other medium of greater viscosity. The equation of motion is then

$$\frac{dv}{dt} = \frac{F - Rv}{m}.$$

If the object begins its motion at time $t = 0$ with velocity $v = 0$, then the solution of this differential equation is

$$v = \frac{F}{R}\left(1 - e^{-\frac{R}{m}t}\right).$$

If we introduce a time constant $\tau = m/R$, then the velocity is given by

$$v = \frac{F}{R}\left(1 - e^{-t/\tau}\right). \qquad (1)$$

Figure 1.47 shows a graph of velocity as a function of time. We see that for a while the velocity increases but then asymptotically approaches a constant value: $v \to F/R$. This limiting value corresponds exactly

Peripatetic Dynamics	*Newtonian Dynamics*
To maintain motion	*To change the state of the motion*
a force is required	*a force is required*
$v \approx F$	$\frac{d}{dt} v \approx F$
when $F = 0$	*when* $F = 0$
then $v = 0$	*then* $v =$ *constant*
Motion is a process	*Motion is a state*

▲ **Table 1.3** Peripatetic and Newtonian dynamics.

to the value determined by Peripatetic dynamics. That is, at least under the given assumption that the frictional force is proportional to the velocity, the terminal velocity that results from today's laws of motion is equal to the velocity derived from Peripatetic dynamics.

For a greater overview, let us consider the extent to which we have to take this asymptotic limit into account in modern practice. This depends, of course, on the numerical value of the time constant $\tau = m/R$. Consider, for example, MILLIKAN's experiment, which was of fundamental significance for modern physics, showing that the charge on the electron can be determined independently of other atomic constants (Section 4.6). The basic phenomenon around which the experiment was constructed was that electrically charged droplets of oil subject to the constant force of gravity and a constant electrostatic force, *move at a constant velocity*. In other words, we can assume that they had already reached their asymptotic (terminal) velocity. To arrive at this velocity, we solve the equation of motion:

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = mg - Rv \quad \left(m = \frac{4\pi r_0^3}{3}\rho;\ R = 6\pi\eta r_0\right), \tag{2}$$

and if we look closely at the time constant, it has the value

$$\tau = \frac{m}{R} = \frac{2}{9}\frac{r_0^2\rho}{\eta} \approx 10^{-5}\ \mathrm{s},$$

so that the oil droplets achieve their terminal velocity practically instantaneously.

Up to now, we have investigated only the asymptotic solution of equation (2), that is, for times $t \gg \tau$. However, the solution for times $t \ll \tau$ is also of interest. Using a series expansion of equation (1), one obtains

$$v \approx \frac{F}{R}\frac{t}{\tau} = \frac{F}{R}\frac{R}{m}t = \frac{F}{m}t.$$

We may summarize the two limiting cases as follows:

$$v = \frac{F}{R}\left(1 - \mathrm{e}^{-t/\tau}\right)$$

$t \gg \tau$ ↓ $v \approx \frac{F}{R}$

The velocity is proportional to the force.

$t \ll \tau$ ↓ $v \approx \frac{F}{m}t$

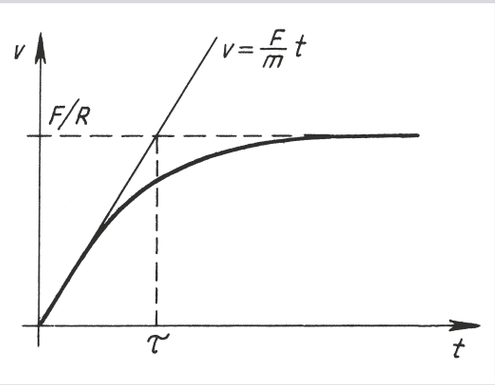
The velocity increases at a constant rate.

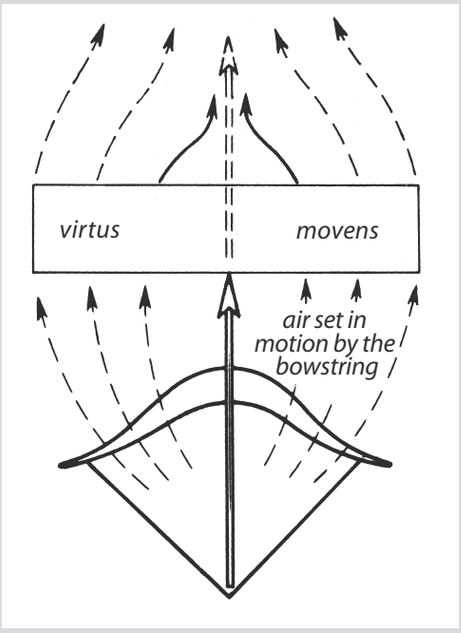
We see that a body starting from rest, at times that are small relative to the time constant τ, will accelerate at a constant rate, whereas at times that are large relative to the time constant, they will move at a constant velocity as required by Peripatetic dynamics. For our further discussion it will be useful to take note of the following peculiarity of the time constant: If the resistance to motion is calculated according to STOKES's law of friction, then one obtains the relation

$$\tau = \frac{2r_0^2\rho}{9\eta}$$

between the time constant and the dimensions and density of a spherical body as well as the viscosity and density of the medium. When we take stones or metal objects that are a couple of centimeters in size and drop them from a height that is easily realized, this τ will be much greater than the time of fall in the experiment. Thus, if we wished to confirm the Peripatetic point of view by means of experiments carried out from the top of a tower, let's take for example the famous Leaning Tower of Pisa, then we would have to use objects that reach their terminal velocity within times that are small in comparison to the total time of fall, which is to say, in fractions of a second; then for spheres having identical external dimensions, their velocities would in fact be proportional to their weights, as required by Aristotelian dynamics. Examples of such objects are hollow thin-walled spheres. If we actually used them in drop experiments, unavoidable disturbances, such as any movement of the air, would surely overwhelm the expected results. It was very fortunate for the history of physics that with free-fall experiments, the kind of motion conforming to the Newtonian description is easier to realize than that which conforms to the Aristotelian conception.

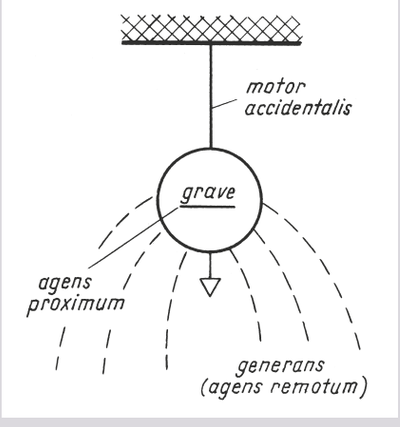
▼ **Figure 1.47** If a constant force acts on an object together with a frictional force proportional to the velocity, then the limiting velocity given by Newtonian dynamics agrees with that obtained from Aristotelian dynamics.

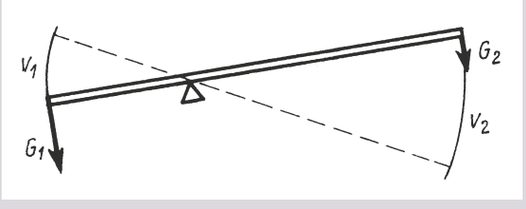




▲ **Figure 1.48** Why does the arrow continue after it has left the bowstring?

▼ **Figure 1.49** The Scholastic answer given by Aristotelian commentators to the question of what moves a body during free fall.





▲ **Figure 1.50** The law of the lever: an idea of Aristotle that is reminiscent of the principle of virtual displacement.

The consistent application of Peripatetic dynamics resulted in a host of difficulties that gave later commentators on Aristotle much to think about. In the end it was the unsatisfactory attempts at interpretation that led to a thorough reevaluation. Thus it was a major problem to specify the cause of the continued motion of an arrow after it has left the bowstring, that is, to answer in this case the question of *motor conjunctus*. According to the Aristotelian conception, the effective cause must stand in immediate contact with the moving body. By the Middle Ages, commentators had arrived at the following explanation: As long as the arrow remains in contact with the bowstring, it is to be understood that the tension in the bow provides the *motor conjunctus*. Here, however, the motive force has three functions:

1. It sets the arrow in motion.
2. It sets the surrounding atmosphere in motion.
3. It transmits to the surrounding air some sort of energy of motion (*virtus movens*).

In the further course of the arrow's motion, the surrounding air takes over the role of *motor conjunctus*, sustaining the arrow in its flight with its *virtus movens*. The motion is passed along to the adjoining masses of air, along with a *virtus movens*. This process is repeated continuously and maintains the motion. The arrow therefore always remains in contact with the medium that causes its motion. This idea is depicted in Figure 1.48.

Even more complicated is the situation of natural motion. Consider, for example, a falling body. It cannot set itself in motion of its own accord because it is not alive, that is, it is not a *res animata*. However, Aristotelian dynamics has no concept of action at a distance. What, then, is the *motor conjunctus,* in this case? The Scholastics had to summon the entire corpus of the Aristotelian conceptual system to bring this process into alignment with Peripatetic dynamics. In their interpretation, the concept of a *generans*—an effective driving cause —enters. Whatever this may have meant, the purpose was to give a mass its weight, its gravitas. This *generans* cannot, however, be seen as the *motor conjunctus* because it does not stand in immediate contact with the falling body. The substantial form created by the *generans* is the proximal cause, the *agens proximum*. The substantial form does not, however, act immediately, but only via the *accidentia*. The *accidens* here, that is, the effective *accidens* (*agens instrumentale*), is actually the *gravitas accidens*, which prescribes the natural condition of a body to be as close to the Earth's center as possible. However, in order to allow motion to come into being, it is necessary for the factors that prevented motion from taking place, the *impedimenta*, to be removed. This task is undertaken by the *motor accidentalis*. Figure 1.49 attempts to clarify this complex conceptual system. However, it must be made clear that any representation using formulas and explanatory diagrams will necessarily falsify the original ideas—all the more so because, in fact, there existed no clear and definite ideas about either the phenomena or the concepts associated with them.

In the following, we discuss two applications of Peripatetic dynamics. Consider first how Aristotle derived the principle of the lever. Let us move the two-armed lever depicted in Figure 1.50 from its equilibrium position. After we turn it through a certain angle, then it is clear that in the corresponding time interval the shorter lever arm describes a proportionally smaller circular arc than the longer. Consequently, its velocity is less.

The velocities stand in relation to each other as the lengths of the lever arms. From this, Aristotle derived the following equations:

$$G_1 v_1 = G_2 v_2, \quad G_1 l_1 = G_2 l_2, \quad \frac{G_1}{G_2} = \frac{l_2}{l_1}. \tag{3}$$

These equations can be interpreted according to Peripatetic dynamics as follows. At first glance, it is surprising that the weight—that is, the force—and the velocity appear *multiplied* on both sides of the equation. However, for Aristotle, the

interpretation is that the weight on one side of the lever attempts to move the weight on the other side, and that the counterweight therefore acts as an impediment to the motion of the former. Thus, the factor G multiplying the velocity in the first formula above plays the role of resistance in the following:

$$\text{velocity} = \frac{\text{effective cause}}{\text{resistance}} \Rightarrow \text{cause} = \text{resistance} \times \text{velocity}.$$

Equation (3) may thus be interpreted as follows: equilibrium means that the cause on the left side of the lever is precisely compensated for by the cause on the right side.

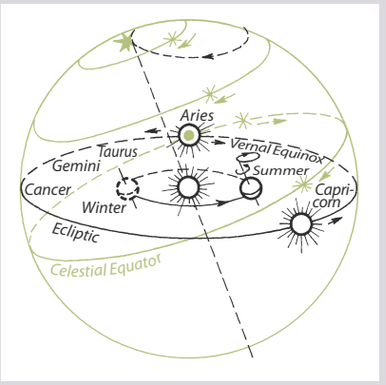
Whatever one may think of this train of thought, it does lead to the correct result. Of much greater significance, however, is the fact that ARISTOTLE investigated a problem in statics by applying the methods of dynamics. This process is indirectly related to investigations into stability, but directly related to the principle of virtual displacement, or, as it was formerly known, the principle of virtual velocity. One may consider ARISTOTLE's treatment as the starting point of this extremely useful method.

From the fundamental principles of Peripatetic dynamics follow the views of Aristotelian physics regarding the nature of a vacuum, which can be expressed as follows: A vacuum cannot exist, since it is both logically and physically absurd. This idea of *horror vacui* ("Nature abhors a vacuum") continued to haunt physics well into the seventeenth century. For ARISTOTLE, the vacuum was a place where nothing was placed (*locus sine locuto*), in itself a self-contradictory idea. He also introduced a number of physical arguments supporting the impossibility of a vacuum, such as the following.

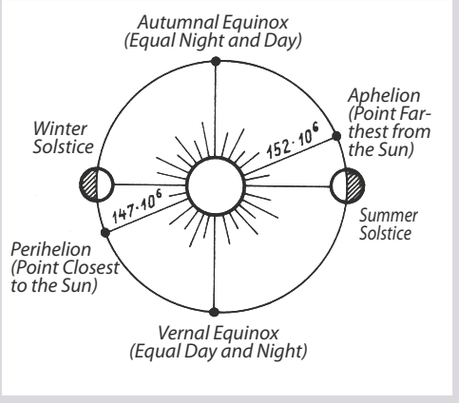
For a body in a vacuum, it would be impossible, due to the lack of bases for comparison, to determine whether it was light or heavy, and consequently, the body would be unable to determine the appropriate direction—up or down—for its natural motion. This is, however, an absurd situation—one that cannot occur. In and of itself, the part of the argument claiming that an object must always be in a uniquely determinable state is no stranger to modern physics, either. When we set it up as a requirement for an oscillator in electrodynamics, or when we state in quantum mechanics that the characteristic quantities have to be uniquely determined, we are following a similar train of thought.

A vacuum creates no resistance to motion, implying an infinite velocity due to the relationship $v = F/R$. This was expressed in Peripatetic physics by saying that a body would require zero time to get from one point to another. In a frictionless medium, the velocity could become infinitely large in Newtonian dynamics as well, although an infinitely long time would be required.

At another point, ARISTOTLE rejects the possibility of a vacuum with a surprising argument. He maintains not that the velocity would be infinite in a vacuum, but that its value would have to remain constant for all time (Quotation 1.25). Here Aristotle gives almost word for word the Newtonian first law of motion, but only as an absurd premise in order to refute what he considers to be an absurd statement, namely the existence of a vacuum. According to our understanding today, this "absurd" premise, which is equivalent to Newton's first law of motion, is one of the most beautiful truths in the history of physics, hence it is no wonder that the opposite of the theorem of the impossibility of vacuum here "proven"—to wit, the existence of the vacuum—also became an important premise of the new physics.



◀ **Figure 1.51** The movements of the heavens must be described from a point of view on Earth, which itself executes a compound motion.

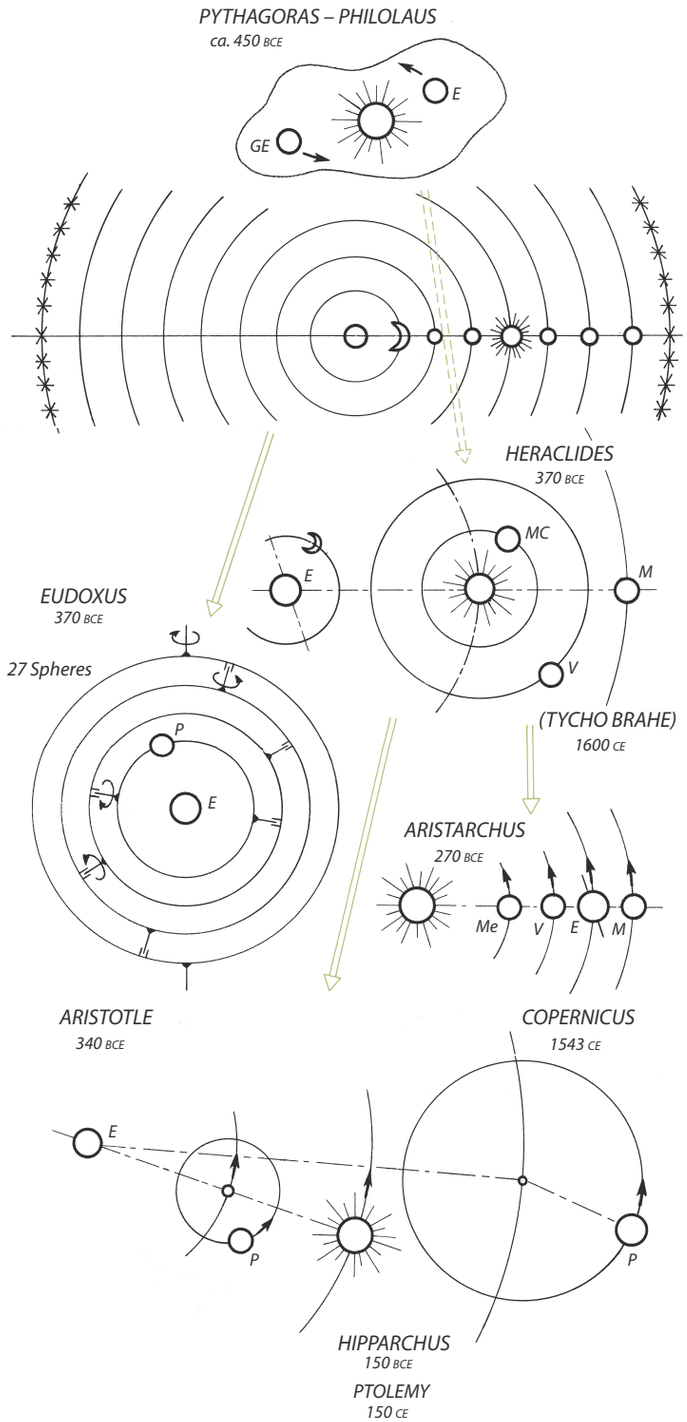


▲ **Figure 1.52** The Earth revolves about the Sun in a nearly circular elliptical orbit, on which we have drawn some significant points. The precise definition of the length of a day or of a year has always been a significant problem, particularly for defining a calendar. The rhythm of life on Earth has been adapted to the Sun: from one noon to the next, from one vernal equinox to the next, these define the solar day (whose length is not constant) and the tropical year (one tropical year = 365.2422 average solar days = 365 days, 5 hours, 48 minutes, 46.98 seconds). The time between two successive culminations of a given fixed star is called a sidereal day; this is shorter by 3 minutes, 56.5 seconds than an average solar day (due to Earth's revolution about the Sun). In contrast, the tropical year is about 20 minutes shorter than the sidereal year, that is, the time for Earth to return to a given point on the ecliptic in relation to the fixed stars (1 sidereal year = 365.2565 average solar days = 365 days, 6 hours, 9 minutes, 9.54 seconds), which is due to the precession of the equinoxes (50.256 seconds of arc per year), which in turn is attributable to the precession of the Earth's axis.

Quotation 1.25

Again, none of these thinkers could say why a body which has been caused to be in motion will stop somewhere; for why should it stop in one place rather than in another? So either it will be resting or it will of necessity be travelling without an end, unless obstructed by something more powerful.

—ARISTOTLE, *Physics,* IV.8 [Apostle, pp. 73]



▶ **Figure 1.53** Ancient conceptions of the cosmos, among which can be found the Copernican system.

> ... since ten is considered to be complete and to include every nature in numbers, [the Pythagoreans] said that the bodies which travel in the heavens are also ten; and since the visible bodies are nine, they added the so-called "Counter-Earth" as the tenth body.
>
> —Aristotle, *Metaphysics* [986a, p. 21]

> Eudoxus held that each of the locomotives of the Sun and of the Moon is in three spheres, of which the first is that of the fixed stars, the second along the circle which bisects the Zodiac, and the third along the circle inclined across the breadth of the Zodiac, but the circle along which the Moon moves is inclined at a greater angle than that along which the Sun moves. The motion of each of the planets is in four spheres, and of these the first and second are the same as those in the previous case (for the locomotion of the sphere of the fixed stars belongs to all the spheres, and that of the sphere next under it which moves along the circle bisecting the Zodiac belongs to all), the poles of the third sphere of each planet are in the circle bisecting the Zodiac, and the locomotion of the fourth sphere is in the circle inclined at an angle to the equator of the third; and the poles of the third sphere are different for different planets, except for Venus and Mercury which have the same poles.
>
> Callippus posited the same position of the spheres as that held by Eudoxus, that is, with respect to the order of the intervals, but while he assigned to Jupiter and Saturn the same number of spheres as Eudoxus did, he thought that two more spheres should be added to the Sun and also to the Moon, if one is to account for the observed phenomena, and one more to each of the other planets.
>
> But if all the spheres combined are to account for the observed phenomena, there must be other spheres for each planet, one less in number than those assigned to it, which would counteract these and restore to the same position the first sphere of the star which in each case is next in order below; for only thus can the motion of the combined spheres produce the motion of the planets. Since, then, the spheres in which the planets are carried are eight for Jupiter and Saturn and twenty five for the others, and of these only those need not be counteracted in which the lowest-order planet is carried, the spheres which counteract those of the first two planets will be six, those of the next four planets will be sixteen, and the total number of spheres which includes both those which carry the planets and the ones which counteract those spheres will be fifty-five. If we are not to add to the Moon and to the Sun the motions we mentioned, all the spheres will be forty-seven. Let, then, this be the number of spheres, and if so, it is reasonable to believe that the immovable substances or principles are also as many. As to what is necessarily the case, this may be left to more competent thinkers.
>
> —Aristotle, *Metaphysics* [1073b–1074a, pp. 207–208]

continued on next page

1.3.3 *Celestial Motion*

In the past, as is the case today, astronomy as a science was expected to answer a host of questions. We shall deal with these in due course but first we need to discuss how far the ancient Greeks progressed in answering them.

The first task of astronomy is to observe and describe all phenomena that occur in the heavens, for example, the rising and setting of the Sun and the Moon, the phases of the Moon, and the motions of the planets and fixed stars. For such descriptions, one may make use of tables, such as those of the Babylonians, or one might develop some sort of geometric model. We can say that the Greeks

succeeded in obtaining a perfect description of the observable celestial motions with the help of geometric models, and therefore, the main portion of this section is devoted to descriptive, or geometric, astronomy.

Of course, we also expect astronomers to provide a physical explanation of observed phenomena. Not only should they give us a mathematical model of celestial motion, as seen by a terrestrial observer, but they should also reveal the physical reality behind what is seen. Thus, some Greek scholars, such as Anaxagoras, Heraclides, and Aristarchus, gave interpretations of this kind that remain arguably true even today, although they were not generally accepted by Greek astronomy.

The Greeks completely ignored the issues of physical interpretation of their world view. In fact, they did not attempt to connect their descriptive theories of celestial motion with a theory of dynamics in any meaningful way.

In contrast, their accomplishments in determining cosmic distances, the radius of the Earth, the distances between Earth and the Moon and Earth and the Sun, as well as the application of their acquired knowledge to the preparation of calendars and maps, are once again worthy of our profound admiration.

In seeking to arrive at a proper evaluation of the Greeks' achievements in geometric astronomy, let us begin by considering exactly what the motions that they had to describe were. Let us for a moment recall our knowledge of the motion of Earth and of the entire solar system (Figures 1.51 and 1.52).

First of all, we note that all motion is observed from the point of view of Earth, to wit, from a frame of reference that rotates about its axis once every 24 hours; that revolves about the Sun with a varying velocity along a path that is not circular, but elliptical; that possesses an axis of rotation that is not perpendicular to the plane of the path of Earth's revolution about the Sun and has, in fact, an angle of inclination that is not constant but undergoes instead a precessional motion with a period of 26,000 years.

It is from within such a coordinate system that the Greeks had to describe the apparent motions of the planets, which themselves traverse similarly complicated paths; the apparent motion of the Sun; and the apparent motion of the background of fixed stars. If we now add the problem that based on various theoretical considerations, the Greeks only allowed for certain kinds of motion for celestial bodies, such as uniform motion in a circular path with Earth at the center, or various combinations of circular paths, which greatly limited the possibilities available to them for the description of motion, we can begin to appreciate the difficulties that they had to overcome and can only marvel at the solutions they did achieve (Quotation 1.26).

Figure 1.53 depicts the various conceptions of how the positions and motions of the planets and the heavens in relation to Earth are to be described, in addition to the time at which each one arose. We have already spoken of the model due to Philolaus (ca. 470–ca. 385 BCE), a follower of the Pythagorean school. Whatever numerological ideas he might have used to create his cosmology, his Earth freely floating in space could now move in any number of ways, and thus his model can serve as the starting point for later models. It is possible that the system of Heraclides, with Earth rotating on its axis, was derived from the Philolaus-system by letting Earth and the counter-Earth melt and merge, and thus enclose the central fire. An additional peculiarity of Heraclides' model is the assumption that the Sun revolves around the Earth, whereas the inner planets—Mercury and Venus—revolve around the Sun. This system is quite similar to the one developed by Tycho Brahe around 1600. It was further developed by Aristarchus into a heliocentric system, which corresponds precisely to the system announced by Copernicus in 1543, except that, unlike Copernicus, Aristarchus did not attempt to refine his system.

Figure 1.53 *continued*

There are some, King Gelon, who think that the number of the sand is infinite in multitude; and I mean by the sand not only that which exists about Syracuse and the rest of Sicily but also that which is found in every region whether inhabited or uninhabited. Again there are some who, without regarding it as infinite, yet think that no number has been named which is great enough to exceed its multitude. And it is clear that they who hold this view, if they imagined a mass made up of sand in other respects large as the mass of the earth, including in it all the seas and the hollows of the earth filled up to a height equal to that of the highest of the mountains, would be many times further still from recognizing that any number could be expressed which exceeded the multitude of the sand so taken. But I will try to show you by means of geometrical proofs, which you will be able to follow, that, of the numbers named by me and given in the work which I sent to Zeuxippus, some exceed not only the number of the mass of sand equal in magnitude to the earth filled up in the way described, but also that of a mass equal in magnitude to the universe. Now you are aware that "universe" is the name given by most astronomers to the sphere whose centre is the centre of the earth and whose radius is equal to the straight line between the centre of the sun and the centre of the earth. ...

But Aristarchus of Samos brought out a book consisting of some hypotheses, in which the premises lead to the result that the universe is many times greater than that now so called. His hypotheses are that the fixed stars and the sun remain unmoved, that the earth revolves about the sun in the circumference of a circle, the sun lying in the middle of the orbit, and that the sphere of the fixed stars, situated about the same centre as the sun, is so great that the circle in which he supposes the earth to revolve bears such a proportion to the distance of the fixed stars as the centre of the sphere bears to its surface. Now it is easy to see that this is impossible; for, since the centre of the sphere has no magnitude, we cannot conceive it to bear any ratio whatever to the surface of the sphere. We must however take Aristarchus to mean this: since we conceive the earth to be, as it were, the centre of the universe, the ratio which the earth bears to what we describe as the "universe" is the same as the ratio which the sphere containing the circle in which he supposes the earth to revolve bears to the sphere of the fixed stars. For he adapts the proofs of his results to a hypothesis of this kind, and in particular he appears to suppose the magnitude of the sphere in which he represents the earth as moving to be equal to what we call the "universe."

—Archimedes, *The Sand-Reckoner* [pp. 221–222]

Quotation 1.26

Just as it is impossible in geometry and music to deduce the consequences of their principles without making hypotheses, so in astronomy it is necessary, in order to speak of the motion of the planets, to establish hypotheses. Above all it is necessary, as everyone agrees, to choose principles serviceable in mathematical studies. The first is that the composition of the world is ordered and

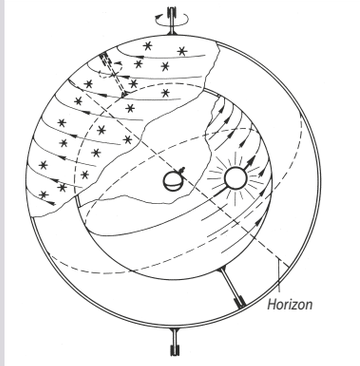
continued on next page

Quotation 1.26, ***continued***

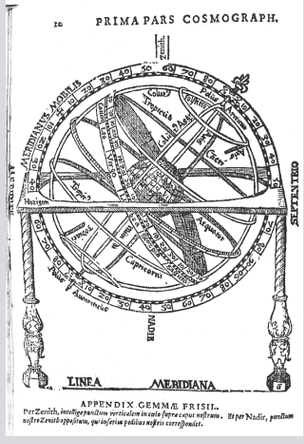
governed by a single principle, that a reality is found to underlie the things which exist or which appear to exist, and that, where the world extends beyond the limits of our vision, we should say that it has its limits rather than that it is the infinite.

The second principle is that the risings and settings of the heavenly bodies are not due to these bodies successively lighting up and going out; if their state were not eternal, there would be no order maintained in the universe. The third principle is that there are seven moving bodies, neither more nor less—a truth known from long observation. The fourth is the following: since it does not stand to reason that all the bodies are in motion or that all are at rest ... it is necessary to inquire as to which is necessarily at rest and which is in movement. It must be believed ... that the earth, Hearth of the Gods, following Plato, is at rest, and that the planets move together with the whole of the celestial vault which envelops them. Next he energetically rejects, as contrary to the foundations of mathematics, the opinion of those who want to make those bodies which appear in motion [the stars] to be at rest, and those bodies immobile by nature or by situation [the earth], to be in motion.

—Eudemus of Rhodes, a student of Aristotle, on astronomical hypotheses [de Santillana, pp. 253–254]



◀ Figure 1.54
The simplest world model that takes into account the annual change in the height of the Sun at noon.



◀ Figure 1.55
The arrangement of Earth with the ecliptic as represented in the geocentric system well into the sixteenth century; see also Figure 2.73. (Petrus Apianus, *Cosmographia*, Library of the Hungarian Academy of Sciences.)

Perhaps it was precisely the system of Heraclides that led Hipparchus to the idea that the motion of the planets about an unmoving Earth can be described as uniform motion along a circular path if the midpoint of a planet's circular orbit itself moves in a circular orbit around the Earth. This idea was refined by Ptolemy into a cosmological system that held sway unchallenged for 1500 years (Section 1.4).

Another chain of ideas led to the Earth-centered spheres of Eudoxus, a friend of Plato. His world system was then further developed by Aristotle. When in the Middle Ages scholars spoke of a simple world system, what they had in mind was a simplified variant of this Aristotelian picture. However, astronomers continued to use the Ptolemaic system. In the late Middle Ages, an attempt was made to combine the two systems into a single physical model.

After this brief overview, we shall next describe the system of Eudoxus and then, in the following section, the Ptolemaic system, which remained in use until it was supplanted by the Copernican system.

Eudoxus tried to implement Plato's principles down to the last detail: The celestial bodies are divine manifestations, and therefore the only possible motion that they can exhibit is one that is perfect, namely, a perfect circle. However, to rescue—as Plato expressed it—the heavenly manifestations, one had to employ a number of circular motions with a common midpoint but with different axes of rotation, whose resultant was then the actual motion. This procedure has been compared with the Fourier series in which a periodic function also has to be expressed as a sum of harmonic functions.

Figures 1.54 and 1.55 show that to describe of the motions of the Sun and the fixed stars, it suffices to introduce two spheres with a common center. The rotation of one of the spheres corresponds to the daily revolution of the stars, while the other corresponds to the motion of the Sun through the zodiac along the ecliptic. The axis of rotation of the second sphere forms an angle with that of the first, resulting from the obliqueness of the ecliptic.

The biggest problem for geometric astronomy is the correct description of the planetary orbits, which appear to be highly irregular. Thus when the orbits of the planets as they revolve about the Sun are observed from Earth (Figure 1.56), they appear, in relation to the background of fixed stars—due to Earth's own orbit about the Sun—at times to slow down or even move backward. Eudoxus described the planets' retrograde motion by introducing two additional spheres. The rotational axis of these two spheres lies in the plane of the ecliptic, forming a very small angle. The angular velocities of the two spheres are equal in magnitude but opposite in direction. The resulting motion is depicted in Figure 1.57. It increases or decreases the average motion along the ecliptic and thus with a suitable choice of axes and angular velocity, the observed motion can be reproduced to within a good approximation. In the end, Eudoxus required 27 spheres to describe the celestial phenomena (Figure 1.58).

1.3.4 *The Aristotelian Worldview*

The characteristic assertions of the Aristotelian worldview are collected in Table 1.4. This unified conceptual edifice has fascinated many of the world's most important thinkers over a very long span of time—more than two thousand years. As we shall see, it was fiercely attacked shortly after its completion; the attacks, however, were always against details; and in the end, because no similarly comprehensive worldview could be built around the solutions to such details, succeeding generations preferred

Cosmos *closed, hierarchical*	*Motion* *a process, not a state*	*Matter* *continuous, not atomic*
All things have their place, which they strive to attain according to their nature		
Celestial spheres, which are enclosed within the sphere of fixed stars	*Motion according to an eternal harmony: uniform circular motion or motion that is the composition of uniform circular motions*	*Unchanging, neither created nor destroyed:* quinta essentia
Sublunar world *Air* *Water* *Earth*	*Natural motion: heavy bodies seek the earth; light bodies, the heavens* *Forced motion: every motion has a motive force that must remain in immediate contact with the moving object*	*The mixing and separation of the elements earth, water, air, and fire comprise the changes in the world*
A vacuum is physically and conceptually impossible		

◀ **Table 1.4** Characteristics of the Aristotelian worldview.

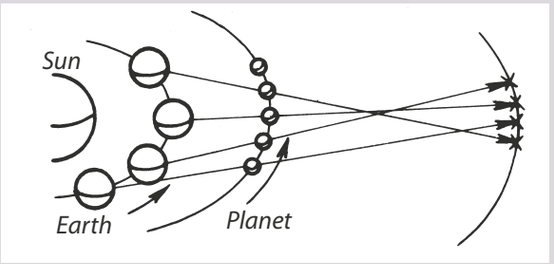
to sacrifice truly significant progress on individual problems in order not to abandon the Aristotelian cosmology, which satisfied the need for a synthesis that provided universally applicable answers. The strength of the Aristotelian worldview stems from the fact that it ultimately derives from everyday common sense, even though it must be said that in many cases the common-sense grasp on everyday phenomena is only superficial.

After Thomas Aquinas harmonized the Aristotelian worldview with Christian ideology in the Middle Ages, allowing the conceptions of Aristotle and Aquinas to be formulated into an official ideology, this worldview no longer rested merely on its own internal logic, but on the authority of Church and state.

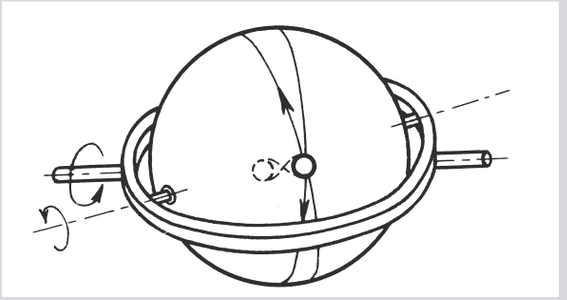
Aristotle's world (Figures 1.59 and 1.60) is that of a cosmos that was to be understood—as was usual in the ancient world—as a structure of perfect order. In such a world, everything has its natural place; all things and all creatures, men and gods, are arranged in a hierarchical order, within which the celestial and terrestrial worlds are sharply separated. The celestial world is the world of perfect order and perfect harmony, in which all is eternal and unchanging. The substance that forms the heavenly spheres is the ether, the quintessence, which is different from the material of which Earth is made. The heavenly bodies move in uniform circles about a central axis or the resultants of several circles. The beings that inhabit the celestial sphere are perfect and divine.

The sublunar world, the terrestrial world, is characterized by change, creation, and dissolution. But here, too, all things, including mankind, have their natural place. The proper place for the heavy sphere of earth is at the bottom, above which lie the sphere of water, above that the sphere of air, and then the sphere of fire. Any deviation from this order is contrary to nature, and every object has an inherent striving to take the place prescribed to it by nature.

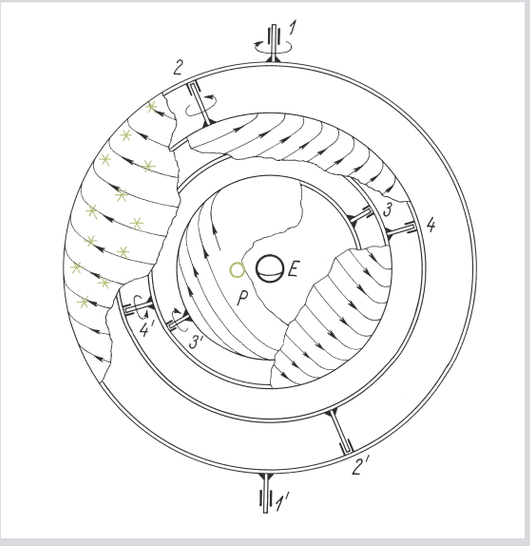
The cosmos is finite and self-contained. There is nothing, not even empty space, beyond the celestial sphere. In the center of this finite universe is Earth because it is made of the heaviest of materials, and the heaviest material has its natural place at the center.



▲ **Figure 1.56** In relationship to the background of fixed stars, the planets describe complicated looping paths.



▲ **Figure 1.57** Eudoxus illustrated the looping motion by means of two spheres with almost parallel rotational axes and opposite directions of rotation.



▲ **Figure 1.58** According to Eudoxus, the spheres required for a correct description of the motion of a single planet.
1 to 1′: rotation in 24 hours (diurnal motion)
2 to 2′: average motion of the planet with respect to the fixed stars
3 to 3′ and 4 to 4′: oppositely rotating spheres for the realization of the looping motion.

Quotation 1.27

The science of pneumatics was held in high regard of old by philosophers and engineers, the former logically deducing its principles, the latter determining them by experimental tests. What we have felt constrained to do in this book is to give an orderly exposition of the established principles of the science and add thereto our own discoveries. We hope in this way to be of service to future students of the subject.

Before we come, however, to the particulars of our exposition there is a general topic to be discussed, namely the nature of the vacuum. Some writers emphatically deny its existence. Others say that under normal circumstances there is no such thing as a continuous vacuum, but that small vacuums exist in a scattered state in the air, water, fire and other bodies. This is the opinion to which we should adhere. We now proceed to show by experimental tests that this is a true account of the matter.

We must first correct a popular illusion. It must be clearly grasped that vessels which are generally believed to be empty are not really empty but are full of air. Now air, in the opinion of the natural philosophers, consists of minute particles of matter for the most part invisible to us. Accordingly if one pours water into an apparently empty vessel, a volume of air comes out equal to the volume of water poured in. To prove this make the following experiment. Take a seemingly empty vessel. Turn it upside-down, taking care to keep it vertical, and plunge it in a dish of water. Even if you depress it until it is completely covered no water will enter. This proves that air is a material thing which prevents the water entering the vessel because it has previously occupied all the available space. Now bore a hole in the bottom of the vessel. The water will then enter at the mouth while the air escapes from the hole. But if before you bore the hole you lift the vessel vertically out of the water and examine it you will see that the interior of the vessel has remained perfectly dry. This constitutes the demonstration that air is a bodily substance. ...

Now let us return to those who absolutely deny the existence of the void. It is of course possible for them to discover many arguments in reply to what has been said and in the absence of any experimental demonstration their logic may appear to have an easy victory. We shall therefore show them, by phenomena which can be brought under observation, two facts: (1) that there is such a thing as a continuous void, but that it exists only contrary to nature, and (2) that in accordance with nature void also exists, though only in small scattered quantities. We will further show them

continued on next page

We may see how difficult it is to change such a comprehensive worldview by attempting to refute any of its concrete assertions either theoretically or experimentally. We cannot, for example, maintain that Earth is a planet like all the others, because that would upset the entire hierarchy. Even the assertion that the Moon is similar in form to Earth would cause severe upset. Let us consider a much simpler problem. We cannot even assert that bodies of differing weights move more-or-less the same way when they fall because, after all, the heavier body strives more to achieve its place assigned to it by nature than the lighter one. We cannot claim that vacuum exist because that would call the finiteness of the cosmos into question, and one could no longer speak of the center of something that is infinite or unbounded. Thus, the Aristotelian worldview must be either accepted or rejected in its entirety. One of the most courageous turns in seventeenth-century scholarship was the rejection of Aristotelian cosmology before having succeeded in supplanting it with a new one, hoping that the wealth of partial results obtained would eventually coalesce into a unified picture.

To the ancients' credit, we should note here that many specific claims of the Aristotelian worldview were forcefully opposed in ancient times, and in a surprisingly modern spirit. The pioneers of the sixteenth and seventeenth centuries would have found fine comrades in arms in their scientific colleagues of two thousand years earlier. We shall mention here only two such scholars: Strato (ca. 340–ca. 269 BCE) and Plutarch (ca. 46–120 CE). Strato gave experimental proofs of the existence of a vacuum (Quotation 1.27). He sharply attacked the "logicians," who argued back and forth until they finally conquered all opponents, but who, so Strato hoped, could not deny experimental facts no matter how they tried. The style that Strato used to describe his experiments is interesting in that it is identical to the style that a modern experiment physicist or a textbook on experimental physics might use to explain the process of an experiment.

Plutarch's observations are of special significance for a different reason. In his work *On the Face of the Moon* (*De facie in orbe lunae*), a conversation between two laymen is presented. This means that such thoughts at that time must already have been the topic of general conversation (Quotation 1.28). It is also of interest that the heliocentric theory of Aristarchus (ca. 310–ca. 230 BCE) appears here. Copernicus knew this work of Plutarch. Look at the metaphor where the motion of the Moon is compared with that of a stone whirled around in a slingshot; this was read by Newton. Although the two laymen still defend the differences between terrestrial and celestial phenomena, if not very convincingly, nevertheless they maintain that the Moon might well be something like a piece of Earth. The Moon might even be habitable! Perhaps it contains life forms that have adapted to the conditions there. The "conversation" ends with a scene in which inhabitants of the Moon ask one another whether Earth, that dark and damp celestial body, full of clouds and dust, might be capable of sustaining life, and these Moon-dwellers could conclude that theirs must be the only habitable celestial body.

1.3.5 *A Selection from Aristotle's Metaphysics*

All men by nature desire understanding. ...

All animals, except men, live with the aid of appearances and memory, and they participate but little in experience; but the race of men lives also by art and judgment. In men, experience comes into being from memory; for many

memories of the same thing result in the capacity for one experience. And experience seems to be almost similar to science and art, but science and art come to men through experience; for, as Polus rightly says, "experience made art, but inexperience, luck."

Now art comes into being when out of many notions from experience we form one universal belief concerning similar facts. For, to have a belief that when Callias was having this disease this benefited him, and similarly with Socrates and many other individuals, is a matter of experience; but to have a belief that this benefited all persons of a certain kind who were having this sickness, such as the phlegmatic or the bilious or those burning with high fever, is a matter of art.

Experience does not seem to differ from art where something is to be done; in fact, we observe that men of experience succeed more than men who have the theory but have no experience. The cause of this is that experience is knowledge of individuals but art is universal knowledge, and all *actions* and productions deal with individuals. The doctor does not cure "a man" universally taken, except accidentally, but Callias or Socrates or someone else to whom also the essence of man happens to belong. If, then, someone without experience has the theory and knows the universal but is ignorant of the individual included under this universal, he will often fail to cure; for it is rather the individual that is curable.

Nevertheless, we regard understanding and comprehension as belonging to art more than to experience, and we believe that artists are wiser than men of experience; and this indicates that wisdom is attributed to men in virtue of their understanding rather than their experience, inasmuch as men of understanding know the cause but men of experience do not. For men of experience know the fact but not the *why* of it; but men of art know the *why* of it or the cause. It is because of this that we regard also the master-artists of a given craft as more honorable, as possessing understanding to a higher degree, and as wiser than the manual workers, since the former know the causes of the things produced, but the latter are like certain inanimate things which act but do so without understanding that action, as in the case of fire which burns. Inanimate things bring about the effects of their actions by some nature, while manual workers do so through habit which results by practicing. Thus, master-artists are considered wiser not in virtue of their ability to do something but in virtue of having the theory and knowing the causes. And in general, a sign of a man who understands is the ability to teach, and because of this we regard art more than experience to be science; for those who have the art can teach, but those who do not have it cannot teach. Again, we do not consider any of the sensations to be wisdom, although these are the most authoritative in the knowledge of individuals; but they do not tell us the *why* of anything, as for example why fire is hot, but only the fact that it is hot.

The first who arrived at any art that went beyond the ordinary sensations was probably admired by men, not only because there was some usefulness in the objects arrived at, but also as being wise and superior to others. As more arts were arrived at, some for the necessities of life and others as the only ends of *activity*, those who arrived at the arts for the latter purpose were always believed to be wiser than those who did so for the former because their sciences were not instrumental to something else. Now when all such arts were already developed, the sciences concerned neither with giving pleasure to others nor

Quotation 1.27, ***continued***

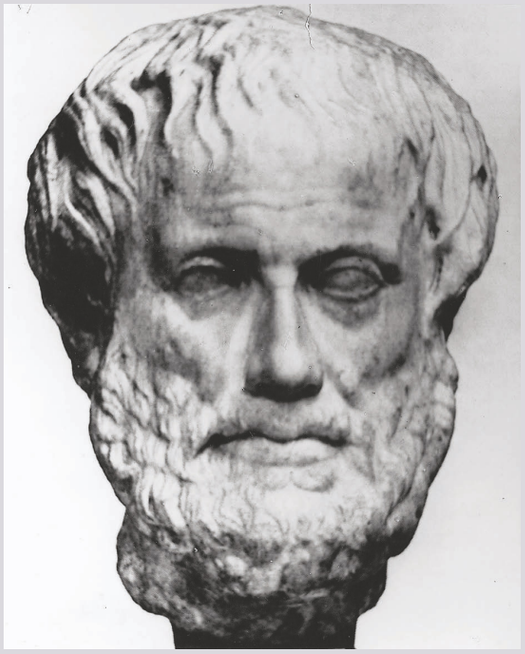
that under pressure bodies fill up with these scattered vacuums. These demonstrations will allow no loophole of escape for these verbal gymnasts.

For our demonstration we shall require a metal sphere, of a capacity of about four pints, made of metal sheeting of such thickness as to resist any tendency to collapse. The sphere must be air-proof. A copper tube, a pipe with a narrow bore, must be inserted in the sphere in such a way that it does not touch the spot diametrically opposite the point of entry but leaves room for the passage of water. The pipe should project about three inches from the sphere. The part of the sphere around the point of insertion of the pipe must be strengthened with tin solder so that the pipe and the sphere present a continuous surface. There must be no possibility that air forced into the sphere by blowing can escape by any crack.

Now let us analyze in detail the implications of the experiment. There is air in the sphere from the beginning as in all vessels popularly called empty, and the air fills the whole of the enclosed space and presses continuously against the containing wall. Now according to the logicians, since there is absolutely no unoccupied space, it should be impossible to introduce water or more air unless some of the air already contained in the vessel be displaced. Further, if the attempt were made to force air or water in, the vessel being full should burst before admitting it. Very well. What in fact happens? One who puts his lips to the pipe can blow a great quantity of air into the sphere without any of the contained air escaping. This happens as often as the experiment is repeated, and it constitutes a clear proof that the particles of air in the sphere are compressed into the vacuums between the particles. ... Again if, after blowing, one stops the pipe quickly with one's finger, the air remains the whole time compressed in the sphere. But on the removal of the finger the air that was forced in rushes out noisily and violently, being expelled by the expansion of the air within owing to its elasticity.

If the reverse experiment be tried, a great quantity of air in the sphere can be sucked out without any other air getting in to replace it .... This experiment conclusively demonstrates that the formation of a continuous vacuum takes place in the sphere.

—Strato (after [Farrington, pp. 31–35])



▲ **Figure 1.59** Aristotle (384–322 bce): born in the Macedonian city Stageira. At 17, he received a substantial inheritance from his father; he went to Athens, where he became the pupil of the 60-year-old Plato. After his teacher's death (347), he left Athens, and in 343 was the teacher of 14-year-old Alexander (the Great). In 334, he returned to Athens, where he taught at the Lyceum, in the grove of Apollo. Perhaps on account of the footpaths in the grove (περιπατοί) or perhaps also because he gave instruction while walking with his students (περιπατέομαι), his school acquired the epithet "Peripatetic." Here he founded a center for science and research unprecedented in scale, where together with his students he gathered knowledge and worked in the most varied subject areas (philosophy, history of philosophy, natural sciences, medicine, history, politics, economics, philology). In 323, after the death of Alexander the Great, he had to flee Athens to avoid trial on charges of godlessness. He moved into his mother's house on the island Euboea, where he died in 322. (The image is that of a Roman copy from the first half of the first century ce; Vienna, Kunsthistorisches Museum.)

All the original manuscripts of his work have been lost. What remain are lecture notes, other notes, and simple collections of material in no particular stylistic arrangement. His school was maintained by two of his favorite pupils, Theophrastus and Eudemus of Rhodes, who preserved the manuscripts. These eventually made their way to Rome, where in the period from 40 to 20 bce they were published under the auspices of Andronicus of Rhodes.

The works on logic are collected in the *Organon* (Όργανον, tool); those on natural philosophy, in the *Physica* (Φυσικά) comprised, among others, of

Physica (Φυσική άκρόασις)

De generatione et corruptione (Περί γενεσέως καί φδοράς)

De caelo (Περί ούρανοϋ)

Meteorologica (Μετεορολογικά)

De anima (Περί ψυχής)

The *Metaphysics* or *First Philosophy* (Πρώτη φιλοσοφία)

with the necessities of life were discovered, and first in such places where men had leisure. Accordingly, it was in Egypt that the mathematical arts were first formed, for there the priestly class was allowed leisure.

—Aristotle, *Metaphysics*, Book A, Chapters 1, 2 (pp. 12–16)

Now that we have established these distinctions, we must proceed to consider causes, their character and number. Knowledge is the object of our inquiry, and men do not think they know a thing till they have grasped the "why" of (which is to grasp its primary cause). So clearly we too must do this as regards both coming to be and passing away and every kind of physical change, in order that, knowing their principles, we may try to refer to these principles each of our problems.

In one sense, then, (1) that out of which a thing comes to be and which persists, is called "cause", e.g. the bronze of the statue, the silver of the bowl, and the genera of which the bronze and the silver are species.

In another sense (2) the form or the archetype, i.e. the statement of the essence, and its genera, are called "causes" (e.g. of the octave the relation of 2:1, and generally number), and the parts in the definition.

Again (3) the primary source of the change or coming to rest; e.g. the man who gave advice is a cause, the father is cause of the child, and generally what makes of what is made and what causes change of what is changed.

Again (4) in the sense of end or "that for the sake of which" a thing is done, e.g. health is the cause of walking about. ("Why is he walking about?" we say. "To be healthy", and, having said that, we think we have assigned the cause.)

....it is the business of the physicist to know about them all, and if he refers his problems back to all of them, he will assign the "why" in the way proper to his science – the matter, the form, the mover, "that for the sake of which".

—Aristotle, *Physics*, Book II, Chapters 3 and Chapter 7

1.4 The Greatest Achievements of the Ancient Sciences

The period of the great syntheses was over; the Hellenistic era brought forth scientific specialists: mathematicians, astronomers, geographers, physicists, physicians, botanists, and even engineers, in the modern sense of the word. The different branches of science develop independently, but the personalities still connect them together. Archimedes is a mathematician, a physicist, and an engineer. Euclid is most famous for his geometry, but writes books on optics and even on music. Apollonius is already a mathematician in a narrow specialty: he is the classical exponent of the theory of conic sections. Even the most universal scholar in this group, Archimedes, is a *specialist* rather than a philosopher, or more precisely, he does not strive to synthesize a unified worldview from his understanding of the details. Thus, in the books on the history of philosophy their names are mentioned at most in passing.

Yet this attitude is unjust because it is silent on the decisive contributions of the Hellenistic specialists to the way we think. Euclidean geometry, besides its immediate influence, for example, on Spinoza and Kant, had immeasurable indirect influence on the spirit of Western thought. It is said that after the Bible, Euclid's *Elements* has had the greatest number of editions. Thus, for more than two thousand years, many generations have learned, with the help of this book, to think logically and to enjoy the beauty of a work constructed entirely of logical elements.

1.4.1 *Archimedes*

Archimedes (see Figure 1.61) was the first to forge a synthesis between mathematics and physics, liberating this connection from Pythagorean mysticism. He thus became a role model for the scientific revolution of the seventeenth century, and therefore our scientific worldview and our entire technological civilization with all its superstructure can ultimately seek their origins in him.

If we were to examine a current physics textbook, say at the high school level, whose subject matter is not the history of physics but an exposition of the physical principles still valid today, the earliest name that we would find is that of Archimedes. The oldest physical law that is still valid in its original formulation is the Archimedean principle of the buoyancy experienced by a body immersed in a fluid.

We meet the name Archimedes in other connections as well. As one of the greatest mathematicians and physicists of the ancient world (if not *the* greatest), he was highly regarded even in his lifetime, with the result that a number of episodes from his life have come down to us (Quotation 1.29).

The Roman architect Vitruvius (fl. first century BCE) relates that Hiero II (king of Syracuse from 270 to 215 BCE) commissioned a crown of pure gold to be made as a gift to the gods. A goldsmith was given the specifications and the requisite amount of gold. On receiving the finished crown, the king declared himself satisfied with the workmanship, but he somehow began to entertain the suspicion—although the weight of the crown and that of the original gold agreed—that the goldsmith had pocketed some of the gold and replaced the missing gold with silver. Archimedes was called in to determine whether the crown was in fact made of pure gold or of a gold–silver alloy, under the condition, however, that the crown not be damaged in any way. Archimedes thought and thought until one day, as he climbed into a tub at the baths and observed the water rising and overflowing the rim, the solution suddenly came to him. He was so excited that he ran naked and dripping wet through the streets of Syracuse to the king, shouting the word that has come down to us today as an expression of the triumph of discovery: *Eureka!* (I have found it!)

According to Vitruvius, Archimedes proceeded as follows: He weighed the crown, and then measured out a lump of pure gold whose weight was equal to that of the crown, and likewise a lump of silver. He then placed the crown in a bucket of water and measured the quantity of water that overflowed. He did likewise for the gold and for the silver (Figure 1.62). In this way, the volumes of the gold, silver, and crown were determined. Obviously, if the crown were made of pure gold, its volume would be equal to that of the lump of gold. And if the crown were made of pure silver, then it would displace a much larger quantity of water, since the density of silver is considerably less than that of gold, and therefore the volume of a quantity of silver is much greater than that of an equal weight of gold. As it turned out, the amount of water displaced by the crown was somewhere between the amount displaced by the gold and that displaced by the silver. Thus, Archimedes had verified that the crown in fact contained an alloy of gold and silver, and the guilt of the goldsmith was convincingly proven.

In fact, the three measured volumes of water allow us to determine the percentage of silver that was in the crown. Let W denote the weight of the crown, which is equal by hypothesis to the weight of the lump of gold and that of the lump of silver: $W = W_{\text{crown}} = W_{\text{gold}} = W_{\text{silver}}$. Let w_{gold} denote the (unknown) weight of the gold in the crown, and w_{silver} the (unknown) weight of the crown's silver. We may then write down the following relationships between the weights of gold and silver: The weight of the crown is the sum of the weights of the gold and the silver that constitute it, namely

$$W_{\text{crown}} = W = w_{\text{gold}} + w_{\text{silver}}.$$

received its name from its place in the edition published by Andronicus (τά μετά τά φυσιχά, which follows physics). In addition, Aristotle's works on ethics, politics, rhetoric, poetics, and economics have been preserved.

In Western Europe, the writings of Aristotle became known through Boethius, who, however, frequently relied on the commentaries on Aristotle written around 275 BCE by the Neo-Platonist Porphyry. At the beginning of the thirteenth century, the study of Aristotle's works was forbidden, but by the end of the century he had become *the* philosopher, due to the efforts of Albertus Magnus and Thomas Aquinas. From the end of the fourteenth century, under the influence of the Renaissance, all the works of Aristotle became available in the original Greek.

> Plato relates himself to the world as a blessed spirit whom it pleases sometimes to stay for a while in the world; he is not so much concerned to come to know the world, because he already presupposes it, as to communicate to it in a friendly way what he brings along with him and what it needs. He penetrates into the depths more in order to fill them with his being than in order to investigate them. He moves longingly to the heights in order to become again a part of his origin. All that he utters is connected to an eternally whole, good, true, beautiful, that he seeks to inspire in every breast. What he devotes himself to regarding worldly knowledge dissolves, one may say, evaporates in his method, in his presentation.
>
> Aristotle, in contrast, stands to the world as a man, an architect. He is only here once and must here make and create. He inquires about the Earth, but not farther than to find a ground. He is indifferent to everything from there to the center of the Earth. He draws a huge circumference for his building, procures materials from all sides, arranges them, piles them up, and climbs thus in regular form, pyramid fashion to the top, whereas Plato, like an obelisk, indeed like a pointed flame, seeks the heavens.
>
> —Goethe, *History of Color Theory* [adopted from Verene 1979, pp.50–51]

Quotation 1.28

Thereupon Lucius laughed and said: Oh, sir, just don't bring suit against us for impiety as Cleanthes thought that the Greeks ought to lay an action for impiety against Aristarchus the Samian on the ground that he was disturbing the hearth of the universe because he sought to save [the] phenomena by assuming that the heaven is at rest while the earth is revolving along the ecliptic and at the same time is rotating about its own axis. We express no opinion of our own now; but those who suppose that the moon is earth, why do they, my dear sir, turn things upside down any more than you do who station the earth here suspended in the air? Yet the earth is a great deal larger than the moon according to the mathematicians who during the occurrences of eclipses and the transits of the moon through the shadow calculate her magnitude by the length of time that she is obscured. ...

continued on next page

Quotation 1.28, ***continued***

Yet the moon is saved from falling by its very motion and the rapidity of its revolution, just as missiles placed in slings are kept from falling by being whirled around in a circle. For each thing is governed by its natural motion unless it be diverted by something else. That is why the moon is not governed by its weight: the weight has its influence frustrated by the rotary motion. ...

All the same, let us assume, if you please, that the motions of earthy objects in the heaven are contrary to nature; and then let us calmly observe without any histrionics and quite dispassionately that this indicates not that the moon is not earth but that she is earth in an "unnatural" location. For the fire of Aetna too is below earth "unnaturally," but it is fire; and the air confined in skins, though by nature it is light and has an upward tendency, has been constrained to occupy an "unnatural" location. ...

If not a single one of the parts of the cosmos ever got into an "unnatural' condition but each one is "naturally" situated, requiring no transposition or rearrangement and having required none in the beginning either, I cannot make out what use there is of providence or of what Zeus, "the master-craftsman" is maker and father-creator. ...

The fact is in brief, my dear Aristotle, that regarded as earth the moon has the aspect of a very beautiful, august, and elegant object; but as a star or luminary or a divine and heavenly body she is, I am afraid, misshapen, ugly, and a disgrace to the noble title, if it is true that of all the host in heaven she alone goes in about in need of alien light, as Parmenides says "Fixing her glance forever on the sun." ...

Though, as you know, Lamprias, I am as eager as any of you to hear what is going to be said, I should like before that to hear about the beings that are said to dwell on the moon—not whether any really do inhabit it but whether habitation there is possible. If it is not possible, the assertion that the moon is an earth is itself absurd, for she would then appear to have come into existence vainly and to no purpose, neither bringing forth fruit nor providing for men of some kind an origin, an abode, and a means of life, the purposes for which this earth of ours came into being, as we say with Plato, "our nurse, strict guardian and artificer of day and night." ...

continued on next page

The volume of the gold in the crown is equal to the volume of a unit weight of gold times the weight of the gold in the crown. That is,

$$v_{\text{gold}} = \frac{V_{\text{gold}}}{W_{\text{gold}}} w_{\text{gold}},$$

where v_{gold} is the volume of the lump of gold. An analogous result holds for the silver:

$$v_{\text{silver}} = \frac{V_{\text{silver}}}{W_{\text{silver}}} w_{\text{silver}}.$$

From this, we obtain the following expression for the volume of the crown:

$$V_{\text{crown}} = \frac{V_{\text{gold}}}{W_{\text{gold}}} w_{\text{gold}} + \frac{V_{\text{silver}}}{W_{\text{silver}}} w_{\text{silver}} = \frac{V_{\text{gold}}}{W} w_{\text{gold}} + \frac{V_{\text{silver}}}{W} w_{\text{silver}}.$$

We now define the ratio of the weight of the silver in the crown to that of the gold as the measure of the goldsmith's perfidy:

$$p = \frac{w_{\text{silver}}}{w_{\text{gold}}}.$$

Using this quantity, we can reformulate the above equations as follows:

$$W = w_{\text{gold}}(1 + p),$$
$$V_{\text{crown}} = w_{\text{gold}} \left(\frac{V_{\text{gold}}}{W} + p \frac{V_{\text{silver}}}{W} \right).$$

A bit of algebraic manipulation yields the following formula for the index of perfidy:

$$p = \frac{V_{\text{crown}} - V_{\text{gold}}}{V_{\text{silver}} - V_{\text{crown}}}.$$

We see here as well that there is no silver in the crown ($p = 0$) if the crown and the lump of gold displace the same volume of water.

The significance of Archimedes' method is that it leads directly to the concepts of specific gravity and specific volume, and furthermore, he offers a method for measuring them—a method that is used even today.

According to other sources, Archimedes solved the problem using what he called the principle of hydrostatics, which is named after him. This principle, which even makes an appearance in a student drinking song, appears as Proposition 16 in Archimedes' work *On Floating Bodies* in the following form: *Any body that is less dense than water attempts, on being completely immersed, to rise with a force that is equal to the difference in the weight of the water displaced by the body and the weight of the body itself. If a body is denser than water, it will sink with a force that is the difference between the weight of the body and the weight of the water displaced.*

Using this principle, Archimedes posed and solved a number of very complicated problems. His most notable results are related to the stability of floating bodies. We shall have more to say about this a bit later.

The principle of the lever was known before the time of Archimedes, but Archimedes brought systematic organization to all of the knowledge then existing and also extended this knowledge, both theoretically and in practice. He may thus be considered to be the founder of theoretical mechanics as a modern science. Archimedes was well aware of the practical significance of his calculations. From him comes the following statement regarding the lever: Give me a fixed place to stand, and I will lift the Earth from its hinges (δός μοι ποϋ στώ καί κινώ τήν γήν). See Figure 1.63.

Archimedes gave the following axioms for his statics:

1. A symmetrically weighted lever is in a state of equilibrium.
2. All of the weight acts at the point of suspension.

Archimedes considers both axioms to be self-evident. In the case of a symmetric load, there is no reason for the lever to tip to one side rather than the other. The second axiom is also easily accepted: all of the weight must be supported somewhere, and the point of suspension is the only possibility.

It is worth noticing that much later, Arab commentators replaced the second axiom by another, giving as authority a Greek scholar named Arsamides, which is apparently nothing more than a misspelling of the name Archimedes. This axiom states that if on a lever in a state of equilibrium equal weights are moved equal distances but in opposite directions, the lever will remain in a state of equilibrium.

Let us now see how the principle of the lever can be derived. This principle asserts that for a two-armed lever in equilibrium, the product of the weight and the length of the lever arm on both sides must be equal. That is, we must have

$$W_1 a_1 = W_2 a_2.$$

Let us load a lever symmetrically, as in Figure 1.64, by placing the same number of objects of equal weight at the same distances from the suspension point. Let the weight of each object be given by W_0, and let the distance between each object be denoted by t. We now divide the collection of objects into two groups, with m objects in the first group and n in the second when there are $m + n$ objects altogether. The total weight of the m objects is given by $W_1 = mW_0$, while that of the n remaining objects is $W_2 = nW_0$. In accordance with Figure 1.64, we replace the m objects by the resultant weight acting at the midpoint of the m objects, and we do the same for the group of n objects. We may now write down the following relationships for the lever arms with weights W_1 and W_2:

$$a_1 = a - \frac{mt}{2}, \quad a_2 = a - \frac{nt}{2},$$

and, because

$$mt + nt = 2a,$$

we obtain the following for the relationship of the lever arms:

$$\frac{a_1}{a_2} = \frac{a - \dfrac{mt}{2}}{a - \dfrac{nt}{2}} = \frac{2a - mt}{2a - nt} = \frac{(mt + nt) - mt}{(mt + nt) - nt} = \frac{nW_0}{mW_0} = \frac{W_2}{W_1}.$$

But this is exactly the principle of the lever.

We shall also give a derivation of the principle of the lever for a special case using the axioms that have come down to us from the Arabs. Figure 1.65 shows a symmetrically loaded two-armed lever, with four objects of equal weight, two attached to the suspension point and one at the end of each lever arm. We now move one of the weights from the suspension point one unit of distance to the right, and the weight at the end of the right-hand lever arm one unit to the left. According to the second axiom, the new configuration is also one of equilibrium. If this procedure is carried out one more time, we obtain, corresponding to the last picture in the figure, a form of the lever principle valid for this particular case: Three weights will balance a single equal weight located three times the distance from the suspension point.

There are even legends concerning Archimedes' death and his grave. In the Second Punic War, the Sicilians were allied with Carthage against Rome. During the Romans' siege of Syracuse under the command of Marcellus (ca. 268–208 BCE), Archimedes brought all his engineering knowledge to bear in the defense of his

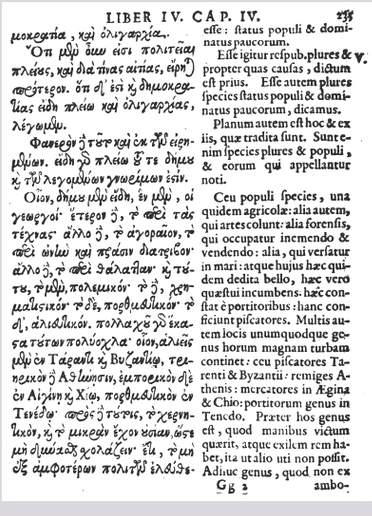
Quotation 1.28, ***continued***

So, for example, in the first place, if the moon is not inhabited by men, it is not necessary that she have come to be in vain and to no purpose, for we see that this earth of ours is not productive and inhabited throughout its whole extent either but only a small part of it is fruitful of animals and plants on the peaks, as it were, and peninsulas rising out of the deep, while of the rest some parts are desert and fruitless with winter-storms and summer-droughts and the most are sunk in the great sea. ...

It is plausible that the men on the moon, if they do exist, are slight of body and capable of being nourished by whatever comes their way. ...

We have no comprehension of these beings, however, nor of the fact that a different place and nature and temperature are suitable to them. Just as, assuming that we were unable to approach the sea or touch it but only had a view of it from afar and the information that it is bitter, unpotable, and salty water, if someone said that it supports in its depths many large animals of multifarious shapes and is full of beasts that use water for all the ends that we use air, his statements would seem to us like a tissue of myths and marvels, such appears to be our relation to the moon and our attitude towards her is apparently the same when we disbelieve that any men dwell there. Those men, I think, would be much more amazed at the earth, when they look out at the sediment and dregs of the universe, as it were, obscurely visible in moisture, mists, and clouds as a lightless, low, and motionless spot, to think that it engenders and nourishes animate beings which partake of motion, breath, and warmth.

—Plutarch, "The Face on the Moon"


LIBER IV. CAP. IV. 278

μοκρατία, καὶ ὀλιγαρχία.

Ὅτι μὲν οὖν εἰσὶ πολιτεῖαι πλείους, καὶ διὰ τίνας αἰτίας, εἰρῆ σθω πρότερον. ὅτι δ' ἐστὶ καὶ δημοκρατίας εἴδη πλείω καὶ ὀλιγαρχίας, λέγωμεν.

Φανερὸν δ' ἐστὶ καὶ ἐκ τῶν εἰρημένων. εἴδη γὰρ πλείω δ' τε δήμου καὶ τῶν λεγομένων γνωρίμων ἐστίν.

Οἷον, δήμου μὲν εἴδη, ἓν μὲν, οἱ γεωργοί· ἕτερον δ', τὸ ἐκ τῶν τέχνας· ἄλλο δ', τὸ ἀγοραῖον, τὸ ὠνίου καὶ πράσεως διαφέρον· ἄλλο δ', τὸ ἐκ θαλάττων· καὶ τούτου, τὸ μὲν, πολεμικόν· τὸ δ', χρηματιστικόν· τὸ δὲ, ποριστικόν· τὸ δ', ἁλιευτικόν. πολλαχοῦ δὲ ἕκαστα τούτων πολύοχλα· οἷον, ἁλιεῖς μὲν ὀλίγαι Τάραντι καὶ Βυζαντίῳ, τεκτονικὸν δ' Ἀθηναίοις, ἐμπορικὸν δὲ ἐν Αἰγίνῃ καὶ Χίῳ, ποριστικὸν ἐν Τενέδῳ· ἐκ δὲ τούτων, τὸ χρηματιστικόν, καὶ τὸ μικρῶν ἔχον ἐσαεὶ, ὥστε μὴ δυνάσθαι σχολάζειν· ἔτι, τὸ μὴ ἐξ ἀμφοτέρων πολιτῶν ἐνδέχε-

esse: ſtatus populi & dominatus paucorum.

Eſſe igitur reſpub. plures & propter quas cauſas, dictum eſt prius. Eſſe autem plures ſpecies ſtatus populi & dominatus paucorum, dicamus.

Planum autem eſt hoc & ex iis, quæ tradita ſunt. Sunt enim ſpecies plures & populi, & eorum qui appellantur noti.

Ceu populi ſpecies, una quidem agricolæ: alia autem, qui artes colunt: alia forenſis, qui occupatur inemendo & vendendo: alia, qui verſatur in mari: atque hujus hæc quidem dedita bello, hæc vero quæſtui incumbens. hæc conflat è portitoribus: hanc conficiunt piſcatores. Multis autem locis unumquodque genus horum magnam turbam continet: ceu piſcatores Tarenti & Byzantii: remiges Athenis: mercatores in Ægina & Chio: portitorum genus in Tenedo. Præter hos genus eſt, quod manibus victum quærit, atque exilem rem habet, ita ut alio uti non poſſit. Adhuc genus, quod non ex

Gg 3 ambo-

◀ Figure 1.60 The *Politics* of Aristotle was published in 1656 in a Greek–Latin parallel-text edition, not just as a work of historical interest, but as a textbook.



◀ Figure 1.61
Archimedes (ca. 287–ca. 212 bce): born in Syracuse. His father, Phidias, had an interest in astronomy. Archimedes had close contact with the ruling family—in particular to King Hiero II and his son Gelon; he may even have been related to them. He studied in Alexandria with the successor to Euclid, perhaps even under Euclid himself. He spent most of his life in his native Syracuse, where he was killed during the occupation of the city by the Romans.

Many works of Archimedes, most in the form of short articles, have been preserved:

1. *On the Sphere and the Cylinder* (Περί σφαίρας καί κυλίνδρου)
2. *On the Measurement of the Circle* (Κύκλου μέτρησις)
3. *On Conoids and Spheroids* (Περί κωνοειδέων καί σφαιροειδέων)
4. *On Spirals* (Περί έλίκον)
5. *On the Equilibrium of Planes* (Περί έπιπέδον ισρροιώννή)
6. *On the Quadrature of the Parabola* (Τετραγωνισμός παραβολής)
7. *On Floating Bodies* (Περί οκουμένων)
8. *On the Sand Reckoner* (Ψαμμίτης); Latin *Arenarius*
9. *On Method* (Έφοδος)

We should mention also the *Collection of Lemmas* and the *Cattle Problem*.

From various texts, it is possible to reconstruct the title or contents of some lost works of Archimedes:

1. *Investigations on Polygons*
2. *On Principles* (Άρχαί)
3. *On the Balance and the Lever* (Περί ζυγών)
4. *On the Center of Gravity* (Κεντροβαρικά)
5. *Optics* (Κατοπτρικά)
6. *On Sphere-Making* (Περί Σφαιροποίας)

The last of these describes a mechanical model that demonstrates the motion of the Sun, the Moon, and the five planets.

The most valuable results of Archimedes were without immediate effect; it was fifteen hundred years before the scientific world reached a level that allowed for their proper understanding and appreciation. The first complete edition of his works did not appear until 1544, and only after the publication of two other great books of historical importance: *De revolutionibus orbium coelestium* of Copernicus and *De humani coporis fabrica* of Vesalius.

native city (Quotation 1.29). Although we may not believe all the stories, according to which he set the Roman ships aflame with the help of mirrors, or lifted ships out of the sea with levers and hooks that reached out over the walls of the city and smashed them against the cliffs, it is nevertheless certain that Syracuse was able to withstand the siege for over a year, a feat that is also ascribed to Archimedes. It was only through treachery that the city was finally taken. Marcellus gave strict orders for the life of Archimedes to be protected at all costs. However, during the plunder and killing that followed the capture of the city, Archimedes was killed. Perhaps he was not recognized. Or perhaps he failed to obey an order. There are several stories about the death of Archimedes. The best-known version has Archimedes deep in thought about the geometric figures drawn in the sand before him, saying to a Roman soldier who approached, "Don't disturb my circles" (*Noli turbare circulos meos*).

Marcellus had Archimedes interred with great pomp and caused to be carved on his tombstone, honoring a wish that the mathematician had expressed, a drawing of a sphere inscribed in a cylinder, for among his numerous results, Archimedes valued the formula for the relationship between these two solids the highest (Quotation 1.29). Cicero (106–43 bce) mentions in one of his writings that during his time as praetor in Sicily, he happened to come upon a neglected gravestone hidden in heavy undergrowth on which was carved this geometric figure, by which he recognized it as the grave of Archimedes.

In his mathematical writings, Archimedes goes a long way toward today's standards of mathematical rigor to such an extent, in fact, that similar precision does not again appear until the nineteenth century. His usual method was to show that the negation of an assertion that he wished to prove leads to a logical contradiction (*reductio ad absurdum*). To do this, of course, one has to be able to state the assertion precisely; in other words, the truth to be proven has to be known beforehand. Let us see how this works with the help of an example. Consider a theorem of Archimedes on the area of a parabolic segment. Archimedes' theorem asserts that the area of such a segment is equal to four-thirds the area of the inscribed triangle, as shown in Figure 1.66. In the proof, he shows that an assertion that the area of the parabolic segment is greater than and an assertion that it is less than this value both lead to a contradiction, and therefore the area can only be equal to the value given. Logically, such a proof is unimpeachable. But naturally the question arises, how did Archimedes know that deviations up and down from this particular value will cause contradictions? Or simply said, how did Archimedes know that the area of a parabolic segment is four-thirds the area of the inscribed triangle? This question is also of pedagogical relevance.

Namely, if we present students with the proof formulated in this manner, they might wonder how they themselves could find the result to be proven in other problems they might face. In the history of science, only a few great personalities have allowed us a glimpse into the birth of a great idea, with all its pains and false starts. Thus we have reason to rejoice in the works of Kepler, who in his writings gives us not only the final results, but all his doubts and erroneous suppositions along the way in full detail. It is far more typical for the public to encounter well-groomed thoughts decked out in formal attire. This is why the surprising discovery in 1906 of Archimedes' work *On Method* was of great significance for the history of science. The Danish historian Heiberg discovered such writings of Archimedes on a palimpsest—a parchment that had been reused by scratching out an

earlier text. In the introduction to this work, contained in a letter addressed to Eratosthenes (ca. 276–194 bce), Archimedes describes the method by which he obtained his solutions (Quotation 1.30). It turns out that he frequently employed mechanical analogies based on the principle of the lever to obtain results in a way that he himself saw as insufficiently rigorous, and then set about to prove them rigorously. Archimedes described his method with clear pedagogical intent, in order to make it possible for others to achieve similar results. This method, which may be called intuitive, is by no means simple. To reconstruct the train of thought often requires a detailed knowledge of the theory of conic sections as well as of solid geometry. To help us in taking the measure of the greatness of this thinker, we will trace the path by which Archimedes arrived at his theorems by an analogous consideration of a mechanical model leading ultimately to a precise proof. Then we will discuss the theorem that Archimedes himself considered his finest result and that is inscribed on his tombstone. We will also consider the problem of the stability of floating bodies. From these examples, it will be evident that Archimedes' mathematical results already contain the germ of integral calculus.

The theorem to be proved is the following: The area of a parabolic segment is equal to four-thirds the area of the inscribed triangle (see Figure 1.66). As we have already mentioned, Archimedes proved this theorem with the method of *reductio ad absurdum*, which means, of course, that he must have known the result beforehand. Therefore, we would like first to obtain the relationship

$$A_{\text{parabola}} = \frac{4}{3} A_{\text{triangle}},$$

using the techniques of *On Method*. To begin, we recall some properties of parabolas, which we shall derive using today's methods, since the details of the derivation are not of significance for this discussion.

Figure 1.67(a) shows a parabola whose axis of symmetry is the *y*-axis. Its equation is given by

$$y = a - bx^2.$$

We stress again that equations of this sort were *completely unknown in antiquity, regarding both content and form*. The figure shows the tangent line to the parabola at point *B*. The equation of that line is

$$y = -2\sqrt{ab}\left(x - \sqrt{a/b}\right).$$

We then draw a line parallel to the *y*-axis that crosses the *x*-axis at point *D*, which is at the arbitrary distance x from the origin. The intersection points of this line with the parabola and the tangent are denoted, respectively, by *E* and *F*. In what follows, we will need the theorem that for an arbitrary point *D*, we have the relation

$$\frac{DE}{DF} = \frac{AD}{AB}.$$

This relationship is easily derived from the given equations for the parabola and tangent, since for the parabola we have

$$y\big|_{x=\xi} = \left(a - b\xi^2\right) = \left(\sqrt{a} + \sqrt{b}\xi\right)\left(\sqrt{a} - \sqrt{b}\xi\right),$$

and for the tangent,

$$y\big|_{x=\xi} = -2\sqrt{ab}\left(\xi - \sqrt{a/b}\right) = 2\sqrt{a}\left(\sqrt{a} - \sqrt{b}\xi\right).$$

Quotation 1.29

And therefore we may not disbelieve the stories told about him, how, under the lasting charm of some familiar and domestic Siren, he forgot even his food and neglected the care of his person; and how, when he was dragged by main force, as he often was, to the place for bathing and anointing his body he would trace geometrical figures in the ashes, and draw lines with his finger in the oil with which his body was anointed, being possessed by a great delight, and in very truth a captive of the Muses. And although he made many excellent discoveries, he is said to have asked his kinsman and friends to place over the grave where he should be buried a cylinder enclosing a sphere, with an inscription giving the proportion by which the containing solid exceeds the contained ...

And yet Archimedes possessed such a lofty spirit, so profound a soul, and such a wealth of scientific theory, that although his inventions had won for him a name and fame for superhuman sagacity, he would not consent to leave behind him any treatise on this subject, but regarding the work of an engineer and every art that ministers to the needs of life as ignoble and vulgar, he devoted his earnest efforts only to those studies the subtlety and charm of which are not affected by the claims of necessity. These studies, he thought, are not to be compared with any others; in them the subject matter vies with the demonstration, the former supplying grandeur and beauty, the latter precision and surpassing power ...

For the art of mechanics, now so celebrated and admired, was first originated by Eudoxus and Archytas, who embellished geometry with its subtleties, and gave to problems incapable of proof by word and diagram, a support derived from mechanical illustrations that were patent to the senses. For instance, in solving the problem of finding two mean proportional lines, a necessary requisite for many geometrical figures, both mathematicians had recourse to mechanical arrangements, adapting to their purposes certain intermediate portions of curved lines and sections. But Plato was incensed at this, and inveighed against them as corrupters and destroyers of the pure excellence of geometry, which thus turned her back upon the incorporeal things of abstract thought and descended to the things of sense, making use, moreover, of objects which required much mean and manual labor. For this reason mechanics was made entirely distinct from geometry, and being for a long time ignored by philosophers, came to be regarded as one of the military arts.

And yet even Archimedes, who was a kinsman and friend of King Hiero, wrote to him that with any given force it was possible to move any given weight; and emboldened, as we are told, by the strength of his demonstration, he declared that, if there were

continued on next page

Quotation 1.29, ***continued***

another world, and he could go to it, he could move this. Hiero was astonished, and begged him to put his proposition into execution, and show him some great weight moved by a slight force. Archimedes therefore fixed upon a three-masted merchantman of the royal fleet, which had been dragged ashore by the great labors of many men, and after putting on board many passengers and the customary freight, he seated himself at a distance from her, and without any great effort, but quietly setting in motion with his hand a system of compound pulleys, drew her towards him smoothly and evenly, as though she were gliding through the water. Amazed at this, then, and comprehending the power of his art, the king persuaded Archimedes to prepare for him offensive and defensive engines to be used in every kind of siege warfare. These he had never used himself, because he spent the greater part of his life in freedom from war and amid the festal rites of peace; but at the present time his apparatus stood the Syracusans in good stead, and, with the apparatus, its fabricator.

When, therefore, the Romans assaulted them by sea and land, the Syracusans were stricken dumb with terror; they thought that nothing could withstand so furious an onset by such forces. But Archimedes began to ply his engines, and shot against the land forces of the assailants all sorts of missiles and immense masses of stones, which came down with incredible din and speed; nothing whatever could ward off their weight, but they knocked down in heaps those who stood in their way, and threw their ranks into confusion. At the same time huge beams were suddenly projected over the ships from the walls, which sank some of them with great weights, plunging down from on high; others were seized at the prow by iron claws, or beaks like the beaks of cranes, drawn straight up into the air, and then plunged stern foremost into the depths, or were turned round and round by means of enginery within the city, and dashed upon the steep cliffs that jutted out beneath the wall of the city, with great destruction of the fighting men on board, who perished in the wrecks. Frequently, too, a ship would be lifted our of the water into mid-air, whirled hither and thither as it hung there, a dreadful spectacle, until its crew had been thrown out and hurled in all directions, when it would fall empty upon the walls, or slip away from the clutch that had held it. ...

At last the Romans became so fearful that, wherever they saw a bit of rope or a stick of timber projecting a little over the wall, "There it is," they cried, "Archimedes is training some engine upon us," and turned their backs and fled. Seeing this, Marcellus desisted from all fighting and assault, and thenceforth depended on a long siege.

—Plutarch, *Lives* [pp. 471–481]

From this, we obtain

$$\frac{DE}{DF} = \frac{a - b\xi^2}{-2\sqrt{ab}\left(\xi - \sqrt{a/b}\right)} = \frac{\xi + \sqrt{a/b}}{2\sqrt{a/b}} = \frac{AD}{AB}.$$

It is even easier to see, with the help of Figure 1.67(b), that we have the following relationship between the areas of triangles *ABC* and *AEB*:

$$ABC_{\text{triangle}} = 4AEB_{\text{triangle}}.$$

On the same figure and as can be read off from the equations of parabolas and their tangents, when $x = 0$, the length of the segment *DF* is twice that of *DE*, and therefore, triangle *BDF* has twice the area of triangle *BDE*, and triangle *AFB* has twice the area of triangle *AEB*. But since the area of triangle *AEB* must then be equal to the area of triangle *FDB*, we have

$$ABC_{\text{triangle}} = 4FDB_{\text{triangle}} = 4AEB_{\text{triangle}}.$$

We come now to the heart of Archimedes' method: How can the principle of the lever be used to determine—even if not completely rigorously—the area of parabolic segments? To do so, triangle *ABC* is hung by its center of mass from the end of a two-armed lever (Figure 1.67(c)). We select a very small vertical slice, denoted by *D'F'* in the figure. The width of the slice is given—using modern notation—by d*x*. According to Archimedes, the total area of the parabolic segment can be expressed as the sum of vertical lines similar to *D'F'*, just as a piece of canvas has area equal to the "sum of the threads." Because of the weakness of this form of argumentation, Archimedes did not use it as a proof, but just as an intuitive aid in understanding. Using the properties of the parabola derived above yields

$$\frac{D'E'}{D'F'} = \frac{AD'}{AB},$$

which becomes, after a bit of manipulation,

$$D'E' \cdot AB = D'F' \cdot AD'.$$

This relationship can be interpreted as saying that the small slice *D"F"* with a lever arm *AD'* balances the slice *D'E'* on the other side of the lever with lever arm *A'B' = AB*. One can also write down analogous relationships for the slices *D"F"*, *D"E"* and *DF*, *DE*. Finally, we obtain that the entire parabolic segment is hanging by the common lever arm *A'B' = AB*. Therefore, the parabolic segment balances triangle *ABC*, since each of its slices, with a lever arm corresponding to its position, contributes to the equilibrium. The contributions of the slices of the triangle can be collected into a single resultant located at the triangle's center of mass and proportional to the area of the triangle. The lever principle then tells us that

$$A'B' \cdot AEB_{\text{parabola}} = A'S' \cdot ABC_{\text{triangle}}.$$

It is well known that the triangle's center of mass is located at

$$A'S' = \frac{1}{3}AB,$$

so that we have

$$AEB_{\text{parabola}} = \frac{1}{3}ABC_{\text{triangle}}.$$

Earlier, we derived a relationship between the area of the triangle *ABC* employed in our considerations of equilibrium and the area of the triangle inscribed in the parabolic segment. If we now make use of the relationship $ABC_{\text{triangle}} = 4AEB_{\text{triangle}}$, we obtain

$$AEB_{\text{parabola}} = \frac{4}{3}AEB_{\text{triangle}}.$$

According to Archimedes, we have now reached the point where we can formulate the theorem to be proved precisely and begin with a rigorous proof. To begin with, we will need two lemmas. The first relates to the areas of polygons that can be inscribed in a parabola that with increasing numbers of sides occupy the entire parabolic segment, as shown in Figure 1.68. To this end, we begin by inscribing the

triangle AEB in the parabolic segment and drawing two lines parallel to the parabola's axis, intersecting the parabola at points C_1 and C_2, thereby determining the triangles AC_1E and BC_2E. From the properties of a parabola it is easy to determine that for the area A_1 of triangle AEB and the two congruent triangles AC_1E and BC_2E, we have the relationship

$$AEB_{\text{triangle}} = A_1 = 4AC_1E + 4BC_2E = 4A_2,$$

where we have defined A_2 to be the sum of the areas of triangles AC_1E and BC_2E. Therefore,

$$A_2 = \frac{1}{4}A_1.$$

These relationships are most easily proved using the equation of a parabola. By dividing the segment AB into eight equal segments, we obtain four new smaller triangles, whose areas total

$$A_3 = A_1 \cdot \frac{1}{16} = A_1 \cdot \left(\frac{1}{4}\right)^2 = A_1 \cdot \left(\frac{1}{4}\right)^{3-1}.$$

Continuing in this way, we find that, in general,

$$A_n = A_1 \cdot \left(\frac{1}{4}\right)^{n-1}.$$

We see, then, that the amount by which the total area of the triangles increases in finer and finer decompositions can be made arbitrarily small.

The second lemma that we need says that summing the areas of all the triangles inscribed in the parabola in Figure 1.68 yields

$$A_1 + A_2 + \cdots + A_n + \frac{A_n}{3} = \frac{4}{3}A_1.$$

The proof of this lemma is simple. We simply sum the finite geometric series

$$1 + q + q^2 + \cdots + q^{n-1} = \frac{1-q^n}{1-q}$$

and obtain, with $q = 1/4$,

$$\frac{1-q^n}{1-q} + \frac{q^{n-1}}{3} = \frac{3\left(1-q^n\right) + \frac{3}{4}q^{n-1}}{\frac{3}{4}\cdot 3} = \frac{4-4q^n+q^{n-1}}{3},$$

and finally, observing that $4q^n = q^{n-1}$ for $q = 1/4$,

$$A_1\left(1 + q + \cdots + q^{n-1}\right) + A_1\frac{q^{n-1}}{3} = \frac{4}{3}A_1.$$

We would like to emphasize here that Archimedes did not work with infinite series and also did not use the notion of a limiting value—at least not explicitly.

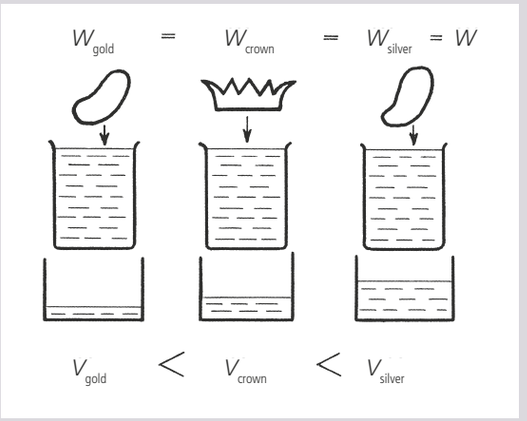
Everything necessary is now in place for us to carry out a rigorous proof. We begin with the theorem, which we have already proved, that

$$A_1 + A_2 + \cdots + A_n + \frac{A_n}{3} = \frac{4}{3}A_1. \qquad \text{(I)}$$

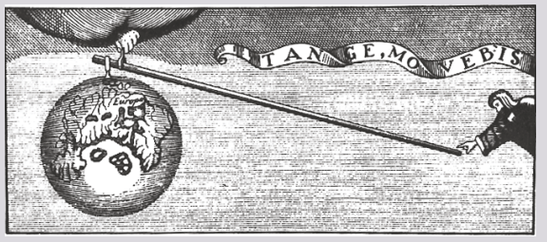
Let us denote the sum of the first n terms by S_n. It is then immediately clear that this partial sum satisfies the inequality

$$S_n < \frac{4}{3}A_1.$$

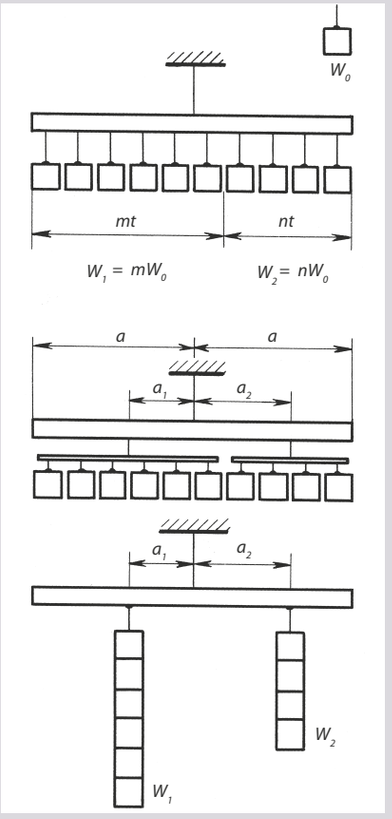
We next observe that the areas of the inscribed polygons, when taken together, are always smaller than the area of the parabolic segment, and furthermore, the difference between these two areas can be made



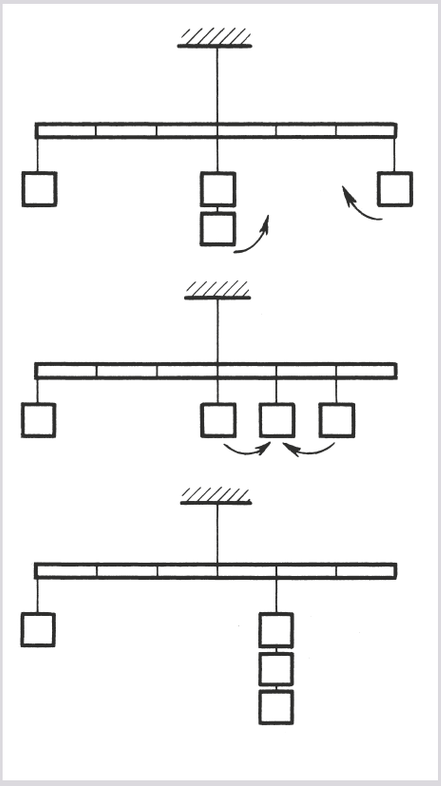
▲ **Figure 1.62** A lump of silver displaces almost twice as much water as a lump of pure gold of the same weight. The amount of water displaced by the crown, also of the same weight, lies between the two values; hence the crown clearly cannot be made of pure gold.



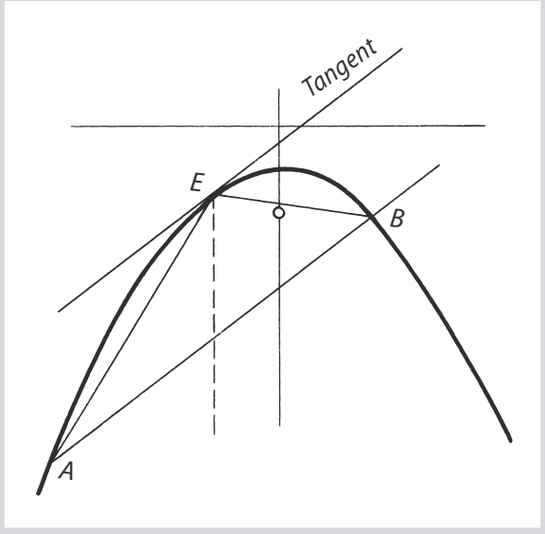
▲ **Figure 1.63** Varignon illustrated in 1687 the oft-quoted saying of Archimedes this way.



◀ **Figure 1.64** Archimedes' derivation of the principle of the lever.

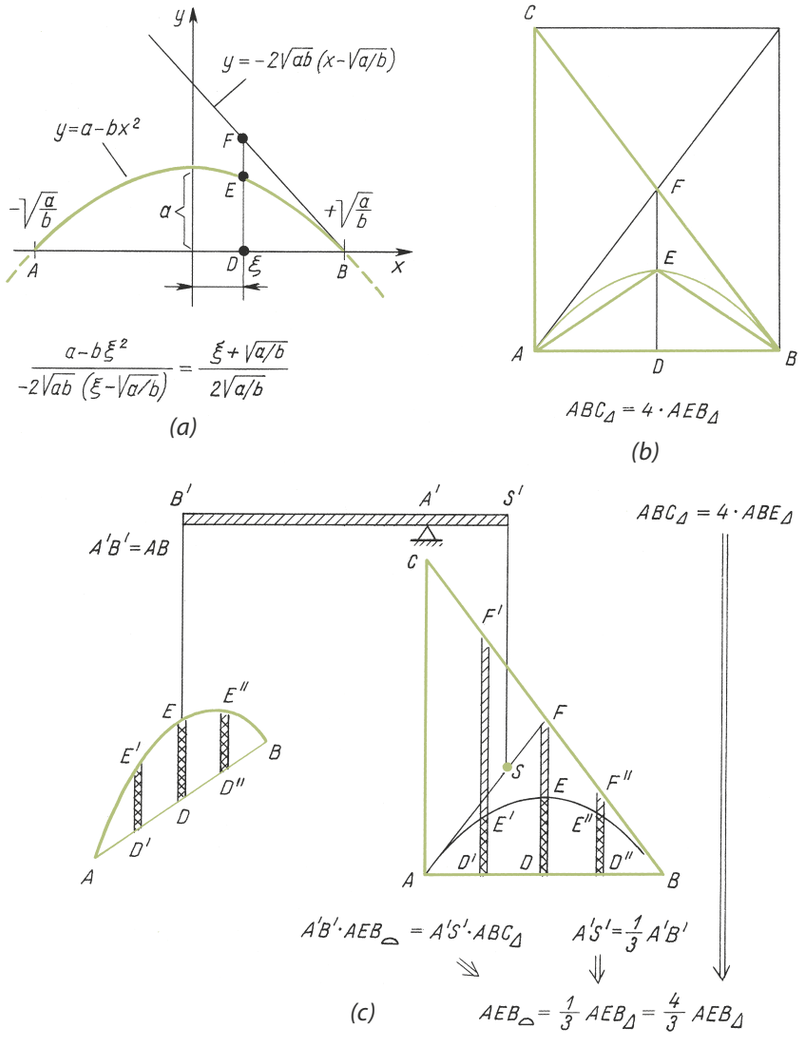


▲ **Figure 1.65** Arab commentators ascribed this derivation to ARCHIMEDES as well.



▲ **Figure 1.66** The ratio between the area of a parabolic segment and the area of the inscribed triangle *AEB* is given by

$$A_{\text{parabola}} = \frac{4}{3} A_{\text{triangle}}.$$



▲ **Figure 1.67** The method by which ARCHIMEDES determined the correct result, which then had to be proved rigorously.

arbitrarily small by taking a fine enough partition. That is, we have

$$S_n < A \equiv A_{\text{parabola}}, \quad \text{but} \quad A - S_n < \varepsilon \quad \text{if} \quad n > m(\varepsilon). \tag{II}$$

The proof of the theorem

$$AEB_{\text{parabola}} = \frac{4}{3} AEB_{\text{triangle}}$$

can now be proven by using our two lemmas and the method of *reductio ad absurdum.*

(a) We assume that the area of the parabolic segment is greater than four-thirds the area of the inscribed triangle; that is,

$$A > \frac{4}{3} A_1, \quad \text{or} \quad A - \frac{4}{3} A_1 = \delta,$$

for some positive number δ. We take a partition fine enough that the triangles with total area S_n fill out the area A up to a difference of ε, with $\varepsilon < \delta$. But that gives us

$$A > S_n > \frac{4}{3} A_1,$$

in contradiction to the lemma that tells us that

$$S_n < \frac{4}{3} A_1.$$

Our hypothesis that the area of the parabolic segment is greater than four-thirds the area of the inscribed triangle therefore leads to an absurd result, and so must be false.

(b) We now assume that the area of the parabolic segment is less than four-thirds the area of the inscribed triangle; that is, that

$$A < \frac{4}{3} A_1,$$

or equivalently,

$$\frac{4}{3} \mathrm{A}_1 - \mathrm{A} = \delta,$$

where again, δ is some prescribed positive number. We may choose the partition to be so fine that the last increase in area is less than this fixed number δ:

$$A_m < \frac{4}{3} A_1 - A.$$

Moreover, we have, of course, the inequality

$$\frac{4}{3} A_1 - S_m = \frac{1}{3} A_m < A_m,$$

and both inequalities together read

$$\frac{4}{3} A_1 - A > A_m > \frac{4}{3} A_1 - S_m,$$

from which we conclude that

$$S_m > A,$$

in contradiction to the inequality $S_m < A$. These two contradictions taken together yield a rigorous proof of the claim that the area of a parabolic segment is equal to four-thirds the area of the inscribed triangle.

As we have already mentioned, Archimedes considered the following theorem, which can be used to determine the volume of a sphere, as his finest achievement.

Figure 1.69 shows a slice through a sphere of diameter $OA = 2r$ and center C, a cylinder of height $2r$ and radius $2r$, and a cone with a base of radius $2r$. The cylinder and cone have the common axis OA; the vertex of the cone is at point O. If a perpendicular is drawn from an arbitrary point T anywhere on the axis OA, then by the Pythagorean theorem, we have for the right triangle OTK the relationship

$$OK^2 = OT^2 + TK^2.$$

Since the triangle OTK' is isosceles, we have $OT = TK'$, and we obtain

$$OK^2 = TK'^2 + TK^2. \qquad (1)$$

In the right triangle OKA, the leg OK is the geometric mean of the hypotenuse and the projection OA of the leg OK onto the hypotenuse:

$$OK^2 = OT \cdot OA = \frac{OT \cdot OA^2}{OA} = \frac{OT \cdot TZ^2}{OA}, \qquad (2)$$

where we have taken into account that $OA = TZ$. If we now combine equations (1) and (2), we obtain

$$TK'^2 + TK^2 = \frac{TZ^2 \cdot OT}{OA}.$$

We now multiply this equation by π and observe that the quantities $\pi TK'^2$, πTK^2, and πTZ^2 are the areas of the circles that arise from intersecting the plane passing through the point T and perpendicular to

Quotation 1.30

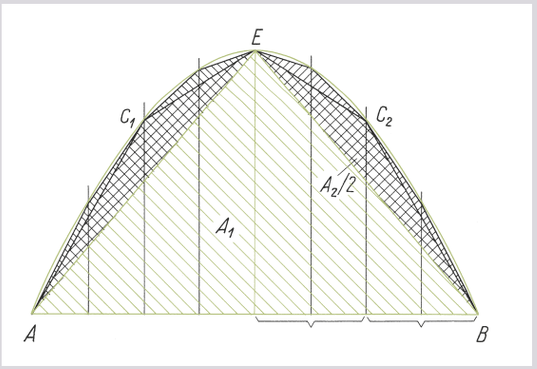
Archimedes to Eratosthenes greeting.

I sent you on a former occasion some of the theorems discovered by me, merely writing out the enunciations and inviting you to discover the proofs, which at the moment I did not give. ...

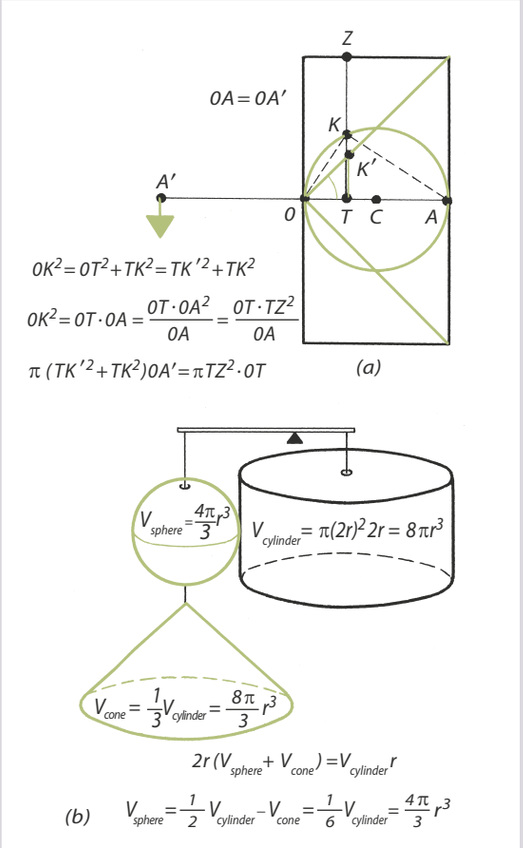
The proofs then of these theorems I have written in this book and now send to you. Seeing moreover in you, as I say, an earnest student, a man of considerable eminence in philosophy, and an admirer [of mathematical inquiry], I thought fit to write out for you and explain in detail in the same book the peculiarity of a certain method, by which it will be possible for you to get a start to enable you to investigate some of the problems in mathematics by means of mechanics. The procedure is, I am persuaded, no less useful even for the proof of the theorems themselves; for certain things first became clear to me by a mechanical method, although they had to be demonstrated by geometry afterwards because their investigation by the said method did not furnish an actual demonstration. But it is of course easier, when we have previously acquried, by the method, some knowledge of the questions, to supply the proof than it is to find it without any previous knowledge. This is a reason why, in the case of the theorems the proof of which Eudoxus was the first to discover, namely that the cone is a third part of the cylinder, and the pyramid of the prism, having the same base and equal height, we should give no small share of the credit to Democritus who was the first to make the assertion with regard to the said figure though he did not prove it. I am myself in the position of having first made the discovery of the theorem now to be published [by the method indicated], and I deem it necessary to expound the method partly because I have already spoken of it and I do not want to be thought to have uttered vain words, but equally because I am persuaded that it will be of no little service to mathematics; for I apprehend that some, either of my contemporaries or of my successors, will, by means of the method when once established, be able to discover other theorems in addition, which have not yet occurred to me.

First then I will set out the very first theorem which became known to me by means of mechanics, namely that *any segment of a section of a right-angled cone (i.e. a parabola) is four-thirds of the triangle which has the same base and equal height,* and after this I will give each of the other theorems investigated by the same method. Then, at the end of the book, I will give the geometrical [proofs of the propositions]

—Archimedes, *The Method* [pp. 11–14]



▲ **Figure 1.68** The "exhaustion" of the parabolic segment: By taking into account triangles with decreasing areas, we are able to approach the area of the parabolic segment more and more closely.



▲ **Figure 1.69** Determining the volume of a sphere according to Archimedes:

(a) The derivation of the relationships among the areas of the circles of radii *TK, TK′,* and *TZ* (cut from the sphere, cone, and cylinder).
(b) The equilibrium of objects made from the same material.

the segment OA with the cone, the sphere, and the cylinder. If we now bring the equation into the form

$$\left(\pi \cdot TK'^2 + \pi \cdot TK^2\right) \cdot OA = \pi \cdot TZ^2 \cdot OT,$$

we can interpret this again as an interesting application of the principle of the lever. If we hang thin slices of a circle cut from the sphere and cone on the left side of a two-armed lever with constant length $OA = OA'$ of the lever arm, these will balance the corresponding circles cut from the cylinder if its lever arm is OT. We now let the point T run along the segment OA, and we sum up the areas of the circular slices. We see, then, that the lever is in balance when the sphere and cone are on one side with lever arm $OA' = 2r$, and on the other side is the cylinder with lever arm $OA = r$ (that is, with these distances from the center of mass, which is the average of the distances OT). The resulting equation of equilibrium is then, with V denoting volume,

$$2r\left(V_{\text{sphere}} + V_{\text{cone}}\right) = rV_{\text{cylinder}},$$

which yields

$$V_{\text{sphere}} = \frac{1}{2}V_{\text{cylinder}} - V_{\text{cone}}.$$

It is well known, however, that a cylinder has three times the volume of a cone with the same base and altitude; that is,

$$V_{\text{cone}} = \frac{1}{3}V_{\text{cylinder}},$$

and so we have

$$V_{\text{sphere}} = \frac{1}{2}V_{\text{cylinder}} - \frac{1}{3}V_{\text{cylinder}} = \frac{1}{6}V_{\text{cylinder}}.$$

Since $V_{\text{cylinder}} = \pi(2r)^2 \cdot 2r = 8\pi r^3$, we conclude that

$$V_{\text{sphere}} = \frac{8\pi r^3}{6} = \frac{4\pi}{3}r^3.$$

Given the severe demands that Archimedes placed on the rigorousness of his proof, it is no surprise that he was not satisfied with the derivation described above. However, once he knew the correct formula to prove, he was able to work out a rigorous proof by using his method of exhaustion. We shall not reproduce that proof here.

Another brilliant work of Archimedes, from the point of view of conjoining mathematics and physics, has as its subject the solution of a physical problem, namely, that of the stability of a floating object, which is a problem of great practical importance. The method of solution given by Archimedes is used even today. When a floating object is deflected from equilibrium, the forces acting on it change in a particular way. What is of interest above all is whether the forces tend to bring the object back to equilibrium or whether the original deflection is magnified. To solve this problem, Archimedes uses a property of the parabola that was already known in his time: the projection of the normal to a parabola onto the parabola's axis is constant. A derivation of this theorem—using current methods—is shown in Figure 1.70(a). Out of the many related theorems, we shall discuss in detail only the following: An object in the form of a paraboloid of revolution floats stably if its height is no greater than 3p/2, where p is the above-defined constant of projection of the normal on the parabola's axis. Figure 1.70(b) shows a floating object deflected from its resting position. If the object is to float stably, the line of action of the buoyant force acting on the paraboloidal segment $VV'T$ at the point S^1 must lie to the left of the line of action S_0S_0' of the gravitational force acting at the object's center of mass, for in this case the torque attempts to turn the object back to its resting position. A sufficient condition for this is that the horizontal tangent of the paraboloid touch a point T lying to the left of the point S_0'. But this condition is satisfied whenever $OG > OS_0$. This is certainly true for $IG \leq OS_0$, where IG is just the projection constant p of the normal onto the paraboloid's axis. However, for a paraboloid of rotation, OS_0 is equal to two-thirds the height. This yields the stability equation

$$p \geq \frac{2}{3}h, \quad \text{or} \quad h \leq \frac{3}{2}p.$$

Finally, we should mention that ARCHIMEDES also considered the case of less-shallow paraboloids, for which

$$h \geq \frac{3}{2} p$$

holds. We shall not go into his investigations in detail, but simply give the result. Let s be the ratio of the specific gravity of the lighter object to that of the fluid in which it the object is immersed:

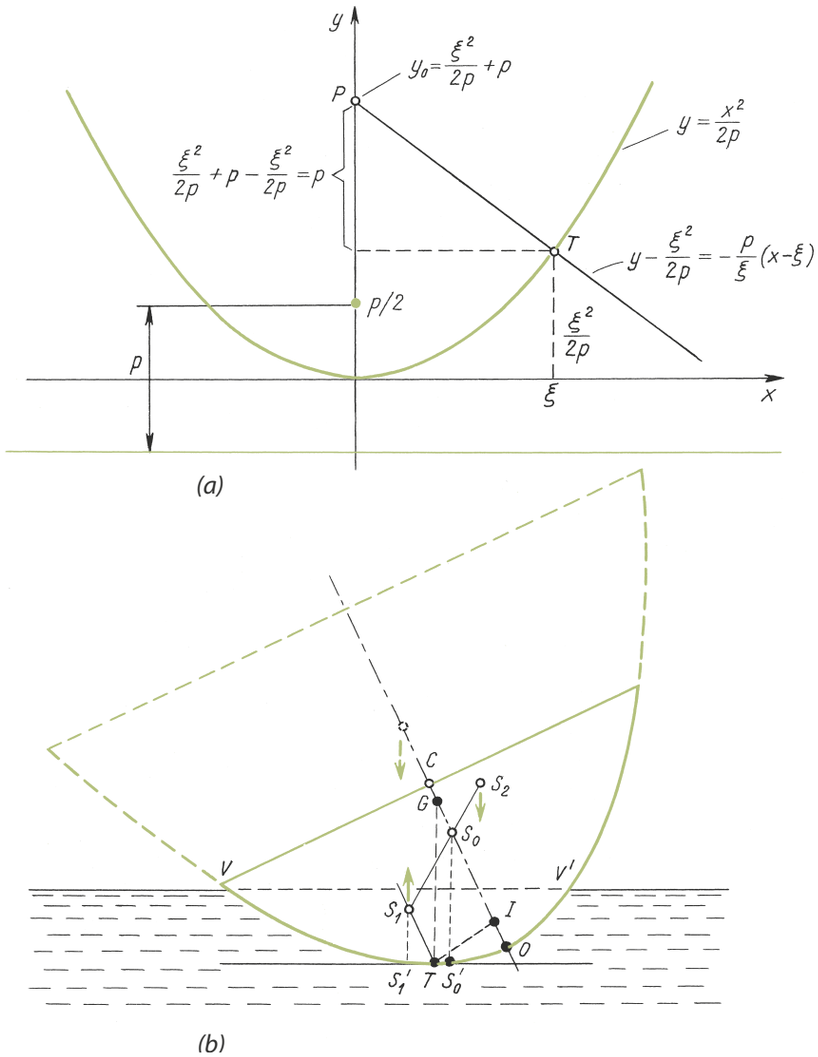
$$s = \frac{\mathrm{sg}_{\text{object}}}{\mathrm{sg}_{\text{fluid}}} < 1.$$

Then the paraboloid of rotation floats stably if

$$1 > s \geq \frac{\left(h - \frac{3}{2} p\right)^2}{h^2}.$$

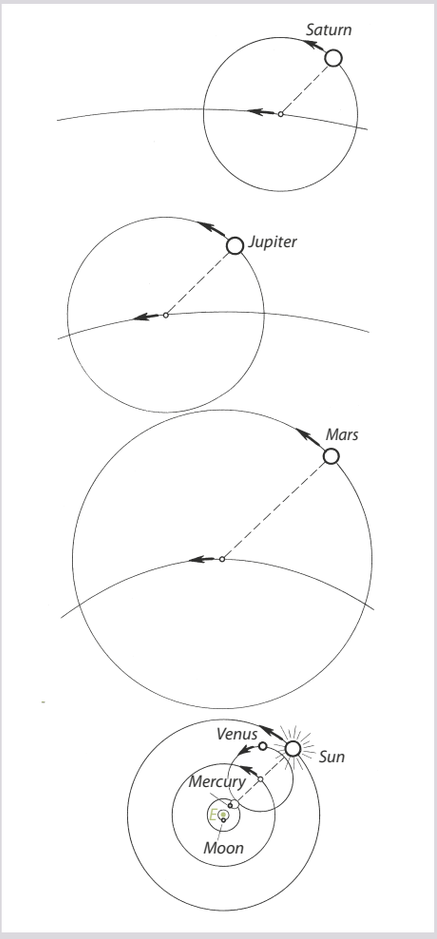
1.4.2 *The Ptolemaic System for Describing Celestial Motion*

A scientific theory that survived for fifteen hundred years while its predictions were compared with observations day by day is worthy of respect. The grand synthesis describing the motion of the celestial bodies—given the name *Almagest* by the Arabs from the Greek *Megale syntaxis* (μεγάλη σύνταξις)—met the needs of



◀ **Figure 1.70** **(a)** The projection of the normal segment *PT* onto the axis of the parabola is constant.

(b) Condition for stability: The line of action S_1S_1' must lie to the left of the line of action S_0S_0'.



▲ **Figure 1.71** Simplified Ptolemaic picture. Since there is no question here of actual distances, the radii of the deferent circles are generally chosen so that the epicycles will not overlap. We have indicated here the special role of the Sun—although this was surmised only in the years just before Copernicus— namely that for the inner planets, the centers of the epicycles lie on the line connecting the Earth to the Sun, and for the outer planets, the radii of the epicycles are parallel to this line [*Harvard* 1970].

> Jupiter, whose orbit is between those of Mars and Saturn, traverses a longer course than Mars, and a shorter than Saturn. Likewise with the rest of these stars: the farther they are from the outermost limits of the heaven, and the nearer their orbits to the earth, the sooner they are seen to finish their courses; for those of them that have a smaller orbit often pass those that are higher, going under them.
>
> For example, place seven ants on a wheel such as potters use, having made seven channels on the wheel about the centre, increasing successively in circumference; and suppose those ants obliged to make a circuit in these channels while the wheel is turned in the opposite direction. In spite of having to move in a direction contrary to that of the wheel, the ants must necessarily complete their journeys in the opposite direction, and that ant which is nearest the centre must finish its circuit sooner, while the ant that is going round at the

continued on next page

more than 50 generations of calendar makers, astronomers, and astrologers. Although this system did not provide a physical model, anyone seeking such a model could invoke the crystal spheres of Aristotle. The Ptolemaic system owes its attractiveness to its adherence to Platonic principles, whereas its precision is the result of its flexibility. This is manifested by the use, in order to "preserve appearances," of types of motion—in addition to the ideal motions that were possible for perfect objects—that paradoxically contradicted the Platonic ideals. Figure 1.71 shows a simplified view of the Ptolemaic picture of the solar system. Earth, resting in the middle, is orbited by the Moon, Mercury, Venus, the Sun, Mars, Jupiter, and Saturn. The Sun and Moon move in perfect circles, executing the ideal form of motion, a circle of constant angular velocity. Circular motion as the fundamental, elementary form of motion is shown separately in Figure 1.72(a). The planets execute more complex, composite motions: They orbit uniform circular epicycles whose midpoints move with constant velocity along the deferent circles (Figure 1.73). Both forms of motion are ideal in the Platonic sense, but in their superposition, the result is one of a variety of paths that depends on the relationship between the angular velocities and the radii of the two circles. Thus, for example, we could arrive at a circular path, as in Figure 1.72(b), whose midpoint does not coincide with the center of the universe, in this case, with the center of the Earth. It can also be a motion in which the midpoint of this just-mentioned circular path—the excenter—itself traverses a circular path. In the most general case, such superpositions can lead to retrograde planetary motion.

Ptolemy, however, required yet another type of motion for his more precise description of observation, one that assaults the Platonic principles: He introduced equants. Consider Figure 1.72(c). We assume that planet P moves along a circle K with midpoint C. That in itself is nothing special. We are simply dealing with the above-mentioned excenter motion. According to Ptolemy, now an additional circular path K' with center C' (*punctum aequans*) is constructed, whereby the radii of the circles K and K' are equal, and we have $EC = CC'$. The planet can now move in its orbit K in such a way that the rotation of the radius $C'P'$ is uniform. The circle K' is called an equant circle (*circulus aequans*). It is of historical interest that it was precisely in the introduction of the equants that Copernicus saw an aesthetic shortcoming, which he took as a sign of the internal contradictions in the Ptolemaic system.

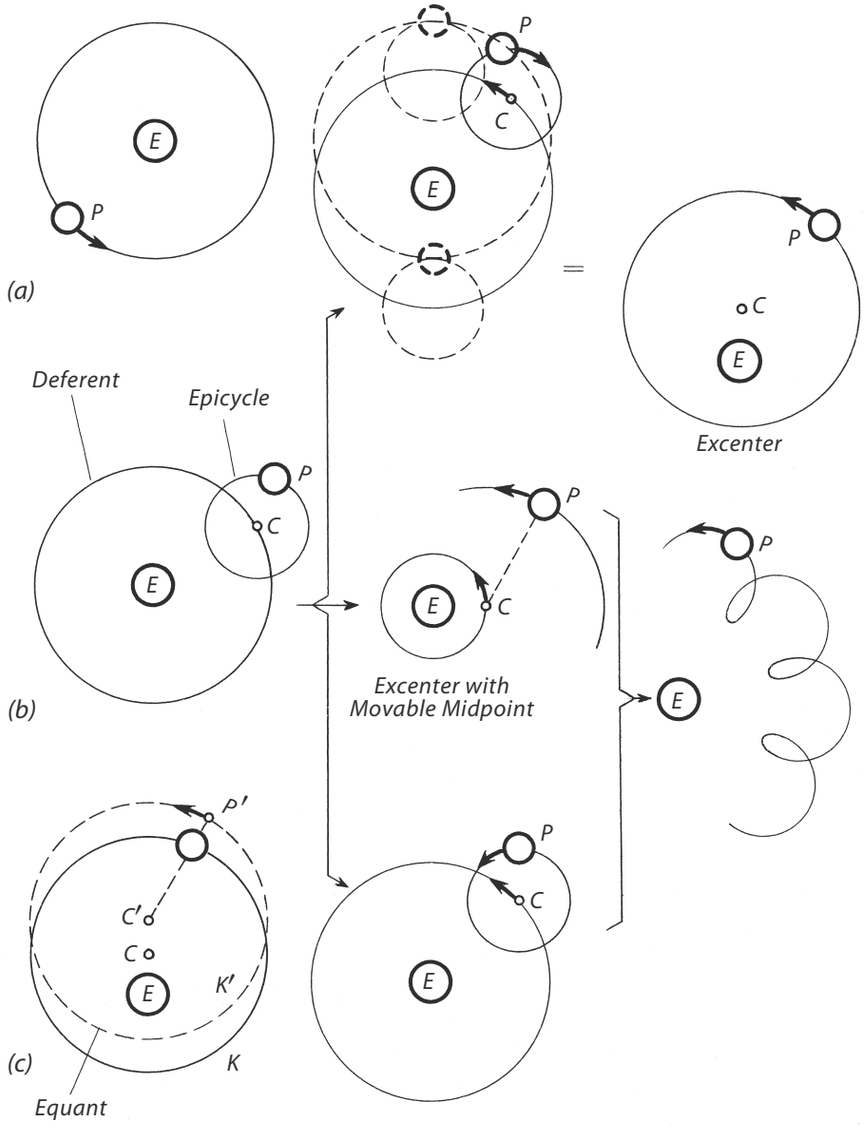
Through a combination of all these types of motion and, if necessary, through the introduction of new epicycles, excenters, and equants, it is possible to predict the motions of the heavens or, by working backward to the time of a given birthday, to cast a horoscope with sufficient accuracy.

It should come as no surprise that no physics of the cosmos could arise from this system. The questions of cause, effective force, and underlying physical law are never even asked. And why should they be, when no physical explanation is needed for the eternal revolutions of the heavenly crystalline spheres?

Posterity's respect for Ptolemy is shown by statues in Gothic cathedrals (see Figure 0.1) as well as the appearance of his likeness and representation of his works in books on astronomy and in decorated manuscripts (Figures 1.74 and 1.75, Plate IV).

1.4.3 *Astronomy and Geography*

Aristarchus of Samos, whose heliocentric system has come down to us indirectly through Archimedes and Plutarch, made a detailed study of the size of the two most conspicuous celestial bodies, the Sun and the Moon, and their distances from



▲ **Figure 1.72** Elementary types of motion in the Ptolemaic system:

(a) Ideal motion: uniform circular motion.

(b) The celestial body *P* traverses the epicycle uniformly; its midpoint *C*, another circle, traverses the deferent, also uniformly.

(c) The planet traverses a circle *K* with midpoint *C*; however, the motion is uniform with respect to the point *C′*.

Earth. His work *On the Sizes and Distances of the Sun and Moon* has survived. In it, Aristarchus gives the diameters of the Moon and Sun as well as the distances Moon–Earth and Sun–Earth in units of Earth diameters. To determine these measurements, he begins with the following measured quantities (Figure 1.76(a)): Let α_M be the angle subtended by the Moon from some point on Earth, and α_S the angle subtended by the Sun. (These angles are not quite constant, since neither the Moon's nor Earth's orbit is an exact circle. The average value of both α_M and α_S is about 0.5°.) If one knew the diameter of the Sun or the Moon, one could use the subtended angles to determine the distances from Earth. Conversely, knowledge of the distance would yield the diameter. Aristarchus used a very clever, but not very precise, method (Figure 1.76(b)) to derive the relationship between the distances Sun–Earth

Figure 1.71 *continued*

outer edge of the disc of the wheel must, on account of the size of its circuit, be much slower in completing its course, even though it is moving just as quickly as the other. In the same way, these stars, which struggle on against the course of the firmament, are accomplishing an orbit on paths of their own; but, owing to the revolution of the heaven, they are swept back as it goes round every day.

—Vitruvius, *The Ten Books on Architecture*, IX.1.14–15 [pp. 255–256]

Aristotle discussed whether the universe is eternal or not as follows:

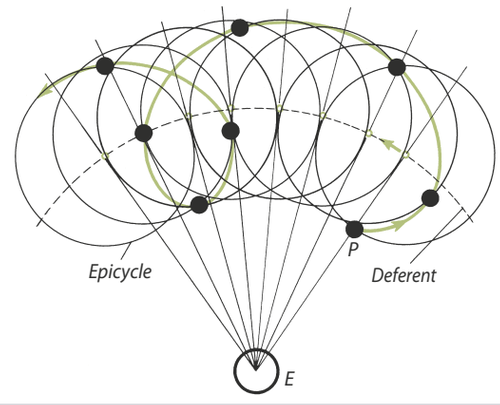
We may now proceed to the question whether the heaven is ungenerated or generated, indestructible or destructible. …

That the world was generated all are agreed, but, generation over, some say that it is eternal, others say that it is destructible like any other natural formation. Others again, like Empedocles of Acragas and Heraclitus of Ephesus, believe that it alternates, being sometimes as it is now and sometimes different and in a process of destruction, and that this continues without end.

Now, to assert that it was generated and yet is eternal is to assert the impossible; for we cannot reasonably attribute to anything any characteristics but those which observation detects in many or all instances. But in this case the facts point the other way: generated things are seen always to be destroyed. …

If the world is one, it is impossible that it should be, as a whole, first generated and then destroyed, never to reappear; since before it came into being there was always present the combination prior to it, and that, we hold, could never change if it was never generated. If, on the other hand, the worlds are infinite in number the view is more plausible. But whether this is, or is not, impossible will be clear from what follows. For there are some who think it possible both for the ungenerated to be destroyed and for the generated to persist undestroyed. (This is held in the *Timaeus*, where Plato says that the heaven, though it was generated, will none the less exist for the rest of time.)

—Aristotle, *On the Heavens*, Book I, Chapter 10 [pp. 463–464]



▲ **Figure 1.73** The construction of a planetary orbit.



▲ **Figure 1.74** Ptolemy, honored for fifteen hundred years as the prince of astronomy, as shown in the picture. ([Reisch 1503], Library of the Hungarian Academy of Sciences.)

We know little of Ptolemy's life (after 83–ca. 168 ce). He was active in Alexandria during the reign of Hadrian. His most important work is *Megale Syntaxis*, in which the Ptolemaic astronomical system is described; it became widely known under its distorted Arabic name *Almagest*.

Ptolemy's geographical works (ΓΓεογραφιχή ύφήγησις) are of almost equal importance. In them he already used circles of latitude and longitude.

In 1175, his work was translated into Latin by Gerard of Cremona. A complete Greek edition appeared in 1538.

Ptolemy also wrote a five-volume work on optics, which has survived only in Arabic translation. Among other things, it investigates the effect of the atmosphere on astronomical observations. In his *Harmonica*, Ptolemy dealt with the theory of music.

Quotation 1.31

Hipparchus can never receive sufficient praise. ... He discovered a nova that came into existence in his lifetime. He was led to wonder with regard to this star, because of the change in the brightness with which it flared up, whether this happened quite frequently and whether the stars we think are fixed actually move. Therefore, he dared a thing that would be astonishing even for God, namely to list the number of stars for future generations and to check the constellations by name, having devised tables to indicate their positions and brightness so that it might easily be discerned not only whether stars perish and are born, but whether they transit and move, and, similarly, whether they grow larger or smaller. Thus Hipparchus left the sky as a legacy to all men—if anyone should be found to claim this inheritance.

—Pliny the Elder, *Natural History* [p. 21]

and Moon–Earth. He first determined that from Earth we see an illuminated half Moon when the line joining the Moon and Sun is perpendicular to that joining the Moon and Earth. When this occurs, if we measure the angle α_{MS} subtended by the lines Moon–Earth and Sun–Earth, then the quotient of the distances Earth–Sun and Earth–Moon can be determined from the triangle *EMS*. However, such a measurement is difficult to obtain with precision because the angle α_{MS} differs only a little from a right angle; indeed, this difference is just 8′. Aristarchus was unable to measure such a small deviation; he could only estimate it.

Figure 1.76(c) shows how the relationship between the diameter of the Moon and that of Earth can easily be determined. One needs to analyze the passage of the Moon through the essentially cylindrical shadow cast by Earth during a lunar eclipse. Now, if we measure the time from when the Moon enters the shadow until it vanishes completely, that time is related to the Moon's diameter. If we also measure the time from the same entry into the shadow until the first moment of the Moon's reappearance, that time is related to the width of the Earth's shadow. The ratio of these two times yields the ratio of the Moon's diameter to that of Earth.

Now the desired quantities can be determined one after the other. From the relationship between the Moon and Earth diameters and the angle subtended by the Moon,

$$\alpha_{\mathrm{M}} d_{\mathrm{EM}} = D_{\mathrm{M}} \rightarrow \alpha_{\mathrm{M}} \frac{d_{\mathrm{EM}}}{D_{\mathrm{E}}} = \frac{D_{\mathrm{M}}}{D_{\mathrm{E}}} \rightarrow \frac{d_{\mathrm{EM}}}{D_{\mathrm{E}}} = \frac{1}{\alpha_{\mathrm{M}}} \frac{D_{\mathrm{M}}}{D_{\mathrm{E}}},$$

we obtain the distance from Earth to the Moon in units of Earth diameters. From the quotient of the distance Earth–Moon and Earth's diameter, we obtain the distance Earth–Sun as

$$\frac{d_{\mathrm{ES}}}{D_{\mathrm{E}}} = \left(\frac{\pi}{2} - \alpha_{\mathrm{MS}} \right)^{-1} \frac{d_{\mathrm{EM}}}{D_{\mathrm{E}}},$$

where we have used

$$\tan\left(\frac{\pi}{2} - \alpha_{\mathrm{MS}} \right) \approx \left(\frac{\pi}{2} - \alpha_{\mathrm{MS}} \right) = \frac{d_{\mathrm{EM}}}{d_{\mathrm{MS}}} \approx \frac{d_{\mathrm{EM}}}{d_{\mathrm{ES}}}.$$

Finally, from the angle subtended by the Sun and its distance from Earth, we use

$$\frac{D_{\mathrm{S}}}{D_{\mathrm{E}}} = \alpha_{\mathrm{S}} \frac{d_{\mathrm{ES}}}{D_{\mathrm{E}}}$$

to obtain the diameter of the Sun in units of Earth diameters. The values obtained by Aristarchus as well as today's values are presented in Table 1.5. We see that he has obtained the correct order of magnitude for the diameter of the Moon; all other quantities are far less than the correct values. In particular, the value obtained for the distance from Earth to the Sun is off by two orders of magnitude. Aristarchus made two mistakes with his initial data: He used 2° for the angle subtended by the Sun and Moon instead of the correct 0.5°; this is surprising, because with the protractors of the time, a much better measurement could surely have been made. It is less surprising that for the angle α_{MS}, he used 87° instead of the correct 89°52′.

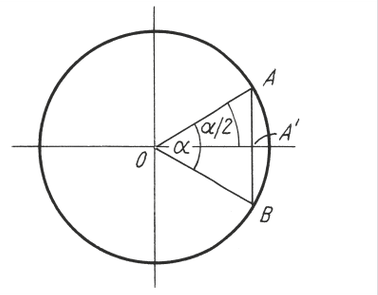
The results that Aristarchus obtained in units of Earth diameters can easily be turned into stadia or kilometers with the help of the measurements of Eratosthenes, who used a simple principle to determine the diameter of Earth (Figure 1.77).

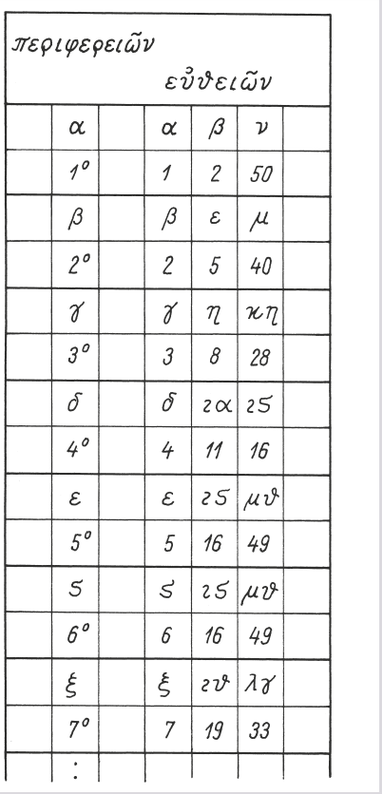
Eratosthenes knew that at the summer solstice, the Sun was at the zenith in Syene (Aswan) because the reflection of the Sun could be seen there in deep wells. He noted further that, at the same time in Alexandria, the angle of the Sun from the vertical was one-fiftieth of a circle (just over 7°). He assumed that Alexandria and Syene lie on the same meridian. From this followed the simultaneity of their noon times, for in locations on the same meridian, the Sun reaches its maximum altitude at the same time. But this assumption was only approximately true; the two cities in fact differ in longitude by 3°. Knowing the distance between Alexandria and Syene now suffices to determine the circumference of Earth, for the distance between the two cities must be one-fiftieth the circumference. Needless to say, the precision of Eratosthenes' terrestrial distance measurement also left much to be desired. He estimated the distance from Alexandria to Syene from the fact that a caravan of camels required 50 days for the journey. On the assumption that a caravan covers about 100 stadia per day, the distance between the two cities must be about $50 \cdot 100 = 5000$ stadia. Multiplying this number by 50 gives us the circumference of Earth at about 250,000 stadia. It is difficult to assess the precision of this figure, since we don't know the conversion factor between a stadium and a meter. There are, in fact, Egyptian, Greek, and also late-Egyptian stadia, whose values varied between 157 and 211 meters. But whichever conversion factor we take, the value found by Eratosthenes is still close to today's value for the Earth's average circumference, about 40,000 kilometers (25,000 miles). The next set of data in Table 1.5 gives the measurements made by Hipparchus (ca. 190–ca. 120 BCE). Each of these values represents a better approximation to reality; above all, the values for the Moon's diameter and its distance from Earth are surprisingly exact. Furthermore, Hipparchus is renowned for his discovery of the precession of the vernal equinox; he arrived at the figure 36″ per year instead of the correct 50″.

Hipparchus also performed a great service to astronomy by cataloguing almost one thousand stars and recording their positions within the various constellations

Name and Date	$\frac{D_M}{D_E}$	$\frac{D_S}{D_E}$	$\frac{d_{ME}}{D_E}$	$\frac{d_{SE}}{D_E}$	πD_E	Remarks
	Today's Values					
	0.27	108.9	30.2	11726	40,000 km	
Aristarchus, 270 BCE	0.36	6.75	9.5	180		$\alpha_M \approx \alpha_S \approx 2°$ (correct value: 30′) $\alpha_{MS} \approx 87°$ (correct value: 89°52′)
Eratosthenes, 230 BCE					252,000 Stadia	Corresponds to 36,000 – 46,000 km 1st: 1 Stadium = Egyptian: 157 m; Greek: 180 m; Late Egyptian: 211 m
Hipparchus, 150 BCE	0.33	$12\frac{1}{3}$	$33\frac{2}{3}$	1245		1 Stadium = 600 Feet
Posidonius, 90 BCE	0.157	$39\frac{1}{4}$	$26\frac{1}{5}$	6550	180,000 Stadia	The discrepancy from the value obtained by Eratosthenes is possibly due only to the difference between the sizes of a stadium
Ptolemy, 150 BCE	0.29	5.5	29.12	605	180,000 Stadia	

▲ **Table 1.5** The measure of the universe; D_M, D_S, D_E, d_{ME}, d_{SE}, denote the diameters of the Moon, the Sun, and Earth, and the distances Moon–Earth and Sun–Earth, respectively. The numbers printed in color are the values known today.





▲ **Figure 1.75** In his astronomical and geographical investigations, Ptolemy made use of the entire mathematical arsenal of his time and indeed advanced it further, notably in the area of spherical trigonometry. We show in the figure a table of trigonometric functions. Ptolemy gives the length of the chord *AB* subtended by the angle α. The relationship between the function used today, $\sin \alpha/2$ (in the figure *AA′* if the radius *OA* is taken to be 1) and the function chord $\alpha \equiv AB$ can be seen on the figure to be

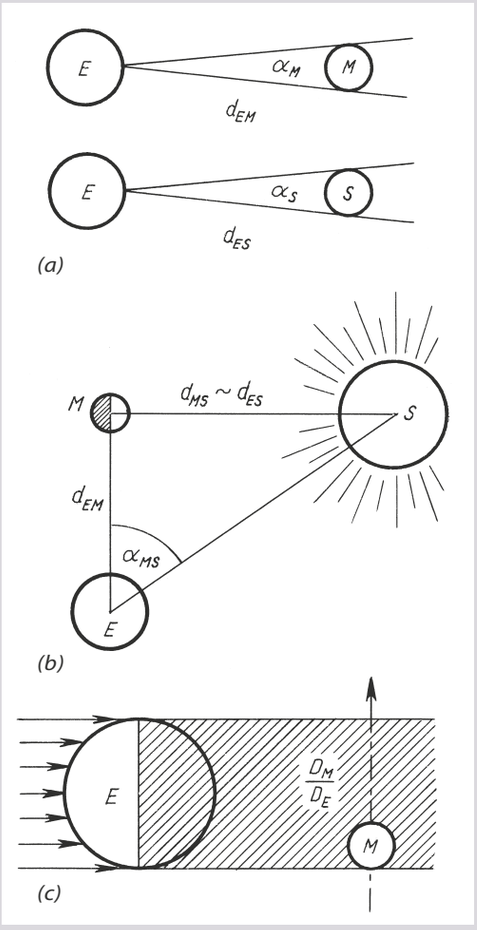
$$\text{chord } \alpha = 2 \sin(\alpha/2).$$

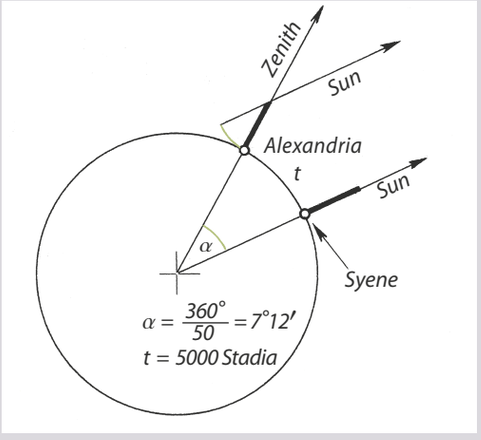
Ptolemy not only knew the addition theorem for the chord function, but also many other theorems relating to trigonometric functions as well as the familiar trigonometric relationships. In his numeric tables, he used a combination of the difficult Greek number system and the Babylonian sexagesimal system. Thus, for example, the symbols 1 2 50 represent the value

$$1 \cdot 60^{-1} + 2 \cdot 60^{-2} + 50 \cdot 60^{-3} = 0.0175$$

Looking up the value for $\sin \alpha/2$ for $\alpha = 1°$ in a function table, we get 0.0088, that is, precisely one-half the Ptolemaic value 0.0175, which must be the case given the relation chord $\alpha = 2 \sin \alpha/2$.

▼ **Figure 1.76** The method used by Aristarchus to measure the relative sizes of Earth, the Moon, and the Sun.





▲ **Figure 1.77** Eratosthenes' principle for measuring the radius of Earth.

> What he [Eratosthenes] says will, however, become clear if the following assumptions are made. Let us suppose, in this case also, first that Syene and Alexandria lie under the same meridian circle; secondly, that the distance between the two cities is 5000 stades; and thirdly, that the rays sent down from different parts of the sun upon

continued on next page

(Quotation 1.31). Table 1.5 contains two additional rows of data, which relate to results of Posidonius (ca. 135–51 bce) and Ptolemy. Posidonius arrived at a value for the distance from Earth to the Sun of the correct order of magnitude as well as the best value from the classical period for the Sun's diameter. As for Ptolemy, we should emphasize that he gave quite precise values for the Moon's diameter and its distance from Earth. His value for Earth's circumference differs from that given by Eratosthenes considerably, but it is possible that this difference is attributable entirely to the use of another definition of a stadium.

The measurement of the heavens proceeded apace along with the mapping of Earth's surface. In his *Geographia*, Eratosthenes considered problems of physical geography and prepared a rather precise map of the Mediterranean Sea and bordering lands. Hipparchus introduced degrees of latitude and longitude. A map produced by Ptolemy shows more than 8000 details with precise locational information. The original has been lost, but a reconstruction is shown in Figure 1.78. This map shows not only the area around the Mediterranean, but the British Isles to the north, India and China to the east, and a large part of Africa to the south. Surprisingly, Ptolemy connected the southern portion of Africa with Southeast Asia, so the Indian Ocean became an inland sea. The Greek cartographers already explicitly considered Columbus's idea that one might reach China or India by a western route across the Atlantic.

If one considers that the Greeks were particularly well versed in geometry, it should come as no surprise that they achieved excellent results in dealing with problems of theoretical cartography. They knew about the conical and stereographic, and even the planispheric projections (Figure 1.79).



▲ **Figure 1.78** A map by Ptolemy. (Székesfehérvár (Hungary) diocesan library.)

1.4.4 *Geometry*

Greek thought and the Greek spirit were perhaps best expressed in geometry. In the world of geometric concepts, the Greeks found what they were vainly seeking in the real world: to be able to reduce everything to fundamental principles and build an ideal "world," even if only a world reduced to geometric forms. This system of certain truth is seemingly complex, yet because of its rigorous logical order, it is clearly understandable, and thus it could become a model of scientific endeavor. In spite of the abstractness of the system, one can see that every one of its concepts is still connected with reality. With geometry as a model, scholars of all eras have been given the opportunity and challenge to describe other realms of objective reality in the manner of geometry (*more geometrico*).

The tight logical construction of Euclidean geometry and its internal coherence are best verified by the fact that its logical structure remained intact even after Bolyai and Lobachevsky (1792–1856) called the unlimited and necessary correctness of its basic axioms into question at the beginning of the nineteenth century (Quotation 1.32).

It is not our purpose in this book to follow the development of Greek mathematics in detail, but we would like to emphasize only a few of its outstanding accomplishments. It is worth noting that Euclid was primarily a compiler and an arranger of existing knowledge. The theorem that every triangle inscribed in a semicircle is a right triangle was found by Thales two to three centuries earlier. Hippocrates of Chios (ca. 470–ca. 410 BCE)—namesake of the physician Hippocrates of Kos (ca. 460–ca. 370)—had already produced a comprehensive book on geometry, which has not survived. Eudoxus (ca. 408–347 BCE), whose contributions to astronomy have already been mentioned, was the first to use the method of exhaustion. Menaechmus (380–320 BCE) did work on the properties of conic sections.

We would like to mention three problems to which the Greeks gave considerable attention. Although they were unable to solve them, their endeavors led to valuable results.

1. *The doubling of the cube.* What is the length of the edge of a cube whose volume is double that of the cube with unit edges? In effect, the problem is to determine the value a such that $a^3 = 2$, or $a = \sqrt[3]{2}$. This is known as the Delian problem because, according to legend, the Athenians—in their efforts to end a plague—had consulted the oracle at Delos, who had set them the task to double the size of the cubical altar. According to Plato, the gods had demanded this not because they needed such an altar, but to reproach the Athenians for not working more intensively on geometric problems. Hippocrates reduced the problem to the search for values x and y satisfying the proportions

$$a : x = x : y = y : 2a.$$

It is worth noting that these proportions may also be represented in the form

$$x^2 = ay, \quad y^2 = 2ax, \quad xy = 2a^2.$$

Thanks to the coordinate geometry of Descartes, we can recognize these as formulas for the conic sections. It has been asserted that it was the Delian problem itself that spurred Menaechmus to his study of conic sections. According to others, however, his motivation came from the problem of constructing sundials.

A method for determining the lengths x and y proposed by Archytas (428–347 BCE) is given in Figure 1.80.

Figure 1.77 *continued*

different parts of the earth are parallel; for the geometers proceed on this assumption. Fourthly, let us assume that, as is proved by the geometers, straight lines falling on parallel straight lines make the alternate angles equal, and fifthly, that the arcs subtended by equal angles are similar, that is, have the same proportion and the same ratio to their proper circles—this having also been proved by the geometers. For whenever arcs of circles are subtended by equal angles, if any one of these is (say) one tenth of its proper circle, all the remaining arcs will be tenth parts of their proper circles.

Anyone who has mastered these facts will have no difficulty in understanding the method of Eratosthenes, which is as follows. Syene and Alexandria, he asserts, are under the same meridian. Since meridian circles are great circles in the universe, the circles on the earth which lie under them are necessarily great circles too. Therefore, of whatever size this method shows the circle of the earth through Syene and Alexandria to be, this will be the size of the great circle of the earth. He then asserts, as is indeed the case, that Syene lies under the summer tropic. Therefore, whenever the sun, being in the Crab at the summer solstice, is exactly in the middle of the heavens, the pointers of the sundials necessarily throw no shadows, the sun being in the exact vertical line above them; and this is said to be true over a space 300 stades in diameter. But in Alexandria at the same hour the pointers of the sundials throw shadows, because this city lies farther to the north than Syene. As the two cities lie under the same meridian great circle, if we draw an arc from the extremity of the shadow of the pointer to the base of the pointer of the sundial in Alexandria, the arc will be a segment of a great circle in the bowl of the sundial, since the bowl lies under the great circle. If then we conceive straight lines produced in order from each of the pointers through the earth, they will meet at the center of the earth. Now since the sundial at Syene is vertically under the sun, if we conceive a straight line drawn from the sun to the top of the pointer of the sundial, the line stretching from the sun to the centre of the earth will be one straight line. If now we conceive another straight line drawn upwards from the extremity of the shadow of the pointer of the sundial in Alexandria, through the top of the pointer to the sun, this straight line and the aforesaid straight line will be parallel, being straight lines drawn through from different parts of the sun to different parts of the earth. Now on these parallel straight lines there falls the straight line drawn from the centre of the earth to the pointer at Alexandria, so that it makes the alternate angles equal; one of these is formed at the centre of the earth by the intersection of the straight lines drawn from the sundials to the centre of the earth; the other is at the intersection of the top of the pointer in Alexandria and the straight line drawn from the extremity of the shadow of the pointer to its base, while the angle at the centre of the earth subtends the arc stretching from Syene to Alexandria. But the arcs are similar since they are subtended by equal angles. Whatever ratio, therefore, the arc in the bowl of the sundial has is 500 stades. Therefore the whole great circle is 250000 stades.

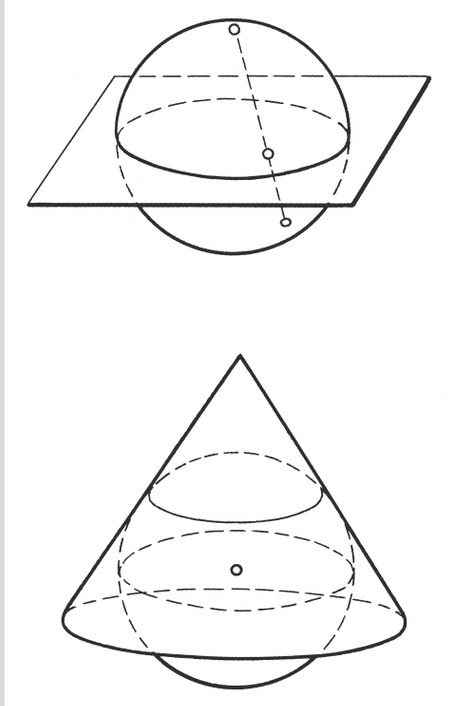
Such is the method of Eratosthenes.

—Cleomedes, *On the Circular Motion of the Celestial Bodies* [pp. 205–207]

Quotation 1.32

The influence of geometry upon philosophy and scientific method has been profound. Geometry, as established by the Greeks, starts with axioms which are (or are deemed to be) self-evident, and proceeds, by deductive reasoning, to arrive at theorems that are very far from self-evident. The axioms and theorems are held to be true of actual space, which is something given in experience. It thus appeared to be possible to discover things about the actual world by first noticing what is self-evident and then using deduction. This view influenced PLATO and KANT, and most of the intermediate philosophers. When the Declaration of Independence says "we hold these truths to be self-evident," it is modeling itself on EUCLID. The eighteenth-century doctrine of natural rights is a search for Euclidean axioms in politics. The form of NEWTON's *Principia*, in spite of its admittedly empirical material, is entirely dominated by EUCLID. Theology, in its exact scholastic forms, takes its style from the same source. Personal religion is derived from ecstasy, theology from mathematics; and both are to be found in PYTHAGORAS.

—BERTRAND RUSSELL, *A History of Western Philosophy*, 1945 [p. 37]



▲ **Figure 1.79** The Greeks had already studied the basic problem of cartography: how to represent the spherical surface of Earth on a flat plane.

2. *The problem of trisecting an angle.* Can an arbitrary angle be divided into three equal angles? This problem was solved by HIPPIAS (fl. fifth century BCE) with the help of a transcendental curve, the quadratrix (Figure 1.81).

A line segment AB rotates at constant angular velocity about the point A while a segment BC is moved parallel to itself, likewise with constant velocity. The intersection point of the two line segments is the quadratrix. Otherwise, the problem of trisecting the angle requires the solution of a cubic equation, due to the relationship

$$\cos\varphi = 4\cos^3\frac{\varphi}{3} - 3\cos\frac{\varphi}{3}.$$

The Greek mathematicians, and those of later times as well, attempted to solve this problem by using only a straightedge and compass. Only in 1837 was WANTZEL (1814–1848) able to prove that the problem is unsolvable, and so all previous attempts were doomed to end in failure.

3. *The problem of squaring the circle.* Here the problem is to find a square whose area is equal to a circle of given radius r, that is, to solve the equation $x^2 = \pi r^2$, or equivalently, $x = r\sqrt{\pi}$, or, according to the original formulation, to construct the number p using only a straightedge and compass. It was only in 1882 that LINDEMANN (1852–1939) showed that π is a transcendental number; that is, it is not the root of any polynomial, no matter the degree, with integer coefficients, and so naturally the problem—in this sense of construction—is unsolvable. It is interesting that at the very beginning, HIPPOCRATES came up with a very promising approach: The area of the crescent-shaped figure (*lunula* = little moon) bounded by circular arcs (Figure 1.82) is exactly equal to that of the triangle AOB.

Struggling with the problem of squaring the circle, both ANTIPHON (b. ca. 480 BCE) and BRYSON (ca. 450–ca. 390 BCE) came up with the idea of squeezing the area of the circle between the areas of inscribed and circumscribed polygons. EUDOXUS and ARCHIMEDES further developed this into the method of exhaustion, which yielded a procedure for determining the value of π to a precision limited only by the effort invested in calculation. Thus after ARCHIMEDES, the value of π could be determined in principle to arbitrary accuracy (Figure 1.83).

The original manuscript of EUCLID's *Elements* (Figure 1.84) has been lost. Around the year 1800, a tenth-century copy, which appears to be authentic, was found in the Vatican library. The work consists of 13 books.

The first book begins with definitions, postulates, and axioms. The definitions have no *logical* function; the concepts abstracted from reality are defined here in order to make the theorems more understandable. Thus, among the definitions we find the following:

1. A point is that which has no part.
2. A line is a length without width.
3. The ends of a line are points.
4. A surface is that which has only length and breadth.
5. The ends of a surface are lines.
6. Parallel lines are those that lie in a plane and never meet even if extended in both directions.

The axiomatic system, consisting of five axioms, contains basic truths that are valid in every science. The five postulates that follow lay down the cornerstones of geometry.

The axioms are these:

1. Things that are equal to the same thing are also equal to one another.
2. If equals are added to equals, the sums are also equal.
3. If equals are subtracted from equals, the remainders are also equal.
4. Things that coincide with one another are equal to one another.
5. The whole is greater than the part.

It is somewhat surprising that the fourth statement appears here rather than among the postulates, since it is clearly a *geometric* concept.

These are the postulates:

1. Any two points can be joined by a straight line.
2. Any straight line segment can be extended indefinitely.
3. A circle can be drawn at an arbitrary center with an arbitrary radius.
4. All right angles are equal to each other.
5. If two lines in the same plane are intersected by a third, then the two lines intersect on that side of the third where the sum of the inner angles is less than two right angles.

This fifth postulate was already considered problematic in antiquity. Was it necessary, or could it be proved from the other four? The answer to this question was finally given by Bolyai and Lobachevsky at the beginning of the nineteenth century: If the fifth postulate is omitted, one can construct a new, consistent, non-Euclidean geometry.

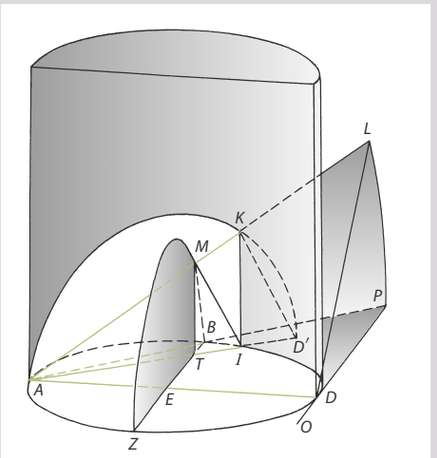
Regarding the truth of these initial statements, the Greeks had quite a modern point of view. Aristotle stated that the truth of the postulates could be verified by the agreement with reality of the consequences derived from them. Proclus (412–485) came even closer to our modern view: All geometric truths are of the form "if–then." What needs to be studied are the consequences that can be drawn from the premises, and it is of little interest whether the premises are correct in the sense that they describe anything real.

Nevertheless, in the two thousand years after Euclid, both the postulates and consequently all the assertions derived from them were considered as nothing less than unassailable truth; they were, in fact, seen as the thoughts of a "geometer God" made accessible to human understanding, and later as a priori constructs of the human mind. Therefore, it seemed simply impossible that a fact might contradict one of the postulates, and thus it comes as no surprise that not even Gauss, who knew better, had the courage to question publicly the correctness of Euclidean geometry.

The next great geometer after Euclid was Apollonius (ca. 262–ca. 190 BCE), who introduced the terms *ellipse*, *hyperbola*, and *parabola* for the conic sections and had so much to say about them that more than fifteen centuries later—up to the time of Pascal (1623–1662)—nothing new was added (Figure 1.85).

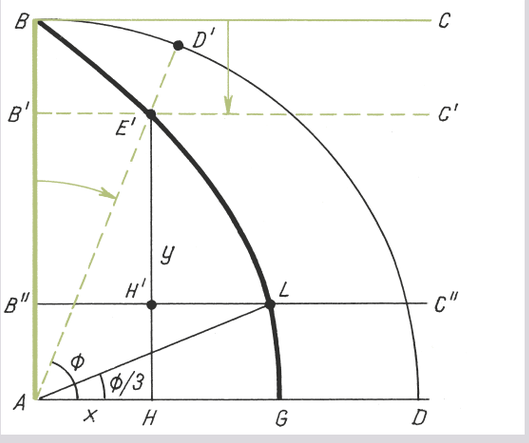
For astronomy, the work of Menelaus (fl. ca. 100 CE) on spherical geometry is of great importance. Ptolemy used this work in his astronomical calculations. It is interesting to note that he used trigonometric functions extensively, although not in the form used today. He determined the chord corresponding to the angle α and worked with that quantity, as shown in Figure 1.75.

Since this whole chapter is about the greatest achievements, it might seem odd to mention here the worst fault of ancient mathematics, namely, the primitiveness of mathematical notation. The whole numbers, as we have said earlier and shown in Figure 1.38, were denoted by letters. To indicate the powers of an unknown, the method displayed in Figure 1.86(a) was employed. Finally, Figure 1.86(b) compares the ancient Greek notation and contemporary notation for polynomial equations.

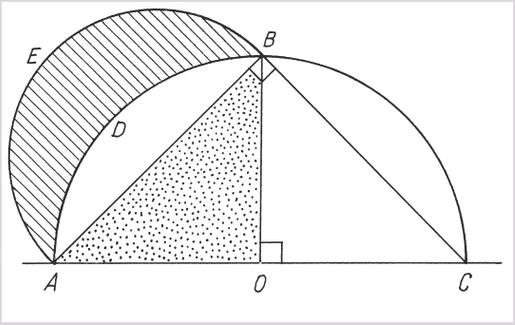


▲ **Figure 1.80** The method of Archytas for doubling the volume of a cube:

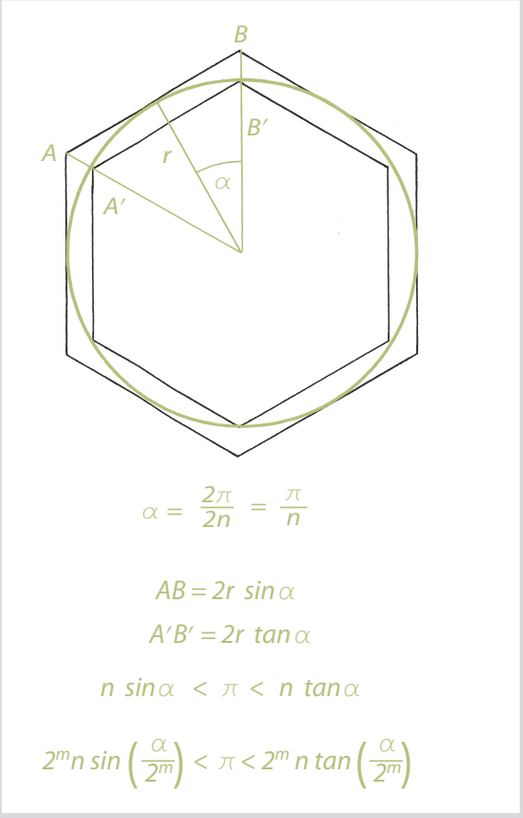
Given two segments $AD = a$ and $AB = b$, if one can find two segments x and y such that $a : x = x : y = y : b$, then it follows that $y^3 = ab^2$, and therefore one has the relation $y^3 = 2$ for $b = 1$ and $a = 2$. The volume of a cube with side length y is therefore twice the volume of a cube with side length $b = 1$. Therefore, we construct on the base circle with diameter AD, a right circular cylinder and a semicircle orthogonal to the circle above the segment AD. When this semicircle is rotated about the fixed point A (maintaining the perpendicular orientation), then the point of intersection of the semicircle with the cylinder's surface describes a curve. One obtains an additional curve from the intersection point of the produced line AB with the cylinder if this is moved in such a way that it generates a cone with axis AD. Let the intersection point of the two circles be denoted by K. According to Archytas, it can rather easily be seen that the desired segment x is given by AK, and y is given by AI (from [van der Waerden 1954] with additions).



▲ **Figure 1.81** Trisection of an angle using a quadratrix. The quadratrix is given by the point of intersection of the line BC, which is moved downward and parallel to itself with constant velocity, and the line AD' rotated about the point A. We obtain one-third of an arbitrary acute angle Φ by dividing into thirds the ordinate of the point lying on the quadratrix belonging to the angle Φ. The point on the quadratrix belonging to one-third of the ordinate is on the angle $\Phi/3$.



▲ **Figure 1.82** The lune of Hippocrates: the area of the lune *ADBE* is equal to the area of the triangle *AOB*. This is easy to see, since the surface of the semicircle *ABE* is equal to the surface of the quarter-circle *AOBD*, and both have a common circular segment *ABD*.



▲ **Figure 1.83** Determination of the area of a circle according to Archimedes: He begins with an inscribed and a circumscribed regular hexagon, and by repeated halving of the central angle, obtains the ratio of radius to side length for regular 12-, 24-, 48-, and 96-gons. Here he uses the inequality, from a source unknown to us,

$$\frac{265}{153} < \sqrt{3} < \frac{1351}{780}$$

for the value $\sqrt{3}$.

Ludolph Van Ceulen (1540–1610), a mathematician and fencing master in Delft, determined in 1596 the value of π to 20 decimal places (later to 35) by using a regular 60×2^{29}-gon.

1.4.5 *Scientific Instruments and Technology*

With a few exceptions, the Greeks did not carry out scientific experiments in today's sense of the word; consequently their measuring instruments were generally rather simple. Hero of Alexandria (ca. 10–ca. 70 ce) describes an important instrument, the diopter, which could be used for both astronomical and terrestrial measurements (Figure 1.87). On a disk that can turn on a vertical axis, two crosshairs are arranged diametrically across from each other and perpendicular to the disk. The crosshairs, together with a connecting bar and a perpendicular pointer, are separately rotatable over the disk, which can be adjusted to be horizontal with the help of a water level. The vertical orientation of the main column is ensured with the aid of a plumb line. Two worm-gear mechanisms are deployed on the column, one allowing the disk to be rotated about the vertical axis and the other for the adjustment of its horizontal tilt.

Among other things, with this instrument one could determine the directions in which the celestial bodies rise and set. Another use, also described by Hero (in *Opera quae supersunt omnia*), is illustrated in Figure 1.88. In digging a tunnel through a mountain with established entrances and with work progressing from both ends, the problem is to determine the direction in which digging should proceed in order that the two shafts meet correctly. To solve the problem, one places the diopter near entrance *B* at a point *E*, chosen for convenience with respect to the profile of the land but otherwise arbitrary. Using the diopter, one finds the point *F* on the line perpendicular to *EB*. Traversing via a sequence of perpendicular lines, one eventually arrives at the line *KL* and on it, the point *M* such that from the tunnel entrance *D*, we have *DM* perpendicular to *KL*. From the distances *DN* and *NB* one obtains the angle α, which specifies the direction in which the shafts are to be dug in relation to the lines *BE* and *MD* so that they meet at a point.

It is likely that Eupalinus used a similar method about 530 bce on Samos to build an aqueduct tunnel more than a kilometer long dug from both directions. When the two tunnels met in the middle of the mountain, the deviation was only ten meters horizontally and three meters vertically.

An apparatus recovered from the sea in 1906 near the island of Antikythera (Figure 1.89) aroused a great deal of interest, and in the past decade it became possible to reconstruct it completely. With its complicated system of gears, it speaks of a level of technological development that had previously not been considered possible in the Hellenistic era. From the inscriptions on it, as well as other objects found with it, the apparatus can be dated to about the year 80 bce. Its restorers called it—with some exaggeration—an ancient computer. Turning an input shaft brings into motion 40 gear wheels, among them a planetary gear train, so that pointers on dials on its external surface show the positions of the Sun, the Moon, and the planets, with all the complexities of planetary motion, including even retrograde motion. It is conceivable that it was moved automatically by a clockwork mechanism attached to the input shaft. The device was likely used for astrological purposes. That it was actually used can be seen from traces of repair work: A gear tooth broke off, due to heavy use, and was replaced with a new one.

Similar apparatuses, although much simpler in their execution, were built by the Arabs about one thousand years later. Since European clocks are derived from such Arabic instruments, the apparatus from Antikythera can be seen as the precursor of all European clocks. Its technical level was not reached again until the eighteenth century.

The construction of various water clocks also attests to a high level of engineering skill in the ancient world. The water clock depicted in Figure 1.90 was reconstructed from a description given by CTESIBIUS (fl. 285–222 BCE). From an artistically rendered gargoyle, water drips into a funnel and then flows through a thin tube into a vertical cylinder that is closed by a piston. The piston moves upward as the water level in the cylinder rises. On the piston rod is mounted a figure that indicates the time with a small stick. In the ancient world, the time between sunrise and sunset was divided into twelve equal segments, making the length of an hour greater in summer than in winter, so that the divisions into hours on the indicator drum were not uniform. When the cylinder was full of water after the course of a day, it emptied itself automatically through a siphon, and the piston was returned to its initial position. The clock thus again indicated the initial hour of the day. The fixed quantity of water that issued at this time from the siphon fell on a waterwheel, turning it by a specific amount. This movement was transferred via a gear system to the indicator drum in such a manner that the hourly division under the pointer would be appropriate to the time of year. Thus the water clock showed not only the hours, but also the day of the year, and all this took place automatically. The only servicing that was necessary was to fill the container behind the gargoyle from time to time, which is analogous to winding a mechanical clock.

The cleverest ideas of Hellenistic technology were frittered away on the construction of amusing machines or automata for impressing visitors to temples (Figures 1.91 and 1.92).

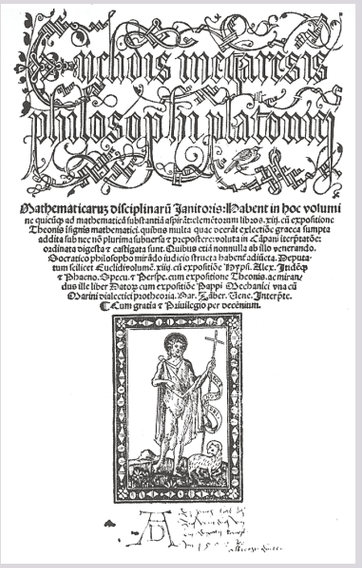
The Romans contributed nothing further to the development of geometry, and in physics they were primarily consumers and transmitters of Greek knowledge. As practical people, they were proud to use their physical and technical capabilities for the construction of public works and not waste time with toys or instruments designed solely for amusement. We find with the Romans the first approaches to a "science in the service of the people." Of particular note from both a social and technological viewpoint, and frequently also for its aesthetic merit, is their extensive water-supply network.

The operation of these facilities was already a complex technological undertaking. Thus, for example, water use had to be measured; the measuring equipment had to be checked by order of the authorities; and theft of water, which was carried out on occasion with clever technical means, had to be detected (Quotation 1.33).

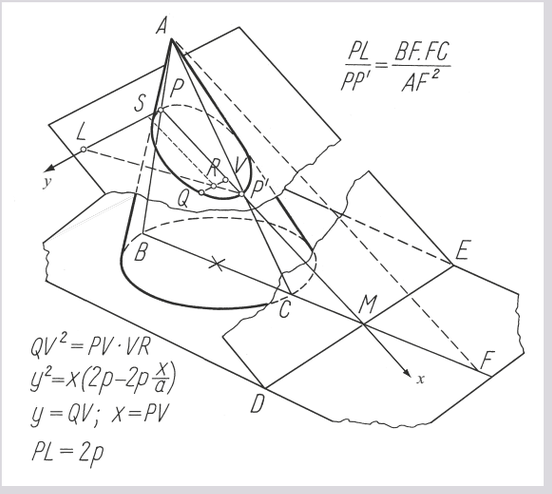
1.5 The Twilight of Hellenism

1.5.1 *Pessimistic Philosophers*

With the establishment of ALEXANDER THE GREAT's empire and its eventual disintegration, the Greek city-states ceased to exist. As a consequence, on the one hand, the feeling of existential and intellectual security that had existed in the heyday of the city-states was lost: physical safety, clarity, and linkage to tradition in political, organizational, and cultural milieus faded away. On the other hand, the Greeks' horizons were broadened as a result of contact with a great variety of religions and philosophical points of view. The results: existential insecurity, spiritual confusion, internalization, and cosmopolitanism. Two thousand years later, POINCARÉ would have this to say about the possible reactions to a cacophony of viewpoints: *To doubt all and to believe all are both convenient behaviors: both remove the obligation to think*. The masses were prepared to accept everything with which they came in contact. Sects flourished, and superstition spread into astrology, magic, and



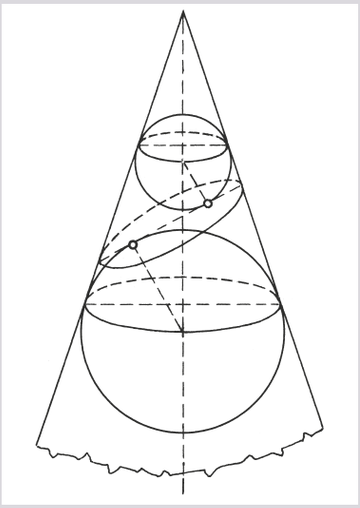
◀ **Figure 1.84** EUCLID's *Elements*: edition with DÜRER's handwritten entries. (Herzog-August-Bibliothek, Wolfenbüttel.)



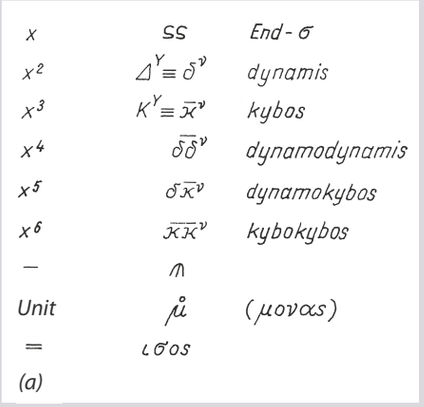
▲ **Figure 1.85 (a)** Derivation of the conic sections after APOLLONIUS: Given a plane, a circle in the plane, and a point *A* lying outside the plane, these objects determine a cone. Let a second plane generating the conic sections cut the plane given above in the line *DE*. A line perpendicular to *DE* through the center of the circle determines the triangle *ABC*. (In the simple arrangement of our picture, the plane of the triangle is perpendicular to the plane of the base circle.) The lines *AB* and *AC* then intersect the conic section at the points *P* and *P′*. Now the line *AF* is constructed parallel to the line *PP′*. If we choose (for an ellipse) the point *L* such that

$$\frac{PL}{PP'} = \frac{BF \times FC}{AF^2},$$

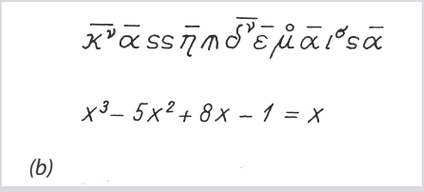
then for every point *Q* on the ellipse we have the relation $QV^2 = PV \times VR$. For an ellipse, the area $PV \times VR$ is always smaller than $PV \times PL$; for a hyperbola, it is always greater; and for a parabola, the two areas are equal. It is from this property that the conic sections get their names (ἔλλειψις = shortage; ὑπερβολή = excess; παραβολή = comparison).

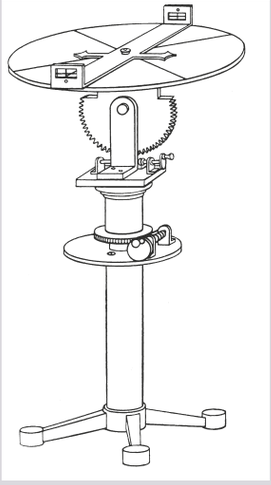


◀ **Figure 1.85 (b)** First Pascal, and later Euler, made further contributions to the theory of conic sections. This figure shows the tangent Dandelin-spheres that locate the foci (G. P. Dandelin, 1794–1847).



◀ **Figure 1.86** Some of the notation used by Diophantus.





◀ **Figure 1.87** Hero's diopter (reconstructed by Schoene) [*Opera,* Vol. III].

augury. In contrast, the philosophical attitude was one of skepticism. It should not be forgotten that even scientists must have doubts: they never know anything with absolute certainty, which still does not prevent them from continually beginning anew to explore the unknown. In fact, as we shall see later, it is through doubt itself that some philosophers would arrive at the secure knowledge of the truth.

The founder of the doctrine of Skepticism as one of the main currents of Hellenism is Pyrrho (ca. 360–ca. 270 BCE). He based his skepticism primarily on the unreliability of information transmitted to us by our senses. Later, the Platonic academy adopted some of his teachings as their own; thereby, the focus of doubt shifted to the uncertainty of fundamental principles of logic, primarily the uncertainty of the principle of deductive proof. After all, a proof is based on accepted, but unproven, statements (axioms), and if those are to be proved, then new axioms must be pressed into service, and so on, ad infinitum, or we could even arrive at the original axioms again in a circular argument. Hence, there is nothing left for a philosopher but to be constantly examining everything (σκέπτομαι, *skeptoma* = examine, hence the designation). In order to arrive at spiritual peace, one therefore had to abstain from holding any definite opinion. For everyday behavior, one might be guided by considerations of probability, custom and tradition, consensus, or natural instinct. A representative of this disposition in the natural sciences was the noted physician Sextus Empiricus (ca. 160–210), who, as his name suggests, based his professional work entirely on an empirical foundation and was completely uninterested in investigating the causes of disease or in proposing any theories about why a treatment might work.

It is of course no accident that the other two significant intellectual currents of Hellenism, namely Stoicism and Epicureanism, appeared at the same time as Skepticism, indeed around the time of the death of Alexander the Great. In contrast to Skepticism ("eschew all opinion"), the other two philosophical schools have a dogmatic character; that is, they attempt to present a worked-out system of teachings and a comprehensive worldview.

It is sometimes asserted that Stoicism is the teaching that least well expresses the Greek way. Its founder, Zeno of Citium (ca. 336–ca. 264 BCE), was a Phoenician, and its most prominent representatives were the Romans Seneca (ca. 4 BCE–65 CE), Epictetus (55–135), and Marcus Aurelius (121–180). The name for this school, *Stoa*, reminds us that Zeno began proclaiming his teachings in the *Stoa poikile* (στοα ποικίλη = hall of painted columns)—teachings that have influenced modern philosophy and speak to modern mankind. The goal of this philosophy is exclusively ethical, to point the way to correct conduct in an uncertain, confused world. The aspects of Stoicism of greatest interest to physics, namely, the physical and epistemological, serve only as tools for finding this correct behavior. According to Zeno, philosophy is like an orchard; logic represents the surrounding walls; physics, the trees; and ethics, the fruit. Its epistemology is empirical: Sensory perceptions awake ideas; of these, those that are useful reappear most frequently and leave behind in every individual lasting impressions (φαντασία καταλεπτική = *phantasia kataleptike*, *consensus gentium*) as criteria of truth. In nature, strictly deterministic laws (ειμαρμένη, *fatum*) prevail. The worldview is materialistic and pantheistic; the godhead (*logos*) permeates all as a fiery breath (*pneuma*); all that exists partakes of it. Thus it is that every living thing is bound with all other living things through a cosmic sympathy—a philosophical basis for a cosmopolitan worldview and a cosmopolitan humanistic ideal. Yet the *fatum* of the Stoics is not cruel: The godhead assumes the responsibility of Providence. How is the individ-

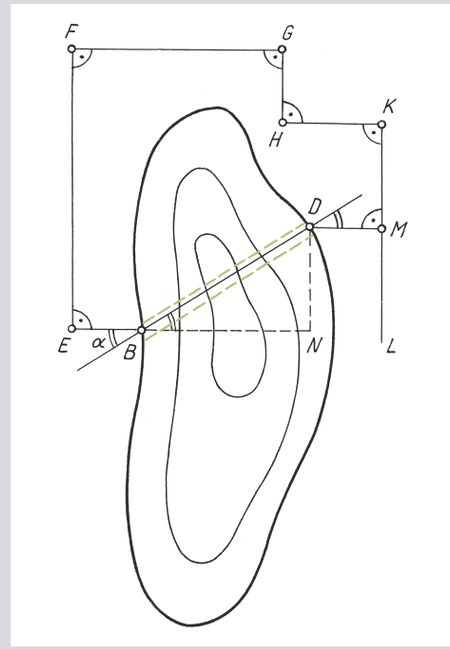
ual to fit into this strictly deterministic world? *Stoa* offers a very modern-sounding answer to this question: "Act in accord with natural law." This is similar to the Hegelian dictum, "freedom is insight into necessity." Seneca says, "If you submit to your fate, it will guide you; if you do not, it will drive you." Thus is it possible to reach a state of complete harmony and freedom from human passions, the state of *apathy* (απάθεια).

It is also possible to derive ideas that are quite modern from the cosmology of *stoa*. In the beginning, all was fire; gradually, air, water, and earth precipitated out. After a time, the world will return to its original state of cosmic fire, and the entire process will repeat.

The notions that individual histories are an integral part of the world's unfolding; the resulting cosmic sympathy that proclaims equality among all peoples, be they Greek, Roman, barbarian, slave, or emperor; the intelligent accommodation of oneself into the teleological order of things; the emphasis on duty, which resonated especially with the Romans; religious tolerance—all these are ideas that continue to this day to affect human thought and action in one form or another (Figure 1.93).

According to the teachings of Epicurus (341–270 bce), it is the task of philosophy to point the way toward a happy life. Virtue is nothing more than the astute search for happiness. Righteousness consists in acting in such a way that one has no need to fear the vengeance of one's fellow man. Phenomenology and physics, the aspects of greatest interest to us, are here again simply aids for finding behavioral norms. According to Epicurus, peace of mind is disturbed primarily by fear. We fear the gods; we fear death and our inevitable fate. The purpose of physics is to free us from such fear. The phenomenology of Epicureanism is based on the senses and is directly linked to that of Democritus (ca. 460–ca. 370 bce). Its worldview is materialistic, but not deterministic, and indeed chance (τύχη, *tyche*) plays a role in nature: The world consists of atoms, which due to their weight fall vertically downward. Yet, from time to time they are deflected in unpredictable ways. It is this phenomenon that allows us to speak of chance. We, too, are subject to the laws of nature, but have a certain degree of free will and thus can master our fate. We need not fear death, for when it comes, we are no more: when our atoms are scattered, the atoms of our souls are scattered as well. Likewise, the wrath of the gods is also not to be feared, for while the gods indeed exist, in their state of bliss they have no concern for man's fate, and thus we do not need to fear their vengeance. Epicurus does not view religion as a comfort for mankind, but as the principal source of its anxieties, from which it must be freed. The ethical goal is therefore pleasure (αταραξία, *ataraxie*), which, however, is not primarily the pleasure of passion: "When we claim that our end goal is pleasure, we are not talking of the pleasure of the dissolute, nor of the delight of the pleasure-seekers, as is taught by some, but rather that our bodies should experience no pain, our souls no distress…. It is impossible to live well and in comfort without practicing righteousness, but it is also impossible without pleasure to live in reason, wellness, and righteousness." Epicurus secured this spiritual peace for himself in his famous garden, debating among his circle of friends, leading a simple life completely removed from the larger society. In this circle, there was a place not only for pupils and good friends, but also for hetaerae, slaves, and children.

Despite any sympathy we may have for the teachings of the Stoics or the Epicureans, it must be said that in the end, progress was not achieved by these schools. The path to the future was indicated much more by religion, irrationalism, and fanaticism (Figure 1.94).

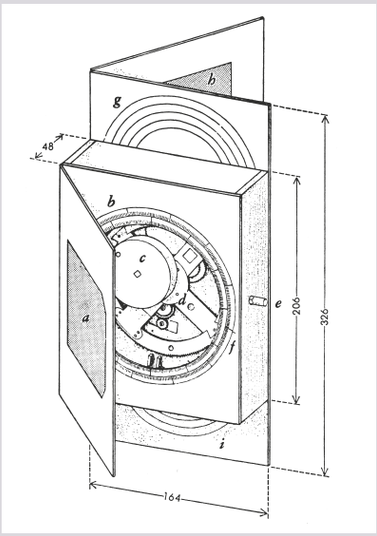


▲ **Figure 1.88** According to Hero, a tunnel can be constructed by digging from both sides of a mountain.

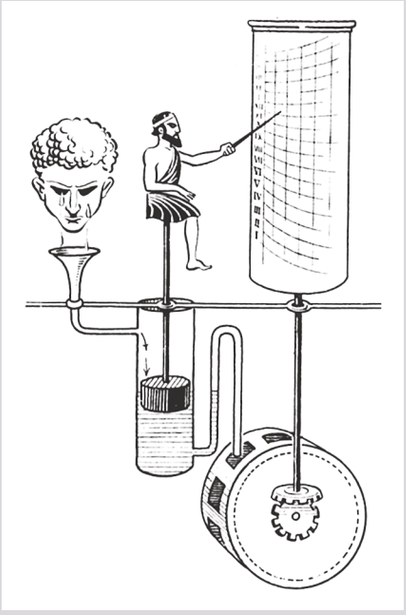
Quotation 1.33

The intake of the New Anio is at the forty-second milestone on the Sublacensian Way, in the district of Simbruvium. The water is taken from the river, which, even without the effect of rainstorms, is muddy and discolored, because it has rich and cultivated fields adjoining it, and in consequence loose banks. For this reason, a settling reservoir was put in beyond the inlet of the aqueduct, in order that the water might settle there and clarify itself, between the river and the conduit. But even despite this precaution, the water reaches the City in a discolored condition whenever there are rains. It is joined by the Herculanean Brook, which has its source on the same Way, at the thirty-eighth milestone, opposite the springs of Claudia, beyond the river and the highway. This is naturally very clear, but loses the charm of its purity by admixture with the New Anio. The conduit of New Anio measures 58,700 paces, of which 49,300 are in an underground channel, 9,400 paces above ground on masonry; of these, at various points in the upper reaches are 2,300 paces on substructures or arches; while nearer the city, beginning at the seventh milestone, are 609 paces on substructures, 6,491 paces on arches. These are the highest arches, rising at certain points to 109 feet. With such an array of indispensable structures carrying so many waters, compare, if you will, the idle Pyramids, or the useless, though famous, works of the Greeks!

—Sextus Julius Frontinus, *The Aqueducts of Rome*, I.15–16 [pp. 357–359]



▲ **Figure 1.89** The Antikythera apparatus (reconstructed by D. Price).



▲ **Figure 1.90** An ancient water clock that runs fully automatically if water is added from time to time (after Hogben).

It is impossible to read the reflections of Pliny the Elder (23–79 ce)—whose ironical tone clearly betrays an atheistic viewpoint—on the relations between men and gods without being affected (Quotation 1.34, Figure 1.95). It is with a certain degree of skepticism that we take note of his assertion that death is the greatest gift of nature, above all when we know how lavishly Pliny enjoyed his life as a wealthy, educated Roman. His discourse on the question of the immortality of the soul—whereby the state of the soul after death corresponds to that before birth, and before birth we begin with nonexistence—finds its echo in the words of Schopenhauer (1788–1860) two thousand years later: *That we are born is not a good omen for immortality*.

In his didactic poem *De rerum natura*, Lucretius attempted to present a complete closed system (Quotation 1.35, Plate V). He, too, considered it important to create a scientific worldview—of both inanimate nature and humankind—in which there are no gods and everything can be explained on its own terms: "because I announce the teaching of great things and eagerly seek to free the spirit from the strait chains of religion."

Lucretius takes quite a different view of man's appropriation of fire from that of Aeschylus in his *Prometheus Bound* of half a millennium earlier. He mentions two possibilities: Lightning striking a tree or perhaps rubbing two sticks together could have started man's first fire. With the sensitivity of a poet, Lucretius describes the power of natural forces and the doubts that arise in men's minds when they are confronted with these powers: "…whether by any chance we have to do with some immeasurable power of the gods, able to make the bright stars revolve with their different movements." It is also interesting to compare the opinion of Socrates in Aristophanes' (ca. 446–ca. 386 bce) *Clouds* regarding the punishments meted out by the gods against those who break their oaths with the following lines from Lucretius: "who is strong enough … to hurl lightning bolts and often to shatter his own temples, and as he passes away into the wilds to cast that bolt in his wrath which often passes the guilty by and slays the innocent and undeserving?"

According to the Skeptics, one cannot know whether gods exist. The Stoics taught that the world and godhead are one, while according to the Epicureans, the gods live their own lives independent of us. It was a man of Jewish origin named Philo (20 bce–50 ce) who was the first to attempt to unite a religion, namely Judaism, and its dogmas with Greek philosophy. In this sense, he could be called the first Scholastic. He even gave a "historical background" for the possibility of harmonization of the two by assuming that Plato was a disciple of Moses.

The creator of the last great system of Greek philosophy, Plotinus (ca. 204–270), the founder of Neo-Platonism, formulated his system at a time when Christianity had already gained considerable ground. His philosophy is permeated with religious ideas without aligning itself with the teachings of any particular religion or with the claim to be providing philosophical support to any religion. Plotinus was attacked by the Christians, but later, many of his ideas were adopted by them. In another direction, Julian the Apostate (331–363), in his attempt to reintroduce the old Roman beliefs, called on Neo-Platonism as a philosophical background to support his views. His metaphysics are exceedingly elaborate and foreign to physics and to the way that physicists think. His main question is this: How can it be possible to create a connection between the material world and the idea of a transcendent god whose properties are unknowable, except perhaps that he can be called "the One, the Good"? One might well compare this concept with the homogeneous, unchanging sphere of Parmenides (fl. ca. 500 bce). Nonetheless,

a physicist might feel sympathy with the thought that in the hierarchy of being—the One → spirit (*nous*) → the soul → the material world—the last link in the chain, representing the object of physical research, also partakes of the beauty that descends from above: In the material world, so to speak, we see the reflection of the beauty of the soul. With PLOTINUS we meet for the first time a form of cognition that was foreign to the Greeks yet will play an important role later: ecstasy.

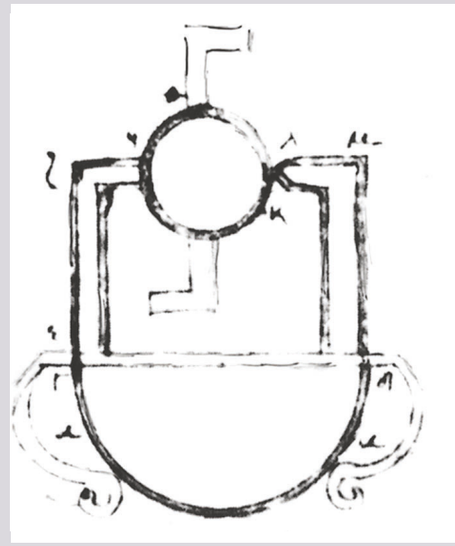
Neo-Platonism is important to us primarily because it is to this system of ideas that we owe the continuity between the Greek and Christian ways of thinking.

"PLOTINUS is both an end and a beginning—an end as regards the Greeks, a beginning as regards Christendom. To the ancient world, weary with centuries of disappointment, exhausted by despair, his doctrine might be acceptable, but could not be stimulating. To the cruder barbarian world, where superabundant energy needed to be restrained and regulated rather than stimulated, what could penetrate in his teaching was beneficial, since the evil to be combated was not languor but brutality" [Russell, 1946, p. 321]. A noted specialist in the history of science, SANTILLANA, formulates the role of Neo-Platonism most succinctly when he says that it was the "carrier wave" that allowed a portion of ancient science to be rescued for the Christian world.

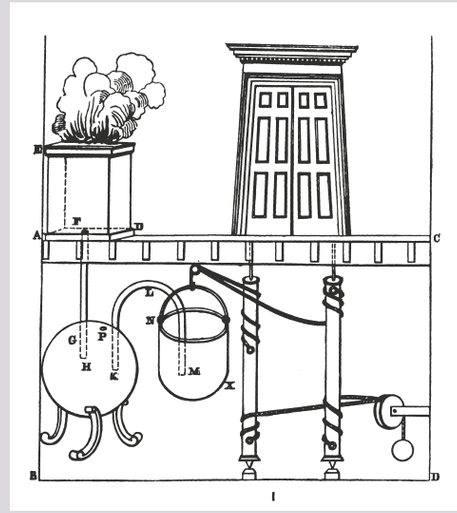
At this point, we should say something about a rather strange mixture of science and mysticism: astrology. This belief system is of significance in the history of science because its practice required a precise knowledge of the movement of the celestial bodies. Thus, even though astrology is a pseudoscience, it served to further true science. The development of astrology was greatly favored by the fact that some earthly phenomena are in fact influenced by the celestial bodies. It was ARISTOTLE who said that the change of seasons on Earth is due to the movement of the Sun, or more precisely, to the inclination of the ecliptic. POSIDONIUS, whose work on astronomy we have already discussed and who was a friend of CICERO, observed during a stay in Spain on the Atlantic coast that the ebb and flow of the tides are directly associated with the orbit of the Moon. An even greater influence of the heavenly bodies on earthly affairs can be derived from the view of the Stoics that man and universe form a unitary whole. This follows as well from a physical model based on atomism. But if everything is interconnected with everything else, it is not such a great leap to conclude that there must be a connection between the fate of men and the movements of the stars (Quotation 1.36).

The belief in auspicious and inauspicious omens—which transcends myth and religious conceptions—is to be found not only in ancient times, but today as well, in the form of more or less seriously accepted superstitions. Particularly in the Near East, superstition was of great significance. Thus on Mesopotamian tablets there have been found more than 5000 such mystical omens. With the penetration of Near Eastern mysticism into the Hellenistic world, astrology became, on the basis of the Greeks' astronomical knowledge, a science. Both according to Roman law and in the Christian world, the practice of astrology was forbidden, but astrologers were nevertheless employed in royal houses well into the seventeenth century.

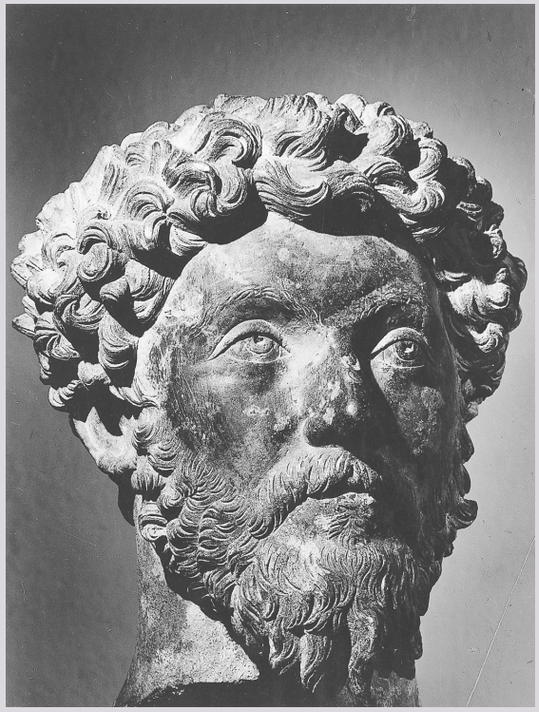
The book of the fictive author HERMES TRISMEGISTUS played an important role in astrology and was considered a treasure trove of secret ancient Egyptian knowledge. In this book, great secrets, among them some of an astrological nature, were revealed in the framework of a conversation between the goddess ISIS and her son HORUS. The world is teeming with gods and demons who are associated with various stars. The fate of men is determined by these demons. A man will be under the influence of the demon who at the moment of the man's birth held the greatest influence according to the configurations of stars in the firmament (Quotation 1.37).



▲ **Figure 1.91** HERO's famous apparatus: the first steam engine; the wheel is driven by the reactive force of a jet of steam; the same principle is used in rockets today.



▲ **Figure 1.92** The air warmed by the sacrificial fire opens the gate to the inner sanctum.



▲ **Figure 1.93** Bronze statue of Marcus Aurelius (121–180), one of the most prominent representatives of Stoic philosophy. Janus-Pannonius Museum in Pécs; the statue was discovered in 1974 in Dunaszekcső (Hungary); photo by K. Nádor.

II.1 When you wake up in the morning, tell yourself: The people I deal with today will be meddling, ungrateful, arrogant, dishonest, jealous, and surly. They are like this because they can't tell good from evil. But I have seen the beauty of good, and the ugliness of evil, and have recognized that the wrongdoer has a nature related to my own—not of the same blood or birth, but the same mind, and possessing a share of the divine. And so none of them can hurt me. No one can implicate me in ugliness. Nor can I feel angry at my relative, or hate him. We were born to work together like feet, hands, and eyes, like the two rows of teeth, upper and lower. To obstruct each other is unnatural. To feel anger at someone, to turn your back on him: these are obstructions.

III.3 Hippocrates cured many illnesses—and then fell ill and died. The Chaldaeans predicted the deaths of many others; in due course their own hour arrived. Alexander, Pompey, Caesar—who utterly destroyed so many cities, cut down so many thousand foot and horse in battle—they too departed this life. Heraclitus often told us the world would end in fire. But it was moisture that carried him off; he died smeared with manure. Democritus was killed by ordinary worms, Socrates by the human kind. And? You boarded, you set sail, you've made the passage. Time to disembark. If it's for another life, well, there's nowhere without gods on that side either. If to nothingness, then you no longer have to put up with pain and pleasure, or go on dancing attendance on this battered crate, your body—so much inferior to that which serves it.

One is mind and spirit, the other earth and garbage.

continued on next page

Today, astrology is again in the ascendant. For the broad public, the face of modernity and scientific thinking is provided by the computer. And so, one may input one's birth data into the computer, which then returns all the good and bad omens that have been handed down since the time of the Chaldeans in all the various astrological works. Thus, the client obtains the predictions and personality profile tailored to the fashion of the times and to individual needs (Quotation 1.38).

Nothing need be added to the criticism of astrology offered by Saint Augustine (354–430) in the fourth century; see Section 1.5.2.

Considered one of the fathers of the Catholic Church, Augustine absorbed all of ancient science in his youth and then led the life of a spoiled, decadent young man. But then he deliberately broke with his past and used his understanding of Hellenistic wisdom to create a forward-looking Christian ideology. It was easy for the apostle Paul to throw aside all worldly wisdom (Quotation 1.39). We may experience Augustine's internal wrestling before taking this step in his *Confessions*, which are a significant contribution to world literature, having served as a model for a long line of works of a similar genre. For the latter, we would like to consider only the following of Augustine's ideas. In the search for truth, beginning with the fact that doubt, thought, and error are certain, Augustine arrived at his *si enim fallor, sum* (for if I err, I am). Thus, he precedes Descartes, although he did not use this truth as the basis of an entire philosophical system.

Even if we decline to accept Augustine's conclusions about the subjectivity of time, it is worthwhile to follow here his thought on the relationship between time, movement, and duration, especially because many great thinkers were inspired by them later (see Section 1.5.3).

In taking our leave of the Hellenistic world in decline, we can do no better than to cite the program that Saint Augustine formulated for the coming epochs:

What should I know?
Nothing but God and the soul.
And nothing further?
No, truly nothing further.

1.5.2 *Augustine on the Absurdity of Astrology*

There was no art by which the future could be foretold ... [but] guesswork was often borne out by mere chance. If a man made a great many predictions, several of them would later prove to be true, but he could not know it at the time and would only hit upon them by chance, simply by opening his mouth to speak. ...

This man, whose name was Firminus, had been educated in the liberal arts and had received a thorough training in rhetoric. He came to consult me, as his closest friend, about some business matters of which he had high hopes, and asked me what prospects I could see in his horoscope, as they call it. ... I did not refuse outright to read the stars for him and tell him what I saw, though I had little faith in it myself. Nevertheless I added that I was almost convinced it was all absurd and quite meaningless. He then told me that his father had studied books of astrology with the greatest interest and had had a friend who shared his enthusiasm for the subject. Each was as intent upon this nonsense as the other, and by pooling their experiences they whetted their enthusiasm to the point that, even when their domestic animals had litters, they would note the exact moment of birth and record the position of the stars, intending to use these observations for their experiments in this so-called art.

Firminus went on to tell me about his own birth. His father had told him that when his mother was pregnant, a female slave in the household of this friend was also expecting a child. Her master was of course aware of her condition, because he used to take very great care to find out even when his dogs were due to have puppies. The two men made the most minute calculations to determine the time of labor of both the women, counting the days, the hours, and even the minutes, and it so happened that both gave birth at exactly the same moment. This meant that the horoscopes which they cast for the two babies had to be exactly the same, down to the smallest particular, though one was the son of the master of the house and the other a slave. For as soon as labor began, each man informed the other of the situation in his house, and each had a messenger waiting, ready to be sent to the other as soon as the birth was announced. As the confinements took place in their own houses, they could easily arrange to be told without delay. The messengers, so Firminus told me, crossed paths at a point which was exactly half way between the two houses, so that each of the two friends inevitably made an identical observation of the stars and could not find the least difference in the time of birth. Yet Firminus, who was born of a rich family, strode along the smoother paths of life. His wealth increased and high honors came his way. But the slave continued to serve his masters. Firminus, who knew him, said that his lot had been in no way bettered.

I believed this story when I heard it, because Firminus was a man whom I could trust. It marked the final end of all my doubts, and my first reaction was to try to redeem Firminus from his interest in astrology. I told him that if I had cast his horoscope and my reading of the stars was correct, I could only have seen in them that his parents were important people, that he belonged to one of the noble families of the town, that he was a freeman by birth, that his upbringing suited his rank, and that his education was liberal. But the slave was born under the very same constellations, and if he had asked me to tell him their meaning, my interpretation of them could not have been true unless I saw in them a family of the meanest sort, the status of a slave, and various other details entirely different from and inconsistent with those which applied to Firminus. This proved that if I were to say what was actually the truth, I should give a different answer to each, though the stars I read were the same; whereas, if I gave the same answer to each, I should be wrong in fact.

—Augustine, *Confessions*, Book VII [Pine-Coffin, pp. 140–142]

1.5.3 *Augustine on Time*

What, then, is time? If no one asks me, I know what it is. If I wish to explain it to him who asks me, I do not know ...

Who is there who will tell me that there are not three times—as we learned when boys and as we have also taught boys—time past, time present, and time future? Who can say that there is only time present because the other two do not exist? Or do they also exist; but when, from the future, time becomes present, it proceeds from some secret place; and when, from times present, it becomes past, it recedes into some secret place? For where have those men who have foretold the future seen the things foretold, if then they were not yet existing? For what does not exist cannot be seen. And those who tell of things past could not speak of them as if they were true, if they did not see them in their minds. These things could in no way be discerned if they did not exist. There are therefore times [future] and times past ...

Figure 1.93 *continued*

IV. 32 The age of Vespasian, for example. People doing the exact same things: marrying, raising children, getting sick, dying, waging war, throwing parties, doing business, farming, flattering, boasting, distrusting, plotting, hoping others will die, complaining about their own lives, falling in love, putting away money, seeking high office and power.

And that life they led is nowhere to be found.

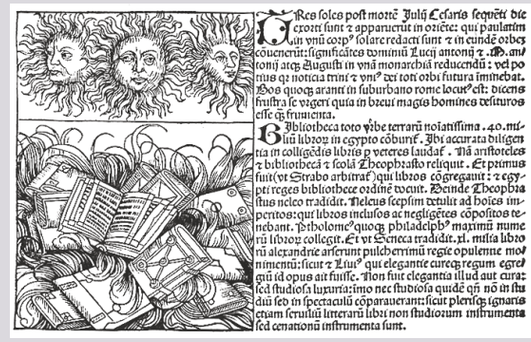
Or the age of Trajan. The exact same things. And that life too—gone.

Survey the records of other eras. And see how many others gave their all and soon died and decomposed into the elements that formed them.

But most of all, run through the list of those you knew yourself. Those who worked in vain, who failed to do what they should have—what they should have remained fixed on and found satisfaction in.

A key point to bear in mind: The value of attentiveness varies in proportion to its object. You're better off not giving the small things more time than they deserve.

—Marcus Aurelius, *Meditations* [pp. 17, 28, 44–45]



▲ **Figure 1.94** Fifteenth-century report on the Library of Alexandria and its destruction (*Schedels Weltchronik*, 1493).

The Library of Alexandria was established ca. 290 BCE by one of Alexander the Great's army commanders, who became king of Egypt under the name Ptolemy I. His successor, Ptolemy II, built the temple to the Muses (μουσείον = *Museion*), of which the library became a part. With its astronomical facilities, medical research laboratory, lecture halls, and copy rooms, it became the scientific center of the Hellenistic world. Several hundred thousand scrolls (500,000, and 700,000, by some reckonings) were housed here; this represents the equivalent of about 100,000 modern volumes. The library burned first in 47 BCE during one of the campaigns of Julius Caesar. The loss was replenished by Mark Antony, who transferred 200,000 scrolls from the library at Pergamum. In 273, in putting down a rebellion, there was another fire, and finally, in 391, the portion preserved in the Serapeum was destroyed by Christian mobs.

Contemporary Arab writers are probably correct in their assessment that there was not much left for Caliph Umar and his soldiers to destroy in 640.

Quotation 1.34

For the imperfections of nature as revealed in man a peculiar consolation is this, that not even God can do everything. He could not, for instance, if he wanted to, commit suicide, which, in trials of our mortal life, is his best gift to men. He cannot make mortals immortal, recall the dead, bring it about that one who has lived has not lived, or that one who has borne office has not borne office. He has no power over the past but oblivion, and, if I may be excused for illustrating our fellowship with God with trivial examples, he cannot make it that twice ten should not be twenty, and so on. By all this is unmistakable revealed the power of nature and the fact that it is this power we call God. I hope I may be pardoned this digression into what I fear have become commonplaces owing to the never-ending debate about God. ...

Beyond the grave lie the empty speculations about the spirits of the dead. For every man it will be the same after his last day as it was before his first. After death neither body nor spirit will have sensation any more than they did before he was born. This vanity of staking a claim on the future and imagining oneself a life in the season of death takes various forms: the immortality of the soul, the transmigration of souls, the life of the shades in the underworld, the worship of the spirits of the dead, even the deification of one who has already ceased to be a man. As if, forsooth, we drew our breath in any way that we could distinguish us from the other animals; as if there were not many creatures who live longer than we do, for whom nobody has imagined a similar immortality. These are the inventions of childish folly, or a mortality greedy of never ceasing to be. Plague take it, what madness is this of repeating life in death? How shall those born ever rest, if sense is to remain with the soul on high or with the ghost below? Nay, this fond fancy destroys nature's chief blessing, death, and doubles the smart of him that is to die by the calculation of what is still to come. If life is to be so sweet, who can find it sweet to have ceased to live? But how much happier, how much more sure, that every man should come to trust himself and take from his proven insensibility of what was before he was born his warrant of the peace that is to be.

—Pliny the Elder, *Natural History* [Farrington, Vol. 2, p. 136]



▲ **Figure 1.95** Gaius Plinius Secundus (23–79), also known as Pliny the Elder, lost his life, according to his nephew, known as Pliny the Younger, while observing the eruption of Vesuvius. His 37-volume work with the title *Historia Naturalis* was the result, according to the author's own report, of having consulted 2000 volumes written by hundreds of authors, and it contains—contrary to earlier practice—an extensive list of sources, 473 authors being listed by name. The work represents a popular account of all of classical knowledge, and may be described as comprehensive, although not particularly critical, and in places even sensationalist. It was widely read until well into the fifteenth century. (Diocesan Library of Székesfehérvár.)

[If] there are times past and future, I wish to know where they are. But if I have not yet succeeded in this, I still know that wherever they are, they are not there as future or past, but as present. For if they are there as future, they are there as "not yet"; if they are there as past, they are there as "no longer." Wherever they are and whatever they are they exist therefore only as present. Although we tell of past things as true, they are drawn out of the

memory—not the things themselves, which have already passed, but words constructed from the images of the perceptions which were formed in the mind, like footprints in their passage through the senses. My childhood, for instance, which is no longer, still exists in time past, which does not now exist. But when I call to mind its image and speak of it, I see it in the present because it is still in my memory.

Whether there is a similar explanation for the foretelling of future events—that is, of the images of things which are not yet seen as if they were already existing—I confess, O my God, I do not know. But this I certainly do know: that we generally think ahead about our future actions, and this premeditation is in time present; but that the action which we premeditate is not yet, because it is still future. When we shall have started the action and have begun to do what we were premeditating, then that action will be in time present, because then it is no longer in time future.

Whatever may be the manner of this secret foreseeing of future things, nothing can be seen except what exists. But what exists now is not future, but present. When, therefore, they say that future events are seen, it is not the events themselves, for they do not exist as yet (that is, they are still in time future), but perhaps, instead, their causes and their signs are seen, which already do exist. Therefore, to those already beholding these causes and signs, they are not future, but present, and from them future things are predicted because they are conceived in the mind. These conceptions, however, exist now, and those who predict those things see these conceptions before them in time present.

Let me take an example from the vast multitude and variety of such things. I see the dawn; I predict that the sun is about to rise. What I see is in time present, what I predict is in time future—not that the sun is future, for it already exists; but its rising is future, because it is not yet. Yet I could not predict even its rising, unless I had an image of it in my mind; as, indeed, I do even now as I speak. But that dawn which I see in the sky is not the rising of the sun (though it does precede it), nor is it a conception in my mind. These two are seen in time present, in order that the event which is in time future may be predicted.

Future events, therefore, are not yet. And if they are not yet, they do not exist. And if they do not exist, they cannot be seen at all, but they can be predicted from things present, which now are and are seen ...

But even now it is manifest and clear that there are neither times future nor times past. Thus it is not properly said that there are three times, past, present, and future. Perhaps it might be said rightly that there are three times: a time present of things past; a time present of things present; and a time present of things future. For these three do coexist somehow in the soul, for otherwise I could not see them. The time present of things past is memory; the time present of things present is direct experience; the time present of things future is expectation.

If we are allowed to speak of these things so, I see three times, and I grant that there are three. Let it still be said, then, as our misapplied custom has it: "There are three times, past, present, and future." I shall not be troubled by it, nor argue, nor object—always provided that what is said is understood, so that neither the future nor the past is said to exist now. There are but few things about which we speak properly—and many more about which we speak improperly—though we understand one another's meaning ...

Quotation 1.35

I love to approach virgin springs and there to drink; I love to pluck new flowers, and to seek an illustrious chaplet for my head from fields whence before this the Muses have crowned the brows of none: first because my teaching is of high matters, and I proceed to unloose the mind from the close knots of superstition; next because the subject is so dark and the lines I write so clear, as I touch all with the Muses' grace. For even this seems not to be out of place; but as with children, when physicians try to administer rank wormwood, they first touch the rims about the cups with the sweet yellow fluid of honey, that unthinking childhood may be deluded as far as the lips, and meanwhile may drink up the bitter juice of wormwood, and though beguiled be not betrayed, but rather by such means be restored and regain health, so now do I: since this doctrine commonly seems somewhat harsh to those who have not used it, and the people shrink back from it, I have chosen to set forth my doctrine to you in sweet-speaking Pierian song, and as it were to touch it with the Muses' delicious honey, if by chance in such a way I might engage your mind in my verses, while you are learning to see in what shape is framed the whole nature of things.

Book I, lines 927–950.

If you hold fast to these convictions, nature is seen to be free at once and rid of proud masters, herself doing all by herself of her own accord, without the help of the gods. For I appeal to the holy hearts of the gods, which in tranquil peace pass untroubled days and a life serene: who is strong enough to rule the sum of the immeasurable, who to hold in hand and control the mighty bridle of the unfathomable, who to turn about all the heavens at one time and warm the fruitful worlds with ethereal fires, or to be present in all places and at all times, so as to make darkness with his clouds and to shake the serene shy with thunder, then to launch lightnings and often to shatter his own temples, and as he passes away into the wilds to cast that bolt in his wrath which often passes the guilty by and slays the innocent and undeserving?

Book II, lines 1090–1104.

That you may not here perhaps be quietly asking yourself the question, it was lightning that first brought fire down to the earth for mortals, and from this all blazing flames have been spread abroad. For we can see many things catch fire implanted with the flames from on high, when the stroke from heaven has given them its heat.

continued on next page

Quotation 1.35, ***continued***

And yet also when a branching tree struck by the winds, swaying and tossed about, leans on the branches of a tree, fire is pressed out by the great force of the friction, at times the burning glare of flame flashes out while branches and trunks are rubbed together. Either of these causes may have given fire to mankind.

Book V, lines 1091–1101.

O unhappy race of mankind, to ascribe such doings to the gods and to attribute to them bitter wrath as well! What groans did they then create for themselves, what wounds for us, what tears for generations to come! ...

Do not nations and peoples tremble, do not proud kings huddle up their limbs smitten with fear of the gods, lest for some base deed or proud word the solemn time of punishment be now brought near at hand?

When also the supreme violence of a furious wind upon the sea sweeps over the waters the chief admiral of a fleet along with his mighty legions and elephants, does he not crave the gods' peace with vows, does he not in his panic seek with prayers the peace of the winds and favoring breezes? But all in vain, since none the less he is often caught up in the furious hurricane and driven upon the shoals of death. So true is it that some hidden power grinds down humanity, and seems to trample upon the noble rods and the cruel axes, and hold them in derision.

Then when the whole earth trembles beneath our feet, when cities are shaken and fall or threaten to fall, what wonder if the sons of men feel contempt for themselves, and acknowledge the great potency and wondrous might of gods in the world, to govern all things?

Book V, lines 1195–1240.

—Lucretius, *De Rerum Natura*

Quotation 1.36

If, then, a man knows accurately the movements of all the stars, the sun, and the moon, so that neither the place nor the time of any of their configurations escapes his notice, and if he has distinguished in general their natures as the result of previous continued study, even though he may discern, not their essential, but only their potentially effective qualities, such as the sun's heating and the moon's moistening, and so on

continued on next page

I once heard a learned man say that the motions of the sun, moon, and stars constituted time; and I did not agree. For why should not the motions of all bodies constitute time? What if the lights of heaven should cease, and a potter's wheel still turn round: would there be no time by which we might measure those rotations and say either that it turned at equal intervals, or, if it moved now more slowly and now more quickly, that some rotations were longer and others shorter? And while we were saying this, would we not also be speaking in time? Or would there not be in our words some syllables that were long and others short, because the first took a longer time to sound, and the others a shorter time? ...

I thirst to know the power and the nature of time, by which we measure the motions of bodies, and say, for example, that this motion is twice as long as that ...

Dost thou command that I should agree if anyone says that time is "the motion of a body"? Thou dost not so command. For I hear that no body is moved but in time; this thou tellest me. But that the motion of a body itself is time I do not hear; thou dost not say so. For when a body is moved, I measure by time how long it was moving from the time when it began to be moved until it stopped. And if I did not see when it began to be moved, and if it continued to move so that I could not see when it stopped, I could not measure the movement, except from the time when I began to see it until I stopped. But if I look at it for a long time, I can affirm only that the time is long but not how long it may be. This is because when we say, "How long?", we are speaking comparatively as: "This is as long as that," or, "This is twice as long as that"; or other such similar ratios. But if we were able to observe the point in space where and from which the body, which is moved, comes and the point to which it is moved; or if we can observe its parts moving as in a wheel, we can say how long the movement of the body took or the movement of its parts from this place to that. Since, therefore, the motion of a body is one thing, and the norm by which we measure how long it takes is another thing, we cannot see which of these two is to be called time. For, although a body is sometimes moved and sometimes stands still, we measure not only its motion but also its rest as well; and both by time! Thus we say, "It stood still as long as it moved," or, "It stood still twice or three times as long as it moved"—or any other ratio which our measuring has either determined or imagined, either roughly or precisely, according to our custom. Therefore, time is not the motion of a body ...

Does not my soul most truly confess to thee that I do measure intervals of time? But what is it that I thus measure, O my God, and how is it that I do not know what I measure? I measure the motion of a body by time, but the time itself I do not measure. But, truly, could I measure the motion of a body—how long it takes, how long it is in motion from this place to that—unless I could measure the time in which it is moving?

How, then, do I measure this time itself? Do we measure a longer time by a shorter time, as we measure the length of a crossbeam in terms of cubits? ...

It is in you, O mind of mine, that I measure the periods of time. Do not shout me down that it exists [objectively]; do not overwhelm yourself with the turbulent flood of your impressions. In you, as I have said, I measure the periods of time. I measure as time present the impression that things make on you as they pass by and what remains after they have passed by—I do not measure the things themselves which have passed by and left their impression on you. This is what I measure when I measure periods of time. Either, then, these are the periods of time or else I do not measure time at all ...

What are we doing when we measure silence, and say that this silence has lasted as long as that voice lasts? Do we not project our thought to the measure of a sound, as if it were then sounding, so that we can say something concerning the intervals of silence in a given span of time? For, even when both the voice and the tongue are still, we review—in thought—poems and verses, and discourse of various kinds or various measures of motions, and we specify their time spans—how long this is in relation to that—just as if we were speaking them aloud.

—Augustine, *Confessions* [Outler]

Quotation 1.38

Earth is a ship that floats on the light, magnetic and other waves emanating from the celestial bodies, and all that lives onboard bathes in this ocean of vibrations. Its passengers, the inhabitants of our planet, react according to their inner natures to these cosmic waves that for six thousand years, astrology has been studying and sorting according to their rhythm, intensity, and frequency in order to determine the consequences and to predict human fate. Indeed, astrology has flowered for sixty centuries and has defied all attacks, critics, prohibitions, curses, and the stake. What religion, philosophy, or science can boast of such a past? [author's translation]

—Mme. Soleil

Quotation 1.39

Where *is* the wise? where *is* the scribe? where *is* the disputer of this world? hath not God made foolish the wisdom of this world?

For after that in the wisdom of God the world by wisdom knew not God, it pleased God by the foolishness of preaching to save them that believe,

For the Jews require a sign, and the Greeks seek after wisdom:

But we preach Christ crucified, unto the Jews a stumbling block, and unto the Greeks foolishness;

But unto them which are called, both Jews and Greeks, Christ the power of God, and the wisdom of God.

Because the foolishness of God is wiser than men; and the weakness of God is stronger than men.

For ye see your calling, brethren, how that not many wise men after the flesh, not many mighty, not many noble, are called:

But God hath chosen the foolish things of the world to confound the wise; and God hath chosen the weak things of the world to confound the things which are mighty.

—I Corinthians 1:20–27

Quotation 1.36, ***continued***

with the rest; and if he is capable of determining in view of all these data, both scientifically and by successful conjecture, the distinctive mark of quality resulting from the combination of all the factors: what is to prevent him from being able to tell on each given occasion the characteristics of the air from the relations of the phenomena at the time, for instance, that it will be warmer or wetter? Why can he not, too, with respect to an individual man, perceive the general quality of his temperament from the ambient at the time of his birth, as for instance that he is such and such in body and such and such in soul, and predict occasional events, by use of the fact that such and such an ambient is attuned to such and such a temperament and is favorable to prosperity, while another is so attuned and conduces to injury? ...

[Many], for the sake of gain, claim credence for another art in the name of this, and deceive the [people], [claiming to be able] to foretell many things, even those that cannot naturally be known beforehand, while to the more thoughtful they have thereby given occasion to pass equally unfavorable judgment upon the natural subjects of prophecy. Nor is this deservedly done; it is the same with philosophy—we need not abolish it because there are evident rascals among those that pretend to it. Nevertheless it is clear that even though one approach astrology in the most inquiring and legitimate spirit possible, he may frequently err, not for any of the reasons stated, but because of the very nature of the thing and his own weakness in comparison with the magnitude of his profession.

—Ptolemy, *Tetrabiblios* 1.2, [pp. 11–15]

Quotation 1.37

Every star has its own demon: good or evil according to its nature, or, better said, according to its actions, since those are what form the nature of the demon. Some demons are good in part and evil in part. All these demons are the champions of all earthly occurrences. They guard the affairs of states and individuals; sometimes they sow confusion. They form our souls in their own way, penetrate into our nerves, brains, blood vessels, and into the depths of our entrails. Each of us, at the moment of receiving life and soul, is taken hold of by the strongest demon and ruled his whole life long; however, the constellation of stars at the moment of birth determined the strongest demon.

—Hermes Trismegistus



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

CHAPTER 2

The Stewards of the Heritage

We make no judgment based on the fame or number of those who hold an opinion, but on the basis of the internal substantiation of a point of view (*De animalibus* 281). Where instruction in faith and custom are concerned, there comes Augustine, where medicine, there Galen and Hippocrates, where the natural sciences, there Aristotle as the greatest authority (II *Sententiarum* 247).

—Albertus Magnus

Every form of science or knowledge is something good. Otherwise, God, in whom nothing can be evil, could not have knowledge of all good and all evil; and therefore the striving for every form of scientific knowledge or knowledge of every sort of thing—whether of a good or of an evil nature—is good in and of itself; yet this striving can turn out to be good or evil depending on additional circumstances: The difference generally depends on the goals.

—Thomas Aquinas, *Quaest. quodlibet*. IV. 9. 16

DOI: 10.1201/9781032697697-3
This chapter has been made available under a CC-BY-NC-ND license.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

The Stewards of the Heritage

2.1 The Thousand-Year Balance Sheet

2.1.1 *Why Did Progress Stall?*

One of the explanations commonly given for the great scientific revolution of the sixteenth and seventeenth centuries, as well as for the industrial revolution that followed, is that by this time Europe had mastered the knowledge of the ancient Greeks. But why did this require a thousand-year wait? Why wasn't there a scientific and industrial revolution right after the blossoming of Hellenistic science? Or in other words, why did Greek science go into decline after having begun with such promising momentum?

In effect, what we need to explore here is how the fifth and the sixteenth centuries differ from each other. That this is a complex question, without an unambiguous answer, is also evident from the fact that the Golden Age of Greek science was *not* in the fifth century: By and large, the best results were obtained as early as 200 BCE. The Greeks, and later the Romans, lived for centuries in relative affluence and with a high degree of social organization, all the while in possession of a scientific treasure that nearly equaled that of the sixteenth century. Furthermore, we can pose a similar question with regard to just the seventeenth century and contrasting Europe with Islam or with the Indian and Chinese cultures. At that time, these Eastern cultures stood at least at the European level, and indeed, in many areas Islamic science exceeded the level of Western Europe. So the question here can be put this way: Why did the scientific–technical revolution start in Western Europe and not in an Islamic center of culture?

It is generally agreed that the basic flaw holding back Greek culture was the utter lack of connection between science and practice. The results of science were not translated into practice, and practice made no demands on science, where by practice we mean of course applications of science in the service of society. It can be said that in ancient society we see no trace of science becoming a productive force. Mechanisms, no matter how ingeniously constructed, served purely for amusement or to impress or strike fear into the hearts of pious pilgrims in the temples. Even an outstanding product of ancient technology, the Antikythera mechanism, with its representation of cycles and epicycles, served only astrological ends in all probability. The production of goods was increased not by the introduction of machinery, but by increasing the number of slaves [Santillana 1961].

This separation of scientific theory from practical application was nothing but a logical consequence of the societal structure, the division of society into citizen and slave (Quotations 2.1 and 2.2). However, the system of production based on the exploitation of slaves eventually was not able to keep up with the increasing demands placed on it by the Roman Empire. The legendary luxury of the aristocracy, constant warfare, the expense of subsidizing tribes along the borders, and the gratification of the plebian masses' desire for *panem et circenses* necessarily led, under such a system of production, to economic catastrophe.

Another frequently-mentioned cause of the stagnation of the ancient methods of production is that, despite sufficient scientific background knowledge, the tech-

Quotation 2.1

The slave is a living tool and the tool a lifeless slave.

—Aristotle, *Nicomachean Ethics*

Quotation 2.2

Slavery no longer paid; it was for that reason it died out. But in dying it left behind its poisoned sting—the stigma attaching to the productive labor of freemen. This was the blind alley from which the Roman world had no way out: slavery was economically impossible, the labor of freeman was morally ostracized. The one could be the basic form of social production no longer; the other, not yet.

—Frederick Engels, *The Origin of the Family*

Quotation 2.3

But all of Europe that is level and has a temperate climate has nature to cooperate with her toward these results; for while in a country that is blessed by nature everything tends to peace, in a disagreeable country everything tends to make men warlike and courageous; and so both kinds of country receive benefits from each other, for the latter helps with arms, the former with products of the soil, with arts, and with character-building. But the harm that they receive from each other, if they are not mutually helpful, is also apparent; and the might of those who are accustomed to carry arms will have some advantage unless it be controlled by the majority. However, this continent has a natural advantage to meet this condition also; for the whole of it is diversified with plains and mountains, so that throughout its entire extent the agricultural and civilized element dwells side by side with the warlike element. But of the two elements the one that is peace-loving is more numerous and therefore keeps control over the whole body; and the leading nations too—formerly the Greeks and later the Macedonians and the Romans—have taken hold and helped. And for this reason Europe is most independent of other countries as regards both peace and war; for the warlike population which she possesses is abundant and also that which tills her soils and holds her cities secure.

—Strabo, *Geography*

Quotation 2.4

Emilian: What did you do to keep your people from falling into the hands of the Teutons?

Romulus: Nothing.

Emilian: What have you done to keep Rome from being violated as I was?

Romulus: Nothing.

Emilian: And how will you justify yourself? You are accused of having betrayed your empire.

Romulus: I didn't betray my empire. Rome betrayed herself. Rome knew the truth but chose violence. Rome knew humaneness but chose tyranny. Rome doubly demeaned herself: before her own people and before the other nations in her power. You are standing before an invisible throne, Emilian; before the throne of all the Roman emperors, of whom I am the last. Shall I touch your eyes that you may see this throne, this pyramid of skulls down whose steps cascade rivers of blood in endless waterfalls, generating Rome's power? What kind of an answer

continued on next page

nology for building mechanisms was lacking. However, one may point again to the Antikythera apparatus, whose creators certainly had the same level of technological skills as the builders of Watt's steam engine.

The decisive causes of the decline of Hellenistic culture, then, are these:

1. Productive work was held in low esteem by society.
2. Slave labor could not be replaced by machines because science and practice were not connected.
3. The civilized world had become financially bankrupt.
4. The Roman Empire could not develop undisturbed, in any case, because it was surrounded by energetic and expansionist barbarian tribes.

We might now attempt an answer to the question of how, with respect to social progress, the fifth and sixteenth centuries were different. Let us consider the above factors one by one.

During the course of the Middle Ages, productive work received increasing recognition. In this, as we shall see, some of the monastic orders played a decisive role.

In the Middle Ages a technical revolution took place, quite independent of science—namely, in the use of water and wind energy, the modernization of agricultural equipment, and the introduction of devices for the improved use of animal power. Thus, the arduous work formerly accomplished by slaves became both more humane and more effective.

The general standard of living rose steadily in the Middle Ages along with the effectiveness of the means of production, except for some lengthy periods of warfare or epidemics (the Hundred Years' War between England and France, the Black Death), during which intellectual development stagnated as well.

After the end of the Islamic expansion and after the expulsion of the conquerors from the north and east (Normans and Hungarians) or after their successful integration, a community of peoples came into being in Europe that was capable of defending itself against further external aggression. Even the internal conflicts that are inherent in development (Investiture Controversy, power struggles among states) did not endanger this community.

If we also take geography into consideration, we can answer the question of why the scientific and technological revolution took place in Western Europe and not in one of the Islamic lands (Quotation 2.3).

Clearly, geographic factors play an important role in the development of the means of production, as well. The variegated European landscape, with its many small rivers and streams suitable for the establishment of water mills, as well as open seashores and plains for the development of wind power, made possible the construction of many mechanisms that worked locally, in small units, and yet were able to replace the work of many people. The cultures of the great river valleys, however, did not have access to such natural resources; also the survival of large numbers of people crowded in the narrow, productive valley areas required a particularly intensive land cultivation, which could only be achieved by means of manual labor. This form of agriculture can be found even today in some countries.

In sum, the technological development and economic structure of the Middle Ages were necessary factors for facilitating the adaptation of the science of antiquity. Furthermore, one of the greatest inventions of this era, the printing press, made this knowledge accessible to much wider segments of the society than could have been imagined in the antique world [Farrington 1949].

2.1.2 *Europe Takes Shape*

During the period of its greatest power, in the second and third centuries, the Roman Empire stretched over an area of four thousand square miles and was divided into two parts by a several-hundred-mile-wide swath between the Adriatic and the Danube (Figure 2.1). It is not surprising that an empire of such size would eventually collapse, but we cannot fail but be impressed by the Romans' organizational ability and their political and military capabilities, which enabled them to build such a vast empire and maintain it for almost half a millennium. The enormous distances made control of the empire extremely difficult even in peacetime, even though the entire Roman Empire was served by a well-maintained network of roads and a highly developed system for the transmission of messages.

Nonetheless, in wartime—above all, in the empire's last century, when several border conflicts arose simultaneously—the organization of troop movements presented great difficulties. Along the entire northern border of the empire, powerful German tribes bent on conquest were biding their time, waiting for the opportunity to penetrate the territory of the empire, whether as allies, enemies, or mercenaries. To the east, there was for a time an uneasy peace with the Parthian Empire. The western German tribes were eyeing the conquest of the wealthier, more fertile, and climatically more favorable Roman provinces, while in the east, the Goths, in flight from the Huns, were seeking admittance to a new home within the borders of the Roman Empire. It is significant that many of the Germans who fought against Rome had served earlier as Roman mercenaries and had thereby learned the Roman arts of warfare. Moreover, they were direct witnesses to the internal weakness of the empire (Quotation 2.4).



▲ **Figure 2.1** The Roman Empire at the time of its greatest extent.

During the last centuries of the Roman Empire, the Roman emperors did all they could to stem the tide of disintegration. Thus Diocletian attempted to replace the despotic monarchy with a tetrarchy, whereby four different rulers were each responsible for a different portion of the empire (Figure 2.2). Constantine made peace

Quotation 2.4, ***continued***

do you expect I can hand down to you, as I sit on top of the colossus that is Roman history? What can be said about your wounds by your Emperor who sits enthroned above the corpses of his own sons and the sons of strangers, above the mound of human sacrifices swept to his feet by the wars of Rome's glory and the gladiatorial games for Rome's amusement? Rome has grown weak, a tottering old hag, but her guilt has not been expiated and her crimes not erased. Overnight the new day has dawned. The curses of Rome's victims are being fulfilled. The axe is put to the trunk, the rotten tree is being felled. The Teutons are coming; we have spilled the stranger's blood; we must now pay back with our own.

—Friedrich Dürrenmatt, *Romulus the Great*



▲ **Figure 2.2** The two emperors (Augusti) Diocletian and Maximian from the San Marco Cathedral in Venice. Behind them are to be found, embracing each other as well, Constantius Chlorus (father of Constantine the Great) and Galerius (Caesares).

with the greatest ideological force of the time, made even stronger by persecution: Christianity. Under the military emperors, a chain of fortresses was established along the Rhine and the Danube. The last emperor of a unified empire, Theodosius (347–395), divided the Roman Empire into two parts, the Western and the Eastern Roman Empires, with Rome and Constantinople as their respective capitals.

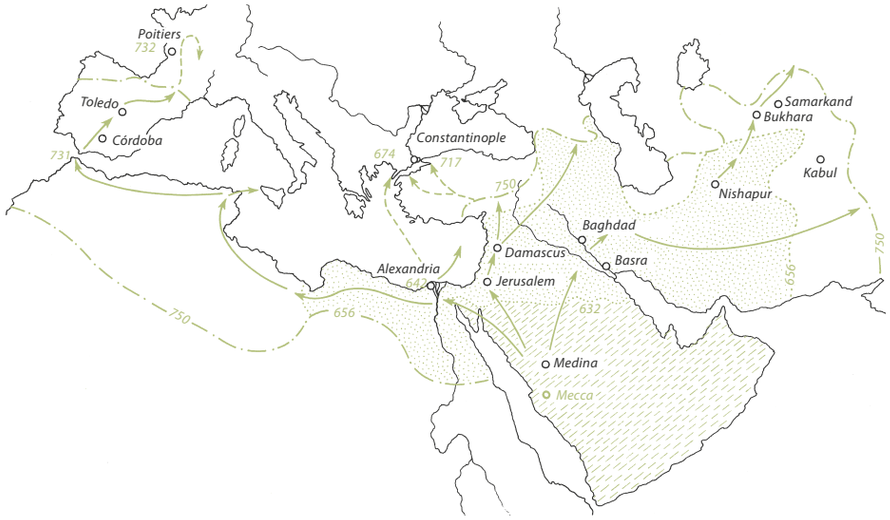
The Western Roman Empire fell in 476 during the war with the Visigoths and Huns. The Germanic warlord Odoacer deposed the last Western Roman emperor, the adolescent Romulus Augustulus. Odoacer in turn perished at the hands of the Ostrogothic king Theodoric the Great. It is revealing that Theodoric occupied Italy as governor of the Eastern Roman Empire and founded a Gothic state there. The Eastern Roman Empire was to survive another millennium before Constantinople finally fell to the Turks in 1453.

While Goths, Vandals, and other Germanic tribes battled one another on the territory of the Western Roman Empire, in the Eastern Roman Empire, Justinian (482–565) sought to renew the empire's former splendor (Plate VI). He put an end to Gothic rule in Italy and destroyed the short-lived Vandal kingdom in North Africa. Two accomplishments of cultural significance are associated with his name: Under his rule, the building of the basilica Hagia Sophia (holy wisdom) was begun, and he systematized all the laws then in effect. The resulting *Codex Justiniani* is the basis of all European legal systems. It was also at his behest that, in 529, the Neo-Platonic Academy in Athens was closed.

The Eastern Roman Empire—or Byzantium, as it was later called—where Greek influence staged a return after the death of Justinian to the extent that Greek supplanted Latin as the official language, is of considerable importance to cultural history for a variety of reasons. Although the ability to develop Greek knowledge further was no longer there, Byzantium continued to be its custodian, not only for itself, but for all of Europe. Furthermore, it served as a bulwark in the east against the Arab expansion that threatened Europe like a pincer from the south and the east. It also had major symbolic significance: even the Germanic tribal leaders felt secure only when they had been acknowledged by the emperor from this distant city—almost the only surviving large city in Europe at that time.

In the centuries that followed right after the fall of the Roman Empire, the Europe to be born was threatened by dangers that, although not deadly, could nonetheless have had grave consequences. But even in its chaotic state, Europe was able to weather these storms. Today only a few names of provinces, such as Catalonia (land of the Goths) and Andalusia (whose name derives from the Vandals), remain from the short-lived states that were created by Germanic tribes. However, the Frankish Empire, which began to take form toward the end of the fifth century, has persisted in more-or-less definite political form up to the present day as France. Europe certainly needed such a strong bastion of defense, because from the Arabian Peninsula, which had previously been known only for its predatory Bedouin tribes, a historically unique military expansion began with the appearance of Mohammed.

Islamic time reckoning starts with the year 622, which marks Mohammed's flight (*hijra*) from Mecca to Medina (*Yathrib*). In the course of half a century, the Arabs had subjugated all of North Africa, and in the northeast they had conquered the Persian Empire and part of Byzantium. Crossing the Straits of Gibraltar, they overran Spain and penetrated into France through the Pyrenees. Here, in 732, they were stopped by the Franks near Poitiers and were eventually pushed back beyond the Pyrenees (Figures 2.3 and 2.4). The following centuries brought new dangers to Europe: the invasions of the Normans and later of the Hungarians.

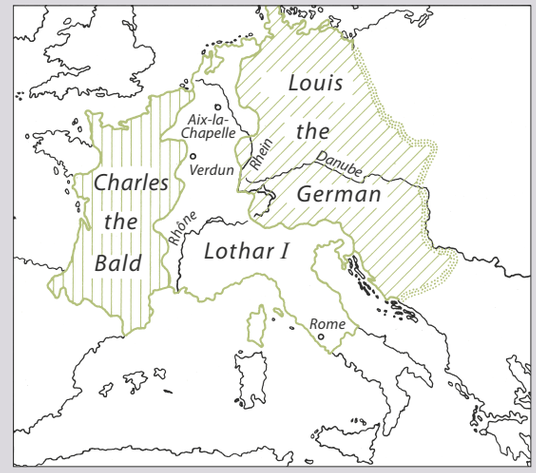


◀ **Figure 2.3** The Arab expansion.

قَالَ رَسُولُ ٱللَّهِ: طَلَبُ ٱلْعِلْمِ فَرِيضَةٌ عَلَى كُلِّ مُسْلِمٍ .

▲ **Figure 2.4** The Koran, the holy book of Islam, not only calls its believers to battle against the infidel, but exhorts them to learning. Here is written,

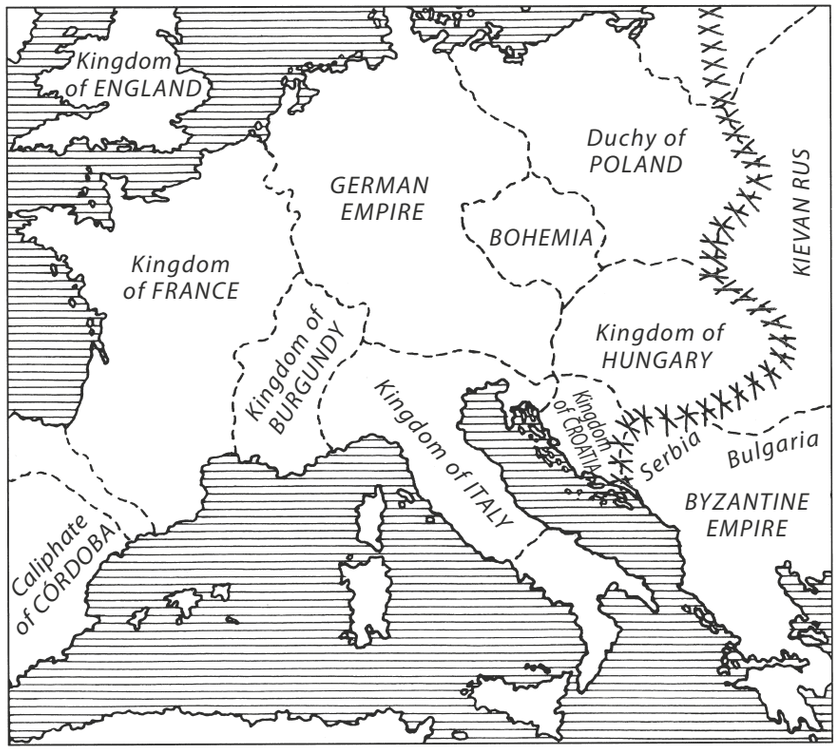
> The holy prophet has spoken: the striving for knowledge is the duty of every Muslim.



▲ **Figure 2.5** A decisive event for the division of Europe: the splitting of the Carolingian Empire according the Treaty of Verdun in 843 among Charlemagne's grandsons, the sons of Louis the Pious (Louis the German, Charles the Bald, and Lothar I). The Treaty of Meerssen of 870 emphasized these divisions, leading to Europe as we know it today.

Following long drawn-out fighting, these two peoples were integrated into the community of Europeans in different ways: Some of the Normans settled in Normandy, while others founded a Norman state in southern Italy and Sicily; the Hungarians established their independent statehood in the Carpathian Basin after adopting Christianity around the end of the tenth century.

Meanwhile in France, Charlemagne, the most talented representative of the Carolingian dynasty, successors to the Merovingians, considered himself to be the heir to the Roman Empire. Having succeeded in pushing back the Avars, he extended his empire to Pannonia (the western part of present-day Hungary). After the conquest of the empire of the Lombards, who had occupied the Po Basin in the seventh century, northern Italy also became part of his empire. In 800, Charlemagne was crowned emperor by the Pope in Rome. Charlemagne's grandsons divided the empire at the river Rhine by the Treaty of Verdun (Figures 2.5, 2.6,



◀ **Figure 2.6** At the turn of the millennium and through the first half of the eleventh century, the outlines of the European states as they appear today began to emerge in Central and Eastern Europe. The Eastern boundary of the Holy Roman Empire of the German Nation was fixed by three states whose power was on the rise.

- Poland: The organization of several West-Slavic tribes into a single state began in the ninth century. In the following century came the acceptance of Christianity. Bolesław I, the Brave (967–1025) ruled over an empire that stretched from the Oder to Kiev. He sought the crown from the Pope at the same time as the Hungarian king Stephen I.
- Hungary: After the settlement of the Carpathian Basin around 896, the Magyars undertook raids throughout Europe before suffering a decisive defeat at the hand of Otto I, the Great, in 955 at Lechfeld. The "father of his country" Stephen I (Saint Stephen) received the crown from Pope Sylvester II and established a strongly centralized feudal state similar to those in the West.
- Croatia: In the ninth and tenth centuries, the South-Slavic tribes became organized into a state. In 910, after the acceptance of Catholicism, Tomislav I was given the title of king; Croatia joined in personal union with Hungary at the end of the eleventh century.

With their connection to the Roman Catholic Church, these states joined the Western cultural community almost automatically due to their common Latin language and alphabet.

continued on next page

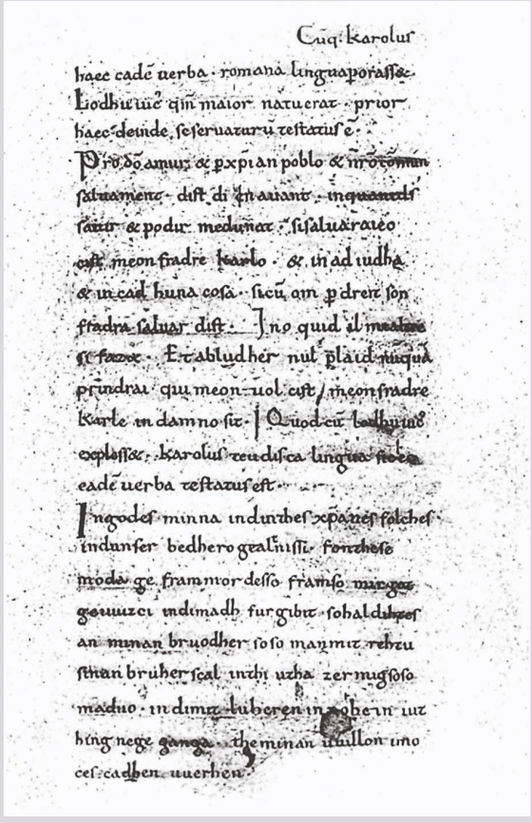
Figure 2.6 *continued*

In southeastern Europe, the Eastern Roman Empire—Byzantium—still existed, and indeed, it remained a great power. To the east, Byzantium battled the Bulgarians, who had settled the lower reaches of the Danube by the seventh century. The Bulgarian tsar Simeon (ruled 893–927) made his country into a major power. After the turn of the millennium, Bulgaria was for a time under Byzantine sovereignty. Byzantium's cultural influence and at times its political power extended throughout the Balkans and to the regions of the northern Danube and the Carpathians. Kievan Rus was the first firmly established state.

The "line of intersection" where the cultural radiation from Rome met that from Byzantium—a broken line to be sure—cannot be sharply delineated, particularly when one considers that parts of the Balkans and Central Europe fell under Turkish dominion.

A unified European culture was able to develop only after the withdrawal of the Turks and the opening of the greatest of the Slavic empires, Russia, in the eighteenth century.

The difference in cultural traditions is evident to this day, sometimes with tragic consequences.



▲ **Figure 2.7** An important cultural document from these times: the Oaths of Strasbourg, pledges of allegiance taken by Charles the Bald and Louis the German on the occasion of their alliance against Lothar I. It is written in Old High German and Early Romance. The Romance text begins "*Pro deo amur*," while the Old High German part reads "*In godes minna*."

and 2.7). This is how, in addition to France, what would later become the Holy Roman Empire and the Germany of today were created. The first ruler of the Holy Roman Empire, Otto I (the Great), was crowned in 962.

We should not forget to mention England, which was to play so great a role in the future. In the fifth and sixth centuries, Angles, Saxons, and other Germanic peoples crossed the English Channel and subdued the remaining Romans and aboriginal Celts in Great Britain. Following the battles with the Danes, England would emerge strengthened under Alfred the Great. The Norman invasion, led by William the Conqueror, was of enormous cultural significance. In 1066, the Anglo-Saxons were defeated at the Battle of Hastings (Figure 2.8). The Normans brought high French culture to the locals, and eventually the two peoples completely assimilated with each other (Figure 2.9).

Thus the cultural roles played by the individual nations can be summarized as follows: Through the Byzantine Empire, about which we shall have much more to say, ancient culture was preserved and transmitted. The city of Byzantium was the model and symbol of civilized urban life.

The legacy of antiquity was preserved longest, aside from Byzantium, in Italy. The Italian cities, although destroyed to a large extent, managed to preserve their cultural heritage practically uninterrupted into the Middle Ages. In the other European states, cities that would become cultural centers had to be built up from nothing. Italy's geographical situation served to its advantage, in particular for trade throughout the Mediterranean region. Even more significant from our point of view is the fact that in Sicily, Italy possessed one of the most important points of contact between the European and Arab worlds.

France was in an advantageous position as well for two reasons: First, it was the oldest independent state in the new Europe, and second, Charlemagne's reign gave impetus, politically as well as economically, to the cultural flowering known as the Carolingian Renaissance (Figures 2.10 and 2.11).

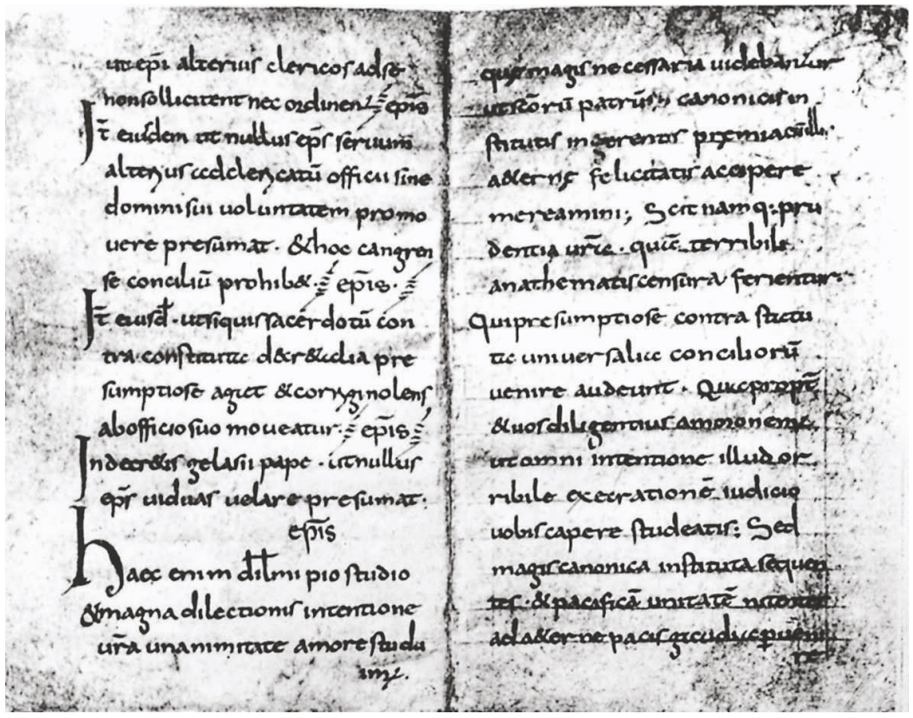
As the new European culture developed, a small country on the very edge of Europe, namely Ireland, played a surprisingly important role. Its very remoteness spared it from turmoil, and in its quiet abbeys, the works of the great writers of the ancient world were studied and preserved.

▼ **Figure 2.8** Matilda, the wife of William the Conqueror, wove this seventy-meter-long tapestry, which today can be admired in Bayeux, as a votive offering for her husband. On it are depicted episodes from his military campaigns. For us, the tapestry is also important because one of the pictures is of a comet that was interpreted as a bad omen for King Harold II. This comet has been identified by astronomers as Halley's comet.



During the decline and fall of the Roman Empire, in a time of mass migration, many found refuge in the Irish islands, which had escaped the ravages of the Huns and the Goths. However, intellectual life blossomed there only after Saint Patrick (of Anglo-Saxon origin) converted the indigenous population to Christianity. Within a short period of time, a number of monasteries were founded (Movilla in 540 and Bangor in 555, to name just the best known). It was from the monastery at Bangor that Columbanus (543–615) set out to found monasteries in Europe, which was still in a state of turmoil. The monks of the Irish monasteries were, of course, concerned primarily with religious questions. However, the effervescent intellectual life inevitably raised scientific questions as well; such discussions attracted foreigners, particularly Anglo-Saxon students. At the beginning of the seventh century, the Venerable Bede wrote,

> At this time there were many in England, both nobles and commons, who, in the days of Bishops Finan and Colman, had left their own country and retired to Ireland either for the sake of religious studies or to live a more ascetic life. In the course of time some of these devoted themselves faithfully to the monastic life, while others preferred to travel round to the cells of various teachers and apply themselves to study. The Irish welcomed them all gladly, gave them their daily food, and also provided them with books to read and with instruction, without asking for any payment.
>
> —The Venerable Bede, *Bede's Ecclesiastical History of the English People* [p. 313]



At the beginning of the ninth century, fate overtook Ireland: the Vikings attacked in 795 and after a few decades set up headquarters in Dublin. Although the Irish put up resistance, and indeed had several great victories over the invaders, the island was not abandoned by the Vikings until the beginning of the eleventh century.

During the period of invasion, the cultural tide had turned: now Irish intellectuals were fleeing into the ordered French territories.



◀ **Figure 2.9** Even at this time, intellectual life had not come to a standstill behind cloister walls. (*Saint Gregory and the Scribes*, ninth–tenth century. Vienna, Kunsthistorisches Museum.)

◀ **Figure 2.10** Two pages from a capitulary of Charlemagne (*Capitularia regum francorum*). This collection of laws exercised great influence on the laws of all the later states of Europe.



▲ **Figure 2.11** Rabanus Maurus (ca. 780–856), *Primus Praeceptor Germaniae*, Alcuin's student, archbishop of Mainz, was one of the most important scholars of the Carolingian Renaissance in the German lands. In addition to theological works, he wrote encyclopedic handbooks of secular knowledge (*De rerum naturis*). Moreover, the universally known hymn *Veni creator spiritus* is attributed to him. (From a ninth-century manuscript from Fulda.)

Although the fermenters of the Carolingian Renaissance at the court of Charlemagne were not Irish—one of the principal architects being Alcuin, an Anglo-Saxon—nonetheless, the Irish played a role here as well. Indeed, the greatest original thinker of the ninth century, Johannes Scotus Eriugena (or John the Irishman), was of Irish origin, as his name tells us twice over—at that time, "Scotus" was the Latin term for Irish or Celtic. He worked in France until late in his life when, threatened by charges of heresy, he fled to the court of Alfred the Great in England.

The degree to which Irish thinkers were respected at that time is indicated by the following lines from the *Gesta Caroli Magni* (composed around 880 in Saint Gall):

> At the moment when Charlemagne had begun to reign as sole King in the western regions of the world, two Scots from Ireland happened to visit the coast of Gaul in the company of some British traders. These men were unrivalled in their knowledge of sacred and profane letters, at a time when the pursuit of learning was almost forgotten throughout the length and breadth of Charlemagne's kingdom and the worship of the true God was at a very low ebb. They had nothing on display to sell, but every day they used to shout to the crowds who had collected together to buy things: "If anyone wants some wisdom, let him come to us and receive it: for it is wisdom which we have for sale."
>
> —Einhard and Notker the Stammerer, *Two Lives of Charlemagne* [p. 91]

To the states of Eastern Europe fell the task of holding back the tide of nomadic peoples from the East and thus the last wave of mass migration. They later became a bulwark of European culture against the expansionist Mongol state.

At the turn of the millennium, Europe appeared as a motley conglomeration of independent states that had been founded by a great variety of peoples. Despite all their differences, these states formed a strong community based on an agreed form of social organization, namely feudalism, and a single ideology: Christianity. This is most obviously attested to by the widespread use of a common language, Latin, making possible cultural exchange of a form that in a particular way is without parallel in history. For example, the University of Paris hosted in succession professors from England (Alexander of Hales), Germany (Albertus Magnus), and Italy (Thomas Aquinas). Students flocked to Paris from all over Europe to hear their lectures, which of course were held in Latin, their second mother tongue, so to speak, and in the intellectual realm, it indeed became their first.

Fighting for survival in bloody struggles against internal and external enemies, each state fashioned a form of rule according to its own situation, but which in the end necessarily were all feudal. In some places the elevation of former slaves and the degradation of landless freemen formed the most exploited laboring class of the feudal system: the serfs. In other lands the defeated population as well as the poorest strata of the conquerors became serfs. The lot of a serf was indeed better than that of a slave; this was true particularly for those bondsmen who were less strongly attached to their feudal lords. Such bondsmen were tenants, and often copyholders, on land provided by the feudal lord; as tenant farmers they owed to their feudal lord and to the Church only determined services. The difference between slavery and the feudal economy is of great significance as far as the advancement of science is concerned. Serfs were directly concerned with increasing production and thus with the development or adoption of improved methods of production. This is one of two decisive factors that led to the technical revolution of the Middle Ages. As we shall see, the other is the direct participation by the educated class, the monks, in productive work (Plate VII).

The feudal lords themselves were not completely independent, being under obligation for service, homage, and fealty (vassalage) as well as tribute to a more powerful feudal lord who had granted them their fiefdom (*feudum*), who in turn was the vassal of a yet more powerful landowner. At the top of this hierarchy stood the king or emperor. This hierarchical structure of feudal society, being someone's senior in one direction and someone else's vassal in the other, was simple only in theory. In reality, the central power often stood only nominally at the top of the pyramid, for in some of the stronger territorial principalities, independent local hierarchies frequently developed. It is therefore no wonder that coalitions across social strata were formed, primarily between the central power and the lower nobility.

Alongside this secular hierarchy there existed an ecclesiastical one with the Pope, as the earthly representative of Christ, at the top. The relationship between the two power structures became the great question of the feudal era. The vicissitudes of the long and complicated Investiture Controversy at times favored one side, as when the Pope excommunicated the emperor, or the other, as when it was the emperor who banished the Pope and appointed his successor (Figure 2.12). It is frequently said that the feudal social order was as dynamic as it was because the centers of gravity of the ecclesiastic and secular hierarchies did not coincide. This prevented any ossification; had the centers been the same, or even if there had been decisive power on one side or the other, the social system could have become static and inflexible.

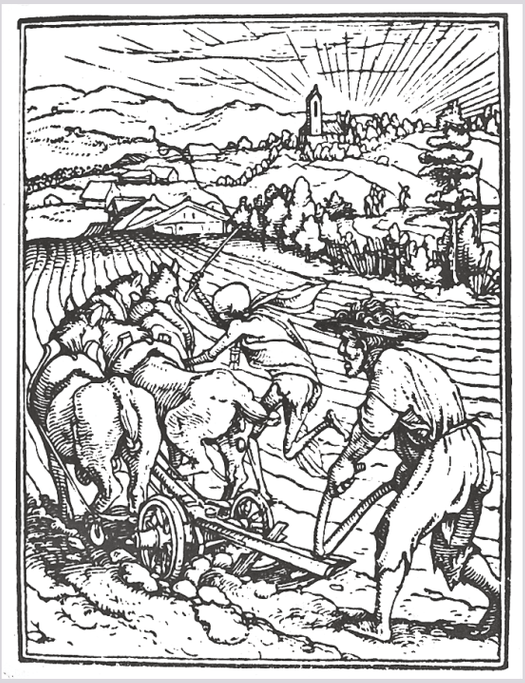
In the eleventh and twelfth centuries a new political force arose with the establishment of cities. The medieval cities—whose beauty continues to amaze us to this day—enabled a new class, the bourgeoisie, to intervene in the power struggles. In increasingly large numbers, craftsmen organized themselves into guilds, elevated their crafts to fine art, and played a major role in technological progress. The construction and decoration of cathedrals occupied large numbers of urban craftsmen and artists, and cities became artistic centers as a result. Moreover, the conditions were favorable for trade as well as for organized scientific instruction. Today, when we admire the overwhelming beauty of the Gothic cathedrals, we must not forget that these were the first great collective creations, the product of free association of free men. That work itself was held in high esteem is clearly demonstrated by the fact that in the building of the cathedrals, people from all social classes took part, even aristocrats and kings, all seeking to further the construction with their own manual labor. Under such conditions, it is understandable that in France, for example, in the period from 1180 to 1270, which is to say within a period of 90 years, 80 cathedrals and large churches were built.



▲ **Figure 2.12** The Pope and emperor at the top of the spiritual and secular hierarchies. Here, in the *Sachsenspiegel*, another collection of legal information of lasting influence (ca. 1224–1231), the two embrace harmoniously; in reality, they fought each other relentlessly for supremacy.

2.1.3 *The Technological Revolution*

Of course, to achieve this level of activity, a certain general standard of living and in turn a corresponding growth in productivity are necessary. We have already mentioned both factors: first, the initiative of the half-free serfs and the free craftsmen, encouraged by material self-interest, and second, the high regard that society placed on work. These conditions made possible a technological revolution in the Middle Ages that had long gone unrecognized by historians and therefore not properly acknowledged. When comparisons of the accomplishments of the "Dark Ages" with those of Greece are made, they are typically *in the realm of theory*. Indeed, mathematics had not reached the level of Archimedes again until only the sixteenth century. But if we consider the state of practical technology, we can see a giant advantage in favor of the Middle Ages. Namely, at the time of collapse of the



▲ **Figure 2.13** One can also follow the development of agricultural practice in works of art. (Hans Holbein, *Dance of Death*.)

▼ **Figure 2.14** A water-powered hammer used for the forming of articles made of steel. (Ruhrland- und Heimatmuseum, Essen [Döbler 1971].)



Roman Empire, except for road construction, the art of warfare, irrigation, and water supply, technology was at the same level as it had been 2000 years earlier.

We now summarize the most important technological innovations of the Middle Ages, where we rely primarily on Klemm [1961]. Many of these methods were already known in China and the Near East; in some cases, one can trace how they arrived in Europe; in others, it appears that it was a matter of an independent European development.

At first glance, the perfection of the horse harness would seem to be of little significance. Since prehistoric times, the ox was regarded as the sole draught animal, and when horses were also used for such work, it was natural simply to adapt the yokes used for oxen. However, this method did not let the horse exert its full power. With the introduction of the horse collar (see Plate VIII), the traction force of the horse was suddenly increased severalfold. This was of great significance not only for carrying burdens for trade and construction, but also for plowing, which became simpler and more efficient. Fields could now be cultivated that earlier had been inaccessible to the old "eight-ox plow."

Also in the early Middle Ages, horses began to be shod with iron, and stirrups were introduced, which revolutionized the use of horses in warfare.

For agriculture, two additional innovations were of great significance. The first, a technological breakthrough, was the invention and implementation of the wheel-supported heavy plow and interchangeable plowshare (Figure 2.13). The second, an agronomical method, was the introduction of the three-field system of agriculture, whereby among the fields to be cultivated, one-third were planted in fall, one-third in spring, and the remaining third left fallow. The result was not only a better yield; in addition, the basis for purposeful collective cultivation was established based on the division of labor.

In many areas, water and wind power began to take the place of human muscles. The water mill had been in existence since late antiquity, but it had not played a significant role. As we have mentioned, the reason for this is not only the availability of cheap slave labor, but also the paucity, in comparison to Western Europe, of suitable geographic and climatic conditions. At the end of the eleventh century there were about 5000 water mills in England. In a single département of France, Aube, there were 200 active water mills. Water power was used not only for grinding grain, but also for a wide variety of tasks, such as the fulling of cloth, the pumping of water from mines, and later, in the smelters, the operation of bellows and trip hammers (Figure 2.14). Of course, the construction of a water mill required a large investment, so it generally belonged to the feudal lord or a monastery. The feudal lord then required his tenants to make use of it—for a fee, of course—indeed, frequently ordering the destruction of all hand mills to prevent circumvention of his order. Here and there, the building of a water mill altered both the economic structure and residential patterns of the surrounding area.

Furthermore, in the eleventh and twelfth centuries, sailing boats were improved with the introduction of the stern-mounted rudder with pintle and gudgeon and the movable sail. It thereby became possible to sail against the wind, and thus to dispense entirely with oarsmen.

A further technological innovation, the compass, greatly facilitated maritime commerce and encouraged in particular the exploration of distant lands. The compass was invented in China, and here we would like to reconstruct in outline the interesting circumstances surrounding its invention. In China, as throughout the world, prophecy and the invocation of magical powers to improve one's condition

either before or after death were of great importance. In China, the following procedure became common for seeking out a favorable gravesite: the balanced spoon depicted in Figure 2.15 was spun on a tablet covered with various pictures and magic symbols. As the spoon came to rest, a "scientific analysis" of the symbols pointed at by its handle determined the place for the grave. To enhance its magical properties, the spoon was fashioned from the rarest and most valuable materials available, such as jade. But magnetite was also used. It was eventually observed that spoons made of magnetite tended to end up pointing in the north–south direction (Figure 2.16), whence the compass. The compass came into general use in Europe during the thirteenth century.

Canal building was of importance in both agriculture and transportation. The first canals equipped with locks stem from the Middle Ages.

Other machines that came into use during the Middle Ages were the foot-driven lathe and many machines used in the production of textiles, such as the foot-operated loom and spinning wheel (Figure 2.17).

The invention of eyeglasses, which came into widespread use in the twelfth and thirteenth centuries, increased the productivity of intellectuals, artists, and artisans. Now they could continue their work as they grew older, whereas earlier this would have been prevented by their worsening eyesight.

Gear-driven clocks operated by weights were developed, and eventually clocks with strike mechanisms, bringing to an end the division of the day into hours of varying lengths based on the varying length of the day from summer to winter.

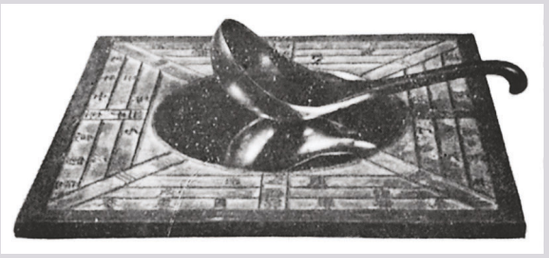
A full account of the progress made in chemical technology would lead us too far afield. We mention only the production of nitric acid, sulfuric acid, and alcohol; the development of oil-based paints; and the manufacture of paper, which laid the foundation for the printing of books.

Later, we shall go into a detailed discussion of the great inventions of the late Middle Ages, such as the development of the blast furnace, the discovery of gun powder, and the invention of the printing press, which are the first indications of the passage from feudalism to capitalism.

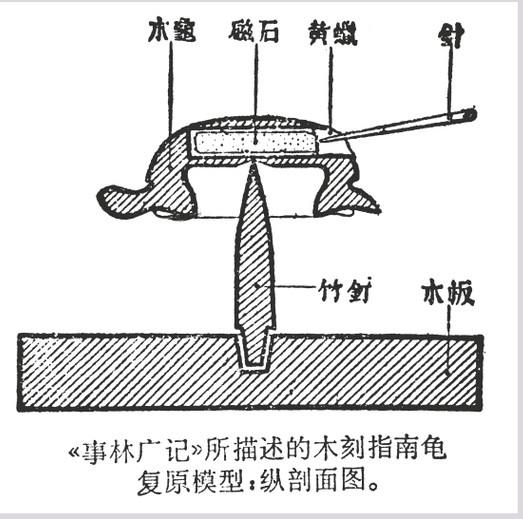
2.1.4 *Monasteries and Universities*

From the very beginning of Christianity, we meet fanatics who alone or in groups withdrew from the community to live according to their own ideas and to follow as closely as possible the commandments of their faith. The idea of living in hermitages or monasteries is not alien to other world religions either. Usually, it signified a withdrawal from worldly life, the renunciation of worldly goods, asceticism, and self-castigation, with the goal of attaining salvation in the hereafter. It is also typical not only of Christianity that monastic orders, despite their withdrawal from the world, have achieved great spiritual—and with it, political—influence.

For the history of science, and in particular for its development in Europe, it was of decisive importance that monastic orders were founded that attempted to combine a contemplative life with an active life (the *vita contemplativa* with the *vita activa*). The active aspect must be taken literally. Thus the Benedictine order, founded by Saint Benedict of Nursia (Plate VII), which can be regarded as a model for all the European monastic orders, had besides the rule that monks are to read books (that is, to develop themselves spiritually), also the requirement that they devote a specific part of each day to physical labor. These Benedictine rules are of significance for two reasons: First, they propose that those engaged in intellectual pursuits should engage



▲ **Figure 2.15** Early form of the magnetic compass: the oracular spoon.



▲ **Figure 2.16** A compass in the form of a turtle. (*Technology and Science in Ancient China.* Beijing, 1978.)



▲ **Figure 2.17** Women weaving. (Bibliothèque Nationale, Paris; after J. Nathan, *Encyclopédie de la littérature française*.)

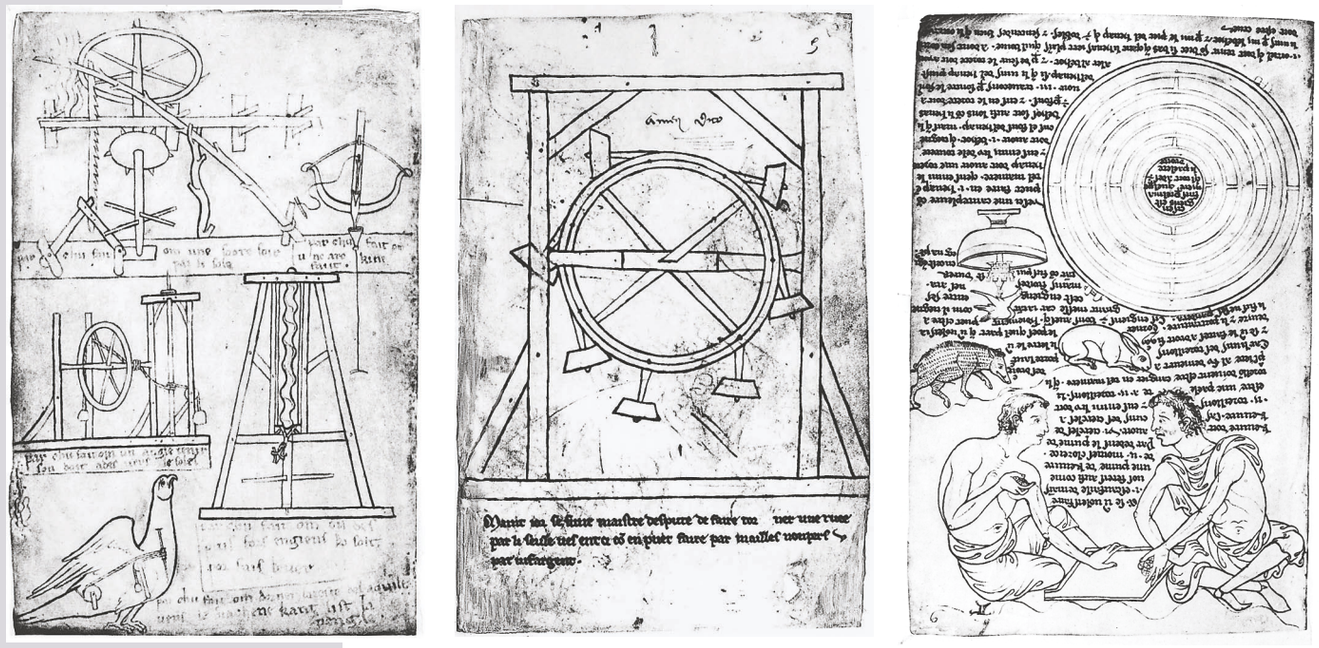


Figure 2.18
Three pages from the sketchbook of Honnecourt. Villard de Honnecourt, French architect of whose work we have knowledge from the decade 1230–1240, produced a sketchbook (*Livre de portraiture*) containing, in the part that survives, 325 ink drawings with accompanying explanations.

in physical labor as well, thereby promoting new techniques of labor and technology, and second, the participation in manual labor by the intellectual elite of feudal society led to a societal recognition of the value of such work. Thereafter, even physical labor served the "glory of God," and now, in addition to mechanical playthings, "inventions" appear that perform useful work, even the first "*perpetuum mobile*," the ostensibly perpetually moving machine (Figure 2.18).

An interesting synchronism of symbolic significance is that in the year 529, when Saint Benedict founded his monastery at Monte Cassino, Justinian ordered the closure of the philosophical schools in Athens.

Of overarching importance for the history of science in the narrower sense are two monastic orders founded at the beginning of the thirteenth century: the Franciscan and Dominican. The founder of the Franciscans, Saint Francis of Assisi, was a man of poetic spirit, with no thought of science. However, the love of nature expressed in his poetry is an important element of scientific research: an interest in nature explored with open eyes and with open heart (Figure 2.19). The rival orders of Dominicans and Franciscans both produced outstanding scholars who held professorships at the University of Paris. We mention here Albertus Magnus and Thomas Aquinas, of the Dominicans, and Robert Grosseteste and Roger Bacon, of the Franciscans.

An institution born in the Middle Ages that exists and flourishes to this day is the university (see Figures 2.20 and 2.21 and Plate IX), which still exhibits many medieval features in its structure. The significance of such an institution of learning can hardly be measured. It was something entirely new, for which there was no model from antiquity. There were indeed schools in Egypt and Mesopotamia, maintained by the state, whose significance for the conservation and transmission of knowledge should not be underestimated; but no sparkling intellectual life flourished within their walls. The schools at Athens gave some indication of how a university might be organized, but they were individual initiatives, and their subject matter was bound to particular teachers or even a single person. The *museion* of Alexandria was an intellectual center, so that it might be seen as a precursor of the university. However, an institution that includes students and faculty with their respective special privileges,



◀ **Figure 2.19** Saint Francis of Assisi. (Painting by Giotto, Louvre, Paris.)

Most high, all-powerful, all good, Lord!
All praise is yours, all glory, all honor
And all blessing.
To you alone, Most High, do they belong.
No mortal lips are worthy
To pronounce your name.

All praise be yours, my Lord, through all that you have made,
And first my lord Brother Sun,
Who brings the day; and light you give to us through him.
How beautiful is he, how radiant in all his splendor!
Of you, Most High, he bears the likeness.

All praise be yours, my Lord, through Sister Moon and Stars;
In the heavens you have made them, bright
And precious and fair.

All praise be yours, My Lord, through Brothers Wind and Air,
And fair and stormy, all the weather's moods,
By which you cherish all that you have made.

All praise be yours, my Lord, through Sister Water,
So useful, lowly, precious and pure.

All praise be yours, my Lord, through Brother Fire,
Through whom you brighten up the night.
How beautiful is he, how gay! Full of power and strength.

All praise be yours, my Lord, through Sister Earth, our mother,
Who feeds us in her sovereignty and produces
Various fruits with colored flowers and herbs.

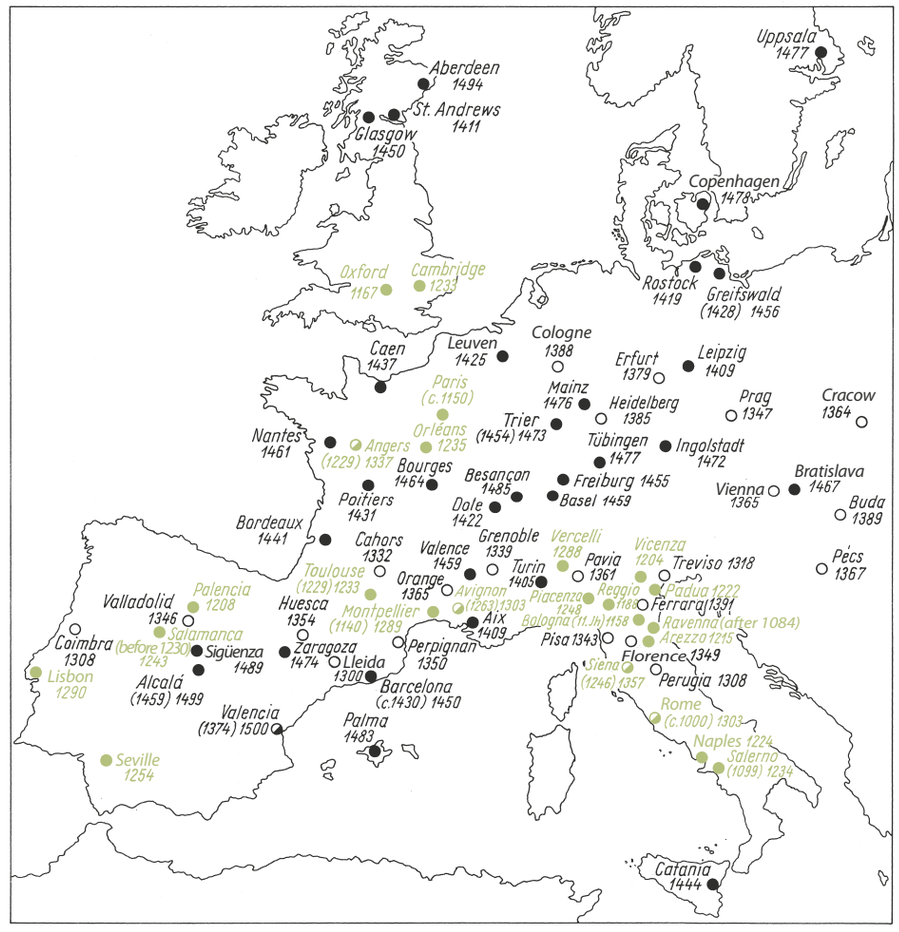
All praise be yours, my Lord, through those who grant pardon
For love of you; through those who endure
Sickness and trial.
Happy those who endure in peace,
By you, Most High, they will be crowned.

All praise be yours, my Lord, through Sister Death,
From whose embrace no mortal can escape.
Woe to those who die in mortal sin!
Happy those She finds doing your will!
The second death can do no harm to them.
Praise and bless my Lord, and give him thanks,
And serve him with great humility.

—Saint Francis, "The Song of Brother Sun"

in which knowledge is transmitted according to a particular plan of instruction, and from which after completion of a certain course of study a degree is awarded, to which certain privileges—as well as formal robes—pertain—the type of teaching extant even in our days—is a product of the Middle Ages.

As an immediate precursor of the Western European universities, one may consider the educational institutions that arose from the schools attached to Islamic mosques. In Western Europe there were developments that resembled this Arab system. Thus the University of Paris arose essentially from a school associated with the Cathedral of Notre Dame. Elsewhere, above all in Bologna, the university, noted for its teaching of law, arose from a cooperative union that had been formed by the students who streamed into Bologna from a variety of localities. The students organized themselves according to their nationalities (*nationes*). In Bologna, there were only two groups: the students from across the Alps (*universitas ultramontanorum*) and those from the Italian side of the Alps (*universitas citramontanorum*). Each university—the word itself then refers to the union of the *students*—was under the leadership of a rector chosen from and by the circle of students. It is noteworthy that, in this organization, the professors did not belong to the university and had to pledge obedience to the rector. In other places, and this was more usual, university meant the union of teachers and students (*universitas*



▶ **Figure 2.20** The expansion of the university. The red dots indicate universities founded in the fourteenth century, whereas the black dots indicate those founded thereafter. The dates in parentheses indicate the first written record of the existence of a university at the given location, and the other dates are those of the establishing documents. Some of the famous universities founded later are those at Königsberg (1544), Leyden (1574), Jena (1558), Graz (1586), Edinburgh (1593), Dublin (1591), Budapest (originally in Nagyszombat, 1635), and Moscow (1755).

The reader should not be surprised to find different dates in other publications. One source of uncertainty is this: At what point can one say that an institution of higher learning that has developed from a monastery school has in fact become a university? Another problem is how to judge continuity in the case of the restructuring of an institution or a change of location. The dates that we have chosen are taken with minor modifications from [Piltz 1981].

Although we cannot share completely in the enthusiasm expressed by J. J. Walsh in the title of his book *The Thirteenth: Greatest of Centuries*, we find his conclusions—especially in the area of education—most impressive.

> Thirteenth is the century of the Gothic cathedrals, of the foundation of the university, of the signing of the Magna Carta, and of the origin of representational government with something like constitutional guarantees throughout the west of Europe. The cathedrals represent a development in the arts that has probably never been equaled either before or since. The university was a definite creation of these generations that has lived and maintained its usefulness practically in the same form in which it was then cast for the seven centuries ever since. The foundation stones of modern liberties are to be found in the documents which for the first time declared the rights of man during this precious period. ...
>
> There are such scholars as St. Thomas of Aquinas, Albertus Magnus, Roger Bacon, St. Bonaventure, Duns Scotus, Raymond Lully, Vincent of Beauvais, and Alexander of Hales, and such patrons of learning as Robert of Sorbonne, and the founders of nearly twenty universities. There were such artists as Gaddi, Cimabue, and above all Giotto, and such literary men as the authors of the Arthur Legends and the Nibelungen, the Meistersingers, the Minnesingers, the Troubadours, and Trouvères, and above all Dante, who is universally considered now to be one of the greatest literary men of all times. ...
>
> It is, however, much more for what it accomplished for the education of the masses than for the institutions it succeeded in developing for the training of the classes, that the thirteenth century merits a place in the roll of fame. This declaration will doubtless seem utterly paradoxical to the ordinary reader of history. We are very prone to consider that it is only in our time that anything like popular education has come into existence. As a matter of fact, however, the education afforded to the people in the little towns of the Middle Ages represents an ideal of educational uplift for the masses such as has never been even distantly approached in succeeding centuries. The thirteenth century developed the greatest set of technical schools that the world has ever known.
>
> —James J. Walsh, *The Thirteenth: Greatest of Centuries,* 1912 [pp. 1, 2, 8]

magistrorum et scholarium). We can view this as a kind of interest group similar to the guilds that craftsmen had grouped in.

After a period of study of four or five years, a student could receive the title *baccalaureatus* (baccalaureate), which survives today in the degree "bachelor of arts," and after another three or four years, the title *magister*, or master of arts.

The first part of the curriculum included grammar, rhetoric, and logic, which formed the *trivium*. Logic entailed Aristotle's *Organon*, with Porphyry of Tyre's *Introduction* (*Isagoge*) to logic as prerequisite. Then came the subjects of the *quadrivium*: arithmetic, geometry, astronomy, and music (Figure 2.22). There were also elective subjects, such as Aristotelian physics.

The curriculum of the *septem artes liberales* served in general as preparation for advanced studies in the faculty of law, medicine, or theology.

Instruction comprised—in today's terminology—attendance at lectures and participation in seminars. The latter form of instruction, again in modern nomenclature, was meant to develop debating ability, to encourage independent thinking. As part of the final examinations, a candidate had to debate continuously with a relay of fresh opponents from six in the morning to six in the evening, with only an hour's break for the midday meal.

Respect for manual labor is evidenced by the fact that Hugh of Saint Victor (1096–1141) extended the seven liberal arts (*septem artes liberales*) to include the seven mechanical arts (*septem artes mecanicae*) (Figure 2.23). These comprised weaving, blacksmithing, architecture, navigation, agriculture, hunting, dramatic arts, and the healing arts.

Figure 2.24 shows how Hugo of Saint Victor's seven mechanical arts fit into the system of liberal arts described above. The figure also shows the functions that such knowledge was meant to serve. The sciences were given to mankind by God as a means to fight against the human weaknesses resulting from original sin: the theoretical sciences against ignorance, the practical sciences, including ethics, against injustice, the curriculum of the trivium against errors in speaking, and finally the mechanical arts against the weaknesses of our bodies.

Let us close with a brief tally of the developments of the Middle Ages: technological skill, more developed than that in antiquity; well-organized and developed institutions of higher learning; a social order that gave broader scope to individual initiative; a more direct connection between theory and practice. However, the theoretical sciences did not reach the level achieved in antiquity even by the late Middle Ages. Nonetheless, as we shall see, in some areas of physics, that level was surpassed.

Table 2.1 presents some important events and dates of significant persons in chronological order.

In the Middle Ages, there were not only Gothic cathedrals and universities in which the latest philosophical questions were debated, and also not only monks letting their voices resound praising God in deep devotion after they came in from the fields, but also dogmatic ossification, fanaticism, witch-burning, torture, feudal exploitation, and the Children's Crusade: this was the other side of the coin, and for many the side that primarily defined the era (Figures 2.25 and 2.26, Quotation 2.5).

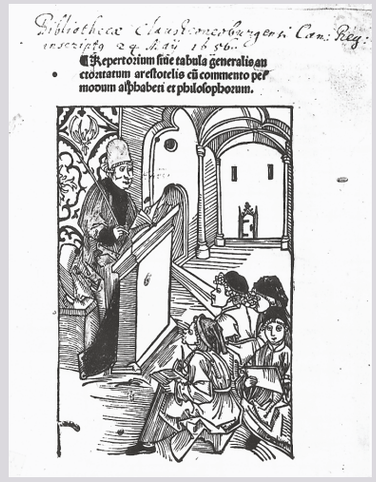
It's true that after groans and tears,
Laments of the most anguished kind,
After sadness, after pangs,
Fatigues and bitter wanderings,
Travail more opened my dull wits,
Acute until then as a ball,
Than had all of Averroes'
Comments on Aristotle done.

Or est vray qu'après plainz et pleurs
Et angoisseux gemissemens,
Après tristesses et douleurs,
Labeurs et griefz cheminemens,
Travail mes lubres sentemens,
Esguisez comme une pelote,
M'ouvrist plus que tous les Commens
D'Averroas sur Aristote.

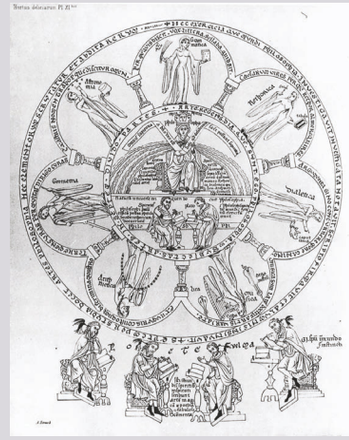
—François Villon, "Le Testament"



◀ **Figure 2.23** The seven mechanical arts (*septem artes mecanicae*) [Klemm 1961].



◀ **Figure 2.21** A university lecture. (The Venerable Bede: Incunabulum, Central Library of the Benedictine order, Pannonhalma, Hungary.)



◀ **Figure 2.22** The seven liberal arts (*septem artes liberales*). The first three images of women personify (clockwise from twelve o'clock) the subjects of the trivium (grammar, rhetoric, dialectic or logic), followed by those of the quadrivium (music, arithmetic, geometry, and astronomy). (Herrade von Landsberg: *Hortus deliciarum*.)

Quotation 2.5

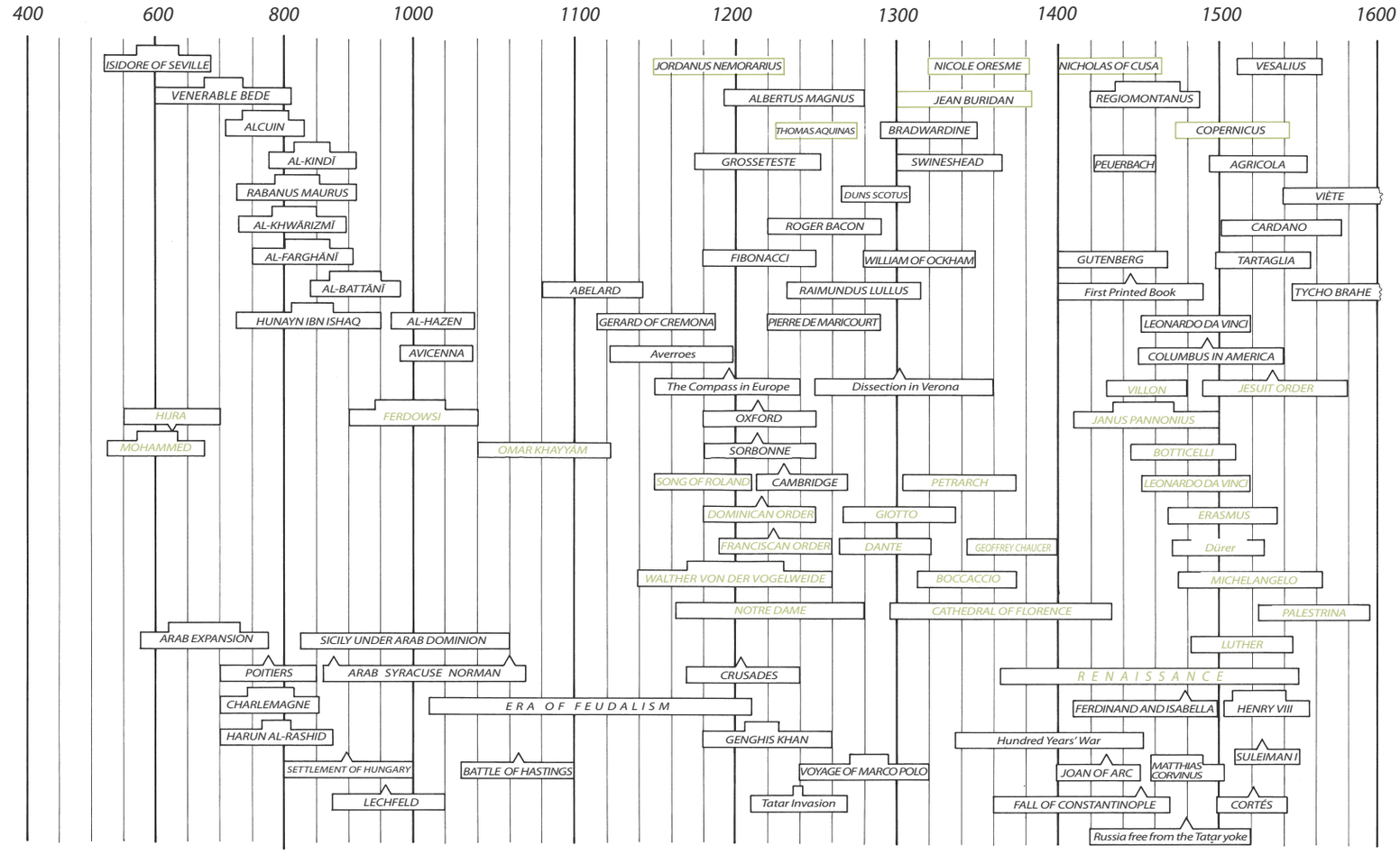
Storming of the fortress Bram in spring 1210

Going on, they came to a castrum known as Bram. Finding it defended against them, they besieged it and captured it within three days, by direct assault and without using siege-engines. They put out the eyes of the defenders, over a hundred in number, and cut off their noses. One man was spared one eye so that, as a demonstration of our contempt for our enemies, he could lead the others to Cabaret.

The Slaughter of the Heretics in Lavaur, 1211

Soon Aimeric, the former lord of Montréal, of whom we spoke above, was led out of Lavaur with up to eighty other knights. The noble Count proposed that they should all be hanged from fork-shaped gibbets. However, after Aimeric, who was taller than the others, had been hanged, the gibbets started to

continued on page 129



▲ **Table 2.1** Chronology of the Middle Ages. The famous historian of science Sarton has this to say about the era:

> It does not follow, as so many ignorant persons think, that the mediaeval activities were sterile. That would be just as foolish as to consider a pregnant woman sterile as long as the fruit of her womb was as yet unborn. The Middle Ages were pregnant with many ideas which could not be delivered until much later.
>
> —George Sarton, *Introduction to the History of Science*, 1947 [p. 15]

The names of the most important personalities of the epoch are printed in green. Those mentioned only fleetingly, if at all, in the main text include Isidore, archbishop of Seville (ca. 560–636), and the English monk known as the Venerable Bede (674–735). Both transmitted to posterity in their like-named books (*De natura rerum*) some of the legacy of ancient knowledge. Isidore relied primarily on Seneca, whereas Bede consulted Pliny. In Ireland, already in the sixth century, the trivium and even the first part of the quadrivium were being taught at the monasteries of Clonard, Bangor, and Iona. In the monasteries founded by Benedict Biscop (in which Bede lived) and in the school at the seat of the bishopric at York, the trivium and quadrivium were in the curriculum. Here Alcuin studied, one of whose pupils was Rabanus Maurus (ca. 780–856), Primus Germaniae Praeceptor, author of the book *De Universo*. This school also produced the first independent thinker of the Middle Ages, the Irishman Johannes (Scotus) Eriugena (ca. 815–877). He wrote a book with the title *De divisione naturale* and represented the point of view *Auctoritas ex vera ratione processit ratio vera nequaquam ex auctoritate* (authority derives from true reason, but true reason does not derive from authority).

However, Arab influences from the south and southwest had an even greater effect than the ferment in the northwest. One of the few scholars of actual Arab descent was al-Kindī (ca. 801–873); he wrote 270 books on all kinds of topics; he created a characteristic Arab philosophy with his interpretation of Aristotle in a neo-Platonic style; and he also clarified the notion of specific gravity and dealt with problems of optics.

Al-Farghānī(Alfraganus, ninth century): One of the first significant astronomers of the Arab–Islamic cultural circle.

Al-Battānī (Albategnius, ca. 858–929): Compiled astronomical tables of improved precision and a handbook; the form of trigonometry used until the sixteenth century can be traced to him.

Al-Farabi (Alpharabius, ca. 873–ca. 950): Theoretician of music.

Hunayn ibn Ishaq (809–877): Nestorian physician, director of the Baghdad translation bureau.

Duns Scotus (ca. 1266–1308): The "doctor subtilis." Scottish Scholastic, pupil of Roger Bacon, denied the possibility of knowledge in questions of faith. His theology has strongly pantheistic overtones.

Agricola (Georg Bauer, 1494–1555): Physician, scientist, wrote on mining and metallurgy; his 1556 book *De re matallica*, Libri XII, was an influential sourcebook in this area for a long time.

Andreas Vesalius (1514–1564): Physician at the courts of Charles V and Phillip II; his 1543 anatomical atlas *De humani corporis fabrica*, illustrated by Titian, represents a turning point in the history of medicine.

Omar Khayyám (1048–1122): Of importance as both poet (*Ruba'i*, quatrains) and scholar (calendar reform, algebra).

Geoffrey Chaucer (1340–1400): English poet and diplomat; his magnum opus is the *Canterbury Tales* (unfinished); in his 1391 book, he explained the astrolabe to an 11-year-old boy.

Poitiers: Here Charles Martel checked the Arab advance during a seven-day battle.

Table 2.1 *continued*

Hastings: William the Conqueror, duke of Normandy, defeated the Anglo-Saxon king Harold II and ascended the English throne.

Interesting synchronicities:

1240–1241: Alfonso X, the Wise, king of Castille, in 1240 (one year before the Tartar invasion of Hungary) called a meeting at which 50 Arab, Jewish, and Christian scholars took part, for the purpose of adapting the Ptolemaic system to new observations by employing new cycles. The result was the Alfonsine tables. Albertus Magnus worked on his philosophical system; Gregory IX forbade the study of Aristotle's natural-philosophical writings until they had been approved by a commission; construction of the cathedral of Notre Dame and the Naumburg Cathedral.

1273: Rudolf of Habsburg became the German king. Thomas Aquinas completed his *Summa theologica*.

We have included Villon's original French text to demonstrate the high level of the poet's educational accomplishment: he was familiar with Averroes, who wrote commentaries on most of the works of Aristotle.

2.2 The Salvage of Ancient Knowledge

2.2.1 *The Direct Path*

The exchange of envoys between Charlemagne and Harun al-Rashid—of *Arabian Nights* fame—at the beginning of the ninth century marked the birth of diplomatic relations between two militarily equal world powers. However, a comparison of the cultural achievements of the two societies shows that France was decidedly inferior. Indeed, while Charlemagne's most pressing cultural concern was how to reduce illiteracy at least among the clergy, scholars at Arab universities and learned societies were discussing the works of Aristotle, Euclid, and Ptolemy. With such a difference in cultural level, it had to happen—one might say by the laws of physics—that a flow began to take place from the higher level to the lower. For this, it is necessary that the less-developed side is receptive, which is to say that it should have the desire and ability to take in the new knowledge. The West was slowly arriving at a stage that made the appropriation, if not yet the further development, of ancient culture possible. Of course, it was not only Arab sources that were available. Figure 2.27 provides a schematic overview of the channels by which ancient works arrived in Europe. Figure 2.28 shows their geographical location and the point of contact where the transmission of knowledge was particularly intensive. The figure shows that it was not only Arab channels that were of crucial importance for such communication, but those through Byzantium as well. Additionally, direct transmission, however modest its content might have been, was still significant because by its continuous presence even in the darkest of times, it assured that the interest in knowledge—and here this means, above all, the works of classical antiquity—was never entirely extinguished.

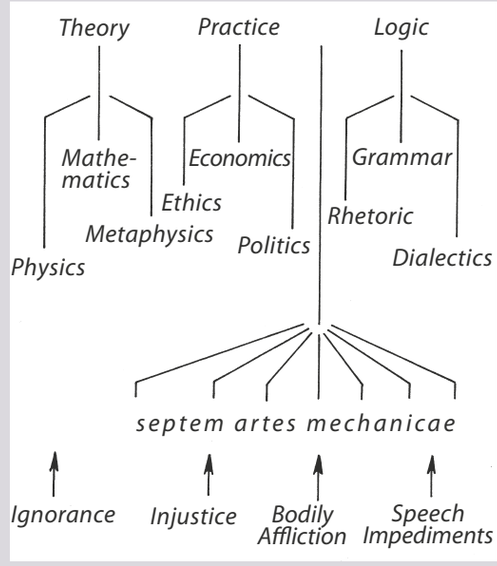
Let us follow this thread first and examine what parts of the ancient legacy the West preserved on its own.

A prominent role in the preservation of ancient works was played by Boethius, who is frequently referred to as the last of the Romans and the first of the Scholastics. Boethius (ca. 480–525) was a confidant of the Ostrogoth king Theodoric the Great, who, amid accusations of participation in a conspiracy, had Boethius incarcerated and eventually condemned to death. While awaiting execution, Boethius wrote his greatest work, *Consolation of Philosophy* (*De consolatione philosophiae*) (Quotation 2.6), which indirectly is also of importance to the history of science. Although today's reader finds in it a pagan Neo-Platonic philosophy, it was viewed by the Church as the work of a martyr to Arian persecution and treated

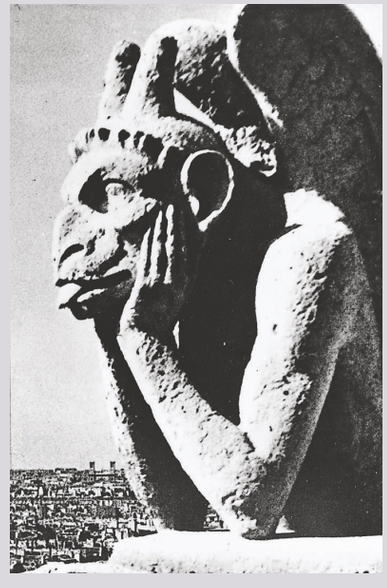
Quotation 2.5, ***continued***

fall down, since through excessive haste they had not been properly fixed in the ground. The Count realized that to continue would cause a long delay and ordered the rest to be put to the sword. The crusaders fell to this task with great enthusiasm and quickly slew them on the spot. The Count had the Dame of Lavaur, sister of Aimeric and a heretic of the worst sort, thrown into a pit and stones heaped on her. Our crusaders burnt innumerable heretics, with great rejoicing.

—Peter of les Vaux-de-Cernay, *The History of the Albigensian Crusade*



▲ **Figure 2.24** The system of knowledge and the human afflictions, resulting from original sin, against which God created the particular branches.



◀ **Figure 2.25** The fear of hell and eternal damnation dominated the imagination of medieval man more than eternal salvation, that is, the possibility of seeing God face to face: *Be sober, be vigilant; because your adversary the devil, as a roaring lion, walketh about, seeking whom he may devour* (I Peter, 5:8).

Quotation 2.6

Once, old Stoic philosophy Brought forth riddling, obscure old men: Sense perceptions and images, They believed, were impressed on minds From the outermost skin of things. A swift stylus does much the same, On a page's smooth surface calm, Imprinting letters now deep-impressed Where there were no such signs before. But if the mind with its proper strength, Proper motions, unfolds no truths— Lies still in its passivity, Just subject to the signs of things, To reproduce, as a mirror does, The empty image of worldly things— Whence this strength in the human mind? Whence this knowledge that sees all things? What force separates what is known? What force gathers the separate parts? What force chooses the twofold path, Thrusts its head in the highest realms, Goes back down to the depths below, Then returns to itself its self, Thus to contradict false with true? See a far more productive case, More far-reaching, more powerful, Than that cause which, as matter does, Accepts passively signs impressed. Nonetheless, there is passive force Which precedes, which excites and stirs Mind's own strength in the body's life, As when light batters at the eyes, Or when voices ring in the ears. Then the strength of the mind, aroused, Draws appearances kept within, Called to motions as like to like, Points them to these external signs, Thus to mingle with images Those true forms that are housed within.

—Boethius, *Consolation of Philosophy*, Book V, Meter 4 [pp. 139–141]

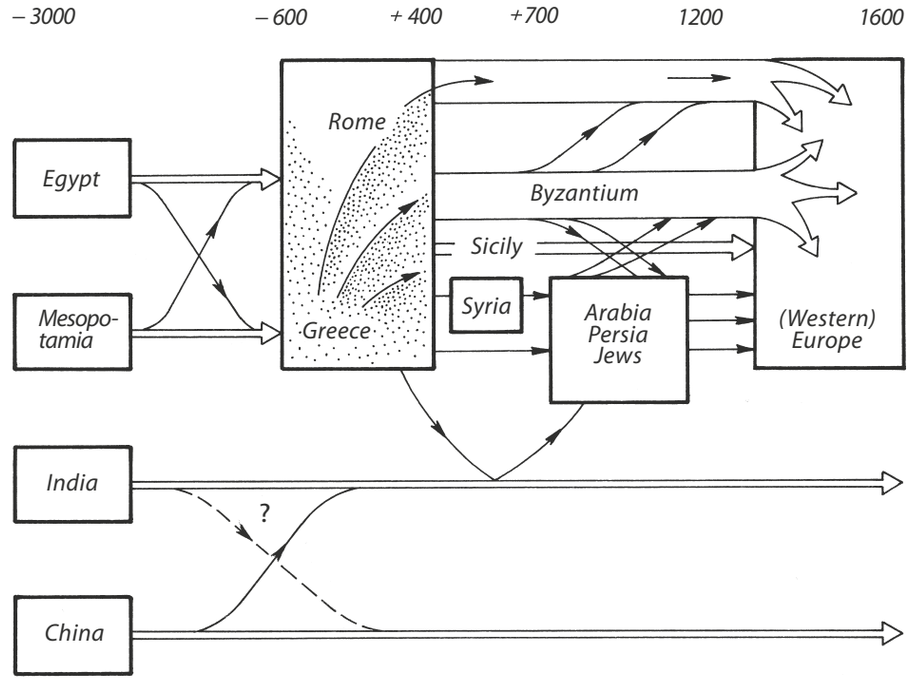


Epitaphe dudit Villon

Freres humains qui apres no⁹ viues
Nayez les cueurs contre no⁹ endurcis
Car se pitie de no⁹ pouures auez
Dieu en aura plustost de vous mercis
Vous nous voiez cy attachez cinq six
Quant de la chair q̃ trop auõs nourrie
Elle est pieca devourree et pourrie
et no⁹ les os devenõs cendres a poulsdre
De nostre mal personne ne se rie
Mais prues dieu que tous nous vueille absoulsdre

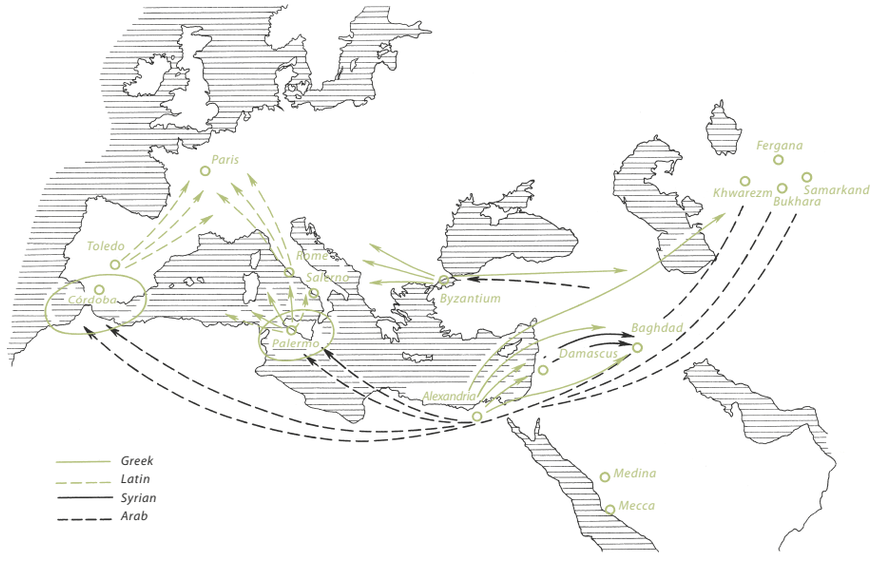
g iii.

◀ **Figure 2.26** Facsimile of a page from Villon's 1489 book of poems.



▲ **Figure 2.27** Schematic representation of the path by which the ancient cultural legacy was transmitted to Europe.

▼ **Figure 2.28** The geography of the cultural bridges from Figure 2.27.



more or less as a religious document. As a result, the other writings of Boethius, as well as his translations and commentaries on the works of ancient pagan authors, also gained acceptance. Together with Cassiodorus, a fellow minister, Boethius strived to make the works of Greek writers accessible to the Latin-speaking world. Thus, he translated the first 53 chapters of Plato's *Timaeus*, some works of Aristotle on logic, and the *Elements* of Euclid. The division of knowledge into the seven liberal arts (*septem artes liberales*) goes back to Boethius and Cassiodorus.

The works of Lucretius, Vitruvius, Seneca, and Pliny were preserved, although not in their entirety. In the seventh century, Isidore of Seville collected all the then common ancient knowledge into an encyclopedia.

Charlemagne invited the Benedictine monk Alcuin (ca. 735–804) of England to his court in order to organize the education program.

To be sure, the knowledge that was disseminated by Alcuin and his school cannot be judged by classical standards. We mention only as a curiosity that in Alcuin's works we first come across the humorous–serious logic problem known to most of us: to transport a wolf, a goat, and a head of cabbage across a river in a boat that can hold only one of the three items in such a way that the wolf is not left alone with the goat nor the goat with the cabbage.

So at the turn of the millennium, Europe was essentially ignorant of the knowledge that had been amassed by ancient Greek scholarship aside from the simplest theorems of Euclid and some fragments of the philosophical works of Plato and Aristotle. Nothing was known of Aristotle's natural philosophy, of the works of Apollonius on the conic sections, or of Ptolemy's astronomy, to say nothing of Archimedes' works on mathematics and physics.

2.2.2 *Byzantium*

As mentioned earlier, the unified Roman Empire was split in two after the death of Theodosius the Great in the year 395. The Eastern Roman, or Byzantine, Empire survived the Western Roman Empire by a thousand years, Constantinople finally falling in 1453.

The confluence of Roman and direct Greek, Hellenistic, Christian, and other Eastern influences created a unique Byzantine culture (Plate VI). The special characteristics that distinguish it from the culture of medieval Western Europe are given in the *Great Soviet Encyclopedia* as follows:

1. a higher level of productivity, at least until the twelfth century;
2. strict adherence to ancient traditions, both in science and in literature;
3. marked individualism (lack of cooperative principles, for example in the quest for salvation), which, however, was not accompanied by personal freedom;
4. a cult in which the emperor was honored as a terrestrial god;
5. homogenization of scientific and cultural life.

Under such conditions, it is understandable that in Byzantine science or philosophy we can find only preciously few works that went beyond what was known in antiquity. Among these, the commentary on Aristotle by John Philoponus (ca. 490–570) is the most significant, in which Philoponus, perhaps based on his own experiments, refutes the theses of Peripatetic dynamics on free-falling objects. In his remarks we can glimpse the germ of the idea of the laws of inertia (Quotation 2.7). He also worked on an astrolabe.

An outstanding creation of applied science is the design and construction of the basilica Hagia Sophia by Isidore of Miletus and Anthemius of Tralles.

In the technological arena, "Greek fire" is worth mentioning; this was a mixture of saltpeter, naphtha, tar, and additional ingredients, the knowledge of which was a closely held secret. Greek fire played an important role in the Byzantine technology of warfare, being used foremost against enemy ships.

In the late Middle Ages, Theodore Metochites (1270–1332) turned against astrology, the practice of which had become very popular. He also questioned the official political ideology, even though he accepted without qualification the ideas of Plato and Aristotle (Plate VI). His student Nicephorus Gregoras (ca. 1295–1360) wrote a detailed study of the astrolabe (Figures 2.29 and 2.30), which contains

Quotation 2.7

Aristotle incorrectly assumes that the times necessary for motion in various materials are in the same relationship as the densities of the materials.

However, that is completely false, and our point of view can be strengthened much more through actual observation than through verbal argumentation. Namely, let two weights fall from the same height, where one is much heavier than the other; we then observe that the relationship of the times necessary for placing in motion do not depend on the weights: The difference between the falling times is very small.

—John Philoponus, *Aristoteles physicorum libri*



▲ **Figure 2.29** The Byzantine astrolabe. The principle underlying this instrument can be traced to Apollonius of Perga, who may have been inspired from his work on stereographic projection. The realization of the instrument itself is attributed to Hipparchus. We have an extensive description of the instrument by Philoponus. By measuring the altitude of a celestial body such as the Sun, time and longitude can be determined.

Its essential components are as follows: A movable diameter (*alidade*) with lines for aiming—not visible in the figure because it is on the other side—by which height can be measured; a movable metal disk (*rete*), showing the locations of the brighter stars and the zodiac; and beneath these an additional, replaceable, disk with engraved vertical circles.

An example of time measurement: The height of the Sun is measured. After finding the location of the Sun in the zodiac given the calendar day, the *rete* is turned until this location meets a curve that corresponds to the measured height. The point determined in this way is projected with a rotating pointer onto a scale on the disk's circumference, and there the time is read off. (After the *Encyclopaedia Britannica* 1929, 1932.)



▲ **Figure 2.30** Measuring with the astrolabe. (Bibliothèque Nationale, Paris. After J. Nathan, *Encyclopédie de la littérature française*.)



▲ **Figure 2.31** Detail of a page from the work *Repetition of the Teaching* (*Mishneh Torah*) by Maimonides (Moses ben Maimon, Maimuni; 1135–1204). (Codex written in 1295. Library of the Hungarian Academy of Sciences.)

The cultural significance of medieval Judaism was heightened by the fact that the Jewish diaspora reached both Arab and Christian lands. Thus the Jews were not only patrons and supporters, but also conveyors of ancient culture between the two cultural areas. Their scientific

continued on next page

an investigation of the projection of spherical curves onto a plane. The particular emphasis placed by Gregoras on the natural sciences is novel. ("To know everything is a natural desire whose legitimacy I cannot deny.") He also predicted solar eclipses. Persian influence can already be seen in the works of Gregoras and Metochites.

However, Byzantium's greatest significance for European culture was by far its role as preserver and transmitter of the culture of antiquity. This is primarily true for literature rather than science, since the greater part of the scientific works found their own way to Europe by way of the Arab world. The Arabs, however, paid no attention to Greek culture beyond the sciences. Therefore, a very large portion of our knowledge of other realms of ancient culture is a result of the passion for collecting and love of order of the Byzantines, who took great pleasure in the compilation of anthologies and lexicons. The tradition of such collections is very old. The first hexameron—a compendium of ancient scientific discoveries and their reconciliation with holy writ—was written by Saint Basil the Great (ca. 330–379). For a long time, this work served as an encyclopedia of the natural sciences, being quoted as late as the seventeenth century.

John of Damascus (ca. 676–749) is a precursor of western Scholastics; he combined Christian theology with Aristotelianism. Perhaps the most notable work of the period is the lexicon of Photios with the title *Myriobiblon* [or *Bibliotheka*]. *Description and Listing of All Books That I Have Read and That My Dear Brother Tarasios Desired to Have a Summary. Altogether That Is Twenty-One Fewer Than Three Hundred Books.* As patriarch of Constantinople, Photios sent Cyril and Methodius to Christianize the Slavs. His famed student Michael Psellos was a polymath; among other things, he discovered some new theorems of arithmetic.

Another important encyclopedia is the *Lexikon Suda* (970), which, in addition to reliable biographies and historical facts, also contains some scientific information.

The flight of Byzantine scholars to Italy in the fifteenth century played an important role in the development of the Italian Renaissance. For example, the establishment of the Florentine *Accademia* by Cosimo de' Medici can be traced back to a proposal of George Gemistus Pletho (1355–1452). Emphasizing the significance of Plato, he helped the fight against the authority of the "Scholastic" Aristotle. He propagated the geography of Strabo, thereby furthering the applied sciences.

For the physicist, it is a surprise and a delight that among the well-known scholars of Byzantine culture one encounters the name of August Heisenberg (1869–1930), a professor of Byzantine studies in Munich and the father of the physicist Werner Heisenberg (see Plate VI).

2.2.3 *The Arab Transmission*

By the middle of the ninth century, the Arab world was already familiar with the entire scientific heritage of the Greeks, available in Arabic translation (Plate X). We must draw attention to the interesting but often overlooked fact that, as has already been mentioned, in the cultural treasure chest of the Greeks, the only thing of interest to the Arabs was the part having to do with science. Although the extraordinary creations of Greek poetry and drama were known, or could have been known to them, there is no trace of any efforts by the Arabs in this direction.

The first contact that the Arabs had with the legacy of Hellenistic culture during their campaigns aimed at world conquest was not particularly auspicious. In 641, Caliph Umar, after taking Alexandria, set fire to the still extensive book collection at the Library of Alexandria with the following words, which have become

infamous: "If the books contain what is in the Koran, then they are superfluous. If they contradict it, they are harmful. In any case, they are to be burned." It is said that for several months thereafter, the religious faithful, who cherished cleanliness, heated their baths with books from the Library of Alexandria.

The Arabs first became acquainted with the Greek legacy through Syrian translations. Syria was home to the Nestorian Christians, who pursued a vigorous scientific and missionary activity. Thus, quite early on they traveled to India and even to China, where Nestorians could be found as late as the eighteenth century. The Nestorians translated the Greek works into their Syriac language, which is closely related to Arabic. The ruling caliphs in Baghdad set up translation bureaus for the purpose of translating these works, first from Syriac into Arabic, and later directly from the Greek. In this way, the Arab world became familiar with the teachings of Hellenistic science. In addition, there was another, more direct route: when JUSTINIAN closed the philosophical school in Athens in 529, some of the philosophers moved to Gundeshapur in Persia. As a result, the Arabs found in their midst a number of Persian–Greek scientific centers that needed only to be further cultivated. Naturally, Arab science did not blossom in the East alone. For Europe, the Arab occupation of Sicily was particularly important, and of even greater importance was its reconquest by the Normans in 1085. It was in this way that the ancient works, both those with Arab commentaries and those that had arrived earlier from Byzantium to Sicily, flowed into the bloodstream of European culture.

The most important area of contact between the Arabs and the West was in Spain, where at that time there were renowned Arab universities in Toledo, Segovia, and Salamanca. These universities also attracted students from the Christian countries of Europe. After their exposure to Greek science, many of them found their life's work in making these teachings accessible to other Christians. Thus, at the beginning of the twelfth century, ADELARD OF BATH, disguised as an Arab, visited these universities and translated all of EUCLID, as well as the works of many Arab scholars, among them MUHAMMAD IBN MUSA AL-KHWARIZMI, into Latin. The greatest achievement of translation was carried out by GERARD OF CREMONA (ca. 1114–1187), who, working in Toledo, translated, among other books, AVICENNA's commentaries on ARISTOTLE from Arabic into Latin. Altogether, he translated a total of some 80 works.

See the caption to Figure 2.31 for information on the cultural significance of Jews in the Middle Ages.

2.2.4 *Return to the Source*

It soon became clear to the scholars of the Middle Ages that the multiple translation of texts that were already difficult to begin with could introduce grave errors and distortions. Consider, for example, writings that were translated from Greek into Syrian, and then through Hebrew and Arabic into Latin, where they were studied at European universities. If we consider as well that the translators' interpretations of the works were strongly influenced by the commentators of the time, then it is easy to understand, for example, that THOMAS AQUINAS came out against the teachings of AVERROES, although AQUINAS had propagated Aristotelian ideas without reservation, and AVERROES referred to ARISTOTLE in everything he wrote.

Thus, a search for original Greek texts began so that they could be translated directly into Latin. One of the locations where such texts were found was Sicily. Alternatively, original Greek manuscripts could be obtained from Byzantium either through diplomatic channels as a gift or by force. A tragicomic episode from

Figure 2.31 *continued*

endeavors were primarily related to religion. However, in the tenth, eleventh, and twelfth centuries, in the "golden age of Spanish Judaism," there were Jewish schools that taught secular science (Córdoba, Toledo). The Jewish communities generally felt more secure in the Arab lands, even though from time to time, they were persecuted there by fanatical Muslims while sometimes they were aided by enlightened Christian rulers.

The Jewish scholar with the greatest influence, MAIMONIDES, was born in Córdoba, but he worked mostly in Cairo. There is but a single contemporary, AVERROES, who could be compared with him. MAIMONIDES wrote his book *Medical Aphorisms* in Arabic, and it was soon translated into Hebrew and Latin. This book summarizes the teachings of GALEN in 1500 brief passages. It served as a foundational work in the Christian world and was often quoted with the remark, "and thus spoke RABBI MOSES." It was thus that Jewish physicians found respect in Christian royal houses.

The quotation that follows is taken from MAIMONIDES' 1190 work *Guide for the Perplexed* (*Moreh Nebuchim*). The themes in this work—the relationship between belief and knowledge—upset not only the Jewish world, but the Christian and even the Islamic worlds as well. These themes became the basis of many strenuous arguments.

> There are three opinions of human beings, namely, of all those who believe that there is an existent deity, with regard to the eternity of the world or its production in time.
>
> The first opinion, which is the opinion of all who believe in the Law of Moses our Master, peace be on him, is that the world as a whole—I mean to say, every existent other than God, may He be exalted—was brought into existence by God after having been purely and absolutely nonexistent, and that God, may He be exalted, had existed alone, and nothing else—neither an angel nor a sphere nor what subsists within the sphere. Afterwards, through His will and His volition, He brought into existence out of nothing all the beings as they are, time itself being one of the created things. For time is consequent upon motion, and motion is an accident in what is moved. Furthermore, what is moved—that is, that upon the motion of which time is consequent—is itself created in time and came to be after not having been. Accordingly one's saying: God "was" before He created the world—where the word "was" is indicative of time—and similarly all the thoughts that are carried along in the mind regarding the infinite duration of His existence before the creation of the world, are all of them due to a supposition regarding time or to an imagining of time and not due to the true reality of time. … Consider this matter thoroughly. For thus you will not be necessarily attached to objections from which there is no escape for him who does not know it. For if you affirm as true the existence of time prior to the world, you are necessarily bound to believe in the eternity of the world. For time is an accident which necessarily must have a substratum. Accordingly it follows necessarily that there existed some thing prior to the existence of this world existing now. …
>
> The second opinion is that of all the philosophers of whom we have heard reports and whose discourses we

continued on next page

Figure 2.31 *continued*

have seen. They say that it is absurd that God would bring a thing into existence out of nothing. Furthermore, according to them, it is likewise not possible that a thing should pass away into nothing; I mean to say that it is not possible that a certain being, endowed with matter and form, should be generated out of the absolute nonexistence of that matter, or that it should pass away into the absolute nonexistence of that matter. To predicate of God that He is able to do this is, according to them, like predicating of Him that He is able to bring together two contraries in one instant of time, or that he is able to create something that is like Himself, may He be exalted, or to make Himself corporeal, or to create a square whose diagonal is equal to its side, and similar impossibilities. … Hence they believe that there exists a certain matter that is eternal as the deity is eternal; and that He does not exist without it, nor does it exist without Him. They do not believe that it has the same rank in what exists as He, may He be exalted, but that He is the cause of its existence; and that it has the same relation toward Him as, for instance, clay has toward a potter or iron toward a smith; and that He creates in it whatever He wishes. Thus He sometimes forms out of it a heaven and an earth, and sometimes He forms out of it something else. The people holding this opinion believe that the heaven too is subject to generation and passing-away, but that it is not generated out of nothing and does not pass away into nothing. For it is generated and passes away just as the individuals that are animals are generated from existent matter and pass away into existent matter. The generation and passing-away of the heaven is thus similar to that of all the other existents that are below it. … This is also the belief of Plato. …

The third opinion is that of Aristotle, his followers, and the commentators of his books. He asserts what also is asserted by the people belonging to the sect that has just been mentioned, namely, that something endowed with matter can by no means be brought into existence out of that which has no matter. He goes beyond this by saying that the heaven is in no way subject to generation and passing-away. His opinion on this point may be summed up as follows. He thinks that this being as a whole, such as it is, has never ceased to be and will never do so; that the permanent thing not subject to generation and passing-away, namely, the heaven, likewise does not cease to be; that time and motion are perpetual and everlasting and not subject to generation and passing-away; and also that the thing subject to generation and passing-away, namely, that which is beneath the sphere of the moon, does not cease to be. I mean to say that its first matter is not subject in its essence to generation and passing-away, but that various forms succeed each other in it in such a way that it divests itself of one form and assumes another. He thinks furthermore that this whole higher and lower order cannot be corrupted and abolished, that no innovation can take place in it that is not according to its nature, and that no occurrence that deviates from what is analogous to it can happen in it in any way.…

continued on next page

the Crusades is the taking of Constantinople in 1204. According to chronicles of the event, shiploads of Greek manuscripts were subsequently brought to the West.

Thomas Aquinas did not know Greek. Perhaps it was for that very reason that he sought a reliable translation of Aristotle. At his instigation, William of Moerbeke (ca. 1215–1286) translated all the works of Aristotle that had not been translated before or of which the translation was poor, together with some works of Archimedes (Figure 2.32). Thus, in the second half of the thirteenth century, Europe came to know, in translations directly from the original Greek, the entire life's work of Aristotle. The translation of the works of Ptolemy took place in the second half of the twelfth century, almost at the same time directly from the Greek and indirectly from the Arabic. It was left to the scholars of the Renaissance to translate the complete available works of Archimedes, a task for which the Middle Ages, it must be said, was not yet ready. Thus in the thirteenth century, the West had only a few minor works of Archimedes, partly translated from the Greek and partly from the Arabic (Figure 2.33).

Table 2.2 shows the times at which particularly significant ancient Greek works became available in Western Europe as well as the language from which they were translated into Latin, and, when known, where the translation was carried out.

▼ **Table 2.2** It is thus that Europe came to know ancient science, in part from Arabic sources, in part from the original Greek texts [Crombie 1952]. [Below is the content of the table]

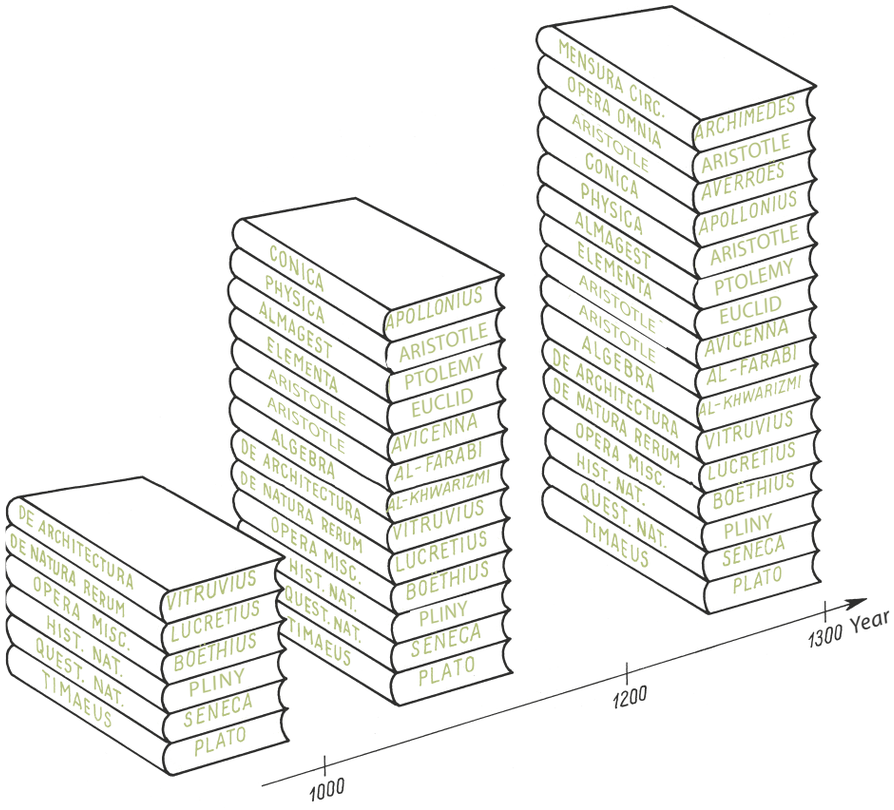
Translated directly into Latin			
Plato	*Timaeus* (chapters 1–53)	Greek → Latin	4th century
Aristotle	some works on logic	Greek → Latin	6th century
Lucretius	*De rerum natura*		
Vitruvius	*De architectura*		
Seneca	*Questiones naturales*		
Pliny	*Historia naturalis*		
Boethius	mathematical and astrological works, Aristotelian commentaries		6th century
Cassiodorus	*Septem artes liberales*		6th century
Bede	*De natura rerum, De temporum ratione*		7th century

Works of Arabic authors			
Al-Khwārizmī	Liber Alchorismi	Arabic → Latin	
	astronomical tables	Adelard of Bath	1126
	Algebra	Robert of Chester	1145, Segovia
Al-Farabi	Distinctio super	Arabic → Latin	12th century
	Librum Aristoteles	Gerhard of Cremona	Toledo
Avicenna	Aristotelian commentaries	Arabic → Latin	Toledo, 12th century
Averroes	Commentaries on the works of Aristotle: *Physica* and *De caelo et mundo*	Arabic → Latin Michael Scot	beginning of the 13th century

continued on next page

Table 2.2 *continued*

Translations of works by Greek authors from the Arabic or Greek			
Aristotle	Meteorologica, Physica, De generatione et corruptione, Metaphysica	Greek → Latin	12th century
Aristotle	essentially the complete works	Greek → Latin William of Moerbeke	1260–1270
Euclid	Elementa	Arabic → Latin Adelard of Bath	1126
Apollonius	Conica	Arabic → Latin	12th century
Archimedes	De mensura circuli, De iis quae in humido vehuntur	Arabic → Latin Greek → Latin William of Moerbeke	12th century 1269
Ptolemy	Almagest	Greek → Latin Arabic → Latin	1160, Sicily 1175, Toledo



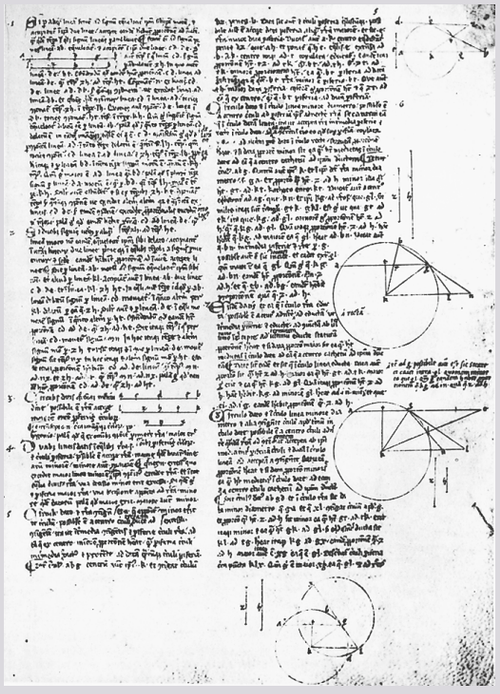
▲ **Figure 2.33** The increasing number of ancient works accessible in Latin as a function of time. For a complete translation of the works of Archimedes, one still had to wait two more centuries.

The books are pictured in their present form. Indeed, since the beginning of the Common Era, papyrus and parchment rolls were being supplanted gradually by pages stitched together, and by the sixth century, they had almost completely disappeared.

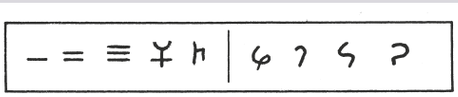
The medieval universities had a great influence on the form of handwritten books. Students regularly took notes, which were then expanded and copied, resulting in a form of book that was cheaper to produce and easier to use, in readable handwriting written with a goose quill (instead of a reed), with page numbers, table of contents, and so on. Books attained their present form only in the fifteenth century (see also Plate XI).

Figure 2.31 *continued*

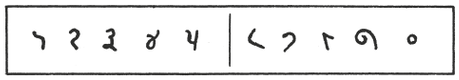
> What I myself desire to make clear is that the world's being created in time, according to the opinion of our Law—an opinion that I have already explained—is not impossible and that all those philosophic proofs from which it seems that the matter is different from what we have stated, all those arguments have a certain point through which they may be invalidated and the inference drawn from them against us shown to be incorrect. Now inasmuch as this is true in my opinion and inasmuch as this question—I mean to say that of the eternity of the world or its creation in time—becomes an open question, it should in my opinion be accepted without proof because of prophesy, which explains things to which it is not in the power of speculation to accede. …
>
> After I have made it clear that what we maintain is possible, I shall begin to make it prevail likewise, by means of speculative proof, over any other affirmations; I refer to my making prevail the assertion of creation in time over the assertion of eternity.
>
> —Maimonides, *The Guide of the Perplexed* [pp. 281–285, 293–294]



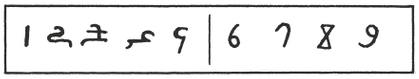
▲ **Figure 2.32** The translation of the works of Archimedes by William of Moerbeke.



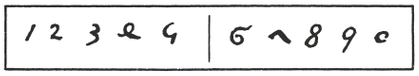
Brahmi



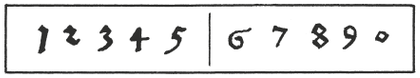
Indic



Arabic

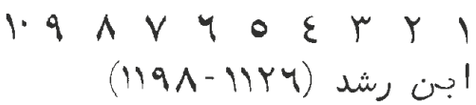


European (15th Century)

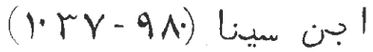


Dűrer

(a)



Averroes (1126 – 1198)



Avicenna (980 – 1037)

(b)

▲ **Figure 2.34**
(a) The development of today's numerals [van der Waerden 1954].
(b) Numerals used in the modern Arabic language (read from right to left) with two examples.

Quotation 2.8

The Greeks, though impure, must be honored, since they were trained in sciences, and therein excelled others. What, then, are we to say of a Brahman, if he combines with his purity the height of science?

—Varahamihira, astronomer (around 500)

2.3 The Indian and Arab Worlds

2.3.1 *The Decimal System*

As interesting as it might be to investigate the development of philosophy and the natural sciences in non-European cultures, many of which show a sharp contrast to the Greek way of thinking, we shall have to content ourselves with a brief summary of Hindu and Arabic physics and mathematics.

India influenced Western science only in the area of mathematics, in contrast to the Arabs, who were well ahead of Europe in both mathematics and physics although their greatest significance to Europe was still as transmitters of the inheritance from antiquity (Plate X).

The golden age of Indian mathematics occurred in the period between 200 and 1200 CE, following the seminal influence of the Greek connection. Before that time, the main concern of Hindu mathematics was in determining the measurements of altars for sacrifices, with the result that the Indians arrived at the formula for the area of a circle quite early, although with a rather poor approximation for the number π (3.09). From the beginning, the Indians showed a predilection for large numbers. Indian literature mentions a host of physical and intellectual competitions in which rival parties sought to outdo each other in the construction of ever larger numbers. Indian mathematicians acknowledged their debt to the Greeks (Quotation 2.8).

The unique contribution of Indian mathematics is the introduction of decimal positional notation and the digit zero. In about the third century, the Indians began to use the so-called Brahmi notation (Figure 2.34). Each numeral up to nine was given its own symbol. For tens, hundreds, and so on, the same symbols were used, with the empty places filled by zeros.

This place-value, or positional, system by itself is, as we have seen, not an independent Indian invention. It probably arrived from Alexandria, but even Babylonian mathematics made use of a positional system, as described earlier. However, the Indians were the first to introduce the digits that we use today, that is, our decimal system. Furthermore, they did not use zero merely to mark a missing place in a number, but they considered it a number in its own right and specified rules for calculating with it. Thus it was determined by Brahmagupta (598–668) that multiplication of an arbitrary number by zero yields zero, and that starting with a given number, addition or subtraction of zero leaves the number unchanged. He even interpreted division by zero with the sort of "rigor" typical of engineering practice; namely, he speculated that the result must be a number so large—today we would say infinite—that adding or subtracting any other number from it would leave its magnitude unchanged. This corresponds to the contemporary interpretation in engineering practice of division by zero if we do not insist on full mathematical rigor.

The Indians also understood the concept of negative numbers. Considering the positive numbers as riches owned, they then regarded the negative numbers as debts, and by applying this analogy, they tried to work out the rules of calculating with them. They correctly arrived at the rules for multiplication, whereby the product of a positive and a negative number is negative, and the product of two negative numbers is positive. They also determined that the square root of a positive number can be either positive or negative. They worked with irrational numbers in the same way as with the natural numbers.

The Indians, and after them the Arabs, owed their progress in the realm of algebra, in which they vastly surpassed the Greeks, to the fact that they were not

fastidious about logical underpinnings and rigorous proof, so that they were not hampered in their formal calculations by the lack of a solid foundation. Earlier, we discussed how the Greeks' backwardness in the domain of algebra was a direct consequence of their overly demanding approach.

2.3.2 *Algebra and Algorithm*

The particular contributions of the Arabs to the development of algebra are easily seen in European languages. Most scientific terms are taken from Greek and less frequently from Latin, and most legal terms are of Latin origin. However, Arab hegemony in algebra can be seen in the fact that the word *cipher* comes from the Arabic *ṣifr*, meaning "empty" and designating the zero.

We should note here that it was not only Arabs who contributed to Arab science, but other people living in Arab lands did so as well. We have already mentioned that after the occupation of Gundeshapur, the Arabs encountered Greek scholars who had fled Athens. Arab science was further enriched as well by Persians, Syrians, Jews, as well as various Central Asian scholars. See Plate X.

The word "algebra" is derived from the title of the book *Hisab al-jabr wa al-muqabala* (which might loosely be translated as *compendium on correct arrangement and balancing*), by al-Khwarizmi. To clarify the terminology used in the book's title, let us consider the equation

$$5x^2 - 6x + 2 = 4x^2 + 7.$$

The word *al-jabr*, from which "algebra" is derived, means reuniting, which is accomplished by bringing the negative term to the other side of the equation:

$$5x^2 + 2 = 4x^2 + 6x + 7.$$

The word *muqābala* means "compensation," indicating the elimination of terms on both sides of the equation:

$$x^2 = 6x + 5.$$

These rules are still used today in solving equations. One of the original meanings of *al-jabr* was the setting of broken bones. In this sense, it was taken into Spanish, where an algebrista is either an *algebraist* or a bonesetter. Thus when Cervantes' Don Quixote unhorses the "Knight of the Mirrors," the latter is handed over to an *algebrista*, who sets his broken bones.

Mohammed Ibn Musa Al-Khwarizmi's book appeared in Latin with the title *Ludus algebrae almucgrabalaeque*. Al-Khwarizmi relied primarily on the writings of Brahmagupta, but he clearly was also familiar with the achievements of the Alexandrian Greeks. It was from al-Khwarizmi's book that Europe learned the Hindu numerals and positional system, which explains why they are called "Arabic numerals" (when they rightly should be called "Indian"), to differentiate them from the Roman numerals, which had long been used and continued to be used alongside the Arabic. Scholars in the Middle Ages were fully aware of the Indian origin of these numerals, as can be seen from the Latin title of a work, *Algoritmi de numeris indorum* (Al-Khwarizmi, on the Indian numerals).

The word "algorithm," which today means a computational procedure, comes directly from the name al-Khwarizmi, which indicates that the Arab scholar Mohammed Ibn Musa came from Khwarezm (Khorasan).

Quotation 2.9

Although I have a handsome face and color,
Cheek like the tulips, form like the cypress,
It is not clear why the Eternal Painter
Thus tricked me out for the dusty show-booth
of earth.

He began my creation with constraint,
By giving me life he added only confusion;
We depart reluctantly still not knowing
The aim of birth, existence, departure.

The cycle which includes our coming and going
Has no discernible beginning nor end;
Nobody has got this matter straight –
Where we come from and where we go to.

When we were children we went to the Master
for a time,
For a time we were beguiled with our own mastery;
Hear the end of the matter, what befell us:
We came like water and went like wind.

—Omar Khayyám, *Ruba'iyat*

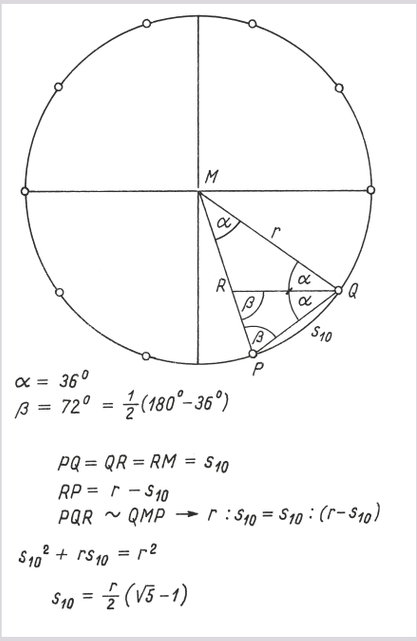


Figure 2.35 The relationships in a regular decagon.

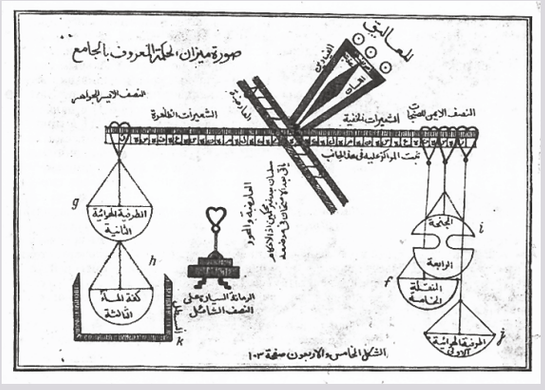


Figure 2.36 Illustration of a balance from AL-KHAZINI's *Book of the Balance of Wisdom*, 1121.

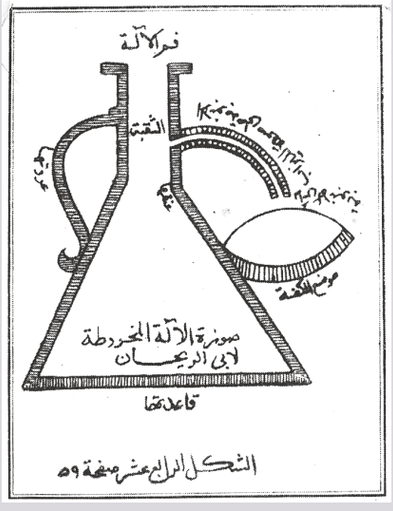


Figure 2.37 The apparatus of AL-BĪRŪNĪ for measuring specific gravity in AL-KHAZINI's book [Dorfman 1974].

The Arabs who had the greatest influence on the Western world in the Middle Ages were AVICENNA (IBN SINA, ca. 980–1037) and AVERROES (IBN-RUSHD, 1126–1198). AVICENNA remained the preeminent authority in medicine for centuries. His philosophy was Aristotelian with a touch of Platonism. His philosophical poems speak to our time as much as those of the poet and scholar OMAR KHAYYÁM, who was two generations younger (Quotation 2.9). AVERROES was referred to in the Middle Ages as "The Commentator." He is even mentioned as such in DANTE's *Divine Comedy*. For him, ARISTOTLE was the philosopher par excellence (Quotation 2.10). We have already spoken of his unique interpretations of ARISTOTLE, which moved THOMAS AQUINAS to take the field against the antireligious views of AVERROES' supporters. Namely, AVERROES had interpreted the Aristotelian natural laws in such a way that the world is predetermined in every detail and was not created; thus, any role for God or Providence in the creation of the cosmos was ruled out.

2.3.3 *Some Outstanding Contributions of Arab Science*

Arab mathematics and physics reached their developmental peak in the fifteenth century. At that time, Europe was surpassed in both disciplines in a number of details. Here we summarize some of the principal achievements of Arab science.

In 1459, the Persian mathematician JAMSHID AL-KASHI calculated the number π accurately to 17 decimal places.

The binomial theorem was known (for positive numbers n) in the form

$$(a+b)^n - a^n = C_{n,1}a^{n-1}b + C_{n,2}a^{n-2}b^2 + \cdots + C_{n,n-1}ab^{n-1} + b^n.$$

The following relationship was known for the binomial coefficients appearing in the above equation:

$$C_{n,m} = C_{n-1,m} + C_{n-1,m-1}.$$

This relationship was used later in Europe to construct Pascal's triangle.

Formulas for the summation of series of natural numbers with various exponents were determined, that is, sums of the form

$$\sum_{a=1}^{n} a^k, \quad k = 1,2,3,4,\ldots.$$

Thus, for example, the sum of the fourth powers of the first n natural numbers was calculated by the formula

$$\sum_{a=1}^{n} a^4 = 1 + 2^4 + 3^4 + \cdots + n^4 = \sum_{a=1}^{n} a^2 \left[\sum_{a=1}^{n} a + \frac{\left(\sum_{a=1}^{n} a\right) - 1}{5} \right].$$

The mathematician AL-KASHI, mentioned above, created a table of sines with a step size of 1′ and a precision of nine decimal places. To create the table, he first calculated the value sin 3° from sin 72° and sin 60° by using the relations

$$\sin(72^\circ - 60^\circ) = \sin 12^\circ = \sin 72^\circ \cdot \cos 60^\circ - \cos 72^\circ \cdot \sin 60^\circ$$

and

$$\sin 3^\circ = \sin(15^\circ - 12^\circ).$$

To determine sin 1°, he used the identity

$$\cos\phi = 4\cos^3\frac{\phi}{3} - 3\cos\frac{\phi}{3}$$

to derive the third-degree equation

$$x^3 + 0.7850393433644006 = 45x,$$

which he solved to an accuracy of 17 places.

With this method, he was able to produce tables with accuracy up to the ninth decimal place. He solved the third-degree equation with a remarkable iteration procedure.

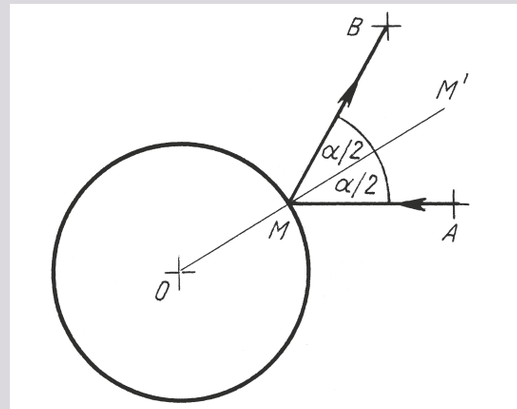
To determine sin 72°, he used the fact that the length of a side of a regular decagon (in which all the central angles equal 36° and all angles between two edges measure 2 × 72° = 144°) inscribed in a circle of radius *r* is equal to

$$\frac{r}{2}\left(\sqrt{5}-1\right)$$

(Figure 2.35).

In physics, the Arabs surpassed the contributions of the Greeks in the field of optics and in the working out of methods for measuring specific gravity (Figure 2.36 and 2.37). In geometry, ALHAZEN (IBN AL HAYTHAM, 965–ca. 1039) was considered to be the greatest authority in Europe for several centuries. He called PTOLEMY's law of refraction into question based on his observations that when light is refracted, the angles of incidence and refraction are not proportional. However, he was unable to derive the correct law. He is the source of theories of parabolic and spherical mirrors, the principle of the camera obscura, and a description of how the eye functions. Moreover, the solution of the problem depicted in Figure 2.38, which at first glance appears to be a geometric one, is generally seen as ALHAZEN's greatest achievement. The task is this: Given two arbitrary points *A* and *B* in a plane external to a given circle, to find the point *M* on the circle's circumference such that the line *OMM'* bisects the angle α formed by the lines *MB* and *MA*. This is in fact closely related to the problem in optics of finding the point on a spherical mirror at which a light ray emanating from a given point *A* must be reflected in order to arrive at a given point *B*. ALHAZEN reduced the problem to a fourth-degree polynomial equation, which he was able to solve.

The Arabs translated PTOLEMY's astronomical works as a matter of course, considering it the *ne plus ultra*, making no effort, at least not with respect to theory, to build upon his results. However, for the astrological work that was so important to them, they required precise astronomical data. Therefore, in the first half of the fifteenth century, they had an observatory built by the Timurid ruler ULUGH BEG. Here AL-KASHI was involved as well. He prepared the aforementioned tables of sines for ULUGH BEG. The largest apparatus belonging to the observatory is an object of wonder to this day: a giant sextant made of marble with a radius of 40 meters covering an arc of 60° (Figure 2.39).



▲ **Figure 2.38** On the problem of ALHAZEN. A ray of light emitted from point *A* should, after reflection in a spherical mirror, arrive at point *B*. According to the law of reflection, the angles *AMM'* and *M'MB* should be equal. (LEONARDO also had a solution to this problem.)

Quotation 2.10

Nature created him [ARISTOTLE] to place before our eyes an example of human perfection. Providence has given him to us so that we might know what can be known. His teaching is the highest truth; his reason, the highest form of human powers of comprehension that can exist.

—AVERROES

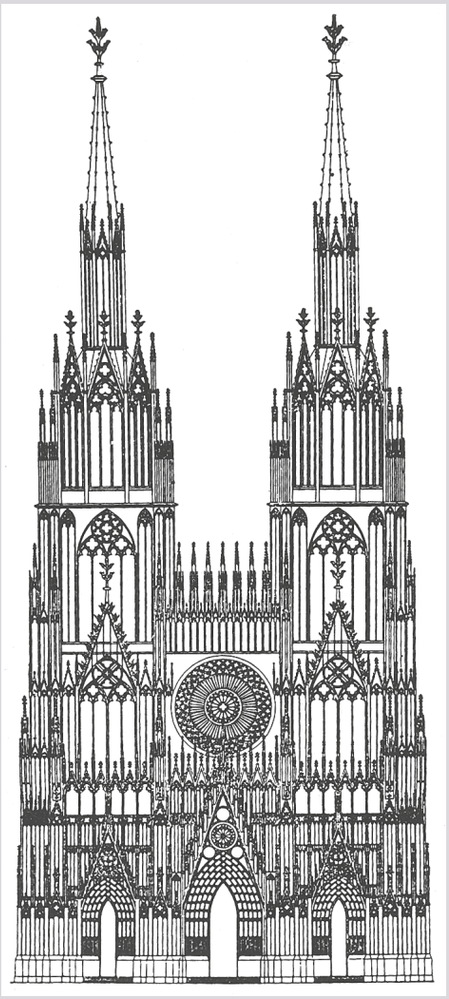


(a)

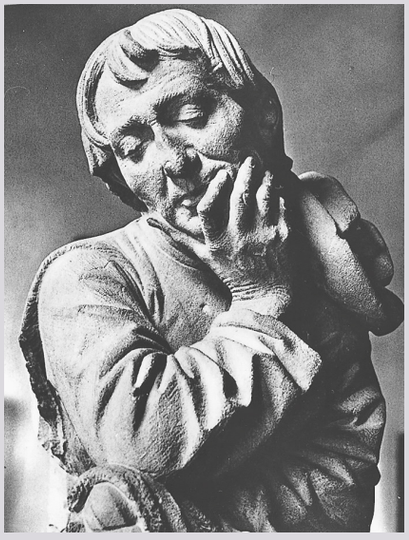


(b)

◀ **Figure 2.39**
(a) The sextant of the observatory built by ULUGH BEG in Samarkand.
(b) ULUGH BEG (portrait from the 1690 book *Prodromus Astronomiae* by HEVELIUS).



▲ **Figure 2.40** The Gothic cathedrals are masterpieces both structurally and aesthetically. The figure shows a plan of the Strasbourg Cathedral by an unknown master (not built).



▶ **Figure 2.41** A builder of a cathedral. (Statue by Nicolaus Gerhaert, probably representing himself; Strasbourg, Musée de l'Oeuvre Notre Dame [Swaan 1969].)

2.4 The West Awakens

The Renaissance broke with the Middle Ages by seeking its ideals of art and of man in antiquity. The intellectual revolutionaries of the seventeenth century had to battle against Aristotelianism, which was under the protection of the Church, as well as against medieval Scholasticism. So it is no wonder that from the sixteenth century on, all medieval philosophies and institutions were seen as obstacles to human progress. Thus, both the arts and sciences of the period were given short shrift.

The Gothic cathedrals, transcending preconceived notions through their imposing size and daring construction, passed on to later generations, consciously or incidentally, an impression of the enormous technical and intellectual accomplishments of the Middle Ages (Figures 2.40 and 2.41). However, the scientific achievements of this period had to be extricated from the large quantity of medieval manuscripts, much in the same manner in which an archaeologist delicately digs up a long-buried city from under deep layers of sand. This work was begun in earnest only at the end of the nineteenth century, with enormous credit due the French scholar Pierre Duhem, whose 1905–1906 book *Les Origines de la Statique* was the first to focus the attention of the general public on the progress made by medieval scholars beyond what had been achieved in antiquity. His work was continued and expanded by Annelise Maier in her book *Die Vorläufer Galileis im 14. Jahrhundert* (Galileo's fourteenth-century precursors). Since then, many scholarly groups have formed throughout the world to research the science of the Middle Ages. Among these we should particularly mention the group at the University of Wisconsin under the direction of Marshall Clagett. We have learned from their work that we can treasure the Middle Ages not only as a particularly rich era in the history of art, but also with regard to the natural sciences. Thus the term "Dark Ages" is completely unjustified.

We have already mentioned that Europe had essentially acquired its modern form by the end of the first millennium. In the eleventh century, Europe was only learning. In the twelfth and thirteenth centuries, it adapted its received knowledge to its own worldview, and in the thirteenth and fourteenth centuries, this knowledge was subjected to critical analysis and was transcended.

Of course, all these remarks apply only to the front lines of scientific research, or, more precisely, to those scholars working at the frontiers of knowledge. The feudal system, with its highly developed and ideologically uniform superstructure, was much too rigid to adapt itself to change. The knowledge achieved by the end of the Middle Ages revealed the first fissures in that unified ideological superstructure. Had the Renaissance thinkers taken notice of this, they could have made use of the breech to tear down the entire ideological edifice (see Section 2.6).

2.4.1 *Fibonacci: The Artist of Computation*

The first significant advance in medieval mathematics is due to Leonardo da Pisa (1170–1250). In the mathematical world, he is known by the patronymic Fibonacci (son of Bonaccio); the Fibonacci numbers, which remain alive in contemporary mathematics, are named for him. His most significant work is the 1202 *Liber Abaci*. It is from this book that Europe learned of Arabic numerals and how to use them. He correctly called them Indian numbers (*figurae indorum*). It is also of interest that while the title indicates that the book will discuss calculations with the abacus, the book, in fact, is a propaganda tract in support of calculating with Arabic numerals.

Indeed, the abacus (Figure 2.42(a)) was at that time—above all as a consequence of the work of Pope Silvester II (Gerbert d'Aurillac, ca. 946–1003)—the most widely used device for carrying out calculations (Figure 2.42(b)). In spite of their practical advantages, it took a long time for Arabic numerals to catch on and gain the upper hand. Even shopkeepers, whose work would have been greatly simplified by the use of Arabic numerals in place of the clumsy Roman numerals and their associated rules of computation, were at first opposed to the new notation. In fact, in 1299, the use of Arabic numerals was forbidden in Florence. The reason cited for this prohibition was simple: It is much easier to falsify business ledgers if these are kept using Arabic numerals because it is easy to insert a zero anywhere, thereby increasing a value by a factor of 10. Such a scale of deception could not be carried out with Roman numerals.

Fibonacci was associated with Frederick III, the famous—one might even say notorious—king of Sicily, who gathered Arab, Jewish, and Christian scholars together at his court. Frederick himself was deeply involved in scientific pursuits and was considered to be a freethinker. It is said that he spoke of Moses, Jesus, and Mohammed only as the "*tres impostores*." As part of the scientific life of the court, public disputations were held on a variety of topics, one of which, a mathematical competition held in 1225 in Palermo, is of considerable historical importance. Fibonacci was challenged to solve the following cubic equation:

$$x^3 + 2x^2 + 10x = 20.$$

He offered the very good approximate solution

$$x = 1 + \frac{22}{60} + \frac{7}{60^2} + \frac{42}{60^3} + \frac{33}{60^4} + \frac{4}{60^5} + \frac{40}{60^6},$$

but without revealing how he had arrived at it.

He showed further that the solution could not be a positive integer, nor even a rational fraction or the square root of a rational number.

Fibonacci was the author of a book on geometry (*Practica geometrica*, 1220), but his real strength was algebra. Like the Indians before him, he interpreted negative numbers as debits and therefore accepted equations whose solutions were negative. He applied algebraic techniques to such problems, which in antiquity had been solved by using geometric methods. He even worked on equations of the fourth degree.

As we have mentioned, his name is familiar in mathematics for a particular sequence of numbers. This sequence, the Fibonacci numbers, begins thus:

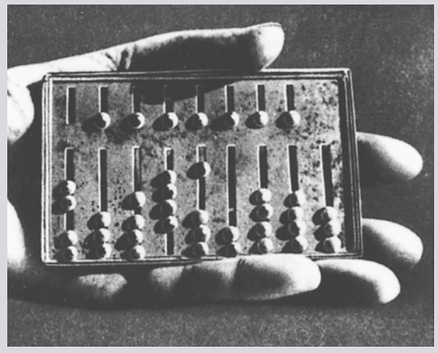
$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55,$$

whereby each number in the sequence (after the first two) is the sum of the two preceding terms. That is,

$$a_n = a_{n-1} + a_{n-2} \quad \text{for } n \geq 2;\ a_0 = 1,\ a_1 = 1.$$

Fibonacci came upon this sequence while investigating the following problem: A farmer owns a pair of breeding rabbits. Such rabbits produce a pair of offspring every month after reaching the age of fertility, which is two months. How many pairs of rabbits will the farmer possess after a given number of months? It is easily seen (Figure 2.43) that the number of pairs is given by the Fibonacci numbers.

▼ Figure 2.42
(a) The abacus (after *Kleine Enzyklopädie Mathematik*).
(b) Roman numerals.



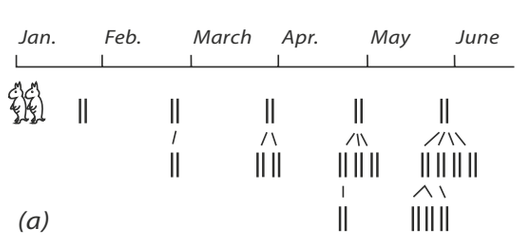
(a)

I	II	III	IV	V	VI	VII	VIII	IX
1	2	3	4	5	6	7	8	9
X	XX	XXX	XL	L	LX	LXX	LXXX	XC
10	20	30	40	50	60	70	80	90
C	CC	CCC	CD	D	DC	DCC	DCCC	CM
100	200	300	400	500	600	700	800	900

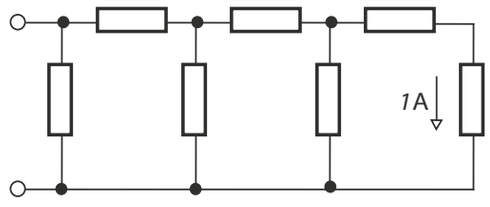
M or CIↃ = 1000

(b)

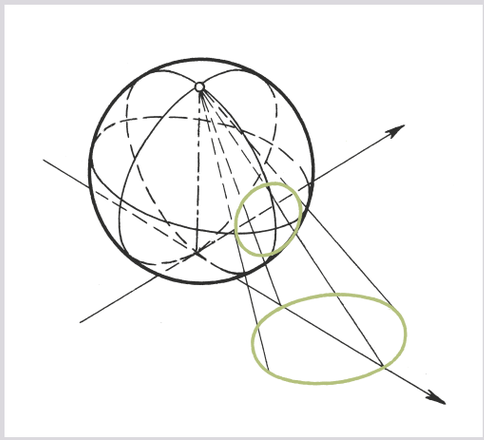
▼ Figure 2.43
(a) On the origin of the Fibonacci numbers. The number of pairs of rabbits due to continuous reproduction by each pair.
(b) The voltages measurable at the resistors of this network are also given by the Fibonacci numbers if all resistors have a resistance of 1 ohm, and a current of 1 ampere flows through the last resistor.



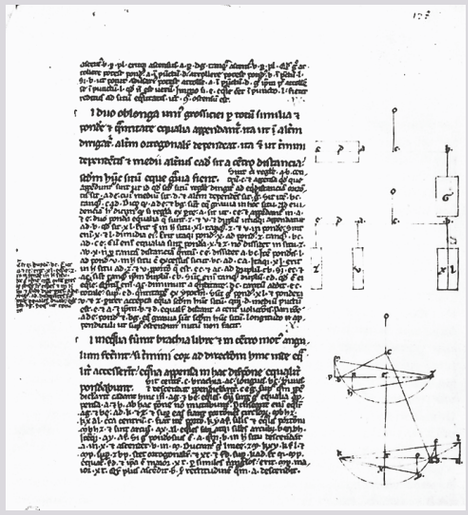
(a)



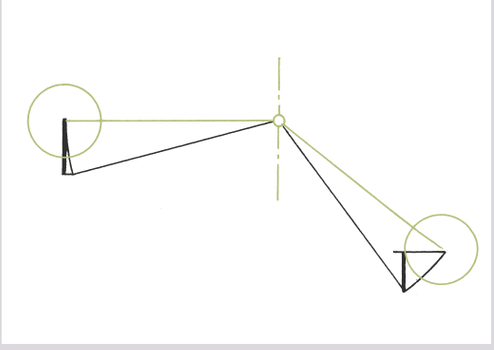
(b)



▲ **Figure 2.44** Jordanus Nemorarius knew that under stereographic projection, a circle is transformed again into a circle.



▲ **Figure 2.45** A page from Nemorarius's book [Clagett 1961].



▲ **Figure 2.46** On the derivation of the principle of the lever, after Nemorarius.

Over the course of history, new and unexpected properties of the Fibonacci numbers were discovered. For example, Luca Pacioli (Plate XIII) gave the following formula for the nth Fibonacci number:

$$a_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$

The number

$$\left(1+\sqrt{5}\right)/2 \approx 1.618$$

is the golden ratio (see Figure 1.40). Later, a relationship was discovered between Fibonacci numbers and the logarithmic spiral. There is even a relationship between the spatial arrangement of the seeds in a sunflower and the Fibonacci numbers.

As can be seen from the above formula, the ratio of two successive Fibonacci numbers approaches the golden ratio as a limiting value:

$$a_{n+1} / a_n \to \left(1+\sqrt{5}\right)/2.$$

2.4.2 *Jordanus Nemorarius: Structural Engineer*

In the Middle Ages, the greatest progress in physics was in the area of mechanics. At the end of the twelfth century, all that were available in this domain were a work attributed to but most probably not written by Aristotle (*Questiones mechanicae*) and one of the works ascribed to Euclid (*Liber Euclidis de ponderibus secundum terminorum circumferentiam*). There is no such work of Euclid known directly from classical Greek sources; it came into Latin through the Arabic. An important role in the development of mechanics in Europe was played by another work, also a translation from the Arabic, with the title *Liber karastonis*, whose author was the Arab scholar Thabit Ibn Qurra. It is not quite clear whether the "karastonis" (Arabic *qarastun*) in the title of the work is a name or refers to the word for scale. At this time, neither the works of Archimedes nor those of Hero and Pappus were known.

The lever problem was treated in a variety of cases in the works mentioned above, and the laws thus derived were applied to simple machines (Section 1.4). Jordanus Nemorarius became interested in this work and developed it further. We know little of his life, in contrast to other scholars, whom we will discuss below. Indeed, there are so few reliable facts that even the period in which he worked is known only to lie somewhere between 1050 and 1350. Most probably, he lived in the thirteenth century. Two works are ascribed to him: *Elementa Jordani super demonstrationem ponderis secundum situs* and *Liber Jordani de ratione ponderis*. The latter work is frequently ascribed to one of Nemorarius's students or else is viewed as a late work of the master because it corrects a number of erroneous assertions from the earlier work. Mathematics also owes Nemorarius a debt of gratitude for several valuable discoveries. For instance, through his investigation of cartographical problems, he determined that under stereographic projection (Figure 2.44), a circle on a sphere is transformed to a circle in the plane. Here, however, we would like to focus on Nemorarius's contributions to statics. His most important discovery determined that if we relate the equilibrium position of levers with unequal lever arms to the movements of the weights applied to the two arms, as it had already been done in peripatetic dynamics (Section 1.3.2), then only the vertical displacement plays a role (Figures 2.45 and 2.46). Equilibrium results when, with

equal weights, the vertical displacements are the same; in every other case, the scale tips in the direction in which the displacement is greater.

Quotation 2.11 is a detailed extract from *Elementa Jordani*. It can easily be deduced from the text that Nemorarius used the following axiom: If an action can lift a weight to a certain height, then it can lift a weight n times as heavy to $1/n$ of that height. This basic law, in which we can see the kernel of the law of conservation of energy, is the *first independent step taken by European science* that goes beyond what was achieved in antiquity and beyond even the Arab refinements of classical science.

This idea of Nemorarius also paves the way for the decomposition of forces into components. Indeed, in his second book, Nemorarius determines the effect of a body placed on an inclined plane in the direction of the slope—today we would say the force component parallel to the slope. His train of thought is given in its entirety in Quotation 2.12. From Figure 2.47, we may say, somewhat simplified, that two weights connected by a string stretched over a pulley are in equilibrium when the vertical projections of the equal motions of the weights on the inclined planes are inversely proportional to the weights. Expressed in today's language and notation, neither of which were in use at that time, Nemorarius's result could be expressed as follows: Equilibrium is obtained when

$$G_1 : G_2 = h_2 : h_1 \quad (s_1 = s_2).$$

This can be transformed into

$$G_1 \frac{h_1}{s_1} = G_2 \frac{h_2}{s_2},$$

from which, finally, for the components of the weight along the slope, we obtain the relation

$$G_1 \frac{h_1}{s_1} = G_1 \sin\alpha_1 = G_2 \frac{h_2}{s_2} = G_2 \sin\alpha_2.$$

Expressed slightly differently, the two bodies are in equilibrium when their weights are in the same ratio as the lengths of the inclined planes.

We emphasize once again that here the diagram, the train of thought, as well as the notation, are purely of our own time.

2.4.3 *Descriptive Kinematics: Nicole Oresme and Merton College*

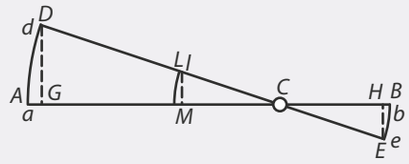
The Middle Ages saw progress in other areas of physics besides statics, which, as we shall see, in a certain sense go beyond even Stevin and Galileo. For instance, there was progress in the descriptive study of motion, that is, in kinematics, that could have made possible a direct connection (although one that was not further used) to the kinematics of the seventeenth century. This progress in kinematics can be traced to Nicole Oresme (ca. 1323–1382), a polymath and bishop of Lisieux, and to Merton College, Oxford. The path leading to their findings is perhaps of greater interest than the results themselves in two important respects. First, the goal of the investigations at the outset was to find numerical values for the relevant qualities, or more precisely, for the *intensities* of the relevant qualities. We have already seen, in our discussion of Aristotle, the important role played in antiquity, and in the Middle Ages as well, by the warm–cold, slow–fast, and light–dark pairs of opposite qualities. In Section 0.2 we described how the quantification of qualities represents

Quotation 2.11

"Omnis ponderosi motum esse ad medium, virtutemque ipsius potentiam ad inferiora rendenti et motui contrario resistendi. Gravius esse in descendendo quando ejusdem motus ad medium rectior. Secundum situm gravius, quando in eodem situ minus obliquus est descendus. Obliquiorem autem descensum in eadem quantitate minus capere de directo."

The movement of every weight is toward the center (of the world), and its force is a power of tending downward and of resisting movement in the contrary direction. That which is heavier descends more quickly. It is heavier in descending, to the degree that its movement toward the center (of the world) is more direct. It is heavier in position when in that position its path of descent is less oblique. A more oblique descent is one which, in the same distance, partakes less of the vertical. One weight is less heavy in position, than another, if it is caused to ascend by the descent of the other. ...

If the arms of a balance are proportional to the weights suspended, in such manner that the heavier weight is suspended from the shorter arm, the weights will have equal positional gravity.

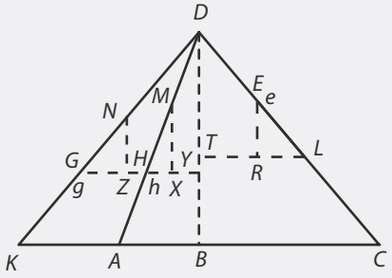


Let the balance beam be *ACB*, as before, and the suspended weights *a* and *b*; and let the ratio of *b* to *a* be as the ratio of *AC*. For let it be supposed that it descends on the side of *B;* and let the line *DCE* be drawn obliquely to the position of *ACB*. If then the weight *d*, equal to *a,* and the weight *e* equal to *b*, are suspended, and if the line *DG* is drawn vertically downward and the line *EH* vertically upward, it is evident that the triangles *DCG* and *ECH* are similar, so that the proportion of *DC* to *CE* is the same as that of *DG* to *EH*. But *DC* is to *CE* as *b* is to *a*; therefore *DG* is to *EH* as *b* is to *a*. Then suppose *CL* to be equal to *CB* and to *CE*, and let *l* be equal in weight to *b*; and draw the perpendicular *LM*. Since then *LM* and *EH* are shown to be equal, *DG* will be to *LM* as *b* is to *a*, and as *l* is to *a*. But, as had been shown, *a* and *l* are inversely proportional to their contrary (upward) motions. Therefore, what suffices to lift *a* to *D*, will suffice to lift *l* through the distance *LM*. Since therefore *l* and *b* are equal, and *LC* is equal to *CB*, *l* is not lifted by *b*; and consequently *a* will not be lifted by *b*, which is what is to be proved.

—Jordanus Nemorarius, *Liber de Ratione Ponderis* [p. 216]

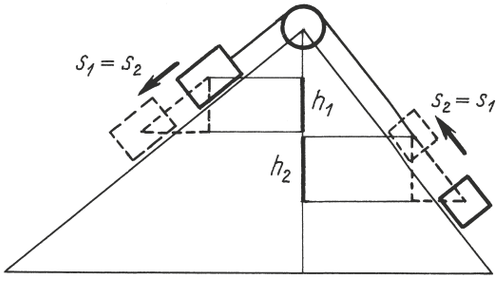
Quotation 2.12

If two weights descend along diversely inclined planes, then, if the inclinations are directly proportional to the weights, they will be of equal force in descending.



Let there be a line *ABC* parallel to the horizon, and let *BD* be erected vertically on it; and from *D* draw the lines *DA* and *DC*, with *DC* of greater obliquity. I then mean by proportion of inclinations, not the ratio of the angles, but of the lines taken to where a horizontal line cuts off an equal segment of the vertical. Let the weight *e*, then, be on *DC*, and the weight *h* on *DA*; and let *e* be to *h* as *DC* is to *DA*. I say that those weights are of the same force in this position. For let *DK* be a line of the same obliquity as *DC*, and let there be on it a weight *g*, equal to *e*. If then it is possible, suppose that *e* descends to *L*, and draw *h* up to *M*. And let *GN* be equal to *HM*, which in turn is equal to *EL*. Then let a perpendicular on *DB* be drawn from *G* to *H*, which will be *GHY*, erect the perpendiculars *NZ* and *MX*; and on *LT*, erect the perpendicular *ER*. Since then the proportion of *NZ* to *NG* is as that of *DY* to *DG*, and hence as that of *DB* to *DK*, and since likewise *MX* is to *MH* as *DB* is to *DH*, *MX* will be to *NX* as *DK* is to *DA*—that is, as the weight *g* is to the weight *h*. But because *e* does not suffice to lift *g* to *N*, it will not suffice to lift *h* to *M*. Therefore they will remain as they are.

—Jordanus Nemorarius, *Liber de Ratione Ponderis* [p. 219]



▲ **Figure 2.47** The derivation of the equilibrium condition for two bodies on inclined planes with differing angles of inclination, after Nemorarius in a modern formulation. In his argument, we can recognize the germ of the idea of the components of a force.

an important milestone in the development of a science. In the Middle Ages, the intensities of qualities, which would eventually have to evolve into the concept of measurability, were rather broadly defined. Thus in the twelfth century, in the maxims (*sententiae*) of Peter Lombard (Petrus Lombardus), which were used as the basis of discussion in theological instruction, the question was raised whether the property of *caritas* (charity) can be of greater intensity in an individual at one time than at another. He gave a very interesting answer to this question. Namely, if a quality existed "absolutely," for example, the divine love that encompassed all *caritas*, then the measure of intensity could be only that fraction of the absolute *caritas* partaken by the person in question.

In the thirteenth and above all the fourteenth centuries, the intensities of qualities of physical interest were frequently described by assigning them numbers with no precise definition and without even the idea that such quantities could or should be measured. Roger Bacon, whom we shall meet again later, went so far as to solve problems of the following type: Given two portions of water with intensity of warmth equal to 6, and one portion of water with intensity 12, how large is the intensity of the warmth of the water after the portions are mixed? Bacon's answer was: the intensity is 8. Consider that Bacon formulated and solved such a problem—which even today is a "touchstone" problem on university entrance examinations—obtaining the correct answer without either of the notions of specific heat and temperature.

The second step of great significance was taken on the path that leads to the fundamental notions of kinematics. An attempt was made not only to characterize the intensities by numbers, but to represent them graphically as the lengths of vertical line segments. It thereby became possible to visualize *changes* in intensity. We see here the first step on the way to the idea of a mathematical function and to the representation of a function in a coordinate system. This idea is exemplified in Quotation 2.13, which contains extracts from the 1350 work of Nicole Oresme called *Tractatus de configurationibus qualitatum et motuum*. Here the values of the intensities, called *latitudines* (*latitudo* = width) are drawn vertically; they rest on different points on the horizontal axis, which is called *longitudo*. For a process of motion, this *longitudo* could be set equal to the distance. Nicole Oresme investigates several forms of motion, which he calls *uniformis*, *uniformiter difformis*, or *difformiter difformis* (Figure 2.48). He states that the scale of measurement for *latitudo* is irrelevant to the character of a change in intensity. Thus, for example, an elliptical path connecting the peaks of the verticals can describe the same change as a circular one (Figure 2.49).

For kinematics, the investigation of uniformly changing motion (*motus uniformiter difformis*) is of particular importance. The question of whether change in velocity is to be understood as a function of time or as a function of position was not an easy one to answer. It gave pause later to Galileo (Section 3.3), and even Descartes. Nicole Oresme and the Merton College scholars Thomas Bradwardine (ca. 1290–1349), William Heytesbury (ca. 1313–ca. 1372), and Richard Swineshead (fl. ca. 1350) determined that in this case, the result (the effect)—that is, the traversed path—depends on the average velocity. This rule is known today as the *Merton rule* because the Merton College scholars knew of it a half-century before the publication of Nicole Oresme's book. In today's notation, the Merton rule can be written

$$v = \frac{v_0 + v_1}{2}.$$

It is one of the basic facts of physics today as well.

2.4.4 *Peripatetic Dynamics Reformed*

Significant work was done as well in the area of dynamics by an Englander and a Frenchman: Thomas Bradwardine and Jean Buridan. In his 1328 work *Tractatus proportionum*, Bradwardine, of Merton College, sharply criticized the fundamental law of Aristotelian dynamics, according to which velocity is proportional to the quotient of force and resistance. Using today's notation, as we have frequently done before, this law can be expressed as

$$v \propto \frac{F}{R}.$$

Bradwardine rejected this relationship because, as he reasoned, it implied the existence of a velocity even when the force was less than the resistance, which is clearly impossible. He rejected as well the frequently used proportion

$$v \propto F - R,$$

according to which the resistance is subtracted from the force. Bradwardine studied and rejected more complicated relationships, for example the proportion

$$v \propto \frac{F-R}{R},$$

and ultimately all such approaches, by relying on direct experience or even by appealing to Aristotle himself. He eventually arrived at a very peculiar law of motion, which we are going to formulate more simply than he did, anachronistically making use of today's concepts. According to Bradwardine, velocity should in fact depend on the quotient of the force and the resistance, that is, we should indeed have $v = v(F/R)$; however, the nth power of the relationship F/R should be associated with the n-fold velocity. This relationship corresponds to the formula

$$nv\left(\frac{F}{R}\right) = v\left[\left(\frac{F}{R}\right)^n\right],$$

which has the following solution, as can be easily checked by substitution:

$$v \propto \log\frac{F}{R}.$$

Of course, Bradwardine did not know of the logarithm function, so he could not have written down the law in this form. However, in this way we can more easily recognize the characteristics of his approach. First, we have a consequence on which Bradwardine placed considerable importance: Velocity is zero when the force is equal to the resistance. If the force is smaller, then a negative value for the velocity results, which, according to Bradwardine, implied that motion was impossible.

There is no reason to explain further why this law of dynamics is totally false. Bradwardine had not even attempted to verify it by empirical means. Nonetheless, his approach is significant because, first of all, *he attacked Aristotelian physics at one of its weak and sensitive points*, and second, new functional relationships, namely the logarithm and its inverse, the exponential, were introduced into physics for the first time. Bradwardine's works obtained a wide readership and with

Quotation 2.13

Every measurable thing except numbers is conceived in the manner of continuous quantity. Hence it is necessary for the measure of such a thing to imagine points, lines, and surfaces—or their properties—in which, as Aristotle wishes, measure or proportion is originally found ... Hence every intension which can be acquired successively is to be imagined by means of a straight line erected perpendicularly on some point or points of the [extensible] space or subject of the intensible thing.

... Therefore, let extension of a quality be called nominally its "longitude" and intension be called its "latitude" or "altitude".

Every velocity endures in time. And so time or duration of that velocity will be the longitude. The intension of the velocity will be its latitude. And although time and a linear magnitude (*linea*) are not [strictly] comparable in quantity, yet there is no ratio found between one time [interval] and another that is not also found between lines, and vice-versa. ... The same thing holds for intension of velocity ... Therefore, we can arrive at a knowledge of the difformities [changes] of velocities by the imagination of lines and also figures.

... The intensive acquisition of a surface quality ... is to be imagined by the ascent of a surface, and by its imagined motion it describes the body by which the quality is designated. And similarly the intensive acquisition of a corporeal quality is imagined by the motion of a surface quality, because by its imagined flux a surface describes a body. And it is not necessary to give a fourth dimension. ...

—Nicole Oresme, *Tractatus* [Clagett 1961]

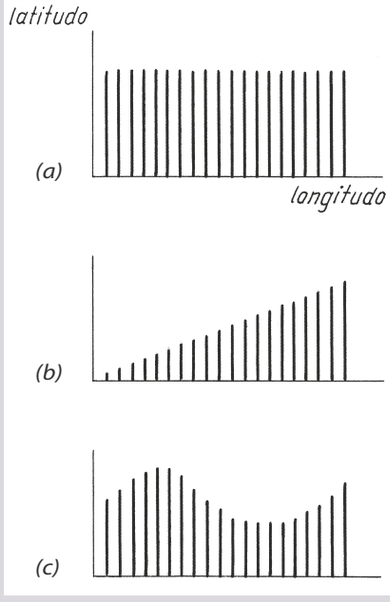
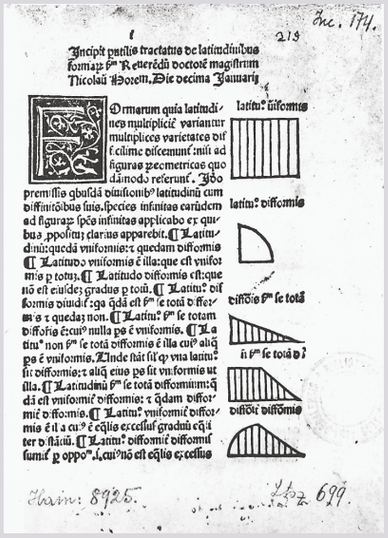
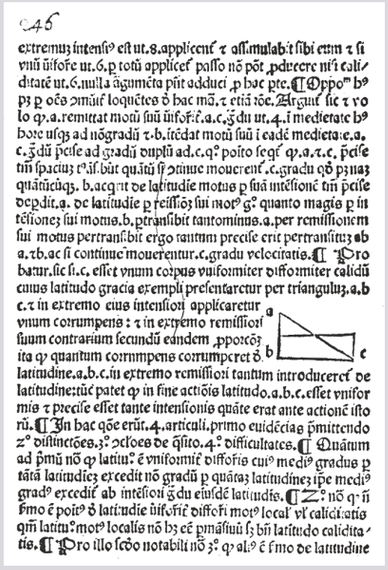
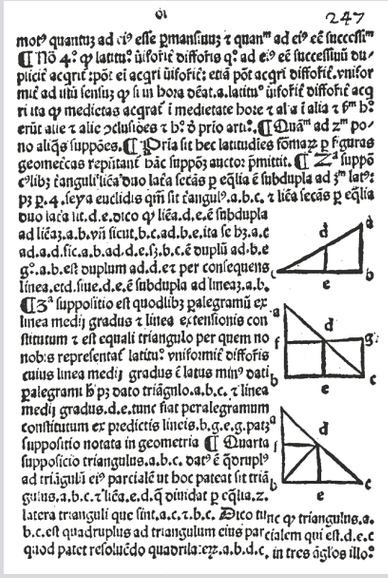


Figure 2.48 Changes denoted by Oresme as **(a)** *uniformis*, **(b)** *uniformiter difformis*, and **(c)** *difformiter difformis*.







◀ ▲ ▼
Figure 2.49
Facsimile of three pages from Oresme's book. (Incunabula, Székesfehérvár Diocesan Library.)

the help of the numerical examples provided by him and his followers, scientist became accustomed to functional relationships of this sort and learned to work with them—at least numerically.

Around 1360, Nicole Oresme formulated a similar relationship; he expressed the connection between force and resistance in terms of a quotient of velocities. In today's notation, his approach becomes

$$\frac{F_2}{R_2} = \left(\frac{F_1}{R_1}\right)^{v_2/v_1}.$$

While Bradwardine's approach led to the function $x = \log y$, Oresme's yields the exponential function $y = a^x$. Here, we again witness a phenomenon that is by no means unique in the history of science: that an idea that is hopelessly wrong for its intended purpose can still be important for the development of physics as a whole even when the original approach leads to a dead end.

2.4.5 *Buridan's Theory of Impetus*

In contrast to the attempts described above, the ideas of Jean Buridan proved fruitful, pointing the way to the further development of dynamics. Born in 1300, Buridan was rector of the University of Paris in 1327. He died in 1358 (Quotation 2.14). In his book *Questiones octavi libri physicorum*, he takes up a problem of Aristotelian dynamics discussed in antiquity and again, especially by the Byzantine philosopher John Philoponus (Section 2.2.2), namely, the question of what makes a body move after it has been thrown. The problem was that, according to Aristotelian physics, a motive force (*motor conjunctus*) was necessary to set an inanimate object (*res inanimata*) in motion, and such a force would have to remain in constant contact with the object. Buridan cites the two possible solutions that were offered by Aristotle:

1. As a moving object leaves a position, a vacuum is left in its place. The air that rushes in as a result of *horror vacui* pushes the object along its course. Aristotle himself rejects this theory with the argument that the streaming air itself would have to leave a vacuum behind it, into which the air behind that would also flow, and so on. As a result, the moving object would end up setting into motion all of the universe behind it, which is clearly impossible.
2. The moving object transfers its capacity to move (*virtus movens*) to the surrounding air, which then continues to maintain the motion. We have already discussed this approach in Section 1.3.

Buridan criticizes both opinions. His critique is noteworthy in two respects: for his *critical stance* against Aristotle; and the fact that Buridan argues on the basis of sound everyday experiences. For these reasons, it is worth presenting his ideas in full detail. Particularly convincing is the argument that if we attached a sharp point to the *trailing* end of a spear, it could be still thrown just as far as a normal spear, even though the air clearly would not be able to push it forward as effectively. Another argument is that a ship can move for a long time after whatever force that originally set it into motion has been discontinued, even though those standing on deck can feel that the air in no way furthers the motion but, if anything, hinders it.

According to Buridan, at the moment of release a certain *impetus* is imparted to an object, which then makes its further motion possible (Quotation 2.15). One can even infer from Buridan's argument—even if it is not clear and concise—that

impetus is proportional to the mass and also to the velocity of the object. Hence the impetus is related, even if only in a rudimentary way, to our current notion of momentum. Another great step taken by Buridan was that he saw the impetus as something permanent, something that did not get used up by itself, but for the *deterioration* of which a particular resistance was required. According to Buridan, an object will eventually lose its impetus because in the course of its motion, it has to overcome continuous resistance. He applied this approach to celestial mechanics and asserted that no supernatural being is required to account for the motion of the heavenly bodies, as had been maintained by medieval theology, which cited Aristotle, because lacking any resistance to motion, they maintain the impetus that was imparted to them at the moment of creation. There are two ways in which these ideas are significant. First, we see here the kernel of the law of inertia; second, accounting for *celestial motion as a special case of terrestrial motion* in the case of vanishing resistant force was a bold idea, because here Buridan *subjects celestial and terrestrial motions to the same law*. This step represents a decisive and indeed a successful attack on the hierarchical ordering of the Aristotelian worldview, in which Heaven and Earth were kept strictly apart.

Buridan also applied his impetus theory to the problem of free fall. According to Max von Laue's 1947 *History of Physics*, the impetus theory is also significant because it not only answers the question for which it was originally devised, to wit why a body moves after having been flung, but it also was of use in solving a host of other problems. Thus in free fall, the *gravitas* continually increases the impetus in such a way that the impetus already acquired is preserved, and it is this increasing impetus that moves the object. If we again place, as did Buridan, this train of thought back in the Aristotelian framework in which the velocity is proportional to the motive force, then the velocity must be proportional to the growing impetus and therefore must itself increase.

Here we see that such a combination of Aristotelian and non-Aristotelian elements can lead to solutions that come very close to today's point of view.

2.4.6 *Physics in Astronomy*

Of course, fourteenth-century researchers were also intrigued by questions relating to astronomy. For astronomical calculations, which were an essential part of astrology, the Ptolemaic system generally provided a satisfactory degree of accuracy. However, leading scholars began to become interested in going beyond the simple plotting of celestial phenomena to asking questions about their physical nature. Thus questions were raised, for example, about the shape of Earth: is it an exact sphere? A method was proposed for finding the answer. Albert of Saxony (ca. 1316–1390), who was rector of the University of Paris in 1353, pointed out in his work *Acutissimae questiones* that Earth's deviation from a perfect sphere could be determined by measuring the distance between neighboring longitudes at different latitudes (Figure 2.50). Also of interest are his investigations of free fall, during which he considered a kind of motion when velocity is increasing proportional to both distance and time. He was alarmed by the implication that after a very long duration of falling, an object would reach an arbitrarily high velocity. He came to the conclusion that during the course of the motion, the resistance must grow more rapidly than the impetus, so that in the end, the velocity will not exceed a certain value.

A further question that arose was whether the centers of Earth and the universe coincide exactly. It was correctly determined that the location of Earth's center

Quotation 2.14

Where is the learned Heloise
For whom they gelded Abelard
Who then turned monk at Saint-Denis? For her love's sake he suffered this. And furthermore, where is the queen
Who commanded that Buridan
Be bagged and cast into the Seine? But what's become of last year's snows?

—Francois Villon, "Ballade des Dames du temps jadis"

Quotation 2.15

It is sought whether a projectile after leaving the hand of the projector is moved by the air, or by what it is moved. It is argued that it is not moved by the air because the air seems rather to resist, since it is necessary that it be divided.

... The motor in moving a moving body impresses in it a certain impetus or a certain motive force of the moving body ... And by the amount the motor moves that moving body more swiftly, by the same amount it will impress in it a stronger impetus.

It is by that impetus that the stone is moved after the projector ceases to move.

... By the amount more there is of matter, by that amount can the body receive more of that impetus and more intensely ... a feather receives such an impetus so weakly that such an impetus is immediately destroyed by the resisting air.

And so also if light wood and heavy iron of the same volume and of the same shape are moved equally fast by a projector, the iron will be moved further because there is impressed in it a more intense impetus, which is not so quickly corrupted as the lesser impetus would be corrupted.

... the Bible does not state the appropriate intelligences move the celestial bodies, it could be said that it does not appear necessary to posit intelligences of this kind, because it would be answered that God, when He created the world, moved each of the celestial orbs as He pleased, and in moving them He impressed in them impetuses which moved them without his having to move them any more except by the method of general influence whereby he concurs as co-agent in all things which take place; "for thus on the seventh day He rested from all the work which He had executed by committing to others the actions and the passions in turn." And these impetuses which He impressed on the celestial bodies were not decreased or corrupted afterwards, because there

continued on next page

Quotation 2.15, ***continued***

was no inclination of the celestial bodies for other movements. Nor was there resistance which would be corruptive or repressive of that impetus. But this I do not say assertively, but [rather tentatively] so that I might seek from the theological masters what they might teach me in these matters as to how these things take place. ...

... From this theory also appears the cause of why the natural motion of a heavy body downward is continually accelerated (*continue velacitatur*). For from the beginning only the gravity was moving it. Therefore, it moved more slowly, but in moving it impressed in the heavy body an impetus. This impetus now [acting] together with its gravity moves it. Therefore the movement becomes faster; and by the amount it is faster, so the impetus becomes more intense. Therefore the movement evidently becomes continually faster.

[The impetus then also explains why] one who wishes to jump a long distance drops back a way in order to run faster, so that by running he might acquire an impetus which would carry him a longer distance in the jump. Whence the person so running and jumping does not feel the air moving him, but [rather] feels the air in front strongly resisting him.

... There is also a fourth doubt [about Aristotle's conclusion] whether, in positing that the earth is moved circularly around its own center and about its own poles, all the phenomena that are apparent to us can be saved (*possent salvari omnia nobis apparentia*).

... But still this opinion is not to be followed. In the first place because it is against the authority of Aristotle and of all the astronomers. But these people [persons wishing to sustain this opinion] respond that *authority does not demonstrate*, and that it suffices astronomers that they posit a method by which appearances are saved, whether or not it is so in actuality. Appearances can be saved in either way; hence they posit the method which is more pleasing to them.

—John Buridan, Commentaries on works of Aristotle [pp. 276, 277, 501]

Quotation 2.16

Then spake Joshua to the Lord in the day when the Lord delivered up the Amorites before the children of Israel, and he said in the sight of Israel, Sun, stand thou still upon Gibeon; and thou, Moon, in the valley of Ajalon.

continued on next page

must change when large masses are shifted on Earth's surface as a result of natural phenomena. Some phenomena that result from Earth's spherical form were considered in accordance with their significance; in other cases, it became something of a sport to enumerate such "effects" that due to their tiny order of magnitude existed only in principle. Thus, it was naturally claimed that two plumb bobs at nearby locations on Earth's surface would not hang parallel to each other because they would both point to the center of Earth. The plumb bobs would cause two towers standing side by side to be inclined a bit away from each other, and the diameter of a well to be smaller at the bottom than at the surface. It even came to a point of legal hairsplitting, with arguments that excavation work at the boundary of a field should be carried out in such a way that digging not be perpendicular to the ground because the result would then be the undermining of the neighboring property. At this point it may be appropriate to note that in the construction of very large structures, such as the Golden Gate Bridge across San Francisco Bay, the deviations from the perpendicular resulting from the curvature of Earth do, in fact, have to be taken into consideration (Figure 2.51).

There was also debate over the arguments for and against the rotation of Earth on its axis. We may mention again Nicole Oresme. In turn, he refuted each of Aristotle's arguments against Earth's rotation and offered a convincing proof that rotation does in fact occur. His argument again has a number of interesting aspects. One is the usual: he has no respect for Aristotle and squarely argues against his opinions. The other, however, is on the positive side: he argues in the refutation that an observer can only perceive relative motion, the result being that the phenomena on Earth and movement of the Heavenly bodies will appear to us exactly the same, whether we imagine that Earth turns about its axis or that the celestial spheres rotate around Earth. If we compare Oresme's ideas with those given 300 years later by Galileo, we are struck by their similarity. Oresme refuted not only Aristotle's objections to Earth's rotation, but those of the Bible as well, above all that found in the book of Joshua (Quotation 2.16). He explained that many of the statements in the Bible are to be taken only as similes; as evidence he makes use of other Biblical quotes. Here also, his argument is similar to that of Galileo (Quotation 2.17). However the fates of the two men were quite different: Oresme rose to a high position in the ecclesiastical hierarchy and died as the bishop of Lisieux. Galileo, in contrast, was forced to recant his teachings.

2.4.7 *Results*

The principal direction of attack against the Aristotelian system and the direction of development of mechanics coincided in the Middle Ages. That is why considerable space is devoted in this book to the concepts and laws of mechanics. We conclude this section by summarizing the results of medieval mechanics after Clagett [1961].

In statics, the principle of virtual work was correctly applied to certain cases that were by no means trivial. Computations were done using torque, even though the concept was not fully explained. For an object on an inclined plane, the component of its weight that is in the direction parallel to the plane was determined.

We have not mentioned to this point that in completing Archimedes' work on hydrostatics as handed down by the Arabs, the concept of specific gravity was clarified.

In the study of motion, laws were worked out for uniform motion and for motion under constant acceleration. The notions of constant velocity, average veloc-

ity, and even instantaneous velocity were introduced in forms identical to those in use today (Quotation 2.18). An outstanding advance in this area was the understanding of relative motion. Velocity under free fall was seen as being proportional both to the duration of the fall and to the distance traversed. (Of course, these two conditions are mutually contradictory.)

With the introduction of the notion of impetus, the constraint of immediate contact with a motive force was abandoned. The motion of the celestial bodies was explained by assuming that the same laws hold true in the heavens as on Earth and that bodies maintain their initial impulse; this insight paved the way to a more precise formulation of the law of inertia. The arguments for and against the rotation of Earth were examined in the spirit of future centuries.

However, it should be emphasized that the possibility that Earth revolved about the Sun was *not* considered. It was exclusively the contribution of Copernicus to have reawakened this classical idea.

2.4.8 Nicole Oresme on the Motion of Earth

1. ... some say the earth is at the center of the universe and revolves and moves circularly around the pole established for this, as written in Plato's *Timaeus*.

   Remark: This was the opinion of one called Heraclides of Pontus who proposed that the earth is moved circularly and that the heavens are at rest. Aristotle does not here refute these opinions; it perhaps seemed to him that they have little [root in] appearance and are well refuted elsewhere in philosophy and astronomy.

   But it seems to me, subject to correction, that one could well support and give luster to the last opinion, namely that the earth, and not the heavens, is moved with a daily movement. Firstly, I wish to state that one could not demonstrate the contrary by any experience. Secondly, [I will show that the contrary cannot be demonstrated] by reasoning. And thirdly, I will put forth reasons in support of it [that is, the diurnal rotation of the earth].

2. As for the first point, one experience [commonly cited in support of the daily motion of the heaven is the following]: We see with our senses the sun and moon and many stars rise and set from day to day, and some stars turn around the arctic pole .... Thus the heaven is moved with a diurnal movement. Another experience is [this]: If the earth is so moved, it makes a complete turn in a single natural day. Therefore, we and the trees and houses are moved toward the east very swiftly, and so it should seem that the air and wind blow continuously and strongly from the east .... But the contrary appears by experience. The third [experience] is that which Ptolemy advances: If a person on a ship moved rapidly eastward and an arrow were shot directly upward, it ought not to fall on the ship but a good distance westward from the ship. Similarly, if the earth is moved so swiftly in turning from west to east, and it has been posited that one throws a stone directly above, then it ought to fall, not on the place it left, but a good distance to the west. But in fact the contrary is clear.

   It seems to me that by [using] what I shall say regarding these experiences, one could respond to all the other [experiences] which might be adduced in this matter

3. ... Again, I make the supposition that local motion can be sensibly perceived only in so far as one may perceive one body to be differently disposed with respect to another. In support of this [I give the following illustration]: If a person is in one ship called *a*, which is moved very carefully [that is, without pitching or rolling]—either rapidly or slowly—and this person sees nothing except another ship called *b*, which is moved in every respect in the same manner as *a* in which he is situated, I say that it will seem to this person that neither ship is moving. And if *a* is at rest and *b* is moved, it will appear and seem to him that *b* is moved.

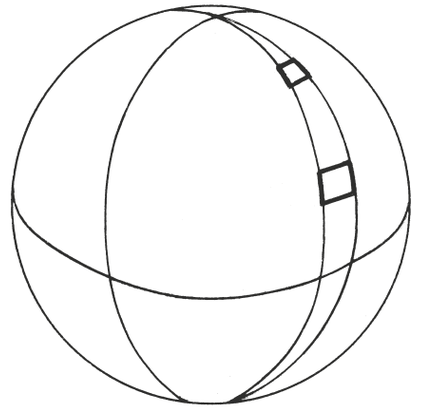
   [On the other hand], if *a* is moved and *b* is at rest, it will appear to him as before that *a* is at rest and *b* is moved ... this is apparent by experience. The reason for this is that

Quotation 2.16, ***continued***

And the sun stood still, and the moon stayed, until the people had avenged themselves upon their enemies. Is not this written in the book of Jasher? So the sun stood still in the midst of heaven, and hasted not to go down about a whole day.

And there was no day like that before it or after it, that the Lord hearkened unto the voice of a man: for the Lord fought for Israel.

—Joshua, 10.12–14



▲ **Figure 2.50** In the Middle Ages, not only was it known that the Earth is spherical, but there were even methods to determine the deviations from perfect sphericity. Four hundred years later, Maupertuis undertook to carry out those measurements. To make the connection, let us jump forward four centuries.

At the end of the seventeenth century, Newton concluded from his equations of motion and theory of gravitation that the Earth is flattened at the poles. The French *Académie* considered this assertion a question of cardinal importance, even a *cause célèbre*, in which the honor of the French nation was at stake because the measurements taken by French scientists, although not definitive, suggested for the Earth a prolate rather than oblate form.

Maupertuis, whom we must imagine as a gallant frequenter of salons, diligently carried out the task of making measurements near the polar circle. In 1737, he presented his findings to the *Académie*: the length of a degree on a meridian is 111.09 kilometers at the polar circle, but 110.46 kilometers measured in France (in today's units). Thus, Newton was correct; the Earth is an oblate spheroid.

Voltaire made sport of Maupertuis's efforts. He remarked ironically, "Why must Maupertuis, beset by wolves in the shimmering polar light, carry out these measurements when it would have sufficed to study Newton's *Principia* in a well-heated room?"

Voltaire likely did not notice the agreement of his argument with that of Simplicio in Galileo's *Dialogo*. Simplicio opined that one can obtain knowledge of nature from the teachings of Aristotle without having to battle the rigors of weather if one knows how to turn the pages of a book (see Quotation 3.24).

Quotation 2.17

Hence in expounding the Bible if one were always to confine oneself to the unadorned grammatical meaning, one might fall into error. Not only contradictions and propositions far from true might thus be made to appear in the Bible, but even grave heresies and follies. Thus it would be necessary to assign to God feet, hands, and eyes, as well as corporeal and human affections, such as anger, repentance, hatred, and sometimes even the forgetting of things past and ignorance of those to come. ... This being granted, I think that in discussions of physical problems we ought to begin not from the authority of scriptural passages, but from sense-experiences and necessary demonstrations; for the holy Bible and the phenomena of nature proceed alike from the divine Word, the former as the dictate of the Holy Ghost and the latter as the observant executrix of God's commands. It is necessary for the Bible, in order to be accommodated to the understanding of every man, to speak many things which appear to differ from the absolute truth so far as the bare meaning of the words is concerned.

—Galileo, "Letter to the Grand Duchess Christina" [pp. 181–182]

Quotation 2.18

For a moving body which starts from rest there can be imagined a range of velocity which increases indefinitely. In the same way can be imagined a range of acceleration or of slowing down (*latitudo intensionis vel remissionis*) according to which a body can accelerate or slow down its motion with an infinitely variable quickness or slowness. This second range is related to the range of motion [velocity] as the motion [velocity] is related to the magnitude [distance] that may be travelled in a continuous manner.

—William of Heytesbury, *De tribus praedicamentis* (1494) [Dugas 1988, p. 67]

There can be imagined one further succession, for every velocity is capable of being increased in intensity and decreased in intensity. Now its continuous increase in intensity is called acceleration, and indeed this acceleration or augmentation of velocity can take place more quickly or more slowly. Whence it sometimes happens that velocity is increasing and acceleration is decreasing, while sometimes both are simultaneously increasing. Similarly, acceleration of this sort sometimes takes place uniformly and sometimes non-uniformly and in diverse ways.

—Nicole Oresme, *The Configuration of Qualities and Motions* [p. 250]

these two bodies, *a* and *b*, are continually changing their dispositions with respect to each other in the same manner throughout when *a* is moved and *b* is at rest as they were conversely when *b* is moved and *a* is at rest. ... Similarly, if a person were in the heavens and it were posited that they were moved with a diurnal movement, and [furthermore] that this man who is transported with the heaven could see the earth clearly and distinctly and its mountains, valleys, rivers, towns, and chateaux, it would seem to him that the earth was moved with a diurnal movement, just as it seems to us on the earth that the heavens move. Similarly, if the earth and not the heavens were moved with a diurnal movement, it would seem to us that the earth was at rest and the heavens moved. ...

4. To the second experience, according to this opinion, the response is [this]: Not only is the earth so moved [diurnally], but with it water and the air, as was said, in such a way that the water and lower air are moved differently than they are by winds and other causes. It is like this situation: If air were enclosed in a moving ship, it would seem to the person situated in this air that it was not moved.

5. To the third experience, which seems more effective, that is, the experience concerning the arrow or stone projected upward, etc., one would say that the arrow is trajected upwards and [simultaneously] with this trajection it is moved eastward very swiftly with the air through which it passes and with all the mass of the lower part of the universe mentioned above, it all being moved with a diurnal movement. For this reason the arrow returns to the place on earth from which it left. This appears possible by analogy: If a person were on a ship moving toward the east very rapidly without his being aware of the movement, and he drew his hand downward, describing a straight line against the mast of the ship, it would seem to him that his hand was moved with rectilinear movement only. According to this opinion [of the diurnal rotation of the earth], it seems to us in the same fashion that the arrow descends or ascends in a straight line ...

6. As to the second point relative to the rational demonstration [of the diurnal movement of the heavens, I first note the following]: It seems to me that this [rational demonstration] proceeds from these arguments which follow and to which I shall respond in such a fashion that, using the same reasoning, one could respond to all other arguments pertaining to it ...

   Again, if the heavens were not moved with diurnal movement, all astronomy (as~e) would be false and a great part of natural philosophy where one supposes this movement of the heavens.

   Also, this seems to be against the Holy Scripture which says [in Eccles. I:5–6]: "The sun riseth, and goeth down, and returneth to his place: and there rising again, maketh his round by the south, and turneth again to the north: the spirit goeth forward surveying all places round about, and returneth to his circuits". And so it is written of the earth that God made it immobile: "For [God] created the orb of the earth, which will not be moved."

   Also, the Scriptures say that the sun was halted in the time of Joshua and that it returned (that is, turned back) in the time of King Ezechias. If the Earth were moved and not the heavens, as was said, such an arrestment would have been a returning and the returning of which it speaks would rather have been an arrestment. And this is against the Scriptures. ...

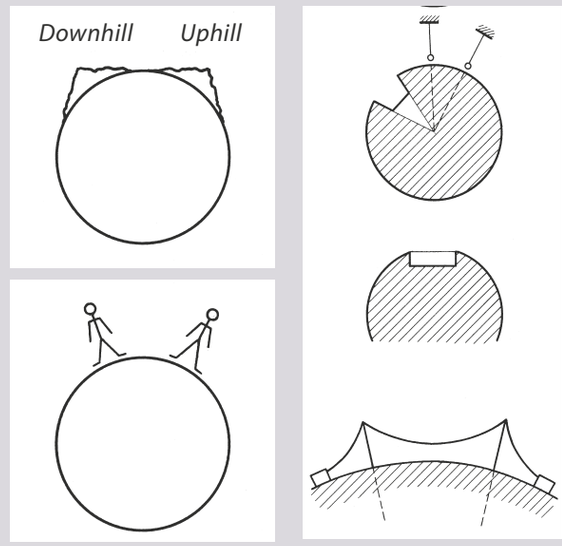
7. To the fifth argument, where it is said that if the heavens would not make a rotation day to day, all astronomy would be false, etc., I answer this is not so because all aspects, conjunctions, oppositions, constellations, figures, and influences of the heavens would be completely just as they are, as is evident clearly from what was said in response to the first experience. The tables of the movements and all other books would be just as true as they are, except in regard to the daily movement one would say of it that it is in the heavens "apparently" (*selon apparence*) but in the earth "actually" (*selon vente*). There is no effect which follows from the one [assumption] more than from the other. ...

8. To the sixth argument concerning the Holy Scripture which says that the sun revolves, etc., one would say that it is in this part [simply] conforming to the manner of common human speech, just as in several places, for example, where it is written that God is "repentant" and is "angry" and "pacified" and other such things which are just as they sound. Also appropriate to our question, we read that God covers the heavens with clouds and yet in reality the heavens cover the clouds. ... To the seventh argument, one would answer in the same way that according to appearances in the time of Joshua the

sun was arrested and advanced or speeded up its movement in the time of Ezechias. It would make no difference as to effect whichever opinion was followed. This latter opinion [supporting the diurnal rotation of the earth] seems more reasonable than the former, as we shall make clear.

9. Again, all philosophers say that something done by several or large-scale operations which can be done by fewer or smaller operations is done for nought. And Aristotle says in the eighth chapter of the first book that God and Nature do not do anything in vain. But if it is so that the heavens are moved with a diurnal movement, it becomes necessary to posit in the principal bodies of the world and in the heavens two contrary kinds of movement, one east-to-west, and others of the opposite kind, as has been said often. With this [theory of the diurnal movement of the heavens] it becomes necessary to posit an excessively great speed. This will become clear to one who considers thoughtfully the height or distance of the heaven, its magnitude, and that of its circuit; for if such a circuit is completed in one day, one could not imagine nor conceive of how the swiftness of the heaven is so marvelously and excessively great. It is so unthinkable and inestimable. Since all the effects which we see can be accomplished, and all the appearances saved, by substituting for this [diurnal movement of the heavens] a small operation, that is, the diurnal movement of the earth, which is very small in comparison with the heavens, and [since this can be done] without making the [number of necessary] operations so diverse and outrageously great, it follows that [if the heaven rather than the earth is moved] then God and Nature would have made and ordained Things for nought. But this is not fitting, as was said.

—Nicole Oresme, *Le Livre du ciel et du monde*, Book II, Chapter 25



▲ **Figure 2.51** A correct physical concept often serves merely as a starting point for hairsplitting logical and legal disputations.

2.5 Medieval Natural Philosophy

2.5.1 *Faith, Authority, and Science*

The place in the universe assigned to medieval man by his faith is on Earth, the "vale of tears," the center of the universe, which was created for mankind and which in both a physical and metaphorical sense revolved about him. During his earthly life, man's task was to strive to be worthy of eternal salvation by coming to know God, to love and serve Him. If he obeys the commandments, he will enjoy everlasting bliss in the hereafter. Divine revelation had made known God's intentions and laws, which were interpreted by the Church Fathers and councils so that they could be put into practice in daily life. So the role of "authority" emerged as a guide to everyday life. In the thirteenth century, alongside the Church authorities, we also find the authority of Aristotle (Quotation 2.19).

The question arises whether any philosophical speculation or even the use of philosophical methods was possible in such a heavily regulated universe (Figure 2.52). Was there a place alongside the authorities in this edifice of ideas for reasoning, common sense, logic, and perhaps even for the alpha and omega of all modern scientific research, from which all knowledge begins and in the end returns: observation and experiment?

In the previous section, we discussed the scientific achievements of the Middle Ages so that we may answer this question in the affirmative without hesitation. We now investigate whether science must inevitably be a foreign body in a world based on faith and whether it can be brought into harmony with the essential elements of the medieval worldview.

We should first point out that as far as the logical structure is concerned, the application of general principles—in this instance, the fixed opinions of the "authorities"—to a concrete case is in complete agreement with the general model of scientific method presented in Section 0.2, in which axioms are used to in-

Quotation 2.19

And raising my eyes a little I saw on high
Aristotle, the master of those who know,
ringed by the great souls of philosophy.

All wait upon him for their honor and his.
I saw Socrates and Plato at his side
before all others there. Democritus

who ascribes the world to chance, Diogenes,
and with him there Thales, Anaxagoras,
Zeno, Heraclitus, Empedocles.

And I saw the wise collector and analyst—
Dioscorides I mean. I saw Orpheus there,
Tully, Linus, Seneca the moralist,

Euclid the geometer, and Ptolemy,
Hippocrates, Galen, Avicenna,
and Averhoës of the Great Commentary.

—Dante, *Inferno*, Canto IV



▶ **Figure 2.52** The medieval cosmos as represented by Dante in his *Commedia divina* (Botticelli's illustration).

Beatrice was looking upward and I at her
when—in the time it takes a bolt to strike,
fly, and be resting in the bowstring's blur—

I found myself in a place where a wondrous thing
drew my entire attention; whereat she
from whom I could not hide my mind's least yearning

turned and said, as much in joy as beauty:
"To God, who has raised us now to the first star
direct your thoughts in glad and grateful duty."

…

Now, as the power of the sun's rays will strip
the wintry ground on which the snow has lain
of the cold and color that held it in their grip,

so you, with mind stripped clean, shall I delight
with such a radiance of living truth
that it will leap and tremble in your sight.

Within the heaven of peace beyond the sky
there whirls a body from whose power arises
the being of all things that within it lie.

The next sphere, that which is so richly lit,
distributes this power to many essences
distinct from itself, yet all contained within it.

The other spheres, in various degrees,
dispose the special powers they have within
to their own causes and effects. All these

great universal organs, as you now know,
proceed from grade to grade. Each in its order
takes power from above and does its work below.

Now then, note carefully how I move on
through this pass to the truth you seek, for thus
you shall learn how to hold the ford alone.

The motion and the power of the sacred gyres—
as the hammer's art is from the smith—must flow
from the Blessèd Movers. It is their power inspires.

And thus that Heaven made loveliest in its wheel
by many lamps, from the deep mind that turns it
takes the image and makes itself the seal.

—Dante, *Paradiso*, Canto II [pp. 403–405]

Quotation 2.20

Therefore, our faith cannot be proved by rational argument, since it transcends human understanding; however, it also cannot be refuted by rational argument in view of its validity. Thus when we discuss the tenets of our belief, we should strive to defend our faith, not to prove it.

—Thomas Aquinas

terpret individual cases—arbitrary axioms that could be based on experiences or could even be the rules of a game. Thus, an important discipline within philosophy, namely logic, with its logical methods and theories of proofs, can play an extensive role. The fundamental problem of logical structures is this: Are the axioms consistent with each other or can they lead to contradiction? Further, the question of completeness is also of great interest: Can all questions be answered with the help of the axioms? Did the thinkers of the Middle Ages exploit all the possibilities available to them? To the full extent! After Abelard (1079–1142) had collected the contradictory pronouncements of the Church Fathers on identical subjects in his work *Sic et non* (*Yes and No*), arguments ensued as to their consistency or inconsistency; they were in effect examining the axiomatic system for contradictions. Furthermore, we have already mentioned that discussions and debates were part of the normal university curriculum, with the aim of deriving all possible logically valid conclusions from a given premise and then either adding them to the existing edifice of ideas or bringing them into harmony with other assertions. In this way, the logic of the classical period was not only reawakened, but also enriched with new elements. The intellectual arsenal of the natural sciences was thereby enriched, whether intended or not.

As a professor at the University of Paris, Pedro Julião (Petrus Hispanus), who became Pope John XXI in 1276, wrote a textbook with the title *Summulae logicales* (Figure 2.53). The book *Perutilis logica*, by Albert of Saxony, also contains a number of modern questions. At the beginning of the fifteenth century, in his *Logica magna*, Paul of Venice (Paulus Venetus) gave a detailed presentation of medieval logic, including a discussion of paradoxes.

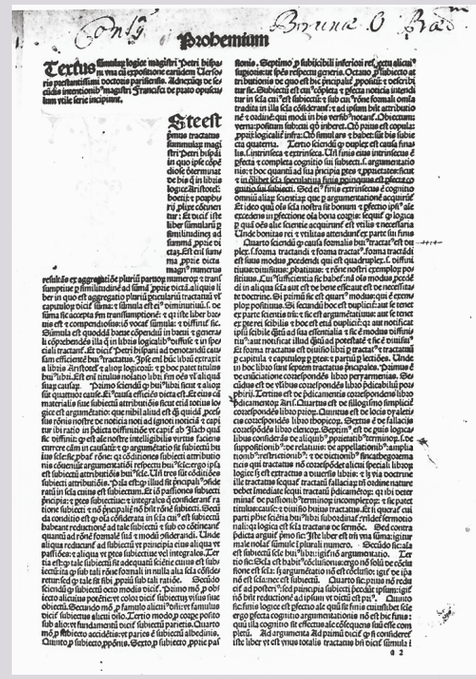
An even wider realm of activity opened for philosophy: to investigate the extent to which the underlying tenets of faith, which is to say dogma, are accessible to the human mind and the extent to which they might possibly be derived with the help of logic, that is, through reason. But philosophy was assigned another role as well, which appears most vividly in the work of Saint Peter Damian (Petrus Damiani, ca. 1007–1072): Philosophy is the handmaiden of theology (*ancilla theologiae*). One

result of the attempt to bring faith and reason into accord is the principle of the *duplex veritas*: both reason and revelation have their distinct domains of use. Representatives of this viewpoint are SIGER OF BRABANT (ca. 1235–ca. 1282) as well as his followers, which include GIORDANO BRUNO and, to a certain extent, also DESCARTES. According to this principle, theology does not have the competency to decide scientific questions, and science is incapable of explaining the truths of revelation. The truths belonging to each realm are so independent of each other that they can even contradict one another. Revealed truth, even if absurd from the point of view of reason, is still acceptable as a tenet of faith. This thought is expressed most cogently by TERTULLIAN's *credo quia absurdum* (I believe because it is contrary to reason).

We should mention here that THOMAS AQUINAS vigorously opposed this line of argument. He supported the view (Quotation 2.20) that although revealed truths cannot always be explained by reason, they cannot contradict common sense. AQUINAS appeals here to AUGUSTINE, according to whom revelation is always to be interpreted in such a way that it does not contradict human reason (Quotation 2.21). Both AQUINAS and AUGUSTINE were imagining debates with "heathen" scholars where one could not appeal to the Bible or to the authority of the Church or its Fathers because, for the faithless scholars these were not fundamental truths. AQUINAS's conception on the improvability of theological theses becomes evident from one of his imaginary discussions in which he gives the advice to the debater to start always from the position that the truths of revelation cannot be contrary to reason, so that their validity surely cannot be *refuted* by reason. In other words, he proposed a defensive posture against the arguments of the faithless: the refutation of the tenet as given by the opponent has to be refuted.

No matter which scholars of the Middle Ages we think of, for them, faith and knowledge were tightly connected, although this connection could take on various forms. Thus, one might take as a motto for medieval philosophy *Auctoritas et Ratio* (authority and reason) or, equally well a saying of ANSELM OF CANTERBURY (1033–1109), *Credo ut intelligam, intelligo ut credam* (I believe in order to understand, and I understand in order to believe).

Thus far, we have spoken only of the relationship between philosophy and theology and have seen how the medieval worldview, based on faith, made accommodation in a very natural way for the purely philosophical, and here primarily logical, investigations. But how does the search for natural laws fit into this picture? In this area as well, there were two very realistic possibilities. The first was the one chosen even by scholars in the seventeenth century; we are thinking here primarily of KEPLER. According to the faith, God created the world as an ordered, harmonious whole. One satisfies the demands of faith precisely when one searches for this order and harmony in the world, and one understands God better when one strives to recognize the beauty of all His creation. In this way, the seventeenth-century scholars could also deflect attacks that considered the pursuit of science harmful because it draws attention away from man's primary duty: the preparation of the soul for eternal salvation. With this ideological justification at hand, science was freed from direct theological goals, and with a slight exaggeration we might say that it could become an end in itself. If we read the gushing pronouncements of THOMAS AQUINAS (Plate XII) about truth, the contemplation of which being the only human activity that brings true happiness, then it is not surprising that AQUINAS held all knowledge, even black magic, to be useful. The reason he gives for this is characteristic of the period: God knows all about good and evil, and because in God nothing can be imperfect, also the knowledge of evil belongs to perfection.



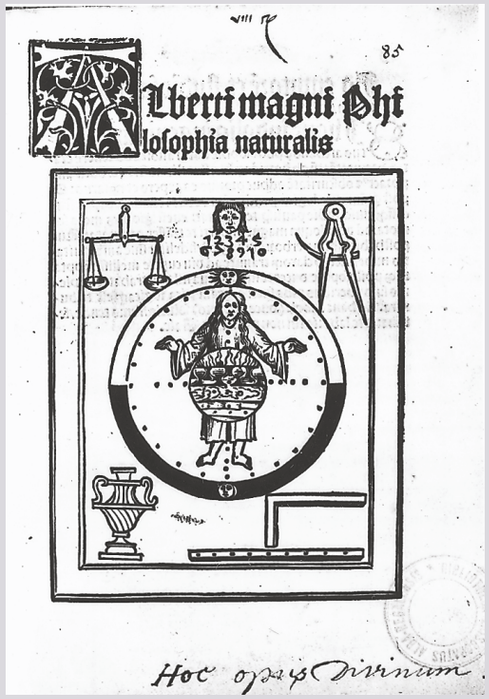
▲ **Figure 2.53** PETRUS HISPANUS: *Summulae logicales*. (Incunabula, 1487, Székesfehérvár Diocesan Library.)



▲ **Figure 2.54** ALBERTUS MAGNUS. (Painting by LUDWIG SEITZ. Rome, Santa Maria dell'Anima.)

The phoenix among teachers,
the incomparable, the prince
of philosophers, the vessel
that poured out the treasures of sacred science,
ALBERT lies here
rich in fame throughout the world,
spoken of by all; confident competitor
in the art of disputation,
greater than PLATO, hardly less than
SOLOMON. Include him, O CHRIST,
in the blessed host of your saints.

According to RHABAN LIERTZ (*Albert der Große als Naturforscher und Lehrer*, 1930), these verses appear on a wooden tablet hung on MAGNUS's grave.



▲ **Figure 2.55** Albertus Magnus: *Philosophia naturalis*. (Incunabula, 1487, Székesfehérvár Diocesan Library.)

Quotation 2.21

It is one thing to inquire which is the true history of creation, another to ask what Moses, who was so good a servant to the family of your faithful, meant those who read or heard his words to understand by them.

As for the first sort of disagreement, I wish to have no dealing with any who think things which in reality are false; and as for the second, I wish to have none with any who think that Moses wrote what was not true.

—Augustine, *Confessions* XII.23 [p. 300]

Quotation 2.22

If I am mistaken, I exist. He who does not exist clearly cannot be mistaken; and so, if I am mistaken, then, by the same token, I exist. And since, if I am mistaken, it is certain that I exist, how can I be mistaken in supposing that I exist? Since, therefore, I would have to exist even if I were mistaken, it is beyond doubt that I am not mistaken in knowing that I exist. And, consequently, neither am I mistaken in knowing that I know.

—Augustine, *The City of God against the Pagans*

However, Aquinas enunciates a condition for the use of knowledge that not only was valid for the Middle Ages, but remains valid to this day: The benefits of knowledge do not depend on what it pertains to, but rather on how it is put to use, as was propounded also by Aquinas's teacher Albertus Magnus.

The other possibility for organically fitting actual philosophical problems into a worldview determined by faith, is connected to questions of epistemology. In investigating the relationship between faith and knowledge, the questions naturally arise as to when the truths of faith are already directly given by revelation, to what extent can man gain knowledge, and how reliable is such knowledge if it is not buttressed by the revealed truths. Discussion of these questions had a long tradition among the Scholastics, going back at least to Saint Augustine, probably the most important of the Church Fathers, who posed the questions of what truth is and how it can be acquired. His ideas, as we saw earlier in Section 1.5.1, are quite modern. Beginning with doubt, he stated that all knowledge, direct or indirect, that comes from our sensory organs is uncertain. Everything is subject to doubt, and one is never free of the possibility of error. From doubt, he arrived at one of his fundamental truths (Quotation 2.22), which may be summarized in the phrase *si enim fallor sum* (I err, therefore I am), which is quite similar to the basic truth formulated by Descartes a millennium later: *cogito ergo sum* (I think, therefore I am). Augustine arrives at certainty through divine inspiration (*illuminatio Dei*). Thus we find another parallel to Descartes in the Augustinian assumption that man is born with the idea of a perfect God.

Aquinas did not adopt Augustine's point of view. In his opinion, knowledge begins with sensory perception, but it goes beyond this. In a nice parable, Aquinas compares Augustine's *illuminatio*, which alone leads to truth, with the Sun, which illuminates everything: The Sun's light is necessary for knowledge, but we must look not only at the Sun, but at the illuminated objects as well.

2.5.2 *Faith and Experience*

The daily contact the monastic orders had with secular life coupled with the reconciliation of faith and science made it possible that empirical observation could slowly work its way into the canon of favored methods for gaining knowledge. It is no coincidence that Albertus Magnus (ca. 1200–1280, Figures 2.54 and 2.55), the *doctor universalis* and teacher of Aquinas, knew nature so well because the rules of the Dominican order permitted its members to travel only on foot. When Albertus Magnus became his order's provincial superior for Germany, he visited all of the Dominican monasteries, walking all the way. This close relationship with nature is palpable in his philosophical works . In Albertus's writings, one frequently encounters the sentence *Fui et vidi experiri* (I was there and saw that it happened in this way). He professes (Quotations 2.23 and 2.24) that mankind's task is to discover what transpires in the natural world as a result of immanent causes. He therefore rejects all conclusions that contradict the evidence of the senses. He defends science against the theologians who are not scientifically minded: Theologians who belittle science are like mindless dogs that bark at the unfamiliar.

The importance of experiment and experience was emphasized by the Oxford School, above all by Grosseteste (ca. 1175–1253) and his student Roger Bacon (Figures 2.56 and 2.57). These two men laid the foundations for the English empirical tradition, which maintained its influence well into the twentieth century. Grosseteste studied the development of general concepts. He maintained that we

are able to arrive at general statements with the help of our sensory organs when we repeatedly perceive the same connection between two phenomena (Quotation 2.25). He instructs how to avoid all irrelevant factors when investigating a particular phenomenon. This is reminiscent of the ideas of Francis Bacon in the seventeenth century. Grosseteste's greatest disciple, Roger Bacon, honored by his contemporaries with the title *doctor mirabilis*, gave the resolute rallying cry that philosophy has no other task than to expound nature and the properties of objects (*tota philosophiae intentio non est nisi verum naturas et proprietates evolvere*). To the two long-acknowledged sources of knowledge, authority and reason, he added a third, namely, experience (*per auctoritatem et rationem et experientiam*). Bacon went even further. All knowledge outside of experience, that is, knowledge based solely on authority and reason, must ultimately be rooted in experience: Everything is to be proved on the basis of experience (*Oportet ergo omnia certificari per viam experientiae*).

With respect to this quotation, one should point out a danger that always arises when one attempts to interpret a quotation of any length removed from its original context. It can be said that for a proper interpretation, one has to consider not only the complete work, but also the intellectual background of the period as well as the other works of the author in question. Naturally, this danger exists for all the quotations in this book. But, let us see in this concrete example how Bacon continues his thoughts after the seemingly very modern statement just cited. Bacon distinguishes two forms of experience: external experience transmitted through the sensory organs, and internal experience, namely, divine inspiration (*divina inspiratio*). The latter is applicable not only to spiritual matters, but also with respect to material bodies and the philosophical sciences (*divinae inspirationes non solum in spiritualibus sed et in corporalibus et scientiis philosophiae*). Bacon distinguishes seven levels of internal experience or divine inspiration, of which the highest is rapture (*raptus*), which corresponds to religious ecstasy. So according to Bacon, science obtained in a state of rapture also qualifies as knowledge rooted in experience.

It was not only in England that empiricism began to develop. Here a book by the Frenchman Peter of Maricourt (Peter Peregrinus) on the properties of magnets (*Epistola de Magnete*, 1269) needs special mention, not for its actual content, which will be discussed in Section 4.4.1, but above all because of the methods that are described therein. Peter refers not only to experiments, but explicitly points out that an experimenter in the natural sciences must acquire the appropriate practical skills; for the first time, we see here stated the necessity of technical ability for scientists.

An extreme offshoot of philosophical empiricism is represented by Nicholas of Autrecourt (ca. 1299–ca. 1369). His critical analysis of knowledge obtained through the senses—reminiscent of David Hume—concludes that a supposedly recognized connection between cause and effect does not give us any guarantees that under the given conditions, the effect will necessarily occur in the future (Quotation 2.26).

Grosseteste already offers a method that should lead to the correct fundamental principles in scientific investigations: the principle of uniformity of nature, whereby objects of the same composition will behave the same way under identical circumstances; and the principle of economy, whereby all things being equal, one should prefer a proof that requires fewer questions to be answered or requires fewer assumptions and axioms to arrive at the derivation.

Here we should mention the principle of Ockham's razor. William of Ockham (also Occam, ca. 1288–ca. 1348), a restless soul, excommunicated Franciscan



◀ **Figure 2.56** Roger Bacon (ca. 1214–1294). Studied in Paris, where he later held his first lectures on Aristotle. After 1247, he studied a wide variety of subjects at Oxford and built himself a laboratory. In 1257, he entered the Franciscan order. Between 1266 and 1268 he wrote *Opus majus*, *Opus minus*, and *Opus tertium*. From 1277, he was imprisoned by his order on charges of heresy.

In general, Bacon's modernity is exaggerated. But in any case, it is correct that he recognized the importance of experimentation. He valued highly the work of Pierre de Maricourt, calling him *dominus experimentorum*. Bacon himself experimented with lenses and mirrors. Above all, he was interested in alchemy. He was the first European to provide a precise recipe for the manufacture of gunpowder (1242). He seriously investigated the problem of flight, making use of a balloon made of thin copper foil with movable wings. He built a camera obscura and actually used it for observations of a solar eclipse. Bacon appreciated the importance of mathematics for the sciences, encouraged the study of languages, and even sketched the outlines of a utopian society.

> Having laid down fundamental principles of the wisdom of the Latins so far as they are found in language, mathematics and optics, I now wish to unfold the principles of experimental science, since without experience nothing can be sufficiently known. For there are two modes of acquiring knowledge, namely, by reasoning and experience. Reasoning draws a conclusion and makes us grant the conclusion, but does not make the conclusion certain, nor does it remove doubt so that the mind may rest on the intuition of truth, unless the mind discovers it be the path of experience; since many have the arguments relating to what can be known, but because the lack experience they neglect the arguments, and neither avoid what is harmful nor follow what is good. For if a man who has never seen a fire should prove by adequate reasoning that fire burns and injures things and destroys them, his mind would not be satisfied thereby, nor would he avoid fire, until he placed his hand or some combustible substance in the fire, so that he might prove by actual experience that which reasoning taught. But when he has actual experience of combustion his mind is made certain and rests in the full light of truth. Therefore, reasoning does not suffice, but experience does.
>
> —Roger Bacon [Newman 1956, p. 1758]

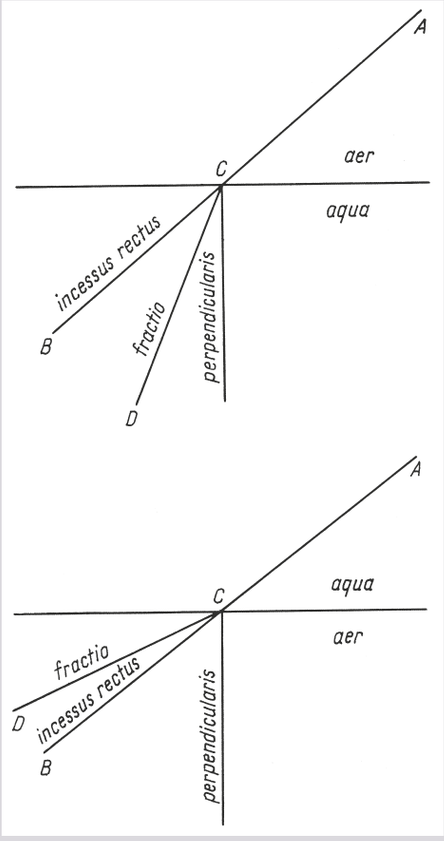
> A fifth part of experimental science concerns the fabrication of instruments of wonderfully excellent usefulness, such as machines for flying, or for moving in vehicles without animals and yet with incomparable speed, or of navigating without oarsmen more swiftly than would be thought possible through the hands of men. For these things have been done in our day, lest anyone should ridicule them or be astonished. And this part teaches how to make instruments by which incredible weights can be raised or lowered without difficulty or labor. … Flying machines can be made, and a man sitting

continued on next page

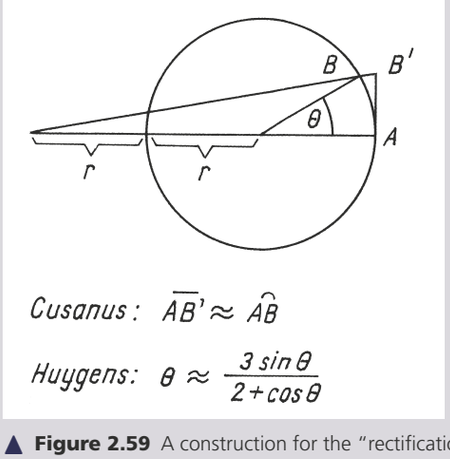
Figure 2.56 *continued*

in the middle of the machine may revolve some ingenious device by which artificial wings may beat the air in the manner of a flying bird. … Also machines can be made for walking in the sea and the rivers, even to the bottom, without danger.

—Roger Bacon, *De secretis operibus artis et naturae* [p. 1010]



▲ **Figure 2.57** Roger Bacon had a clear conception of the phenomenon of refraction, yet he was unable to formulate a quantitative law.



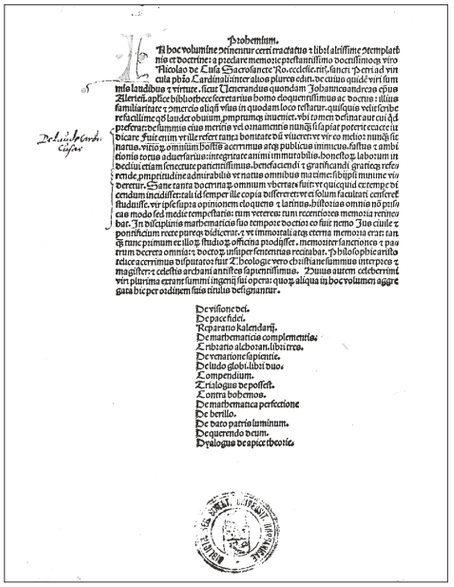
▲ **Figure 2.59** A construction for the "rectification" of the arc *AB* according to Nicholas of Cusa (Cusanus). The approximation is surprisingly good. It so happens that two centuries later, Huygens discovered and used the same approximation.

friar, supporter of nominalism, chose this as his motto for philosophical investigation: *Entia non sunt multiplicanda praeter necessitatem*; that is, entities should not be multiplied beyond necessity, or greater effort should not be expended to accomplish something that can be done with less effort.

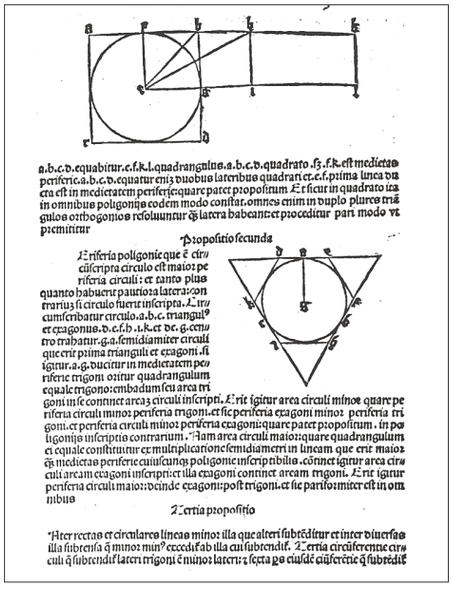
Some are of the opinion that these statements contain the germ of the minimality principle of physics (Maupertuis) or of Mach's principle of economy.

If one removes its religious and indeed mystical elements, the book *De docta ignorantia* (Quotation 2.27), by Nicholas of Cusa (1401–1464), can be seen as revolutionarily in its modernity. Here for the first time we come across the idea of an infinite universe. Neither Copernicus nor Galileo dared to endorse such a thesis. Giordano Bruno was the first follower of Cusa,; finally Descartes duly acknowledged the great cardinal as the one who first postulated an infinite universe (Figures 2.58 and 2.59).

Quotation 2.28, from the book *Idiota de staticis experimentis*, is evidence of Cusa's enthusiasm for experimentation; Quotation 2.29 reports Giordano Bruno's opinion of Cusa.


Prohemium.

Delusione dei.
De perfectis.
Reparatio kalendarij.
De mathematicis complementis.
Liber de alchoran. libri tres.
De ternario sapiente.
De ludo globi. libri duo.
Compendium.
Trialogus de possest.
Contra hobemos.
De mathematica perfectione
De beryllo.
De beato patris luminum.
De quaerendo deum.
Dyalogus de apice theoriae.

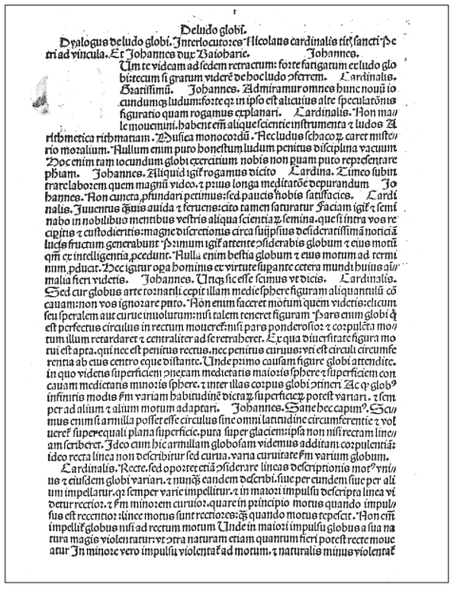

a.b.c.d. equabitur. e.f.k.l. quadrangulus. a.b.c.d. quadrato. f.g.f.k. est medietas periferie. a.b.c.d. equatur enij duobus lateribus quadrati. et. e.f. prima linea data est in medietatem periferie quare patet propositum. Et sicut in quadrato ita in omnibus poligonijs eodem modo constat. omnes enim in duplo plures triangulos orthogonios resolumuntur ex latera habeant: et proceditur pari modo vt premittur

Propositio secunda

Et riferia poligonie que e circumscripta circulo est maior periferia circuli: et tanto plus quanto habuerit pauciora latera: non tamen circumscribatur circulo. a.b.c. triangulo et exagonus. d.e.f.h.i.k. et de. g. centro trahatur. g.a. semidiameter circuli que erit prima trianguli et exagoni. si igitur. a.g. ducitur in medietatem periferie trigoni. oritur quadrangulum equale trigono: cenbadum seu area trigoni in se continet area 3 circuli inscripti. Erit igitur area circuli minor quare periferia circuli minor periferia trigoni. et sic periferia exagoni minor periferia trigoni. et periferia circuli minor periferia exagoni quare patet propositum. in poligonijs inscriptis contrarium. Nam area circuli maior: quare quadrangulum et equale constituitur ex multiplicatione semidiametri in lineam que erit maior ex medietate periferie cuiuscunq; poligonie inscriptibilis. cōnnet igitur area circuli aream exagoni inscripti: et illa exagoni continet aream trigoni. Erit igitur periferia circuli maior: deinde exagoni: post trigoni. et sic pariformiter est in omnibus

Tertia propositio

Inter rectas et circulares lineas minor illa que alteri subtenditur et inter diuersas illa subtensa q̄ minor: minor ecce dicat ab illa cui subtendit. Tertia circūferentie circuli q̄ subtendit latera trigoni ē minor: latera: 3 secta p̄ eius dē cūferētie q̄ subtendit


Deludo globi.

Dyalogus de ludo globi. Interlocutores Nicolaus cardinalis tit. sancti Petri ad vincula. Et Johannes dux Baioharie.

Johannes.

Cum te videam ad sedem tuam rectum: forte fatigatum: et ludo globi tuum si gratum videre de ludo globi certum. Cardinalis.

Baarissimi. Johannes. Admiramur omnes hunc nouū iocundumq; ludum: forte q; in ipso est aliquius alte speculationis figuratio quam rogamus explanari. Cardinalis. Non mediocriter me mouemini. habent enī aliquē scientie instrumenta: 3 ludos Arithmetica rithmanam. Musica monocordū. Hec ludus schaco 3 carent misterio moralium. Nullum enim puro honestum ludum penitus disciplina vacuum. Doceminī tam iocundum globi exercitium nobis non quam puto representare phiam. Johannes. Aliquid igit rogamus dicito Cardina. Timeo subintrare laborem quem magnū video. 3 prius longa meditatōe depurandum Johannes. Non cuncta profundari penitus: sed paucis nobis satisfacies. Cardinalis. Juuenus tuus auida 3 feruens: cito tamen saturatur. Faciam igit 3 seminabo in nobilibus mentibus vestris aliqua scientia 3 semina. que si intra vos recipiens 3 cultu dilectionis magne dilectionis circa sapientiā desideratissimā notitiā lucis fructum generabunt. Primum igit attente p̄siderabis globum 3 eius motū qui est intelligentia. pcedunt. Nulla enim bestia globum 3 eius motum ad terminum pduxit. Nec igitur opa dominis et virtute superante cetera mundi huius animalia fieri videtis. Johannes. Uraq; sic esse scimus vt dicis. Cardinalis. Sed cur globus arte tornanti. cepit illam medie sphere figuram aliquantulū cōcauam: non vos ignorare puto. Non enim faceret motum quem videtis celicum seu sperlem aut curtie unoquorum: nisi talem teneret figuram. Pars enim globi q̄ est perfectus circulus in rectum moueret: nisi para ponderosior 3 corpulētior motum illum retardaret: 3 centraliter ad se retraheret. Et qua diuersitate figura motui est apta. qui nec est penitus rectus. nec penitus curuus: vti est circuli circumferentia ab eius centro eque distante. Unde primo causam figure globi attendite. in quo videtis superficiem sue eam mediocritie maioris sphere: 3 superficiem concauam mediocritas minoris sphere. 3 inter illas corpus globi positum. Sic q̄ globus infinitis modis fm variam habitudinē dicra; superficiaz; potest variari. 3 tem per ad alium 3 alium motum adaptari. Johannes. Sane hec capim⁹. Sacramus enim si armilla posset esse circulus sine omni latitudine circumferentie 3 voluere: super equali plana superficie. pura super glaciem: ipsa non nisi rectam lineam scriberet. Ideo cum hic armillam globosam videmus additam: copulentia: ideo recta linea non describitur sed curua. varia curuitate fm varium globum.

Cardinalis. Recte. sed apponere etā p̄siderare lineas descriptionis mathematicas: vnius 3 eiusdem globi variari. 3 nunq̄ eandem describi. siue per eundem siue per alium impellatur. q; semper varie impellitur. 3 in maiori impulsu descripta linea velocior erit. 3 fm minorem curuitatem. quare in principio motus quando impulsus est rectus: hinc motus sunt rectiores. q̄ quando motus repescit. Non est impellit globus nisi ad rectum motum. Unde in maiori impulsu globus a sua natura magis violentatur vt extra naturam etiam quantum fieri potest recte moueatur. In minore vero impulsu violentat ad motum. 3 naturalis minus violentat

◀ ▲ **Figure 2.58** Facsimile pages from Nicholas of Cusa's book. University of Budapest library.

2.6 The Renaissance and Physics

2.6.1 *Art, Philology, and Science*

It is natural that in concert with historians one regards the Renaissance as one of the most brilliant epochs in human history. During this time, man, freed from the chains of dogma, turned his eyes from the heavens earthward, as his concerns shifted from life in the great beyond to life in the here and now. In place of the relationship between man and God, the concern was now with the relationships between man and man and between man and nature, and the world became anthropocentric, humanized.

In a good many modern works on the history of science, however, the Renaissance is accused of intellectual servility and an almost superstitious respect for antiquity. It is frequently added that the height of the Renaissance, that is, the half-century beginning around 1500, the age of Leonardo, Michelangelo, Raphael, and Dürer, could well deserve the title of the most barren 50 years in the history of science (Quotation 2.30).

In what follows it shall not be our intent to try to reconcile these two viewpoints; nor shall we take the easy way out by asserting that the truth must lie somewhere in the middle. In this book, we show that both points of view, sharp as they are, are valid despite the apparent contradiction. Of course, we restrict ourselves to the issues of the history of science—more precisely, to the history of mathematics and physics. In the case of the arts, we accept all the superlatives without reservation.

From the viewpoint of the history of science, the negative features of the Renaissance can be summed up as follows:

1. Human and artistic ideals were sought exclusively in antiquity. This worship of antiquity necessarily led, both in the arts and the sciences, to a rejection of the immediate past, which was scorned as Gothic barbarism.
2. In all respects, this age favored dynamic personalities. The greatest intellectual energies were absorbed by the arts, exploration, and the Reformation.
3. In the late Middle Ages, the hubs of scientific life were at Paris and Oxford. Therefore, the devastation brought about by the Hundred Years' War between England and France presented an objective hindrance to further advancement.
4. Due to the respect granted to ancient texts, philology assumed an exaggerated importance. This led, on one hand, to a draining off of intellectual energy and, on the other, to a false ordering of the importance of individual sciences. Scientific work, even that of the natural sciences and above all that of medicine, exhausted itself in an endless search for additional ancient texts or in an ever more precise interpretation of known works (Table 2.3).

In contrast, the following factors may be cited as positive and progressive:

1. As follows from the preceding point, the activities of the philologists indirectly held things back. However, their work had some direct advantages as well, because as a result of their efforts scholars involved in the study of the natural sciences had access to all the ancient works in more reliable and faithful translations.
2. Art and artistic creation were directly or indirectly joined to the sciences and advanced their development. A direct linkage existed when scientific knowledge was required in the processes of artistic creation. Thus, the design and

Quotation 2.23

Since reason, the source of all knowledge, belongs to the soul, since even what is holy rests on the soul's testimony, the laws are innate, since the soul is the form and being of the body, so the knowledge of the soul contributes not only to knowledge of all natural things, but it also promotes the sacred science by showing the noblest object of divine reason. *Therefore, knowledge of the soul is the greatest of the natural sciences*.

—Albertus Magnus, *De anima*, III, 3

Quotation 2.24

The sublime God rules over the natural world and directs it through natural causes, and these are what we seek here [in the book of the natural science], since we cannot so easily discover the divine causes, because they are not so obvious. (*Met*. p. 118) ... We do not have to investigate in nature how God, the creator, according to his free will, uses his creatures to bring about miracles, whereby he demonstrates his omnipotence, but even more, what can occur in nature according to natural causes in a natural way.

—Albertus Magnus, *Lib. de coelo et mundo*

Quotation 2.25

This therefore is the way by which the abstracted universal is reached from singulars through the help of the senses. ... For when the senses several times observe two regular occurrences, of which one is the cause of the other or is related to it in some other way, and they do not see the connection between them, as, for example, when someone frequently notices that the eating of scammony [plant] happens to be accompanied by the discharge of red bile, and does not see that it is the scammony that attracts and withdraws the red bile; then, from the constant observation of these two observable things, he begins to form a third unobservable thing, namely, that scammony is the *cause* that withdraws red bile. And from this perception repeated again and again and stored in the memory, and from the sensory knowledge from which the perception is built up, the functioning of the reasoning begins. The functioning reason therefore begins to wonder and to consider whether things really are as the sensible recollection says, and these two lead the reason to the experiment, namely, that he should administer scammony after all other causes purging red bile have been isolated and excluded. When he has administered scammony many times with the sure exclusion of all other things that withdraw

continued on next page

▶ **Table 2.3** Important scientific and philosophical works in the library of Matthias Corvinus.

Aristotle: *Libri Physicorum VIII*
Aristotle: *Opera cum commentariis Averrois (Organon. Physica. De caelo et mundo. De generatione et corruptione …) Averroes: De Substancia orbis*
Ptolemaeus Claudius: *Geographica*
Ptolemaeus Claudius: *Magnae compositionis libri (seu Almagest)*
Strabo: *Geographia*
Seneca: *Varia opera philosophica (…De septem liberalibus artibus. De questionibus naturalibus)*
Lucretius Carus: *De rerum natura Libri IV*
Plinius Secundus: *Historiae naturalis Libri XXXVII*
Beda Venerabilis: *De natura rerum*
Thomas Aquinas: *Commentaria li Librum de coelo et mundo Aristotelis*
Regiomontanius: *Epitome Almagesti*
Leon Battista Alberti: *De re aedificatoria*
Ficinus Marsilius: *Epistolarum Libri III et IV*

Quotation 2.25, ***continued***

red bile, then there is formed in the reason this universal, namely, that all scammony of its nature withdraws red bile, and this is the way in which it comes from sensation to a universal experimental principle.

—Robert Grosseteste, Commentary on the *Posterior Analytics* of Aristotle, Book I, Chapter 14 [Crombie 1952, pp. 221–222]

Quotation 2.26

If anyone maintain that what occurs is the natural consequence of something else, I ask him, what do you mean by natural consequence when you say that what has in the past up to the present day led to the same result will in the future lead to the same result whenever it is applied. However, since we do not know the premises, we cannot assert that something brought about in the past arbitrarily often will do the same in the future. And even more: If we assume certain conditions as the causes for particular effects, then we may not state with certainty that when these same conditions are satisfied, the expected effect will take place.

—Nicholas of Autrecourt [Weinberg 1948, p. 31]

Quotation 2.27

The world has no circumference, because it if had a center and a circumference, and thus had a beginning and end in itself, the world would be limited in respect to something else, and outside the world there would be something other, and space, things that are wholly lacking in truth. Since, therefore, it is impossible to enclose the world between a corporeal centrum and a circumference ...

... though the world is not infinite, yet it cannot be conceived as finite, since it has no limits between which it is confined. The earth, therefore, which cannot be the center, cannot be lacking in all motion;

... But we cannot discover motion unless it be by comparison with some fixed ... poles or the centers and assuming them in our measurement of the motions [as being at rest]

continued on next page

construction of the enormous cathedrals, above all their domes, would have been unthinkable without knowledge of statics; and in the fine arts, the artists needed the knowledge of anatomy and the laws of perspective. The freely soaring human spirit with its indirect effects, with its choice of themes, freed from its former bonds, has inspired all areas of culture, including the sciences.

3. The dynamism of Renaissance man manifested itself most clearly in the geographical discoveries that expanded the boundaries of the known world. By breaking the limits, by surpassing the accomplishments of the ancients in this field, the man of the Renaissance was finally able to break free of the voluntary conviction that the ancient world was intellectually superior and take the first steps toward complete intellectual independence.
4. While the Reformation did not directly free the human spirit from the tyranny of authority, for the Bible still held its place as the highest authority, the ferment of the Reformation set wide-ranging intellectual processes in motion, which eventually lead to mankind awakening to the recognition that it was possible to subject matters of faith, including questions of authority, to rational discussion. The success of the Reformation proved that, in this realm, authorities other than the Catholic Church, which previously had been revered as the sole repository of truth, could also hold sway.

Before we turn to the discoveries made in the domains of mathematics and physics during the Renaissance, we need to say something about the relationship between the humanists and the natural scientists as evidence of the negative influence of the Renaissance. Not only did the humanists fail to make any contribution to the development of the sciences, they in fact belittled and ridiculed the scientists themselves—at times perhaps deservedly (Quotation 2.31). Above all, they criticized their Latin, condemning it as barbaric, strident, coarse, and uneducated. The humanists did not attempt to refute scientific findings, instead arguing that in works on the natural sciences one can find sentence constructions that do not appear in Cicero. Thus, in his 1514 work *De caelo et mundo*, Agostino Nifo ridiculed a group of scientists called *Calculatores* by referring to them as *Captiunculatores* (equivocators). He gives Albert of Saxony the belittling names Albertus Parvus and Albertutius,

and another humanist calls Heytesbury the greatest of all sophists, and opposes his arguments by appeal to the "pure and noble voice of Aristotle."

2.6.2 *Progress in Mechanics*

It should not be left unmentioned that for all their adoration of antiquity, the natural scientists of the Renaissance did not forget their medieval predecessors. Thus, Cardano (1501–1576), a notable mathematician, gave a list of scholars in order of importance, beginning with the greatest: Archimedes, Aristotle, Euclid, Duns Scotus, the calculator Swineshead, Apollonius, Archytas, etc. This list is noteworthy in that it was only around this time that Europe had become familiar with the entire lifework of Archimedes and was thus receptive to his work and truly capable of discovering his significance (Figure 2.60, Plate XIII).

In mechanics, we find very little progress as compared to the achievements of the fourteenth century and, instead, much recourse to the original ideas of Aristotle. In a number of cases, the Renaissance scholars did a great service in part by preserving some ideas for posterity and in part by providing a clearer formulation of them. Particular mention is due the Spanish Jesuit Domingo de Soto (1495–1560), who was the confessor of the Holy Roman emperor Charles V.

In Domingo de Soto, we find for the first time a connection between continuously changing motion and free fall. As we saw earlier, each of these was a topic of considerable discussion in the Middle Ages, but a clear recognition that the *uniformiter difformis* motion is actually realized in nature under free fall was lacking (Quotation 2.32). de Soto does not claim that this discovery was his own. Indeed, the opinion is often advanced that the scholars at Oxford and Paris schools saw this connection as so obvious that they did not consider it worthwhile to explain it.

Tartaglia (ca. 1500–1557), known primarily for his mathematical work, and his contemporary Cardano, likewise a mathematician, were also involved in problems related to mechanics. From Tartaglia, we have the works *Nova scientia* (1537) and *Quesiti et inventioni diversi* (1546); and from Cardano, *De subtilitate* (1551) and *Opus novum* (1570). Tartaglia worked on the problem of projectile motion: he divided the trajectories of the object into three segments (Figure 2.61). The first segment is a straight line, followed by a circular arc, and finally, the path ends in vertical fall. This last segment corresponds to Aristotelian natural motion, and the first two were conceived at times as homogeneous forced motion, at times as forced and mixed motion.

Giambattista Benedetti (1530–1590) sharply criticized the Aristotelian conception of free fall, particularly the notion that a body in a vacuum should fall with infinite velocity. It was, of course, well known that Aristotle used this inference to support his telling argument that a vacuum could not exist. Benedetti revised the law of motion, subtracting the resistance from the driving force instead of dividing the force by the resistance. As we saw earlier, Bradwardine had already formulated such a law of motion but then he himself rejected it. Based on his new law, Benedetti argued that, due to the absence of a resistive force in a vacuum, the motion in free fall would be determined by gravitation alone. Building on this, he proved that objects composed of substances of the same density but with arbitrary mass must fall at the same velocity in a vacuum. This conclusion should hold approximately for free fall in air; but it stands in pointed contradiction to

Quotation 2.27, ***continued***

... But for us it is clear that this earth really moves, though it does not appear to us to do so, because we do not apprehend motion except by a certain comparison with something fixed. Thus is a man in a boat, in the middle of a stream, did not know that the water was flowing and did not see the bank, how would he apprehend that the boat was moving? Accordingly, as it will always seem to the observer, whether he be on the earth, or on the sun or on another star, that he is in the quasi-motionless center and that all the other [things] are in motion, he will certainly determine the poles [of this motion] in relation to himself; and these poles will be different for the observer for those on the moon and Mars, and so on for the rest. Thus, the fabric of the world (*machina mundi*) will *quasi* have its center everywhere and its circumference nowhere, because the circumference and the center are God, who is everywhere and nowhere.

... Still it does not seem that, according to the order of nature, there could be a more noble or more perfect nature than the intellectual nature which dwells here on this earth as in its region, even if there are in the other stars inhabitants belonging to another genus: man indeed does not desire another nature, but only the perfection of his own.

—Nicholas of Cusa, *De docta ignorantia (Learned Ignorance)* [Koyre 1957, pp. 11–19]

Quotation 2.28

Layman: ... For example, if a piece of wood is weighed and if after it is burned its ashes are weighed, then there is known how much water was in the wood. For only water and earth have heavy weight. Similarly, from the differences of weight of a piece of wood in air, in water, and in oil, we know how much heavier or lighter is the water that is in the wood than is pure water from a spring; [we know], likewise, how much air [was in the wood]. Similarly, [we know] from the difference of weights of the ashes how much fire [was in the wood]. And in these ways the elements are ascertained by a quite accurate surmise, although preciseness is always unattainable. And just as has been stated about wood, so [something similar could be said] about herbs, meats, and other objects.

Orator: No pure element is said to be observable. [So] how do we experience an element by means of a weight-scale?

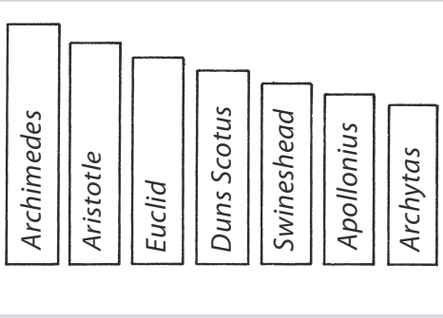
Layman: Suppose that after a hundred pound of earth have been placed in a cask someone were to cull the one hundred pounds successively of herbs or of seeds that were strewn amidst that earth and that

continued on next page

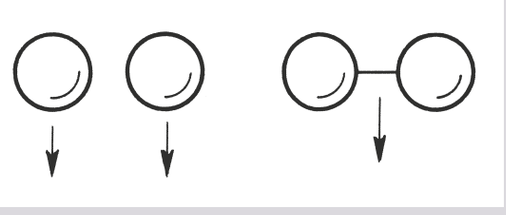
Quotation 2.28, ***continued***

were previously weighed [together with that earth]. And suppose that [thereafter] he were to weigh the earth again. He would find that the earth had diminished in weight only a little. From this fact he would know that the collected herbs have weight mainly from water. It follows that thickened waters in the earth have attracted earthenness and that by the activity of the sun the waters have condensed into an herb. If these herbs were burned, wouldn't you—through surmise on the basis of the difference of all the weights—reach [a conclusion about] however much earth you would find over and above one hundred pounds? And isn't it evident that water produced this [additional] earth? Elements are, in part, transformed one into another. For example, in the case of a plate-of-glass placed in the snow, we experience that air on the glass is condensed into water, which we find as a fluid on the glass.

—Nicholas of Cusa, *Idiota de Staticis Experimentis* [pp. 614–615]



▲ **Figure 2.60** The hierarchy given by Cardano: Archimedes precedes Aristotle.



▲ **Figure 2.62** A simple thought experiment by Benedetti. If two bodies are joined by a massless rod, then nothing should change in the properties of their free fall. Therefore, the doubled mass will fall with the same velocity as the individual masses.

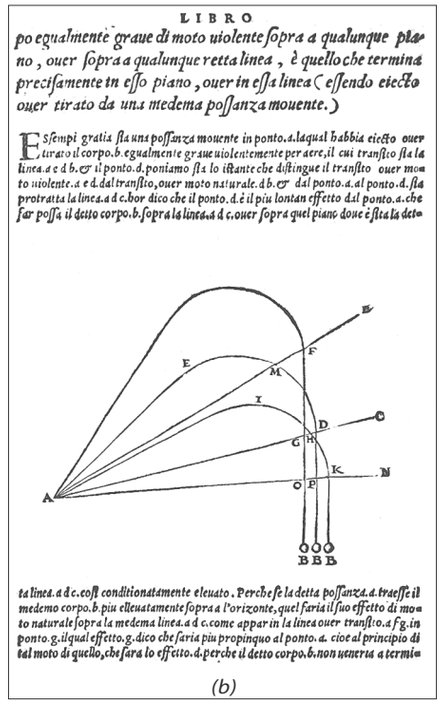
the Aristotelian formulation. Let us—thus, Benedetti—drop two spheres of the same weight and density at the same time (Figure 2.62). They then will fall alongside each other at the same velocity. But if now the spheres are joined by a thin rod of negligible mass, then nothing essential has been altered in the state of motion, although the falling body now has double the mass of the individual spheres. However, according to Aristotle, the velocity of an object of twice the weight should fall at a velocity twice as great.

As Benedetti continues his argument, he draws incorrect conclusions: according to him, the velocity should be proportional to the density.

Benedetti's argument, according to which bodies of the same density but different masses should fall at the same velocity, or—as we would formulate it today—with the same acceleration, seems quite compelling even today—indeed, almost logically necessary. However, if we look more deeply into the physical content of this assertion, we notice that—to be sure, in the view of modern physics—the equivalence of inertial and gravitational mass is implicitly assumed. This equivalence, however, is not a logical necessity but a physical fact. In principle, it is not unthinkable that two bodies of the same mass and density could fall with different accelerations. With such an experimental result, one would have later discovered in physics that the bodies had different "gravitational charges" (see Section 5.2.8).



(a)



(b)

▶ **Figure 2.61** **(a)** Euclid opens up for us all the sciences. At the entrance to philosophy, Plato is standing guard and warns, "Let no one ignorant of geometry enter here" (*Huc nemo geometriae expers ingrediatur*)

(b) According to Tartaglia, the path of a projectile consists of three segments: a slanted line, a circular arc, and finally a vertical line.

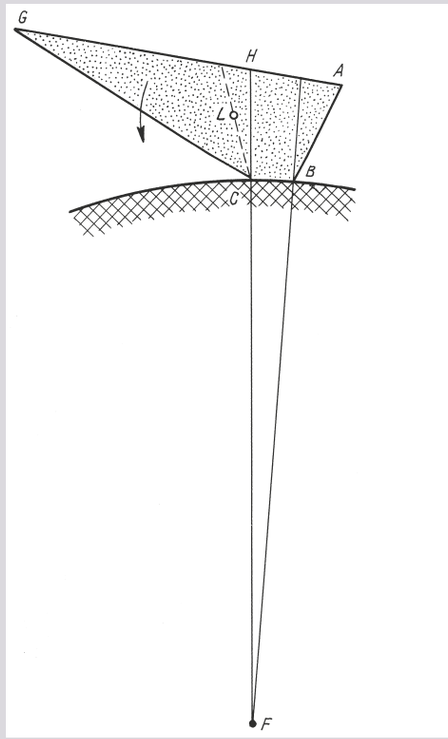
Tartaglia. *Nova scientia*, 1550. Library of the Hungarian Academy of Sciences.

The works of Villalpando (1552–1608) on statics deserve particular mention. He was the first to investigate the stability of an object resting on the ground, or—to formulate it more simply—the question of when objects tip over (Figure 2.63). A body of arbitrarily complex form is stable with respect to tipping over if the vertical through the center of mass passes through the horizontal supporting surface of the body. From the figure, it is clear that Villalpando, with excessive exactitude, even took into account the curvature of Earth. Here again we have a problem and a solution that are today a traditional part of the school curriculum.

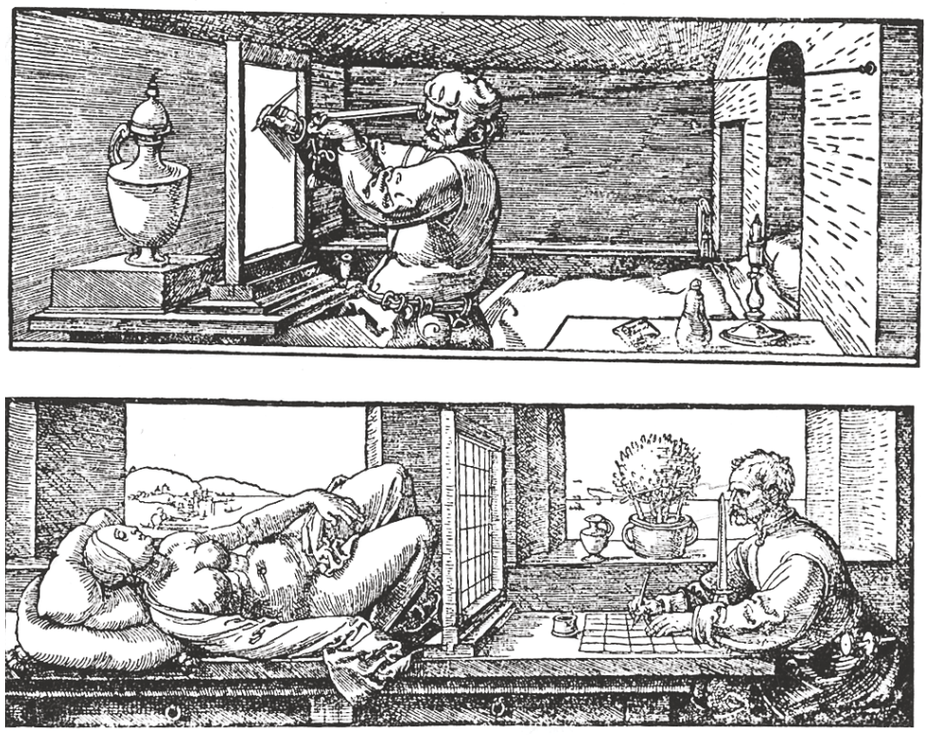
2.6.3 *The Science of Artists*

We have already mentioned that many Renaissance artists directly contributed to the advancement of science. Thus with his book *De Prospettiva Pingendi*, Piero della Francesca (ca. 1420–1492) earned the title of the greatest geometer of his day. Brunelleschi (1377–1446) was an excellent mathematician; according to Vasari, he began painting only in order to find applications for geometry. The work of Leon Battista Alberti (1404–1472) must be given special emphasis. In his works *Della Pittura* (1435) and *Ludi mathematici* (1450), he describes the laws of perspective and the role of binocular vision in depth perception. He draws a clear and simple connection between a painted image and reality, and in this way he also formulates the associated geometric problems. For the creation of an image, we should imagine a ray of light that leads from the painter's eye to the object that is to be depicted. The point where the ray crosses a transparent plane located in front of the painter corresponds to the point in the image that will represent the given point on the object.

This rather primitive-seeming method of image construction leads directly to the following question: If the painter, while maintaining the image plane, views the object from another vantage point, then he, of course, will obtain a different picture. But both pictures represent the same reality, so they must stand in a relationship determined by the laws of geometry. What is common to both pictures from the viewpoint of geometry will be studied via a new branch of mathematics: projective geometry, which, however, was not established until quite a bit later, with the work of Desargues and Monge. Dürer, who was well versed in mathematics, gives construction procedures similar to those of Alberti (Figure 2.64). His 1525 treatise *Underweysung der Messung mit dem Zirckel und Richtscheyt, in Linien, Ebnen und ganzen Corporen* (Instruction in measuring with compass and straightedge in lines, planes, and solids) is devoted to these kinds of questions.



▲ **Figure 2.63** A drawing from the book *Apparatus Urbis et Templi Hierosolyminati* (1653) by Villalpando on the stability of bodies with regard to toppling. The perpendicular from the center of gravity must lie within the supported region.



◀ **Figure 2.64** Image construction after Dürer.

Quotation 2.29, ***continued***

You will not believe (O most learned auditors) that I am falsely courting your favor if you will only inspect more inwardly your riches, you who are perceptive beyond others. Indeed, ever since that time when sovereignty first devolved upon your countrymen, the majority of the arts and known human capacities have been found among you, with whom none among foreign nations would be found equal. Who at that time was comparable to Albertus Magnus Suevus? Did he not stand forth as superior in many respects to that prince Aristotle, to whom he was devoted to a degree that was (insofar as he wore a hood) most unseemly, given the condition of his time? Good God, and if he should be compared to that man from Cusa, who, to the extent that he is greater, is accessible to fewer?

continued on next page

Quotation 2.29, ***continued***

Provided it were not disturbing to see him in old fashioned clothing, I would not avow this man's capacity to be equal to that of Pythagoras, but rather would I declare it far superior to that of Pythagoras. Do you even consider Copernicus such that he was not merely a mathematician but also (which is a miracle) a physicist, as it were in passing? He is revealed to have understood more in two chapters than Aristotle and all the Peripatetics in their entire contemplation of nature.

—Giordano Bruno, Farewell address to the Academy of Wittenberg, 1588, translated by Paul Franz

With Dürer we will also find two new geometric concepts: the *spatial curve* such as the *helix*, and the representation of solid objects via projection onto two mutually perpendicular planes, which would later become a fundamental technique in representational geometry.

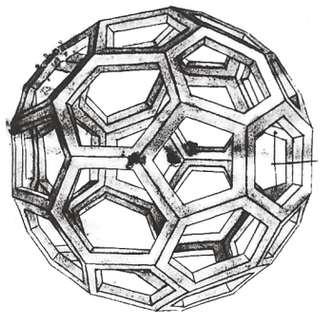
Navigation on the high seas required maps of ever greater precision. In the second half of the fifteenth century, the primary centers of map production were Antwerp and Amsterdam. The best-known cartographer was Gheert Cremer (1512–1594), who is better known by his Latinized name Mercator, a name that has been preserved in the Mercator projection.

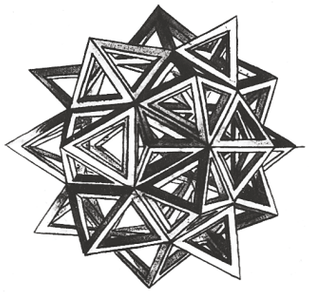
2.6.4 *Leonardo da Vinci*

The most famous scholar of the Renaissance was Leonardo da Vinci (1452–1519). His name is frequently invoked as the usual counterexample to the barrenness of the Renaissance in science (Figures 2.65 and 2.66). In a book on mechanics written at the end of the eighteenth century, Lagrange summarized the history of mechanics, in which Archimedes is followed directly by Galileo, as though in the intervening 2000 years nothing had been contributed. However, in the nineteenth century, the complete diaries of Leonardo became known, and Ernst



▶ **Figure 2.65** Self-portrait by Leonardo da Vinci.





◀ **Figure 2.66** Polyhedra; drawings by Leonardo. Such figures can be found today in books on solid-state physics. Leonardo's polyhedra are quite similar to the Brillouin zones, because they both illustrate symmetry relationships.

In fact, the top polyhedron is one of the semiregular Archimedean polyhedral mentioned earlier. The carbon atoms in the C_{60} fullerene molecule are arranged in such a manner (see Figure 1.46).

Mach, in his 1883 book *Die Mechanik in ihrer Entwicklung* (The development of mechanics), singles out Archimedes and Leonardo as the two prominent figures marking the two ends of the hiatus. This already pushed the beginning of the development of mechanics back by 100 years. Today, with our wealth of information on science in the Middle Ages, this starting point must be moved back by yet another 200 years. We have already mentioned (Section 2.4) that Pierre Duhem performed an invaluable service with his unraveling of the findings of medieval science, which one could almost characterize as an archeological excavation. What is of interest here is that it was his search for the sources upon which Leonardo had drawn that led Duhem to medieval mechanics. He was guided in his search by the hunch that no one working essentially in a vacuum, without any predecessors and with only the power of his own genius, could have raised so many questions and been able to answer them more or less correctly.

It is relatively easy to evaluate Leonardo's importance in the history of science by his influence on posterity. Although, in his diaries, Leonardo often writes about a work in progress, he generally succeeded in producing only disconnected sketches, and no work of his was published in his lifetime. Thus it is that many of his ideas, written down here and there, did not become known until the middle of the nineteenth century, and so there can be no talk of any direct influence. Regarding Leonardo's originality and genius as a scholar, opinions diverge widely, ranging from "the most brilliant scholar of the era" to "half-educated amateur." Duhem, who perhaps can be accused of partiality for the Middle Ages, fails to find in Leonardo's notes on mechanics a single significant idea not taken from some medieval scholar.

In any case, the accusation is justified that Leonardo did not recognize the necessity for an organizing principle for science, whereby connections could be made between apparently unrelated phenomena. It is not without good reason that today's physicists say that a bad theory is better than no theory at all. It is, therefore, not surprising that the value of the brilliant fragments in Leonardo's notebooks is called into question (Quotation 2.33), because an assertion is scientifically relevant only when the words in which it is written are fixed as scientific concepts, which is again possible only within the framework of a theory. For example, there is no doubt that it is in Leonardo's diary that the coupling of uniformly changing motion with free fall appears for the first time (Quotation 2.34)—an association that we mentioned earlier when discussing the work of Domingo de Soto. It is unlikely that de Soto took the idea from Leonardo. But Leonardo's inventiveness in this regard can be attacked on the basis that it was his habit to assume a simple proportional relationship between quantities. It so happens that a number of relationships in nature are linear, that is, proportional, although most of them are not. Leonardo investigated countless phenomena, always using the simple assumption of linearity, so naturally he was able to arrive at the correct results in some cases.

However, if one approaches Leonardo's entire lifework with the biases of a layman, then one is immediately struck by its breadth and by the variety of problems, topics, and questions that he investigated. Even if we consider Leonardo merely as the compiler of the knowledge of mechanics of his time, despite of the lack of organization in his work, we would still have to grant him a prominent place among his contemporaries. His quantitative questions and his emphasis on the importance of experience and experiment are entirely modern characteristics. In addition, a host of ideas—arguably of greater or lesser consequence—originated with him, such as the study of motion with the help of an inclined plane (Figure 2.67).

Quotation 2.30

Historians of letters had meanwhile created the myth of the "Renaissance." According to this myth, in the Middle Ages man hibernated beneath a pall of scholastic repetition, borrowed from Aristotle and enforced by the Church; the Renaissance, casting this all aside, opened its eyes and discovered man and the world by personal sensation. While for the arts this theory may be tenable as an expression of taste, in science other than anatomy it is not: The "early" Renaissance, 1450–1500, stands in the front rank of competition for the most sterile half-century of Western mathematics and physics, and the only exact science from the "late" or "high" Renaissance, 1500–1550, is algebra, which grew, not from open-eyed wonder at the world, but rather from bookish study of Arabian mathematics. Where is the empirical science which should have crowned the "rebirth" of knowledge?

—C. Truesdell, *Essays in the History of Mechanics* [p. 25]

Quotation 2.31

One sees raised a vast edifice of assertions and contradictory propositions touching uniform motion, non-uniform motion, motion uniformly varied and non-uniformly varied. There are innumerable people who discuss, without getting anywhere, that which cannot present itself in nature.

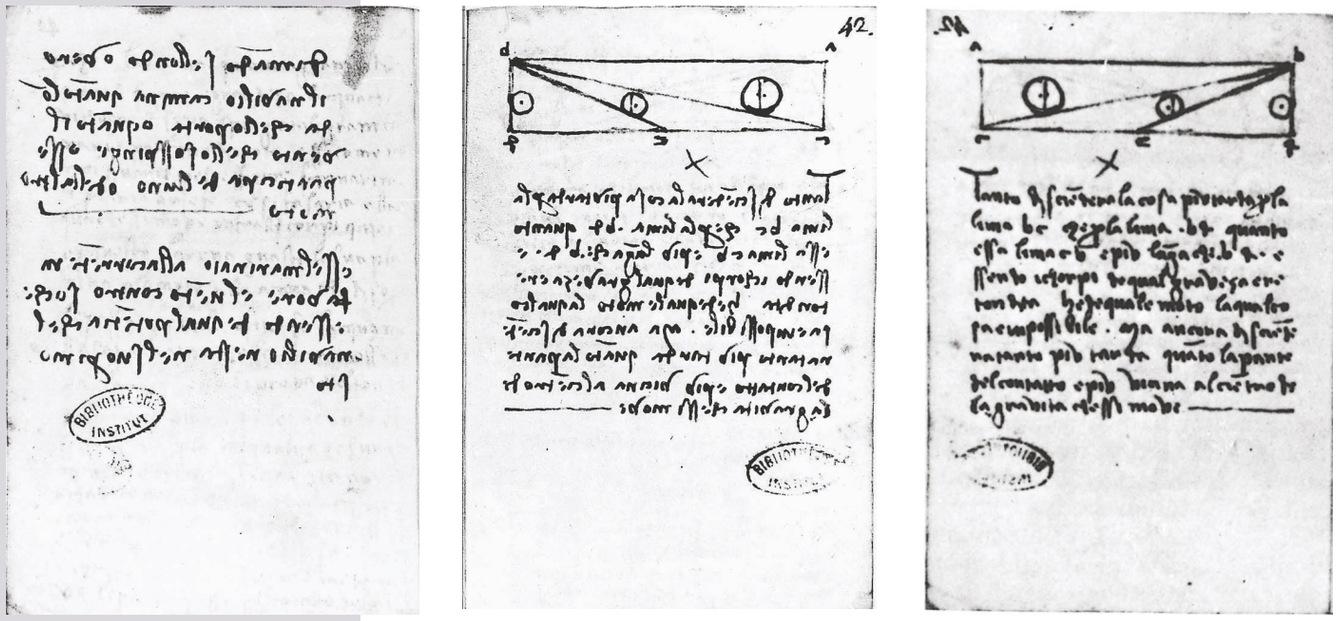
—Jean Luis Vives [Mason 1953, p. 94]

Quotation 2.32

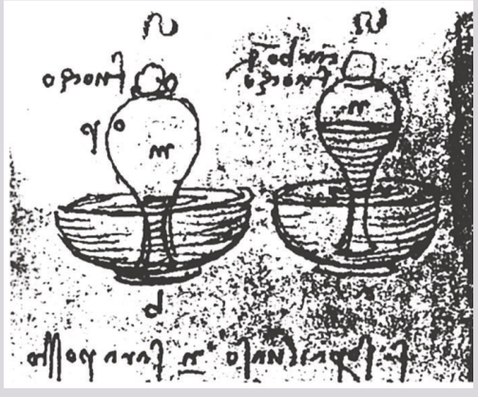
Movement uniformly nonuniform *(uniformiter difformis)* as to time is nonuniform in such a manner that if it is divided according to time (i.e., according to before and after), the middle point of any part at all exceeds [in velocity] the least velocity of that part by the same proportion that the mean is exceeded by the greatest velocity [of that part]. This species of movement belongs properly to things which are moved naturally and to projectiles. For when a body falls through a uniform medium, it is moved more quickly in the end than in the beginning. On the other hand, the movement of projectiles [upward] is less quick in the end than in the beginning. And so the first is uniformly increased, while the second is uniformly decreased.

... [Consider] a body falling through an hour which accelerates uniformly from zero degree of velocity to a velocity of eight. Then it would traverse ... just as much space as another body moving uniformly through the hour with a speed of four.

—Domingo de Soto, *Questiones super octo libros Physicorum Aristotelis* [Clagett 1961, pp. 555–556]



▶ **Figure 2.67** We do not know whether LEONARDO arrived by experimentation or by logical deduction at this result that he could not have learned from earlier medieval teaching: the travel time of a body moving on inclined planes of the same height but different angles of inclination is proportional to the length of the planes. (LEONARDO wrote his diaries in mirror writing.)



▲ **Figure 2.68** LEONARDO discusses—essentially with himself—the following problem: How does water reach a mountaintop, where it appears as the source of a stream? Or more generally, how can water climb from a lower to a higher level? The setup pictured here shows an idea for a possible experiment (*Codex Leicester*).

Finally, if we add to all of this the fact that for LEONARDO the goals of knowledge, the goals of the search for natural laws, were practical applications and that he himself made great efforts in this direction (Figures 2.68, 2.69, and 2.70), we are justified in raking him among the greats in the history of science.

2.6.5 *The Professional Astronomers Take the Stage*

In the development and in the perfection of research methods in physics, and also in the discovery of concrete physical laws, the scope and importance of astronomy were without parallel up to the time of NEWTON. The reason is clear: As we have frequently pointed out, the conditions under which the celestial bodies move would be difficult to reproduce on Earth, even only approximately. In other words, only in the heavens does reality coincide with the abstract model.

It may be stated without exaggeration that at the end of the Middle Ages, astronomy was the only area of the natural sciences that in effect followed today's norms of research. The starting point was an axiom, namely that of uniform circular motion of the celestial bodies; based on this axiom, a detailed (geometric) model was developed for each planet, which made possible the prediction of its future location in the heavens and the determination of its location at any time in the past. Measuring instruments were available that allowed theoretical predictions to be verified. All of this was primarily in service of a practical requirement, the preparation of calendars, notably to the determination of holidays. And even more important was providing astrologers with data for the casting of horoscopes. These were the functions of astronomy inherited from the past. A new challenge to astronomers came from seafarers: how could one determine the position of a ship on the high seas? In the development of navigation, one can trace how the spirit of the Renaissance promoted progress but at the same time impeded the natural sciences.

The astrologers were also aware of the necessity of studying the ancient astronomical works in better translations—or ideally, in the Greek original; the most important of these was the *Almagest* of PTOLEMY. They considered their quest for better Greek texts itself an advancement of science. At the same time, however,

they noticed in their daily practice that even when using these more precise texts, they still were not able to arrive at a more precise agreement between their predictions and their experimental observations. Thus, for example, for the year 1432, a conjunction of all the planets in the sign of Libra was foreseen, which should have resulted in a global catastrophe. Not only did the world not come to an end, but the predicted conjunction also failed to take place.

It is thus not surprising that the great breakthrough of the next century was helped the most by the *professional astronomers* of the fifteenth century, who can be seen as the immediate predecessors of Copernicus. We are thinking here primarily of two characteristic figures of this epoch: Georg von Peuerbach (Purbachius, 1423–1461) and Regiomontanus (Johannes Müller, of Königsberg, 1436–1476).

Peuerbach was a typical Renaissance man; he studied at the University of Vienna, where later he taught Roman literature and where his Latin poetry brought him renown. He monitored solar and lunar eclipses with the aid of the Alfonsine tables and compared the results with his own observations. He was also involved in the construction and perfection of astronomical instruments. However, Peuerbach's greatest influence was as a teacher. Based on the Ptolemaic worldview, he developed a clear, simplified theory of planetary motion. His textbook, written in Latin, was the most important textbook of the period, going through 56 printings and translated into Italian, Spanish, French, and Hebrew (Figures 2.71 and 2.72). In his book, he drew attention to the fact that the movements of the planets in their epicycles are somehow related to the motion of the Sun. Cardinal Bessarion (1403–1472), who had fled Byzantium and devoted his life to the revival of ancient science, urged Peuerbach to produce a new translation of Ptolemy's great work; however, Peuerbach's untimely death prevented him from finishing this task.

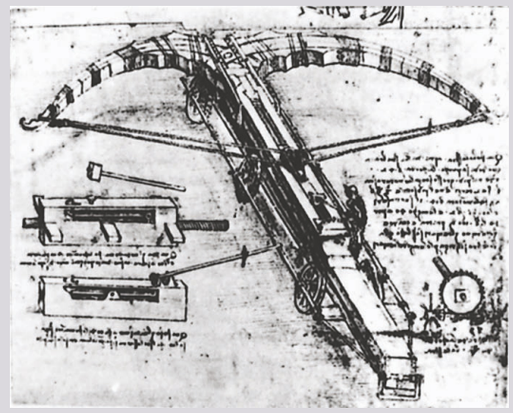
Regiomontanus, Peuerbach's student and later friend, was a child prodigy. At the age of 12, he calculated the orbits of the planets, and two years later, the Holy Roman emperor Frederick III assigned to him the task of casting his bride's horoscope. He spent a couple of years at the court of the Hungarian king Matthias Corvinus and was also called to the newly founded University of Pozsony (Bratislava). At court, he studied the Greek manuscripts that King Matthias had brought back as booty from campaigns against the Ottomans. Regiomontanus took up the work of Peuerbach, although instead of a complete translation of the *Almagest*, he was ultimately only able to produce a digest (Figure 2.73).

Regiomontanus carried out his astronomical observations in Nuremberg, where in 1471 he had also founded a printing house with the aim of publishing a series of scientific works of ancient and recent times. The publisher's plan and several published books have survived to the present day.

Quotation 2.35 should give us an idea of just how modern a thinker Regiomontanus was. It may only have been his untimely death at the age of 40 that prevented him from taking the decisive step to a heliocentric theory: In a surviving fragment of a letter, he writes that due to the motion of Earth, the motion of stars changes; also in a marginal note he made in a book by Archimedes, he refers to the importance of Aristarchus of Samos, the ancient Greek proponent of a heliocentric theory.

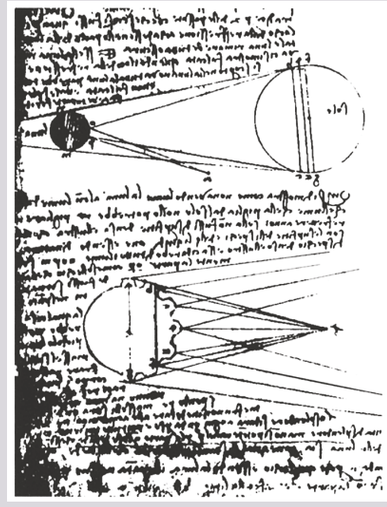
2.6.6 *The Role of the Printing Press*

The rapid spread of scientific ideas is intimately connected with the most significant technical invention of the Renaissance: the printing press (Johannes Gutenberg, ca. 1400–1468).



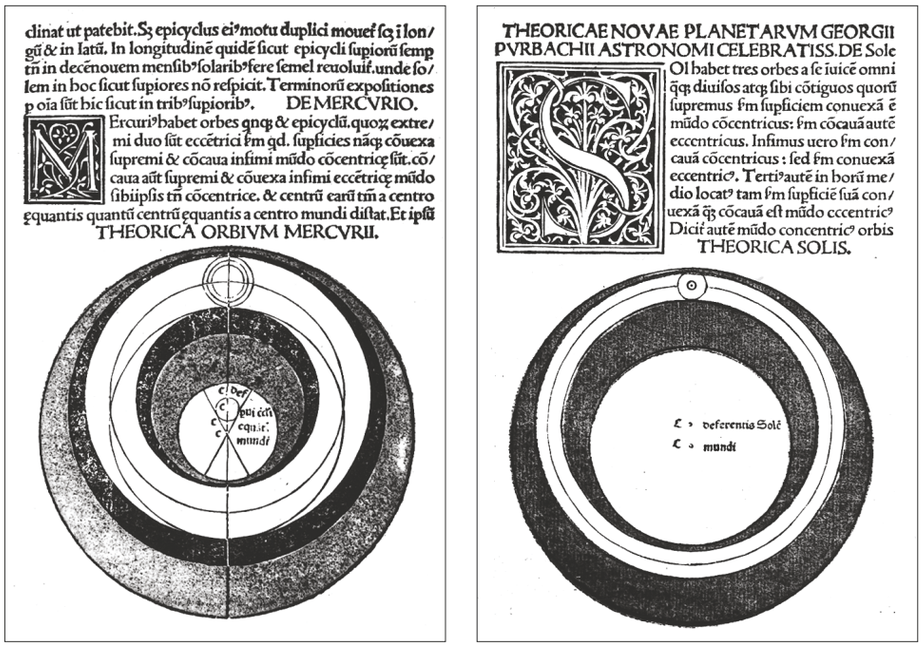
▲ **Figure 2.69** Leonardo's "supercannon."

> Most illustrious Lord … having fully studied the work of all those who claim to be masters and artificers of war … I will lay before your lordship my secret inventions, and then offer to carry them into execution at your pleasure. …
>
> An extremely light and strong bridge. An endless variety of battering rams. A method of demolishing fortresses built on a rock. A kind of bombard, which hurls showers of small stones and the smoke of which strikes terror into the enemy. A secret winding passage constructed without noise. Covered wagons, behind which whole armies can hide and advance.
>
> —Leonardo da Vinci, *Codice Atlantico* [Clark 1952, p. 37]

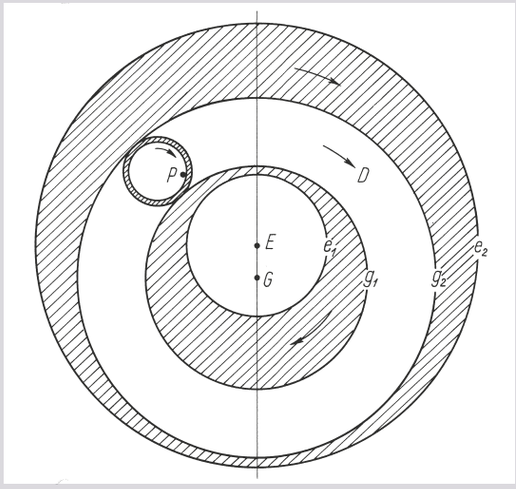


◀ **Figure 2.70** A page from the *Codex Leicester*. Drawings and comments on the relative sizes of Earth, Sun, and Moon. Leonardo assumes (correctly) that the Moon does not produce its own light but only reflects that of the Sun (and sunlight reflected from Earth). In response to his unknown opponent (*avversario*), Leonardo maintains (incorrectly) that this reflection is from waves of water. The opponent objects that even if water were to found on the Moon, there could not be waves, since there is no wind on the Moon.

The opinions of the *avversario* in Leonardo's commentaries generally represent those of Leonardo himself and point to a critical uncertainty regarding the position he has taken.



▶ **Figure 2.71** Left: The representation of the movement of the Sun in Peuerbach's book *Theoricae Novae Planetarum*. Right: Explanation of the movement of planet Mercury, which is, even in this simplified form, very confusing. (Central Library of the Benedictine Order, Pannonhalma.)



▲ **Figure 2.72** Ptolemy himself—and indeed in one of his later works, the *Hypotheses Planetarium*—attempted to provide a physical interpretation for his geometric constructions of planetary orbits. These ideas were more fully worked out by Arab scholars. In the simple case pictured here, we see two spherical surfaces e_1 and e_2 about Earth's center *E*, and two others, g_1 and g_2, about a different center *G*. The striped regions represent invisible yet solid material. The planet *P* is attached to the surface of yet another sphere. The outer sphere e_2 executes the (daily) motion of the heavens and carries the ring *D* along with it. The ring *D* moves independently, and within it turns, again independently, the sphere carrying the planet *P*. In this way, deferent and epicycle are realized physically, so to speak.

Quotation 2.33

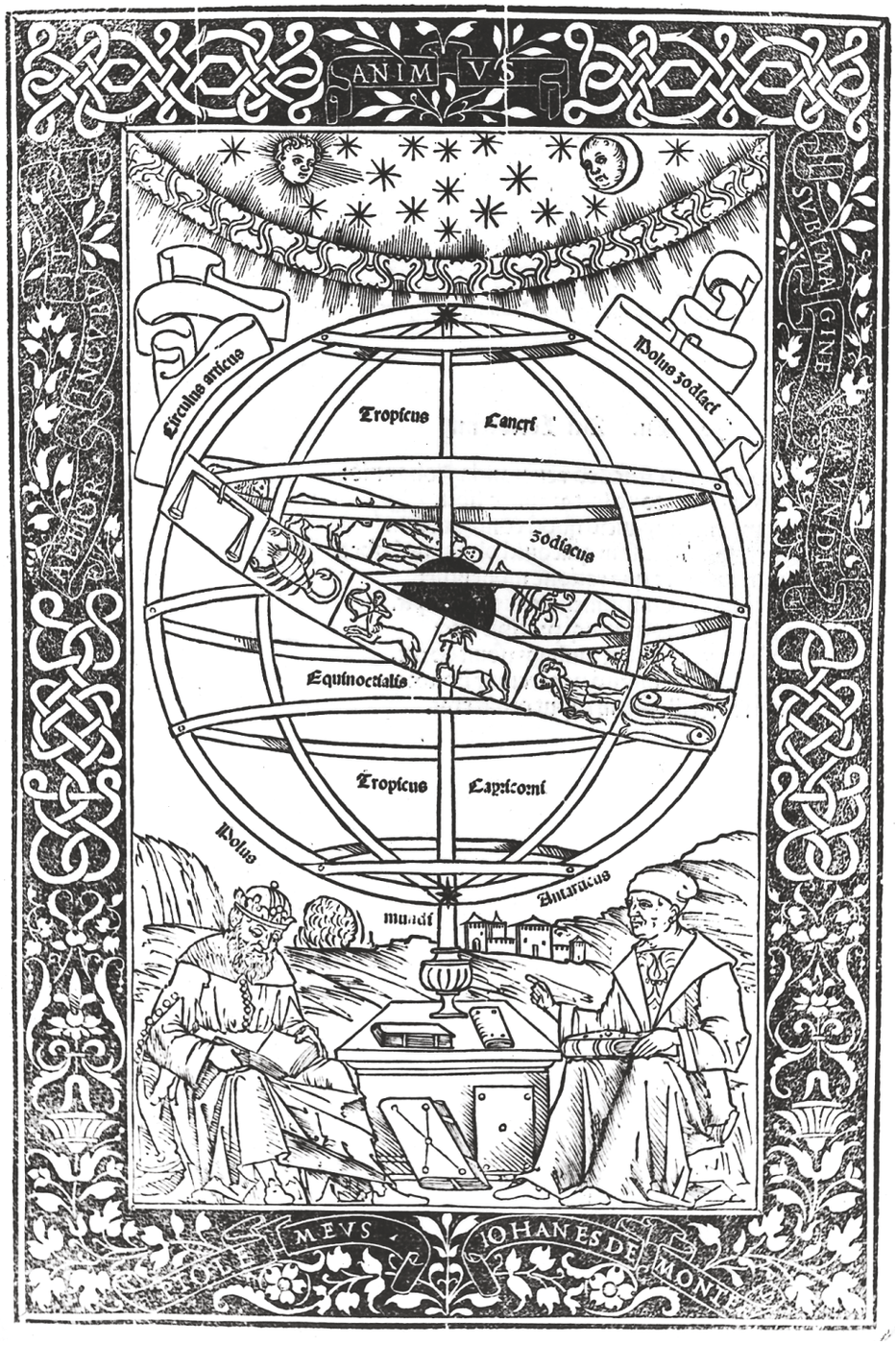
To me it seems that those sciences are vain and full of error which are not born of experience, mother of all certainty, first-hand experience which in its origins, or means or end has passed through one of the five senses. And if we doubt the certainty of everything which passes through the senses, how much more ought we to doubt things contrary to these senses—*ribelli ad essi sensi*—such as the existence of God or of the soul or similar things over which there is always dispute and contention. And in fact it happens that whenever reason is wanting men cry out against one another, which does not happen with certainties.

—Leonardo da Vinci, *Trattato Della Pittura* [p. 84]

For the point of view of the history of science, it is interesting to look at the first printing houses, which, for economic reasons would have published only books that had expectations of selling well. Table 2.4 allows us to draw some interesting conclusions about the cultural interests of people in the second half of the fifteenth century. From this table, we can see the number of first editions of books with scientific content up to the year 1500 (incunabula) with respect to the epochs in which the authors lived as well as who was most prolific. We see that in addition to antique and Arab authors, medieval Scholastics are represented as well. This is not surprising because at the universities, and above all among the theological faculties, the books of these authors were used as textbooks (Plate XIII).

At the end of the sixteenth century, on the eve of the great scientific revolution of the seventeenth century, we witness an odd phenomenon. In the late Middle Ages and the early Renaissance, the exuberant intellectual life in the Church and at the universities produced bold new ideas—we need think only of the University of Paris and Oxford University or of the ideas of Cardinal Nicholas of Cusa—but by the end of the sixteenth century, both institutions had become ossified and indeed were outspoken bastions of reactionary orthodoxy (Figure 2.74). At the universities it was believed that with the acquisition of the knowledge of antiquity, all scientific questions had been answered, and one had only to locate the relevant book and a suitable citation in order to solve any given problem. It is precisely because of this attitude that the Renaissance has been accused by historians of science of "servility." The reason why the Catholic Church dug in its heels can be seen, above all, in the success of the Reformation. The Church believed that a strict adherence to authority—the Bible, Aristotle, and the Church Fathers—was the only defense against the reformers.

At the same time, life itself was becoming more dynamic, the tensions between orthodox ideology and everyday reality grew ever greater—all indicating the necessity of a breakthrough in the natural sciences.



▲ **Figure 2.73** Regiomontanus and Ptolemy. Title page of the incunabulum *Epytoma Joannis de Monte Regio in Almagestum Ptolomei*, 1496. (Library of the Hungarian Academy of Sciences)

The armillary sphere (the word *armilla* means bracelet), symbol of astronomy, was used both as a didactic model and an important measuring instrument (Figure 1.54), along with the astrolabe, into the seventeenth century. The first description of an instrument with nine rings, such as the one pictured here, is known from Alexandria (second century CE). Hipparchus and Eratosthenes mention even earlier instruments of this kind. Ptolemy called it "astrolabe," a name that was later associated with a similar instrument (see Figure 2.29).



(a)



(b)

▲ **Figure 2.74** Despite the many scattered germs of original and progressive ideas, the medieval hierarchical order prescribed the intellectual horizons of mankind until well into the sixteenth century:

(a) In nature: the hierarchy earth, water, air, fire, followed by the celestial spheres and then the realm of the chosen. (From Petrus Apianus, *Cosmographia*.)

(b) In science: theology at the top; Peter Lombard as the highest authority. His book *Sententiarum libri IV* served for several centuries as the basis of theological instruction.

Then come Aristotle (*physica*), Seneca (*ethica*), and beneath them, Pythagoras (*musica*), Euclid (*geometria*), Ptolemy (*astronomia*). In the next row, Aristotle (*logica*), Cicero (*rhetorica, poetica*), Boethius (*mathematica*). (From Gregor Reisch, *Margarita philosophica*, Library of the Hungarian Academy of Sciences.)

Quotation 2.34

Though I may not, like them, be able to cite the authors, it is a much greater and worthier thing to cite by citing experience, mistress to their masters. They go blown up and pompous, dressed and ornamented not by their own labors, but by others'; and the do not allow me my own; and if they scorn me, an inventor, how much more will they, not inventors but trumpeters and reciters of others' works, be blamed. ...

No effect occurs in nature without a reason. Understand the reason and you need no experience. ...

If two unequal spherical bodies, heavy and of the same material and descent, fall from high to low, the one will fall as much more swiftly than the other as the diameter of the one goes into the other. ...

Therefore, in every doubled quantity of time, [a falling body] doubles the length of its descent and the speed of its motion. ...

The thing that moves in natural motion acquires in every degree of motion degrees of velocity, which degrees will have the same proportion, the last to the next-to-last, as has the second with the first. ...

The thing will go down by the line *bc* that much slower than by the line *bd* as that line *bc* is longer than *bd*, the bodies being of equal heaviness and roundness. ...

—Leonardo da Vinci [Truesdell 1968]

Quotation 2.35

Believe me that I carry a number of thoughts and even more within myself that I would love to communicate for your views, had I sufficient free time, thoughts about whose correctness I am thoroughly convinced and thoughts that still hover in the air and awake longing in the soul to investigate them. To begin my letter with the firmament—this was the topic of our earlier conversations—I can only wonder at the intellectual inertia of our astronomers, who like superstitious old women see all explanations and tables that they find in their books as divine and unalterable truths; they rely on the authors and do not concern themselves with truth.

—Regiomontanus, Letter to Bianchini

▼ **Table 2.4** The number of writers whose works appeared as incunabula, ordered by historical epoch. The great interest in medical books is indicated by the fact that three physicians are among the authors: the Greek Hippocrates (ca. 460–ca. 370 BCE), the Arab Zakarīyā Rāzi (865–925), and the Greco-Roman physician Galen (129–200). Johannes de Sacrobosco (John of Holywood, ca. 1195–ca. 1256) was a minor astronomer; Raimundus Lullus (1232–1315) was a Spanish Scholastic who offered prescriptions for laymen in his book *Ars magna et ultima* on how to discover truth [Crombie 1952].

before the Common Era	25
first to fifth centuries	50
sixth to seventh centuries	7
eighth to twelfth centuries	46
thirteenth century	40
fourteenth century	64
first half of the fifteenth century	51
second half of the fifteenth century	378

The following authors are represented with the most works:

Hippocrates	Regiomontanus	Ptolemy
Aristotle	Thomas Aquinas	Raimundus Lullus
Zakarīyā Rāzi	Avicenna	
Sacrobosco	Galen	
Albertus Magnus	Pliny	
…	Averroes	
	…	

CHAPTER 3

Demolition and the Construction of a New Foundation

Though why is it necessary to reckon the value of divine things in cash like victuals? Or what use, I ask, is knowledge of the things of Nature to a hungry belly, what use is the whole of the rest of astronomy? Yet men of sense do not listen to the barbarism which clamors for these studies to be abandoned on that account. We accept painters, who delight our eyes, musicians, who delight our ears, though they bring no profit to our business. And the pleasure which is drawn from the work of each of these is considered not only civilized, but even honorable. Then how uncivilized, how foolish, to grudge the mind its own honorable pleasure, and not the eyes and ears. It is a denial of the nature of things to deny these recreations. For would that excellent Creator, who has introduced beauty and power to delight, have left only the mind of Man, the lord of all Nature, made in his own image, without any delight? Rather as we do not ask what hope of gain makes a little bird warble, since we know that it takes delight in singing because it is for that very singing that the bird was made, so there is no need to ask why the human mind undertakes such toil in seeking out these secrets of the heavens. For the reason why the mind was joined to the senses by our Maker is not only so that Man should maintain himself, which many species of living things can do far more cleverly with the aid of even an irrational mind, but also so that from those things which we perceive with our eyes to exist we should strive towards the causes of their being and becoming, although we should get nothing else useful from them. And just as other animals, and the human body, are sustained by food and drink, so the mind of Man, which is something distinct from Man, is nourished, is increased, and in a sense grows up on this diet of knowledge, and is more like the dead than the living if it is touched by no desire for these things. Therefore as by the providence of Nature nourishment is never lacking for living things, so we can say with justice that the reason why there is such great variety in things, and treasuries so well concealed in the fabric of the heavens, is so that fresh nourishment should never be lacking for the human mind, and it should never disdain it as stale, nor be inactive, but should have in this universe an inexhaustible workshop in which to busy itself.

—Johannes Kepler, *Epitome astronomiae copernicane (Epitome of Copernican Astronomy)* [p. 55]

DOI: 10.1201/9781032697697-4
This chapter has been made available under a CC-BY-NC-ND license.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Demolition and the Construction of a New Foundation

3.1 The World in 1600

The beginning of the seventeenth century both in Europe and in other parts of the world saw a shift not only in politics and economics but in human consciousness as well. After the fireworks of the Italian Renaissance, the lands on the Atlantic coast with a door wide open to the world began to take center stage. Spanish hegemony was already on the wane. The Low Countries seceded (1581), the Spanish Armada had been destroyed (1588), and Philip II was dead (1598), although in the midst of this military and political decline a golden age of Spanish literature and art flourished.

In France, it had not even been a generation since the horror of the St. Bartholomew's Day Massacre of 1572, and yet Henry IV (Henry of Navarre) was able to state, "*Paris vaut bien une messe*" ("Paris is well worth a mass"), which led to the Edict of Nantes (1598) and therewith to an official demonstration of religious tolerance. Moreover, he was in the position to realize his goal that "Come Sunday, every farmer should have a chicken in his pot." With this, the religious war was now effectively at an end in France, but its denouement was still to come in Germany.

England was still living in the Elizabethan "Golden Age." Shakespeare was writing his plays. Central Europe, and especially Germany, was absent from the political and cultural arenas for several generations due to the religious wars and peasants' revolts as well as the increasing attention given to maritime trade (Quotation 3.1).

Changes of world-historical importance were taking place in Eastern Europe. The Russian Empire was beginning to take shape, with Boris Godunov ruling over an already strongly centralized state, relying on a standing army. Yermak's Cossacks had already conquered almost the whole of Siberia.

In the sixteenth century, the Ottoman Empire was already in decline after having reached its zenith and no longer posed a serious threat to the West. With the defeat of the Ottoman fleet at the Battle of Lepanto (1571), an equilibrium was finally restored, to the advantage of the West (Quotation 3.2).

Beyond the horizons of Europe, India—with a high level of cultural, political, and military development, and strong economic influence under the rule of Akbar the Great—could stand up to any comparison with Europe. Akbar not only increased India's military might, but also encouraged a flowering of the arts (in particular, miniature painting). His policies of religious tolerance were far ahead of his time.

In China, Wanli, an emperor of the Ming dynasty ruled, and despite signs of political and economic decline, the arts and sciences were flourishing there as well.

Europe was already expanding. The first Europeans were visiting India and China—in India, the English and the Dutch, and in China, the Portuguese. In Macao, the first European colony was established, and the Jesuits began to gain influence over the spiritual life there.

These indications show that European culture possessed an expansive power the likes of which had not been seen before in world history. By this time, the greater part of the globe had already been traversed by European navigators (Figure 3.1). Not only had Spaniards and Portuguese discovered America and circumnavigated the globe; they had also completed the colonization of the Philippines (1593).

The dynamism and expanding horizons of Europe resulted in a self-reinforcing cycle. Even the development of religion—both the Reformation and the

Quotation 3.1

Anno 1539, on the evening of 18 January at around 6 o'clock a comet was seen here in Wittenberg; it was dark, but very long, at 20 degrees, its tail reaching toward the sign of the fish, from evening toward midnight, till just before sunrise, when it stands at its highest. This comet was seen by D. Martinus Luther, D. Jonas, M. Philippus Melanchthon, D. Milich, and M. Erasmus, the mathematician, with great wonderment. Then D. M. L. said, "I would predict to Germany, not from the stars, but I proclaim God's anger from theology and the word of God; for it is impossible that Germany shall go unpunished; it must suffer a great setback, since it cannot be otherwise; for God will every day be provoked to destroy us; those blessed by God and the godless shall go and be killed.

"Let us only pray that we not despise God and his word. Whether or not we are great sinners, we nonetheless have forgiveness of sin and eternal life, as is promised us by the Word, furthered by Turk and emperor. They should not harm us, but support, I only grieve for our posterity, who shall be again brought from light into darkness."

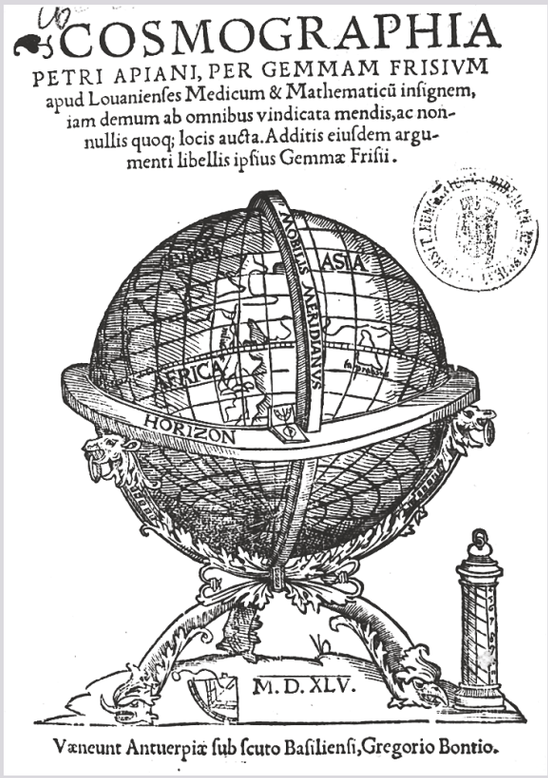
Then M. Ph. spoke: "... that in 1505 a comet was also seen before the Venetian war, and before Modona was conquered."

—Martin Luther

Quotation 3.2

White founts falling in the Courts of the sun,
And the Soldan of Byzantium is smiling as they run;
There is laughter like the fountains in that face of all men feared,
It stirs the forest darkness, the darkness of his beard;
It curls the blood-red crescent, the crescent of his lips;
For the inmost sea of all the earth is shaken with his ships.
They have dared the white republics up the capes of Italy,
They have dashed the Adriatic round the Lion of the Sea,
And the Pope has cast his arms abroad for agony and loss,
And called the kings of Christendom for swords about the Cross.
The cold queen of England is looking in the glass;
The shadow of the Valois is yawning at the Mass;
From evening isles fantastical rings faint the Spanish gun,
And the Lord upon the Golden Horn is laughing in the sun.
...

continued on next page





◀ ▲ **Figure 3.1** The most obvious indicator of an expanding horizon: an increase in geographical knowledge. The globe, earlier an abstract possibility, has now become an aid in instruction. (Petrus Apianus: *Cosmographia*. University Library, Budapest.)

Quotation 3.2, ***continued***

St. Michael's on his Mountain in the sea-roads of
the north
(Don John of Austria is girt and going forth.)
Where the grey seas glitter and the sharp tides
shift
And the sea-folk labour and the red sails lift.
He shakes his lance of iron and he claps his wings
of stone;
The noise is gone through Normandy; the noise is
gone alone;
The North is full of tangled things and texts and
aching eyes,
And dead is all the innocence of anger and surprise,
And Christian killeth Christian in a narrow dusty
room,
And Christian dreadeth Christ that hath a newer
face of doom,
And Christian hateth Mary that God kissed in
Galilee,—
But Don John of Austria is riding to the sea.
Don John calling through the blast and the eclipse
Crying with the trumpet, with the trumpet of his
lips,
Trumpet that sayeth ha!
Domino gloria!
Don John of Austria
Is shouting to the ships.

continued on next page

Counter Reformation—led to an expanding horizon linked to a growing sense of self. The diminishment of the authority of the Church, the direct reading and interpretation of the Bible, and the religious disputes arising from theological questions written in national languages and waged through pamphlets were key factors leading to far-reaching ideas and the development of a critical attitude.

Contact with less-developed peoples in the newly discovered parts of the world gave Europeans even more self-confidence and encouraged them to start to doubt the ideas handed down from antiquity and idolized by the Renaissance since, after all with their new discoveries, they had gone beyond the *ultima thule*, the ends of the word. This kind of flamboyant dynamism and self-confidence found its artistic expression in the flowing lines and grand style of the Baroque.

All things considered, life at the beginning of the seventeenth century was quite modern: The first joint-stock companies had already been formed; the first newspapers had appeared; and letters could be sent through the post. It was only within the walls of the universities, where Aristotle continued to hold sway as the arbiter of all questions, that intellectual life seemed to have come to a halt (Quotation 3.3). Yet, it would not be long before the self-confident toolmakers and engineers would have their say (Quotation 3.4, Figure 3.2). The words of Giordano Bruno at his trial after he was condemned to be burned at the stake give a hint that the expression of philosophical truth could not be held back much longer: "Perhaps you, my judges, pronounce this sentence against me with greater fear than I receive it."

When we read Giordano Bruno's dialogues, we are not surprised at the inquisitors' judgment against him nor are we surprised at their fears:

> *Burchio:* Where then is that beautiful order, that lovely scale of nature rising from the denser and grosser body which is our earth, to the less dense [sphere] which is water and on to the subtle [sphere] which is vapour, to the

yet subtler which is pure air, on to the subtlest which is fire and finally to the divine which is the celestial body? ...

Fracastorio: You would like to know where is this order? In the realm of dreams, fantasies, chimeras, delusions. ...

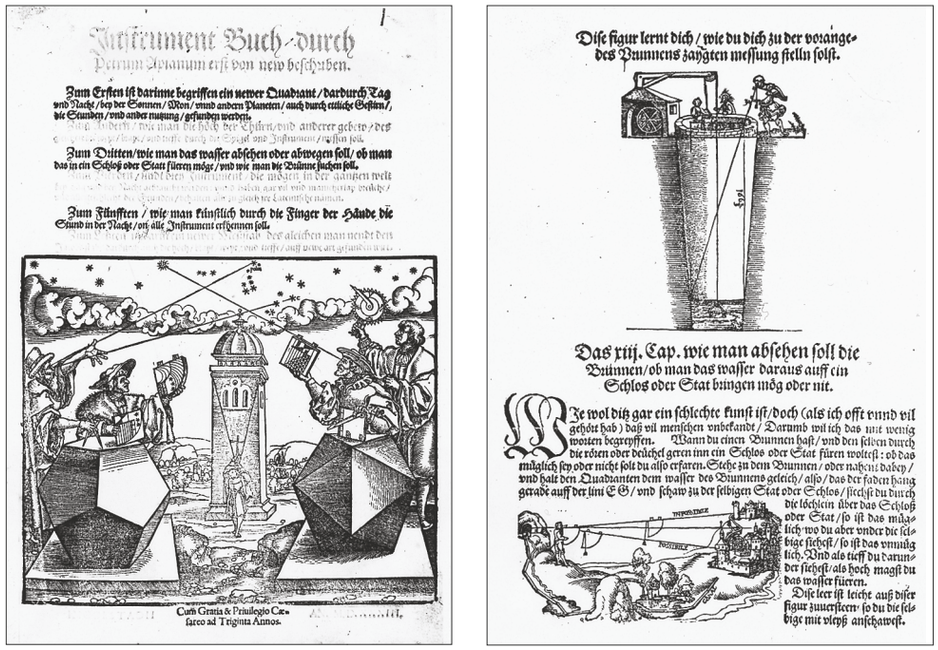
Burchio: Then you deny the famous distinction of the elements?

Fracastorio: I deny not the distinction of the elements, for I leave everyone at liberty to distinguish as he pleaseth concerning natural things. But I deny this order, this disposition that the earth is surrounded and contained by water, water by air, air by fire, fire by the heaven. ... The famous and received order of the elements and of the heavenly bodies is a dream and vainest fantasy, since it can neither be verified by observation of nature nor proved by reason or argued, nor is it either convenient or possible to conceive that it exist in such fashion. But we know that there is an infinite field, a containing space which doth embrace and interpenetrate the whole. In it is an infinity of bodies similar to our own. No one of these more than another is in the centre of the universe, for the universe is infinite and therefore without centre or limit.

Burchio: Do you think Plato is ignorant, Aristotle is a fool and all of those who followed their track are simple, stupid, and fanatic?

Fracastorio: My friend, I do not say that ... I think they are heroes of the Earth: but I do not believe without reason, nor want to accept teaching whose opposite is so obviously true that even you would have understood were you not completely blind and deaf?

—Giordano Bruno, *On the Infinite Universe and Worlds* [pp. 313–324]



▲ **Figure 3.2** The positions of the stars, heights of towers, and depths of mine shafts are measured with simple instruments and ingenious methods. (Petrus Apianus. Instrumentenbuch. University Library, Budapest, 1533.)

In reference to a 1545 catalogue of the books of the Swiss scholar Conrad Gesner, Figure 3.3 provides an overview of the categories of books that had been published in Latin, Greek, or Hebrew in the various branches of human knowledge. We see that theology still has the greatest presence as the number of books on mechanics, arithmetic, geometry, music, and astronomy taken together is not even one-third

Quotation 3.2, ***continued***

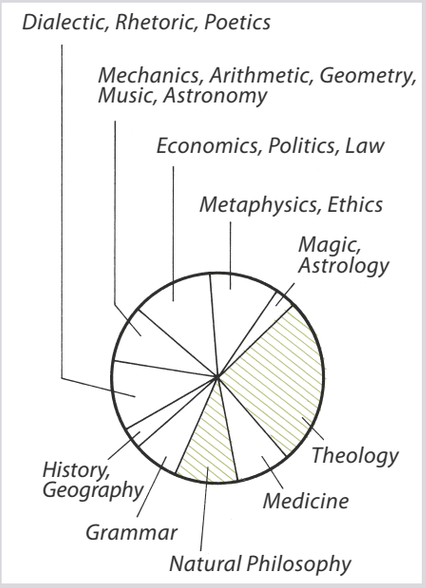
King Philip's in his closet with the Fleece about his neck
(Don John of Austria is armed upon the deck.)
The walls are hung with velvet that is black and soft as sin,
And little dwarfs creep out of it and little dwarfs creep in.
He holds a crystal phial that has colours like the moon,
He touches, and it tingles, and he trembles very soon,
And his face is as a fungus of a leprous white and grey
Like plants in the high houses that are shuttered from the day,
And death is in the phial and the end of noble work,
But Don John of Austria has fired upon the Turk.
Don John's hunting, and his hounds have bayed—
Booms away past Italy the rumour of his raid.
Gun upon gun, ha! ha!
Gun upon gun, hurrah!
Don John of Austria
Has loosed the cannonade.
...

Don John pounding from the slaughter-painted poop,
Purpling all the ocean like a bloody pirate's sloop,
Scarlet running over on the silvers and the golds,
Breaking of the hatches up and bursting of the holds,
Thronging of the thousands up that labour under sea
White for bliss and blind for sun and stunned for liberty.

Vivat Hispania!
Domino Gloria!
Don John of Austria
Has set his people free!

Cervantes on his galley sets the sword back in the sheath
(Don John of Austria rides homeward with a wreath.)
And he sees across a weary land a straggling road in Spain,
Up which a lean and foolish knight for ever rides in vain,
And he smiles, but not as Sultans smile, and settles back the blade. ...
(But Don John of Austria rides home from the Crusade.)

—G. K. Chesterton, *Lepanto*



▲ **Figure 3.3** The distribution of scientific books published before 1545 according to Conrad Gesner.

Quotation 3.3

I speak thus to my listeners: Look, this is taught by Aristotle, that is the opinion of Plato; so speaks Galen, and thus Hippocrates. Who listens to me must concede: The works of Borrius command complete trust, for not one of them is his own; rather, the most illustrious minds speak through his mouth ... If thoughts come to me that I don't find in their works, I discard them at once as suspicious or put them aside until they become old and die out before they have seen the light of day.

—Girolamo Borrio, Pisa, 1676, in Lalande, *Lectures sur la philosophie des sciences*

▶ **Figure 3.4** The stages along the path to the unification of celestial and terrestrial physics.

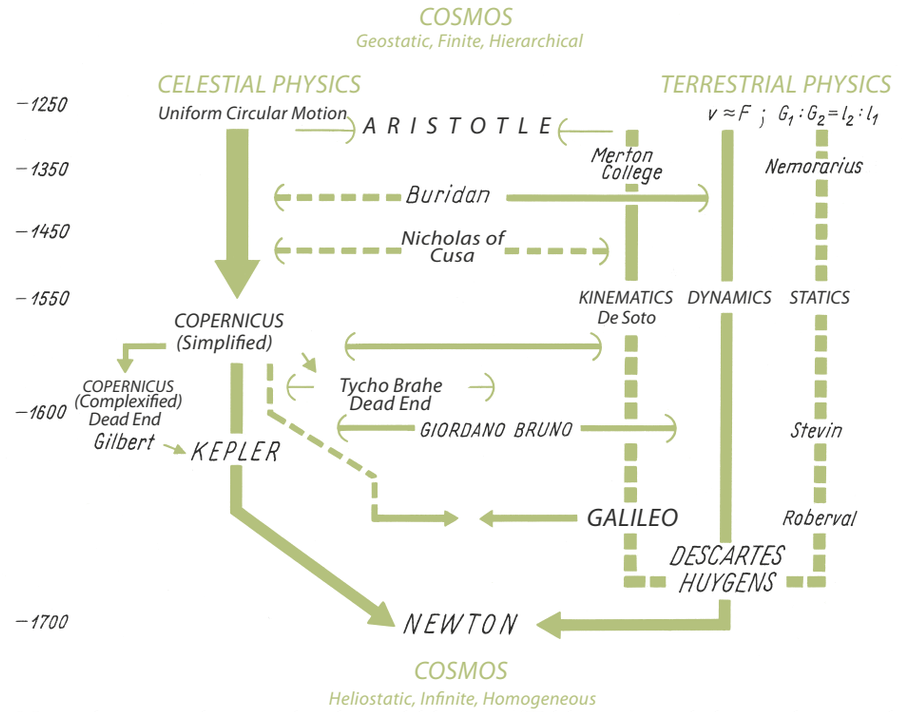
Quotation 3.4

And albeit, it may be said by the learned in the Mathematicalles, as hath beene already written by some, that this is no question or matter for a Mechanitian or Mariner to meddle with, no more than is the finding of the Longitude, for that it must bee handled exquisitely by Geometrical

continued on next page

that of books on theological subjects. At first glance, natural philosophy appears to be adequately represented, since it comprises about one-third of the total number of theological texts. However, when we take a closer look at what books fall into this category, the proportion no longer seems so favorable, since books on plants, animals, the soul, as well as curiosities of nature are included here. Indeed, books on natural curiosities enjoyed great popularity even in the seventeenth and eighteenth centuries because they apparently satisfied a universal human craving for sensation.

What, then, were the tasks standing before the natural sciences at the beginning of the seventeenth century? A very general answer to this question may easily be given: to upend the Aristotelian worldview and replace it with a new one. We see in Figure 3.4 what natural scientists would have to accomplish, whether they knew it or not: They needed to replace the cosmology of antiquity and the Middle Ages, which was geocentric, finite, and hierarchical, with a model of a heliostatic (that is, with the Sun at rest), infinite, and homogeneous cosmos governed by a universal set of laws. To accomplish this goal, the main task would then to be to unify the physics of the heavens with the physics of the Earth. Figure 3.4 gives the names of the most important figures who were involved in this unification. A complete synthesis was eventually achieved with Newton's laws of motion and his law of universal gravitation. After Newton, the physics of the heavens again diverged from terrestrial physics, no longer from philosophical necessity, but as a step in the practical specialization of the sciences.



New ideas were born in books and libraries (Table 3.1), and the modern world began to take shape. However, we should not forget that this process took place against the background of a declining guild structure and the rise of manufacturing with continually increasing productivity, accompanied, however, by a commoditization of manual labor. In short, the pros and cons of capitalism were already becoming apparent. The concomitant technical problems needing to be solved represented an enormous challenge for physics (Figure 3.5).

We must not overlook the fact that ideological arguments were in no way limited to verbal exchanges, particularly in cases where economic and ideological issues were

Table 3.1 Significant ideas in significant books of the time.

1543	Copernicus	*De revolutionibus orbium coelestium*
1585	Giordano Bruno	*Del infinito, universe e mondi*
1585	Stevin	*Weeghconst*
1600	Gilbert	*De magnete*
1609	Kepler	*Astronomia nove aitiologethos seu Physica coelestis*
1610	Galileo	*Sidereus Nuncius*
1619	Kepler	*Harmonices mundi …*
1620	Bacon	*Novum Organum*
1623	Galileo	*Saggiatore*
1632	Galileo	*Dialogo sopra I due massimi sistemi del mondo*
1638	Galileo	*Siscorsi e dimonstrazioni matematiche intorno a due nuove scienze*
1637	Descartes	*Discours de la method*
1644	Descartes	*Principia philosophiae*
1647	Pascal	*Expériences Nouvelles touchant le Vide*
1660	Boyle	*Touching the Spring and Weight of Air*
1672	Guericke	*Experimenta nova Magdeburgica*
1673	Huygens	*Horologium oscillatorium*
1687	Newton	*Philosophiae naturalis principia mathematica*
1690	Huygens	*Traité de la lumière*
1690	Locke	*An essay concerning human understanding*

Quotation 3.4, ***continued***

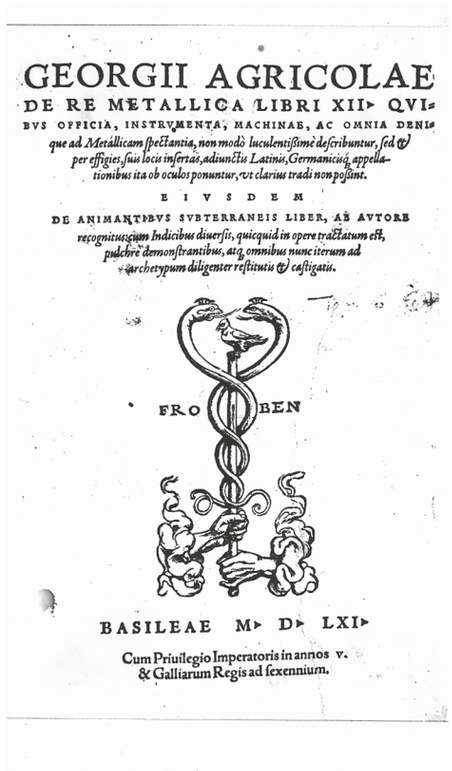
demonstration, and Arithmeticall Calculation: in which Artes, they would have all Mechanitians and Sea-men to be ignorant, or at leaste insufficientlie furnished to performe such a matter. ... But I doe verily thinke, that notwithstanding the learned in those Sciences, being in their studies amongst their bookes, can imagine greate matters, and set downe their farre fetcht conceits, in faire showe, and with plawsible wordes, wishing that all Mechanicians were such, as for want of uttereance, should be forced to deliver unto them their knowledge and conceites, that they might flourish upon them, and applye them at their pleasures: yet there are in this land divers Mechanicians, that in their severall faculties and professions, have the use of those Artes at their fingers endes, and can apply them to their severall purposes, as effectually and more readily, then those that would most condemne them.

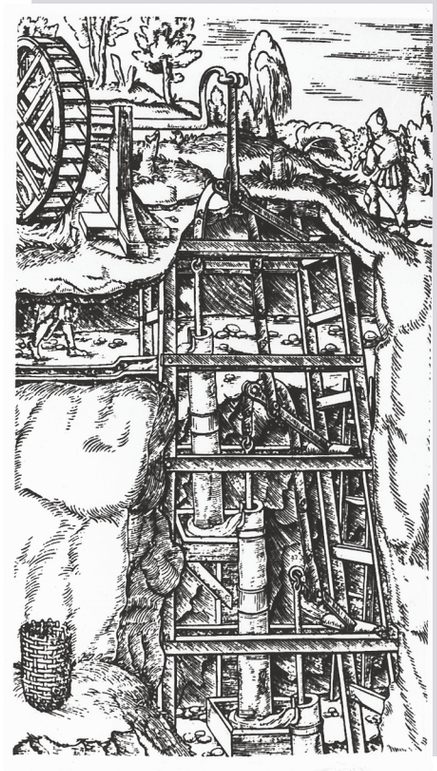
—Robert Norman, *The Newe Attractive* [p. ii]

closely connected. Among the events that finally led to the vision of a homogeneous universe—or, more concretely, to the replacement of the relation $\mathbf{F} \propto \mathbf{v}$ by $\mathbf{F} \propto \Delta\mathbf{v}$—belong also the burning of heretics at the stake and the horrors of the religious wars (Figure 3.6).

Figure 3.5 The book *De re metallica*, by Agricola (Georg Bauer, 1494–1555), was an epoch-making event in mining and metallurgy. Of particular interest for us are the methods of measurement described and, above all, the construction of multistage pumps. The Scholastic problem of *horror vacui* has here become a technical problem.









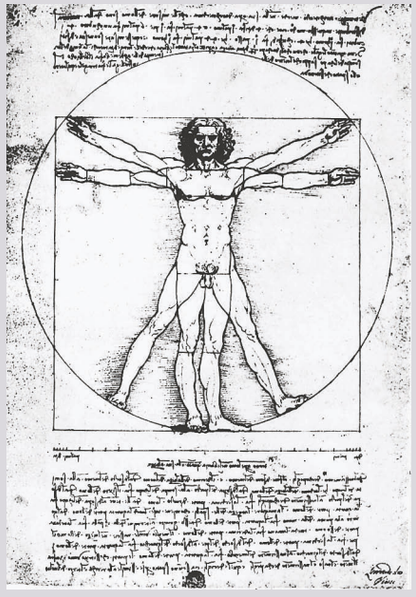
▶ **Figure 3.6** Theoretical arguments were not the decisive factor in the formation of a new worldview. (Hans Holbein the Younger: *Bad War*.)

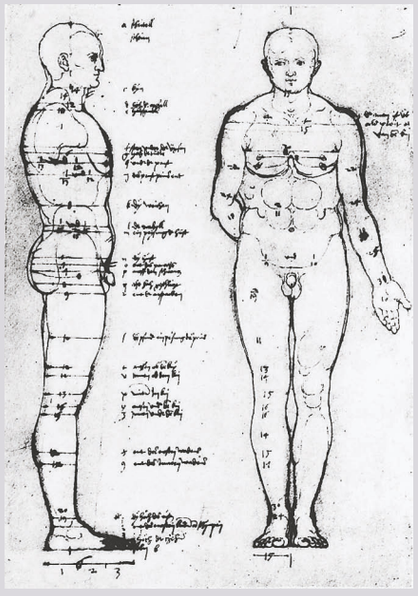
Quotation 3.5

Following Plato and the Pythagoreans, the greatest mathematicians of that divine age, my teacher thought that in order to determine the causes of the phenomena circular motions must be ascribed to the spherical earth.

—Rheticus, "Narratio Prima" [pp. 147–148]

▼ **Figure 3.7** Canon of proportions for the human figure by Leonardo (top) and Dürer (bottom).





3.2 Numerology and Reality

3.2.1 *Back to Plato in a New Spirit*

Earlier, we saw that during the Renaissance, the arts and classical philology became the focus of intellectual activity. Only two important aspects of the natural sciences were preserved and further developed. First, an attempt was made to describe artistic ideals in terms of numbers and geometric figures, which to some extent reached beyond Aristotle all the way to Plato and Pythagoras. It is enough to observe that just about every artist had formulated his own canon of harmonious proportions for the human figure (Figure 3.7). Attempts were made to bring numeric or geometric harmony not only to the human figure, but also to buildings and even the proportions of printed letters.

The second important aspect of the natural sciences was a new emphasis on nature. This expressed itself in a meticulous description of natural phenomena and objects that at times could be considered scientific in today's sense of the word. Here we are not thinking of Leonardo, in whose person the artist and natural scientist are so inseparably combined that it is often difficult to say whether his drawings are of greater significance for art or for science.

The dominant characteristic of modern science is the close interplay between observation and mathematical description; hence, in the above aspects we see the emergence of the basic conditions for the new science. In contrast with the starting point espoused by Pythagoras and Plato, we can see that a great step forward had already been taken. As we have seen, for the idealistic philosophers of antiquity, harmony was the starting point (Section 1.2.3). Nevertheless, the shadow-world of phenomena had to be "rescued" at the same time, but they did not care much whether this rescue was complete or not. After all, everyday reality could hope to mirror the world of Platonic ideals only in rough outline, so that, for example, a triangle drawn on a piece of paper could reproduce the geometric idea of a triangle only in some approximate sense.

Seventeenth-century thinkers were still convinced of the mystical role of numbers, but they also had an unconditional respect for *facts*. As we shall see, for Kepler, the slightest discrepancy between theory and observation was cause enough for him to cast aside previous notions and to search for new harmonies, regardless of how strongly he believed in the construction of the universe according to Platonic ideals.

The rejection of medieval Aristotelianism through the embracing of Greek science before Aristotle can be seen as a step in the right direction. At the same

time, the complete dismissal of medieval achievements interrupted the development of the natural sciences and, above all, of physics, which had begun in the late Middle Ages. The Renaissance turned a blinding light on ancient science, but this very action obscured the accomplishments of the so-called Dark Ages.

3.2.2 *The Retrograde Revolutionary: Copernicus*

The first and most important achievement of the Neo-Pythagorean point of view was the elaboration of the Copernican heliocentric system. A precursor of Copernicus was Regiomontanus, who perhaps only because of his untimely death was unable to take the decisive step (Section 2.6.5).

It was in the first years of the sixteenth century, while visiting the Italian universities, that Copernicus (Figure 3.8, Plate XIV) most likely became acquainted with the ideas of Aristarchus of Samos, according to which the Sun occupies the central position in the universe, and the planets—including our own Earth—revolve about the Sun in circular orbits. In a strange way, what especially pained Copernicus, as he described it in the dedication to his magnum opus, was the necessity of introducing *equants* into the Ptolemaic system, which amounted to an obvious betrayal of the Platonic conception that the planets, in their perfection, could only move along perfect—that is, circular—paths *with constant velocity*, and that at the outset, one should have to allow combinations of such perfect paths in order to achieve agreement with observation. As we might put it today, Copernicus did not attack the Ptolemaic system from the left, but from the right (Quotations 3.5 and 3.6).

The scientific world awaited the publication of Copernicus's works with great expectation. In his 1514 manuscript *Commentariolus*, Copernicus at first sketched out only the basic idea of his system. Then, in a 1540 book, *Narratio prima*, Rheticus described Copernicus's ideas in detail. Copernicus himself saw his own great work, *De revolutionibus orbium coelestium* (On the revolutions of the celestial spheres), only on his deathbed in 1543, a year that is frequently cited as marking the beginning of modern science; see Figure 3.9.



Quotation 3.6

People give ear to an upstart astrologer who strove to show that the earth revolves, not the heavens or the firmament, the sun and the moon, as if someone moving by carriage or by ship thought himself to be stationary and the land and trees to be moving. But the case is more like this: One who wants to appear clever cannot go along with what everyone seems to observe; he must come up with something different. So this clever soul, who wants to turn the whole of science of astronomy upside down. Yet even in those difficult matters I put my faith in Sacred Scriptures, for Joshua commanded the sun to stand still, and not the earth.

—Martin Luther, June 4, 1539 [Deely 2001, pp. 495–496]

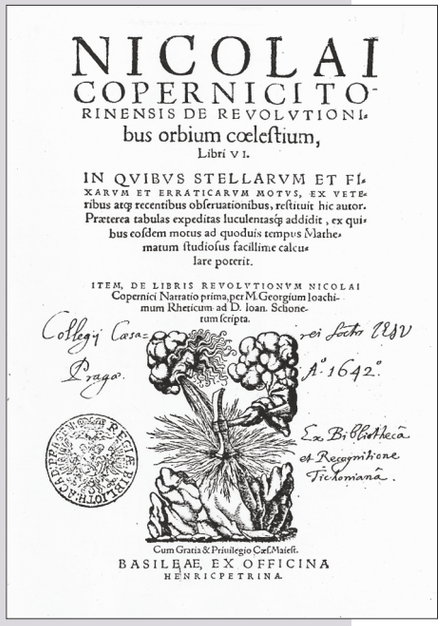


◀ ▼ **Figure 3.8** According to Hevelius, the predecessors of Copernicus were Ptolemy, Albategnius (al-Battani), and Regiomontanus. Prince William IV of Hessen (1532–1592) became known for his catalog of stars. The picture is from Hevelius's book *Uranographia* (1687).

Nikolaus Copernicus (Mikołaj Kopernik): born 1473 in Toruń, son of a merchant, raised after his father's death by his uncle, the bishop of Warmia. Studied in Kraków, then in Bologna, Padua, Ferrara (where in 1503 he received a doctorate in ecclesiastical law), and Rome. Also studied medicine and served as his uncle's personal physician. From 1512, without ordination, was canon at the Frombork cathedral. Died in Frombork, 1543. Published his ideas about astronomy first in 1512 in his *Commentariolus*, which was distributed in manuscript as a pamphlet. An extensive description of his cosmology is contained in the 1540 article by the Wittenberg professor Rheticus: *Narratio Prima*. His magnum opus, *De revolutionibus orbium coelestium*, was published in Nuremberg shortly before his death.

First of all I wish you to be convinced … that this man whose work I am now treating is in every field of knowledge and in mastery of astronomy not inferior to Regiomontanus. I rather compare him with Ptolemy, not because I consider Regiomontanus inferior to Ptolemy, but because my teacher shares with Ptolemy the good fortune of completing, with the aid of divine kindness, the reconstruction of astronomy which he began, while Regiomontanus—alas, cruel fate—departed this life before he had time to erect his columns.

—Rheticus, *First Narration (Narratio Prima)* [Boas Hall 1962]

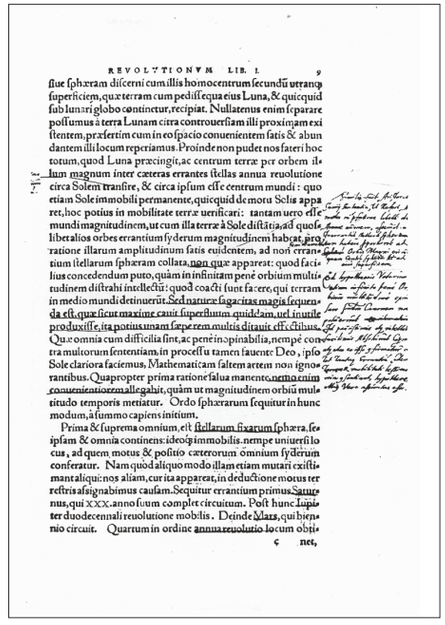

NICOLAI
COPERNICI TO-
RINENSIS DE REVOLVTIONI-
bus orbium cœlestium,
Libri VI.

IN QVIBVS STELLARVM ET FI-
XARVM ET ERRATICARVM MOTVS, EX VETE-
ribus atq; recentibus obseruationibus, restituit hic autor.
Præterea tabulas expeditas luculentasq; addidit, ex qui-
bus eosdem motus ad quoduis tempus Mathe-
matum studiosus facillime calcu-
lare poterit.

ITEM, DE LIBRIS REVOLVTIONVM NICOLAI
Copernici Narratio prima, per M. Georgium Ioachi-
mum Rheticum ad D. Ioan. Schone-
rum scripta.

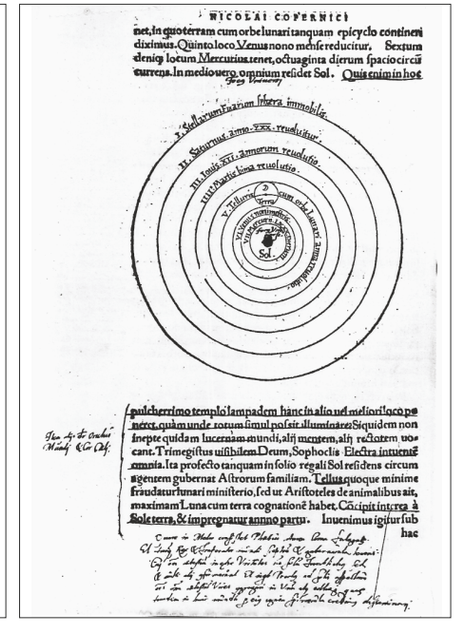
Collegij Casa- rei Doctor 1580
Pragæ. A.° 1642.
Ex Bibliotheca
et Recognitione
Ticho Brahe.

Cum Gratia & Priuilegio Cæs. Maiest.
BASILEAE, EX OFFICINA
HENRICPETRINA.


REVOLVTIONVM LIB. I 9

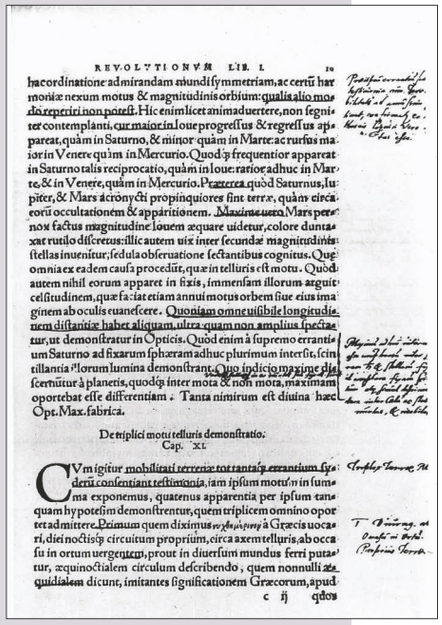
siue sphæram discerni cum illis homocentrum secundũ uerang
superficiem, quæ terram cum pedisequa eius Luna, & quicquid
sub lunari globo continetur, recipiat. Nullatenus enim separare
possumus à terra Lunam citra controuersiam illi proximam exi
stentem, præsertim cum in eo spacio conuenientem satis & abun
dantem illi locum reperiamus. Proinde non pudet nos fateri hoc
totum, quod Luna præcingit, ac centrum terræ per orbem il-
lum magnum inter cæteras errantes stellas annua reuolutione
circa Solem transire, & circa ipsum esse centrum mundi: quo
etiam Sole immobili permanente, quicquid de mora Solis appa
ret, hoc potius in mobilitate terræ uerificari: tantam uero esse
mundi magnitudinem, ut cum illa terræ à Sole distantia, ad quos
liberalios orbes errantium syderum magnitudinem habeat, pro
ratione illarum amplitudinum satis euidentem, ad non errant-
tium stellarum sphæram collata, non quæ appareat: quod facti
us concedendum puto, quàm in infinitam penè orbium multi
tudinem distrahi intellectũ: quod coacti sunt facere, qui terram
in medio mundi detinuerũt. Sed naturæ sagacitas magis sequen-
da est, quæ sicut maxime cauit superfluum quiddam, uel inutile
produxisse, ita potius unam sæpe rem multis dituit effectibus.
Quæ omnia cum difficilia sint, ac penè in spinabilia, nempè con-
tra multorum sententiam, in processu tamen fauente Deo, ipso
Sole clariora faciemus, Mathematicam saltem aetem non ignos
rantibus. Quapropter prima ratione salua manente, nemo enim
conueniat iocm allegabit, quàm ut magnitudinem orbiũ mul
titudo temporis metiatur. Ordo sphærarum sequitur in hunc
modum, à summo capiens initium.

Prima & suprema omnium, est stellarum fixarum sphæra, se
ipsam & omnia continens: ideoq; immobilis, nempe uniuersi lo
cus, ad quem motus & positio cæterorum omnium syderum
conferatur. Nam quod aliquo modo illam etiam mutari existi
mant aliqui: nos aliam, cui ita appareat, in deductione motus ter
restris assignabimus causam. Sequitur errantium primus Satur-
nus, qui xxx. anno suum complet circuitum. Post hunc Iupi
ter duodecennali reuolutione mobilis. Deinde Mars, qui bien-
nio circuit. Quartum in ordine annua reuolutio locum obti-
net,


NICOLAI COPERNICI

net, in quo terram cum orbe lunari tanquam epicyclo continen
diximus. Quinto loco Venus nono mense reducitur. Sextum
denicq; locum Mercurius tenet, octuaginta dierum spacio circũ
currens. In medio uero omnium residet Sol. Quis enim in hoc

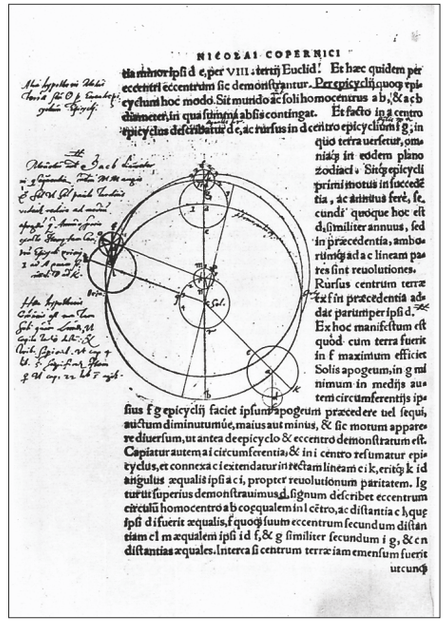
pulcherrimo templo lampadem hanc in alio uel meliori loco po
nere, quàm unde totum simul possit illuminare? Siquidem non
inepte quidam lucernam mundi, alij mentem, alij rectorem uo
cant. Trimegistus uisibilem Deum, Sophoclis Electra intuens
omnia. Ita profecto tanquam in solio regali Sol residens circum
agentem gubernat Astrorum familiam. Tellus quoque minime
fraudatur lunari ministerio, sed ut Aristoteles de animalibus ait,
maximam Luna cum terra cognationẽ habet. Cõcipit interea à
Sole terra, & impregnatur anno partu. Inuenimus igitur sub hac


REVOLVTIONVM LIB. I 10

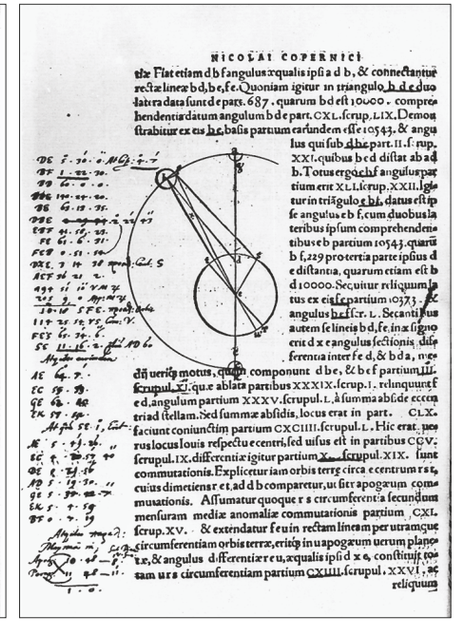
hac ordinatione ad mirandam mundi symmetriam, ac certũ har
moniæ nexum motus & magnitudinis orbium: qualis alio mo
do reperiri non potest. Hic enim licet animaduertere, non segni
ter contemplanti, cur maior in Ioue progressus & regressus ap
pareat, quàm in Saturno, & minor quàm in Martẽ: ac rursus ma
ior in Venere quàm in Mercurio. Quod & frequentior appareat
in Saturno talis reciprocatio, quàm in Ioue: rarior adhuc in Mar
te, & in Venere, quàm in Mercurio. Præterea quòd Saturnus, Iu
piter, & Mars ἀcronyctὶ propinquiores sint terræ, quàm circa
eorũ occultationem & apparitionem. Maxime uero Mars pere
now factus magnitudine Iouem æquare uideatur, colore dunta
xat rutilo discretus: illic autem uix inter secundæ magnitudinis
stellas inuenitur, sedula obseruatione sectantibus cognitus. Qug
omnia ex eadem causa procedũt, quæ in telluris est motu. Quod
autem nihil eorum appareat in fixis, immensam illorum arguit
celsitudinem, quæ facit etiam annui motus orbem sue eius ima
ginem ab oculis euanescere. Quoniam omne uisibile longitudi-
nem distantiæ habet aliquam, ultra quam non amplius spectae
tur, ut demonstratur in Opticis. Quòd enim à supremo errantiũ
um Saturno ad fixarum sphæram adhuc plurimum interfit, scin
tillantia illorum lumina demonstrant. Quo indicio maxime ipsa
scernũtur à planetis, quodq; inter mota & non mota, maximam
oportebat esse differentiam. Tanta nimirum est diuina hæc
Opt. Max. fabrica.

De triplici motu telluris demonstratio.
Cap. XI.

CVm igitur mobilitati terreniæ tot tantaq; errantium sy-
derũ consentiant testimonia, iam ipsum motum in sum-
ma exponemus, quatenus apparentia per ipsum tam
quam hypotesim demonstrentur, quem triplicem omnino oportet admittere. Primum quem diximus περὶ κέντρου à Græcis uocari, dicinoctiq; circuitum proprium, circa axem telluris, ab occa
su in ortum uergentem, prout in diuersum mundus ferri puta
tur, æquinoctialem circulum describendo, quem nonnulli ἰσημερινὸν quidialem dicunt, imitantes significationem Græcorum, apud


NICOLAI COPERNICI

tia minor ipsi d e, per VIII. tertij Euclid. Et hæc quidem per
eccentri eccentrum sic demonstrantur. Per epicyclũ quoq; epi
cyclum hoc modo. Sit mundo ac soli homocentrus a b, & a b
diameter, in qua finitima absis contingat. Et facto in a centro
epicyclus describatur d e, ac rursus in d centro epicyclium f g, in
quo terra uersetur, omnia
q; in eodem plano
zodiaci. Sint epicycli
primi motus in succedẽ
tia, ac æquitus ferẽ, se
cundi quoque hoc est
dissimiliter annuus, sed
in præcedentia, ambo
rumq; ad a c lineam pa
res sint reuolutiones.
Rursus centrum terræ
ex situ præcedentia ad
dat parumper ipsi d. Ex hoc manifestum est
quòd cum terra fuerit
in f maximum efficiet
Solis apogeum, in g mi
nimum in mediis au
tem circumferentijs ip-
sius f g epicycli faciet ipsum apogeum præcedere uel sequi,
auctum diminutumue, maius aut minus, & sic motum appare
re diuersum, ut antea de epicyclo & eccentro demonstratum est.
Capiatur autem a i circumferentia, & in i centro referatur epi
cyclus, et connexa c i extendatur in rectam lineam c i k, eritq; k id
angulus æqualis ipsi a c i, propter reuolutionum paritatem. Ig
turut superius demonstrauimus d, signum describet eccentrum
circuli homocentro a b eo qualem in i cẽtro, ac distantia c k, quę
ipsi d fuerit æqualis, f quoq; suum eccentrum secundum distan
tiam c l m æqualem ipsi i d f, & g similiter secundum i g, & c en
distantias æquales. Interea si centrum terræ iam emensum fuerit
utcunq;


NICOLAI COPERNICI

tiæ. Fiat etiam d b angulus æqualis ipsi a d b, & connectantur
rectæ lineæ b d, b e, f e. Quoniam igitur in triangulo b d e duo
latera data sunt de parts. 687. quarum b d est 10000. compre
hendentia datum angulum b de part. CXL. scrup. LIX. Demon
stratur ex eis hac basis partium earundem esse 10543. & angu
lus qui sub d b e part. II. f. rup.
XXI. quibus b e d distat ab a d
b. Totus ergo b f angulus par
tium erit XLI. scrup. XXII. Igi
tur in triangulo e b f, datus est ip
se angulus e b f, cum duobus la
teribus ipsum comprehenden
tibus e b partium 10543. quarũ
b f, 229 pro tertia parte ipsius d
e distantia, quarum etiam est b
d 10000. Sequitur reliquum la
tus ex eis f e partium 10373. &
angulus b f e. r. l. Secanti hus
autem f e lineæ b d, f e, in æ signo
rum erit d x æ angulus sectionis, dife
rentia inter f e d, & b d a, mo
dũ uerũ motus, quem componunt d b e, & b e f partium III.
scrupul. XI. quæ ablata partibus XXXIX. scrup. I. relinquunt f
e d, angulum partium XXXV. scrupul. L, à summa absidis eccen
triad stellam. Sed summæ absidis, locus erat in part. CLX.
faciunt coniunctim partium CXCIIII. scrupul. L. Hic erat mo
rus locus Iouis respectu ecentri, sed usus est in partibus CCV.
scrupul. IX. differentiæ igitur partium x. scrupul. XIX. sunt
commutationis. Explicetur iam orbis terræ circa ecentrum r s t,
cuius dimetiens r et, ad b comparetur, ut fiat apogæum com
mutationis. Assumatur quoque r s circumferentia secundum
mensuram mediæ anomaliæ commutationis partium CXL.
scrup. XV. & extendatur f e uin rectam lineam per utramque
circumferentiam orbis terræ, eritq; in apogæum uerum plane
tæ, & angulus differentiæ r u, æqualis ipsi d x e, constituit to
tam u r s circumferentiam partium CXIIII. scrupul. XXVI. ac
reliquum

▶ **Figure 3.9** The book *De revolutionibus orbium coelestium*, one of the most important works in the history of European culture, was published in 1543 in Nuremberg. The book contains a brief foreword directed to the reader that is attributed not to Copernicus, but to Osiander. Following the foreword is a letter from Cardinal Schönberg to Copernicus, and finally a dedication to Pope Paul III.

The entire work consists of six books. The first book describes a simple model of a heliocentric solar system and contains, moreover, a summary of the necessary mathematical theorems of trigonometry, plane and spherical. In the following five books, Copernicus offers a description of the orbits of Earth, the Moon, and the planets using tables, epicycles, and deferents.

The pages reproduced here from the second edition (Basel, 1566) are taken from the exemplar at Charles University, Prague, whose unique feature is that it contains Tycho Brahe's handwritten notes. The Basel edition also contains the book *Narratio Prima* by Rheticus (facsimile edition, Prague, 1971).

> I have no doubt that certain learned men, now that the novelty of the hypotheses in this work has been widely reported—for it establishes that the Earth moves, and indeed that the Sun is motionless in the middle of the universe—are extremely shocked, and think that the scholarly disciplines, rightly established once and for all, should not be upset. But if they are willing to judge the matter thoroughly, they will find that the author of this work has committed nothing which deserves censure. For it is proper for an astronomer to establish a record of the motions of the heavens with diligent and skilful observations, and then to think out and construct laws for them, or rather hypotheses, whatever their nature may be, since the true laws cannot be reached by the use of

continued on next page

In the history of human culture, the displacement of Earth by the Sun as the midpoint of the cosmos is referred to as the *Copernican revolution*, and any important idea could hardly be given higher praise than to be compared with it. Thus, for example, Kant speaks with great pride of his "Copernican revolution" in his investigation of the role of reason in the process of understanding nature.

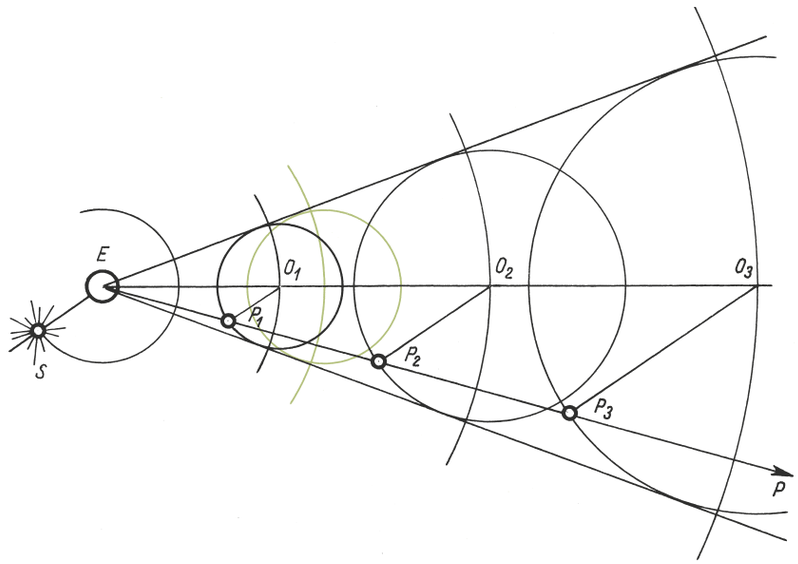
In what follows, we will first look at the Copernican system and then investigate the extent to which it represents progress with respect to methods of calculation and to physical and ideological foundations.

Copernicus describes his system in the first part of his book (Quotation 3.7). First let us make clear that the simplified Copernican system and the simplified Ptolemaic system (Figure 3.10) provide identical descriptions of the motions of the Sun and planets from the point of view of a terrestrial observer. In both cases, the observer is viewing the movement from Earth and can follow the paths of the

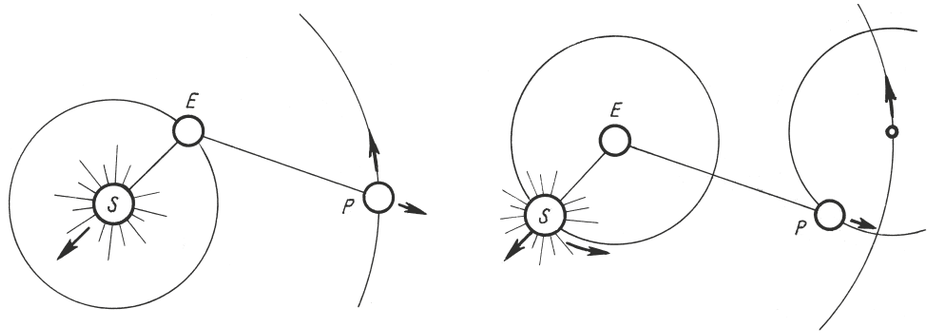
celestial spheres based on their relative positions with respect to the fixed stars. The two systems, *from this viewpoint,* are equivalent when both the Copernican and Ptolemaic observers see the same angular coordinates for the Sun and planets in relation both to one another and to a selected fixed star, and therefore both observers can record the same trajectory on their star maps.

We begin the proof of this equivalence with the observation that in the Ptolemaic system, only the *ratio* of the radius of the deferent and that of the epicycle have any significance; the *radius* of the latter can therefore be chosen *arbitrarily* (Figure 3.11). For the convenience of the proof, let us choose it to be equal to the radius of Earth's orbit (Figure 3.12). The following illustrations demonstrate the equality of the relevant angles to a terrestrial observer in the cases of an inner (Figure 3.13) and an outer (Figure 3.14) planet.

Let us carry out the proof in somewhat more detail in the case of an outer planet. We begin with the Copernican system and draw in Figure 3.14, in black, the resulting paths. At a given point in time, suppose that Earth is located at the point E and a planet at point P. The directions are then shown in which an observer on Earth views the Sun (S) and the planet. Now we let Earth assume the role played by the Sun and move the point E into the place previously occupied by S, resulting in the point S being mapped to S', so that the Sun is again viewed from Earth at the same



▲ **Figure 3.11** In the Ptolemaic system, the radii of the epicycles and deferent circles can be chosen arbitrarily, as long as their ratio is kept fixed. The line joining the center of an epicycle to its planet (for the outer planets) is, however, always parallel to the line connecting Earth and the Sun.

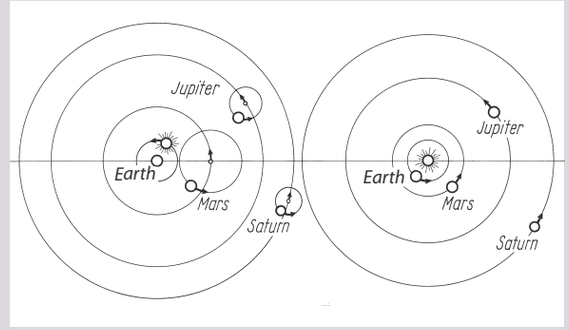


▲ **Figure 3.12** We have here chosen the deferent circle of the Ptolemaic system in such a way that the epicycle's radius is the same as that of the orbit of Earth. It is clear that the observable angles are equal.

Figure 3.9 *continued*

reason; and from those assumptions the motions can be correctly calculated, both for the future and for the past. Our author has shown himself outstandingly skilful in both these respects. Nor is it necessary that these hypotheses should be true, nor indeed even probable, but it is sufficient if they merely produce calculations which agree with the observations. That is, unless anyone is so ignorant of geometry and optics that the epicycle of Venus seems to him probable, or he thinks that it is in accordance with its law that it is sometimes ahead of the Sun and sometimes lags behind it by forty degrees or more. For who does not see that from that assumption it necessarily follows that the star's diameter appears more than four times greater, and its area more than sixteen times greater, at perigee than at apogee, to which all the experience of the ages is opposed. There are other things also in this discipline which are no less absurd, which it is quite unnecessary to examine for the present purpose. For it is clear enough that this subject is completely and simply ignorant of the laws which produce apparently irregular motions. And if it does work out any laws—as certainly it does work out very many—it does not do so in any way with the aim of persuading anyone that they are valid, but only to provide a correct basis for calculation. Since different hypotheses are sometimes available to explain one and the same motion (for instance eccentricity or an epicycle for the motion of the Sun) an astronomer will prefer to seize on the one which is easiest to grasp; a philosopher will perhaps look more for probability; but neither will grasp or convey anything certain, unless it has been divinely revealed to him. Let us therefore allow these new hypotheses also to become known beside the older, which are no more probable, especially since they are remarkable and easy; and let them bring with them the vast treasury of highly learned observations. And let no one expect from astronomy, as far as hypotheses are concerned, anything certain, since it cannot produce any such thing, in case if he seizes on things constructed for any other purpose as true, he departs from this discipline more foolish than he came to it. Farewell.

—Osiander, "To the Reader Concerning the Hypotheses of This Work" in Copernicus, *On the Revolutions of the Heavenly Spheres* [pp. 22–23]



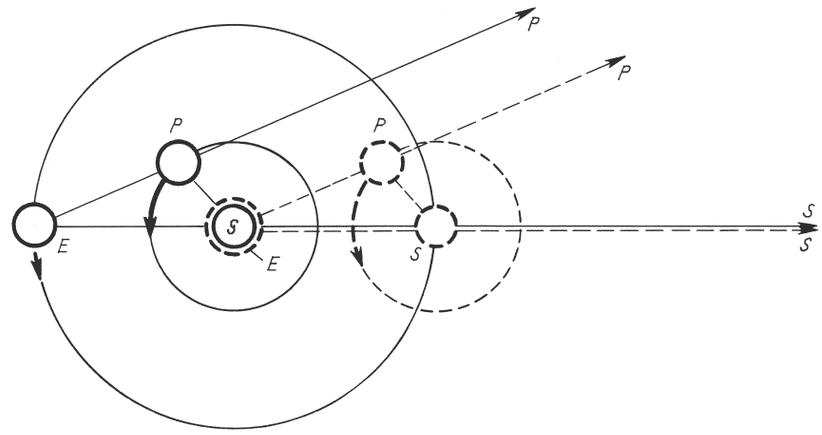
▲ **Figure 3.10** The Ptolemaic and Copernican systems in their simplest representations.

Quotation 3.7

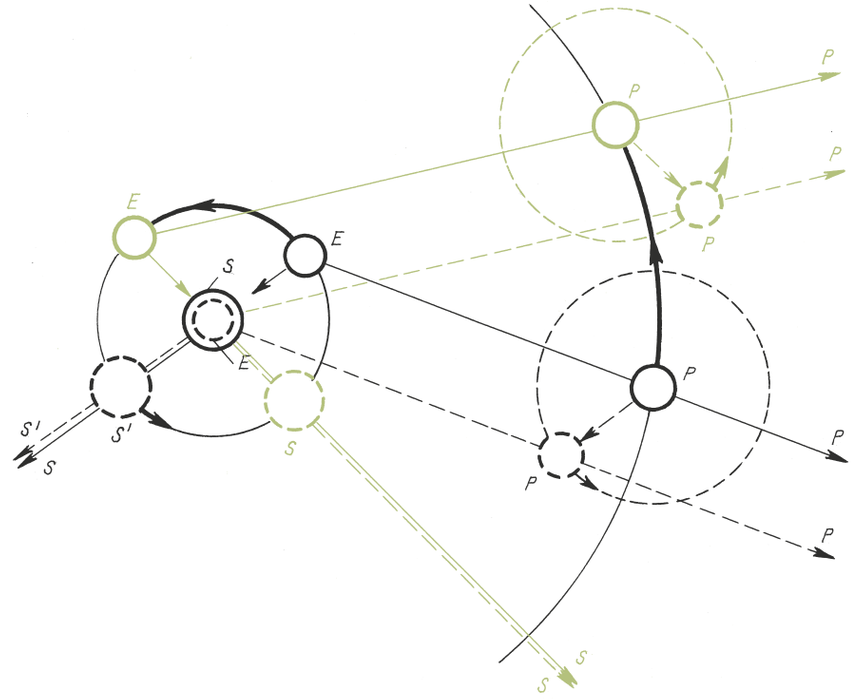
The universe is globe-shaped, either because that is the most perfect shape of all, needing no joint, an integral whole; or because that is the most capacious of shapes, which is most fitting because it is to contain and preserve all things; or because the most finished parts of the universe, I mean the Sun, Moon and stars, are observed to have that shape, or because everything tends to take on this shape, which is evident in drops of water and other liquid bodies, when they take on their natural shape. There should therefore be no doubt that this shape is assigned to the heavenly bodies. ...

The first and highest of all is the sphere of the fixed stars, which contains itself and all things, and is therefore motionless. It is the location of the universe, to which the motion and position of all the remaining stars is referred. For though some consider that it also changes in some respect, we shall assign another cause for its appearing to do so in our deduction of the Earth's motion. There follows Saturn, the first of the wandering stars, which completes its circuit in thirty years. After it comes Jupiter, which moves in a twelve-year-long revolution. Next is Mars, which goes round biennially. An annual revolution holds the fourth place, in which as we have said is contained the Earth along with the lunar sphere which is like an epicycle. In fifth place Venus returns every nine months. Lastly, Mercury holds the sixth place, making a circuit in the space of eighty days. In the middle of all is the seat of the Sun. For who in this most beautiful of temples would put this lamp in any other or better place than the one from which it can illuminate everything at the same time? Aptly indeed is he named by some the lantern of the universe, by others the mind, by others the ruler. Trismegistus called him the visible God, Sophocles' Electra, the watcher over all things. Thus indeed the Sun as if seated on a royal throne governs his household of Stars as they circle round him.

—Copernicus, *On the Revolutions of the Heavenly Spheres* [pp. 36, 49–50]



▲ **Figure 3.13** Equivalence in the case of the inner planets: The midpoint of the epicycle always lies on the line connecting Earth and the Sun. The radius of the epicycle is equal to that of the orbit of the planet, and the deferent coincides with the orbit of Earth.



▲ **Figure 3.14** For an outer planet, we show here the agreement of the two world systems when the planet moves.

angle as before. If we also require the planet to be seen from Earth at the old angle, then it, like Earth, must be moved parallel through the same distance. We thereby obtain the configuration shown in the figure with dashed lines: the Ptolemaic view.

We now consider the situation at a later point in time and enter into the figure the configuration according to the Copernican system, again with solid lines, and that for the Ptolemaic system with dashed lines, but this time in color. We see that at this later time as well, the positions of the Sun and the planet are in the same relationship in both systems, provided the following two conditions are satisfied: in the Ptolemaic system, the Sun must orbit Earth in a circle corresponding to the orbit of Earth, and secondly the planet moves on a circular epicycle whose

midpoint runs along the Copernican orbit, with the ray connecting the midpoint to the planet being at every moment parallel to the ray from Earth to the Sun. For an inner planet, the epicycle's midpoint moves in such a way that it is always on a line connecting the Sun and Earth. We saw this line in Section 1.4 (Figure 1.72), and it now can be understood what was only surmised, namely that the Sun plays some distinguished role in connection with planetary motion.

A glance at Figure 3.10 shows that the Copernican system is the simpler of the two. Furthermore, we find that in addition to the angles, the distances also acquire meaning, and it becomes a simple matter to determine the distance from each planet to the Sun. We recall here once again (Section 1.4) that in the Ptolemaic system, the radii of the deferents are chosen for the individual planets in such a way that after bringing in the epicycles, there is no overlap. Even the order of the planets was arbitrarily chosen according to the principle, later proven correct, that the more distant the planet, the smaller its mean velocity with respect to the background sphere of fixed stars.

The distance to one of the inner planets can be determined in the Copernican system by using Figure 3.15. We have merely to find the maximal angle at which the planet is seen with respect to the Sun, as measured from Earth. From the illustrated triangle and from the Earth-Sun distance *ES*, the distance of the planet from the Sun can be determined using the relation

$$PS = ES \sin\alpha .$$

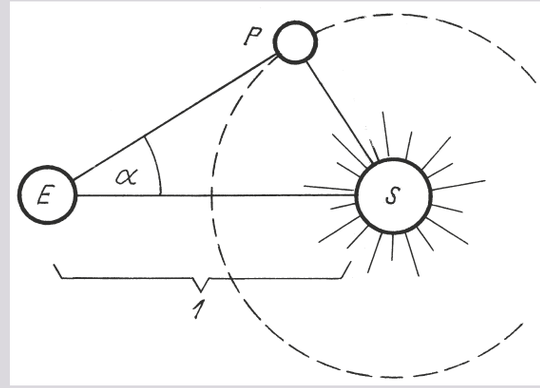
In the Copernican system, one can calculate a planet's orbital period about the Sun. With the help of Figure 3.16, let us investigate how this orbital period can be determined for an inner planet. We select a specific position for the planet, for example the position that we discussed above for the determination of distance, in which the planet is observed at a maximal angle with respect to the Sun. Let us now determine the time it takes for the planet to return to the same position with respect to an observer on Earth, of course. We have denoted this position of the planet by P'. We see that during this time Earth has moved in its orbit through a certain angle from its initial position while the inner planet has completed one revolution about the Sun and a part of a second revolution. Let us express the movement of Earth in its orbit by the angular velocity ω_E and that of the planet by ω_P, and let t_0 denote the time that the planet takes to achieve the same position with respect to a terrestrial observer. Then the angle through which Earth moves is $\omega_E t_0$, while that for the planet is $\omega_P t_0$. As can be seen from the figure, the two angles are connected by the relation

$$\omega_E t_0 = \omega_P t_0 - \alpha + \alpha - 2\pi .$$

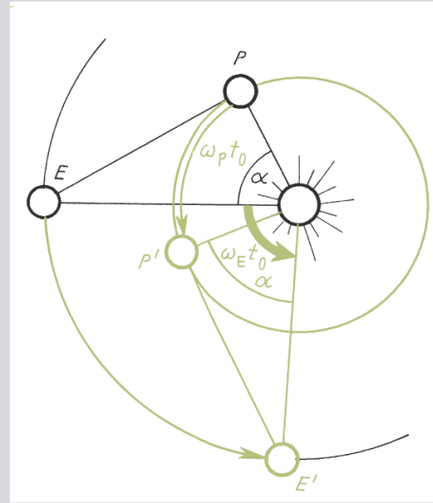
Let us denote the orbital period of the Earth by T_E and that of the planet by T_P, then the angular velocities and periods are related as follows:

$$\omega_E = \frac{2\pi}{T_E}, \quad \omega_P = \frac{2\pi}{T_P}.$$

We now continue our observations over time intervals in which the configuration under consideration occurs *N* times, apply these to the first equation where



▲ **Figure 3.15** Determination of the orbital radius of an inner planet in terms of the unit distance from Earth to the Sun.



▲ **Figure 3.16** Determination of the orbital period of an inner planet according to Copernicus.

Figure 3.17 Determination of the orbital period of an outer planet.

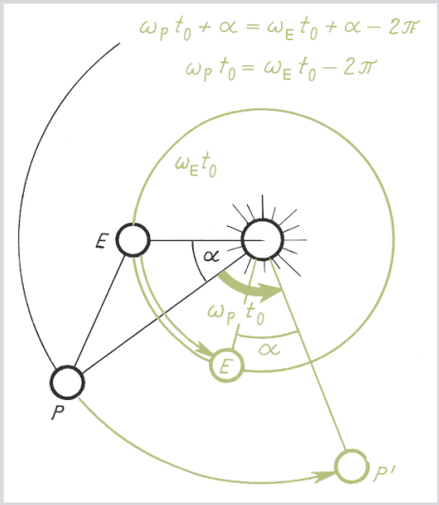
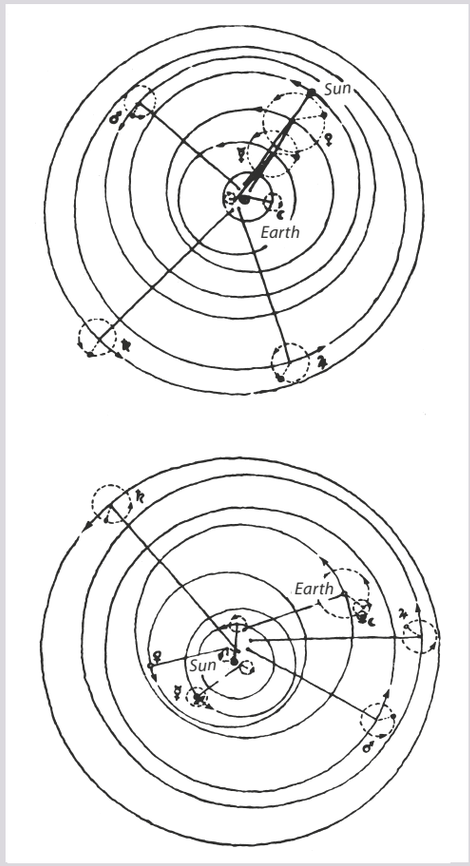


Figure 3.18 The difference between the Ptolemaic (top) and Copernican (bottom) systems largely disappears when, in order to obtain agreement with observations, epicycles and deferents are brought in [Cohen 1960, p. 58].



we also substitute the angular velocities in terms of the orbital periods:

$$\frac{2\pi}{T_E}(t_0 N) = \frac{2\pi}{T_P}(t_0 N) - 2\pi N.$$

If we measure the time in years—that is, if we use the Earth's orbital period as the unit of time, thereby replacing T_E by 1—then from

$$T_E = 1, \quad t_0 N = T \quad \text{(in years)},$$

we obtain the relationship

$$T = \frac{T}{T_P} - N,$$

and finally,

$$T_P = \frac{T}{T+N}.$$

This relationship gives the orbital period of an inner planet expressed in Earth years.

With the help of Figure 3.17, we may now, using analogous reasoning, determine the orbital period of an outer planet.

The simplicity of the pictures and the fact that one can give a physical meaning to the periods as well as the distances provide a strongly convincing argument for the validity of the Copernican system. However, it is precisely its equivalence with the Ptolemaic system that reveals the inadequacies of the Copernican picture sketched here. The simplified Copernican system is equivalent to the simplified Ptolemaic system. But we also know that for a precise description of planetary motion in the Ptolemaic system, many more epicycles and equants must be introduced, and it is therefore obvious that the Copernican cosmology as sketched above, despite its convincing simplicity and physical interpretation, cannot correctly describe the observed phenomena.

It is at this point that difficulties began to arise for Copernicus. As we know today, the planets do not move in circular orbits, but elliptical ones, and the assumption of circular trajectories was inadequate for reproducing the observations even with the precision achieved in the sixteenth century or in antiquity. This could be one of the reasons—the others are perhaps ideological—that Aristarchus rejected this whole theory in his time. True to the task that he set for himself, Copernicus wanted to use in his theory only uniform circular motion, but the result was that the beauty and the transparency of his entire theory was lost. Thus Copernicus was also forced to admit epicycles into his system, and because he wished to avoid equants, in place of the few circles he could eliminate by moving the point of reference for planetary motion from Earth to the Sun, he promptly had to employ many others due to his steadfast support of Platonic principles. In the end, to describe the motions of the planets in accord with observation, Copernicus had to use more than 30 circles in a wide variety of combinations.

Figure 3.18 shows—still with some simplifications—the Copernican and Ptolemaic systems side by side. Without reading the figure caption, at first glance it is difficult to distinguish which is which. In both systems, the planets move along epicycles whose midpoints move along a deferent circle. In Figure 3.19, the motion of Earth around the Sun is shown in detail. While we see that Earth moves in a circular orbit about the Sun, the center of this circle, however, is not the Sun but a fictive point in space that itself moves on another small circle, whose center, in turn, moves on a circular path around the Sun. The Copernican system has to be this complex in order to agree with observed facts. From the point of view of physics, the Sun here

has no role other than that of an illuminating lantern for the planets. Therefore, in many works on the history of science, the Copernican system is not identified as *heliocentric*—that is, a system with the Sun at the center—but as *heliostatic*, with the Sun at rest, in contrast to the geostatic system, in which Earth is at rest.

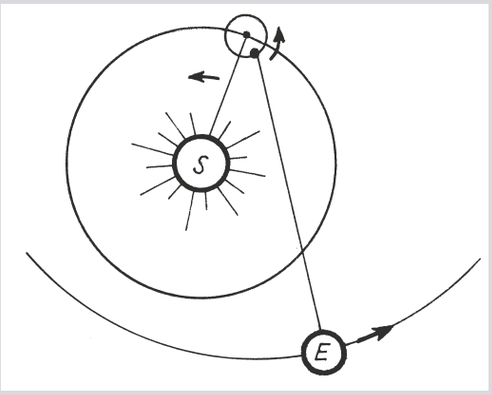
Thus, in the fully elaborated Copernican system, simplicity and aesthetics are lost. The physically distinguished role of the Sun is also lost, while the fault that is usually brought up in connection with the Ptolemaic system remains, namely that the planets travel in their epicyclic orbits about a point that exists only in our imagination as a fictive mathematical point of reference; when we seek a physical explanation, an effective force, we do not get any support from the theory for finding the origin of such a force. In the simplified Copernican system, in contrast, the Sun resides at the midpoint, and so it is obvious that the effective force is connected somehow with the Sun. But what are we to make of a Copernican system tamed to observation, in which this central position of the Sun has been obscured?

It is a much debated question, whether the foreword to Copernicus's book—which is due, according to historians, not to Copernicus himself but to Osiander—was in fact written in order to protect the book from the persecution of the clergy. In this foreword it is stated that the ideas presented in the book are nothing more than mathematical hypotheses, useful for simplifying astronomical calculation, and that they should not be taken as a representation of physical reality. But taking the book as a whole, this statement should not, in fact, be seen as a mollification directed at the Church, for the *fully elaborated* Copernican system can in fact be considered *only* as mathematical hypothesis, insofar as it admits no possible physical interpretation. Apparently, the contemporaries of Copernicus and the generation of astronomers that followed saw things this way, and perhaps this explains the fact that although the preliminary announcement of the Copernican system was received with great interest and expectation, nevertheless, after the appearance of the complete work, hardly anyone—with the exception of a few professional astronomers—took notice of it for the next several generations (Figure 3.20). This is particularly true of philosophers and clerics; the new system was not taken as a provocation or as a challenge, but merely as a variation on the Ptolemaic system with relatively few alterations.

To support this view, we cite in Figure 3.21 the title of an English translation of *De Revolutionibus* in 1576 suggesting that Copernicus had revived the old Pythagorean doctrines. At the same time, the Jesuit priest Clavius, whom the Pope had charged with reforming the calendar, used the Copernican system in his work. Many historians of science would place the Copernican system in the second or third century, rather than the sixteenth, if the judgment were to be based only on the assertions made within the system itself or—as a literary theorist would put it—if one considered only the stylistic elements.

Indeed, Copernicus was quite conservative in many respects. He spoke of the revolutions of the celestial orbits, meaning that he still believed in the crystalline spheres into which the planets were affixed like gems in their settings, so that they turned together. He also imagined the system of fixed stars as belonging to such a crystalline sphere. An interesting further development in the theory can be found in the English edition mentioned above, in which the stars are no longer arranged on the surface of a sphere, but rather in space. This is the first time that we see a picture in which the solar system is surrounded by an infinite sea of stars.

In some places in his writings, Copernicus answers arguments against the movement of Earth in a way that corresponds even to our present-day science. In other places, he still argues in the Aristotelian spirit. A good example for this is his

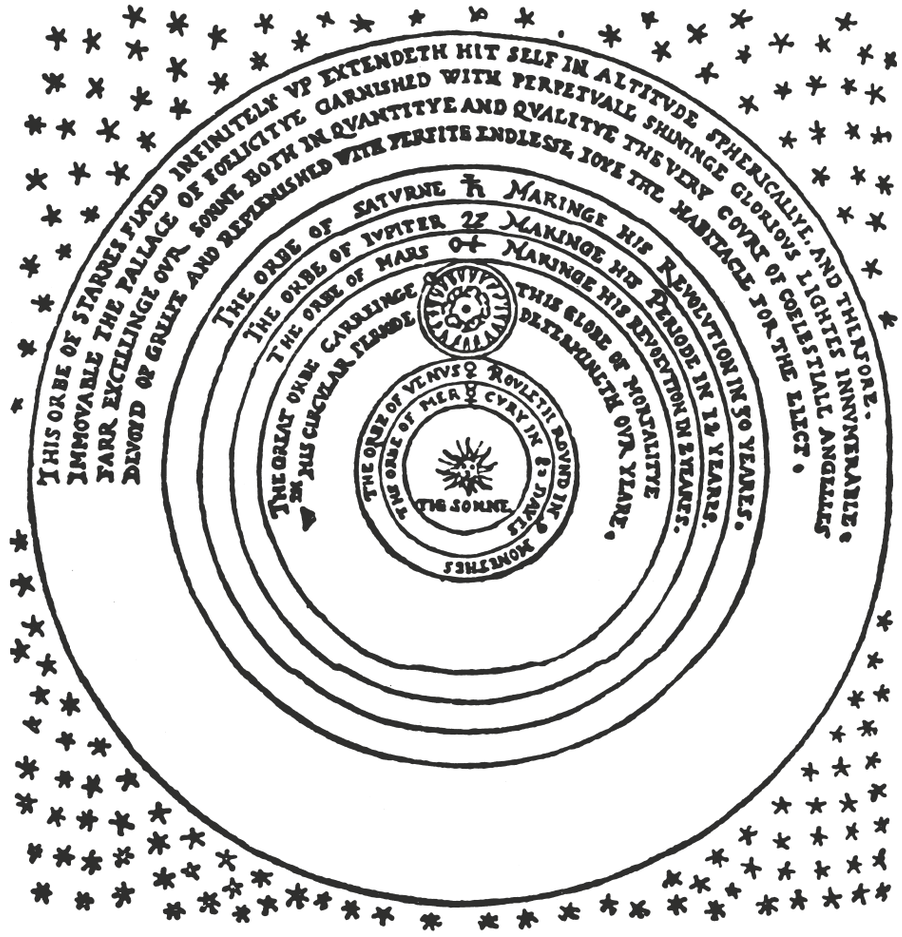


▲ **Figure 3.19** Copernicus did not use equants. However, to "rescue the phenomena," he had to introduce complex double epicycles. Earth's motion is also represented by such a complicated system [Hall 1965, p. 62].

> Then Mercury runs on seven circles in all; Venus on five; the earth on three, and round it the moon on four; finally Mars, Jupiter, and Saturn on five each. Altogether, therefore, thirty-four circles suffice to explain the entire structure of the universe and the entire ballet of the planets.
>
> —Copernicus, *Commentariolus* [p. 90]

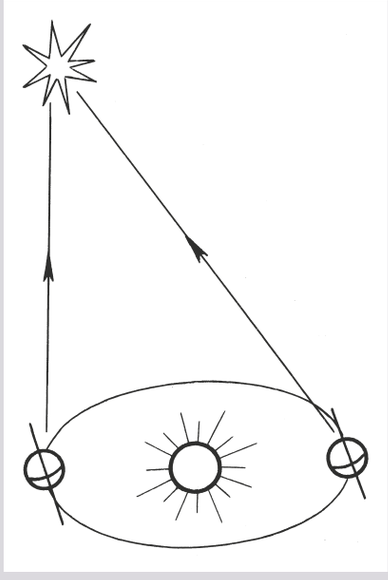


◀ **Figure 3.20** In his 1589 book, pictured here, Maginus made use of Copernicus's astronomical data but rejected the Copernican hypotheses regarding the motion of Earth as absurd. (Central library of the Benedictine order, Pannonhalma.)



▶ **Figure 3.21** A representation of the Copernican system in Digges's book. Digges accepted the Copernican system and even goes a step further: The stars are no longer arranged on crystalline spheres, but distributed in space. The text in the outer ring of the picture, in modern English, reads thus:

> This orb of stars fixed infinitely up extends itself in altitude spherically, and therefore, immovable the palace of felicity garnished with perpetual shining glorious lights innumerable, far excelling our sun both in quantity and quality the very court of celestial angels, devoid of grief and replenished with perfect endless love the dwelling place for the elect.
>
> —Thomas Digges, *Perfit Description of the Caelestiall Orbes according to the most ancienne doctrine of the Pythagoreans lately revived by Copernicus and by Geometricall Demonstrations approued,* 1576



▲ **Figure 3.22** If Earth orbits the Sun, then the stars must appear at different angles from different points on the orbit of Earth—or the stars must be very far away.

answer to the question of why bodies fall toward the center of Earth. Recall that Aristotelian physics held that all objects have their natural place, the heavy bodies below and the light above, and every body strives to assume the place assigned to it by nature. But if Earth no longer has a central, special place in the cosmos, then the notions "above" and "below" become meaningless. Copernicus explains the falling of objects as the striving of all things to unite with and make whole the place where they belong. However, we can also interpret this explanation—with a large measure of goodwill—as saying that every object has in it a center of force, or if you like, that it gravitationally attracts.

To the clever objection of why the fixed stars, in rotating about Earth in their orbits around the Sun, are not seen at different angles at different times of the year, or why—as we would put it today—no parallax is observed (Figure 3.22), Copernicus offered a clever answer: The crystalline sphere of the stars is so far away that Earth's orbit shrinks to a point by comparison. From today's point of view, such an answer is completely correct, but at that time, it did not appear convincing because it implied a size of the universe that was beyond the comprehension of that era. Moreover, the argument could be countered on considerations of parsimony, namely, that the vanishing of a logically positive phenomenon (parallax) had to rest on an additional hypothesis (the immeasurably large extent of the sphere of stars).

We summarize the significance of Copernicus's work as a negative result and two positive *possibilities*:

1. He was the first in history to give a logically complete account of the heliocentric system on the basis of Pythagorean and Platonic principles. It turned

out, however, that neither the attainable precision nor the physical interpretation was able to go significantly beyond the Ptolemaic system.

2. By placing Earth among the planets and thereby raising (or perhaps degrading) it to the rank of a celestial body, Copernicus put aside the sharp distinction between terrestrial and celestial phenomena and thus attacked Aristotelian physics in one of its fundamental theses, thus threatening to undermine the until-then well-founded hierarchical cosmology.
3. Copernicus's system assigned a special role to the Sun, if not in the fully elaborated variant, but at least in the basic system that could then serve as the starting point for the foundation of physical astronomy. It is no coincidence that every progressive proponent of the Copernican cosmology would seize on the simplified version, whether the reason was their limited understanding of mathematics and astronomy, sufficient for reading only the first chapter of *De revolutionibus* and not the other five, or a deliberate disregard of those last five chapters in order to move the basic Copernican system in another direction.

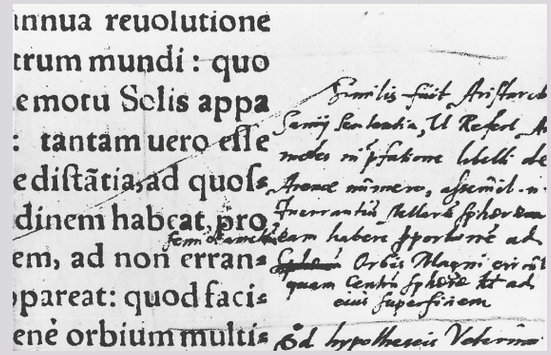
We must, of course, add that the power of any idea depends to a very great extent on—indeed, is often completely determined by—whether it is expressed at the opportune moment in history. In Copernicus's time, his system was not seen as something radically new. Figure 3.23 shows Tycho Brahe's handwritten marginal notes in a copy of *De revolutionibus*, in which he remarks that the ideas expressed there were already known to Aristarchus. However, Copernicus revived Aristarchus's ideas at the right moment and with the best mathematical tools available at the time, and he worked them out extensively, even if he did not make sufficiently critical use of the experimental data. Thus the Copernican system became an enormously powerful weapon in the hands of those who were to finally undo the Aristotelian cosmology and thereby it created the foundations for the building of a new natural science (Quotation 3.8).

Astronomy after Copernicus developed in two directions: One branch involved philosophers and physicists who were not concerned with the astronomical details and accepted the simplified Copernican variant. With the help of the theory, a new poetic vision of the world was created (Giordano Bruno), dangerous sociopolitical conclusions were drawn (Campanella), and thus with the introduction of both physical and philosophical arguments, this variant led finally to success (Galileo). In view of its immediate historical effects, this direction is the more important of the two.

Concrete physical results, however, were achieved primarily in the other branch, which led the most systematical and precise astronomical measurements in the history of physics until then (Tycho Brahe) to a refinement of the Copernican system, namely the elliptical orbits (Kepler), to which finally, although it took almost another century, a physical interpretation could be attached (Newton). The philosophical branch will be discussed in Section 3.3, let us now follow the fate of the Copernican system in the hands of the professional astronomers.

3.2.3 *A Compromise: Tycho Brahe*

Frederick II, king of Denmark, was able to offer his court astronomer, Tycho Brahe (1546–1601), an opportunity to pursue his work in a way that was unique in the seventeenth century (Figure 3.24)—at least in Europe. It seems that only in ancient Alexandria and in the seventeenth century only outside of Europe—for example, in Samarkand at the court of Ulugh Beg—were astronomers able to work free of material want and with ample scientific equipment at their disposal.



▲ **Figure 3.23** A marginal note by Tycho Brahe in the exemplar of the book *De revolutionibus* depicted in Figure 3.9.

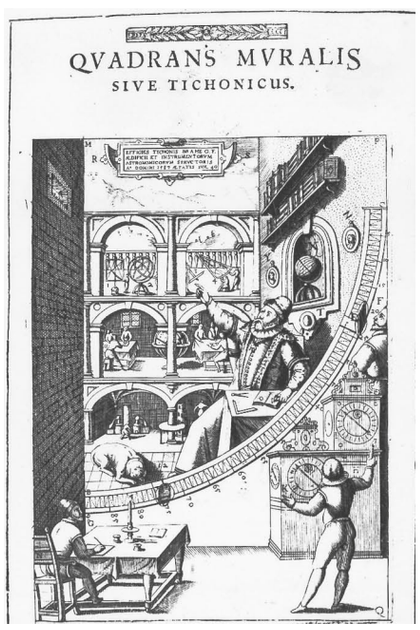
Quotation 3.8

Salviati: ... Nor can I ever sufficiently admire the outstanding acumen of those who have taken hold of this opinion and accepted it as true; they have through sheer force of intellect done such violence to their own senses as to prefer what reason told them over that which sensible experience plainly showed them to the contrary. For the arguments against the whirling of the earth which we have already examined are very plausible, as we have seen; and the fact that the Ptolemaics and Aristotelians and all their disciples took them to be conclusive is indeed a strong argument of their effectiveness. But the experiences which overtly contradict the annual movement are indeed so much greater in their apparent force that, I repeat, there is no limit to my astonishment when I reflect that Aristarchus and Copernicus were able to make reason so conquer sense that, in defiance of the latter, the former became mistress of their belief.

—Galileo, *Dialogue Concerning the Two Chief World Systems—Ptolemaic & Copernican* [p. 328]



(a)



(b)



(c)

▶ Figure 3.24
(a) Tycho Brahe was able to have such a palace constructed for his astronomical investigations. **(b)** The great mural circle. **(c)** The new star (nova) in Tycho's book. (Széchényi National Library, Budapest.)

Quotation 3.9

Last year, on November 11, after sunset, when I was contemplating the stars, as is my custom when the sky is calm, I observed, near the zenith, a new and extraordinary star, conspicuous in its brightness beyond all others. And when it was sufficiently clear to me—for I have almost from boyhood known perfectly all the stars of the sky (for there is no great difficulty in this science)—that no star had ever been visible in that part of the sky before, or perhaps only a very small one, to say nothing of one of such conspicuous splendor, I was brought to such a state of admiration of the thing that I had no shame in doubting my faith, which must be delimited by my own eyes. When I considered that the star truly shone there, and could, once its location had been pointed out, indeed be seen by others, my doubt was excited no further. Among all the miracles that have appeared in the whole of nature since the beginning of the world, this is certainly either the greatest or at least the equal of that in which the Sun's course was held back by the prayers of Joshua—or perhaps the eclipse which, as the sacred oracles attest, occurred at the moment when the heavenly Sacrifice was completed. For it is agreed upon by all the Philosophers, and the matter itself does not appear obscure, that there is no alteration, either of generation or corruption, in the ethereal region of the celestial world; that the sky and the ethereal bodies contained within it neither increase,

continued on next page

Tycho Brahe was what one might call a genuine observer. Not only did he spend almost 20 years observing and recording the motions of the planets, but he also investigated how to get the most precise results using the equipment available at the time. Thus, he concluded, for example, that the precision of astronomical measurements could not be improved beyond a certain point merely by making the instruments larger because the instruments could not be constructed with a sufficient degree of precision and also because they would not be rigid enough. Shortly before the invention of the telescope, Tycho Brahe achieved the highest degree of precision yet, with an error of only two minutes of arc (one-thirtieth of one degree) in angle measurements.

It is a fortunate coincidence that a nova (1572) and a comet (1576) were observed during his lifetime. According to his own notes, Tycho Brahe was deeply shaken at the appearance of a new star; he first put the event down to illusion or delusion and was convinced only after lengthy observations that the new star did indeed belong to the firmament along with the other, familiar stars. In the face of experimental facts, after some hesitation he finally declared (Quotation 3.9), that not only was the sublunar world subject to the laws of change—creation and destruction—but also the world of the celestial bodies, which had always been thought to be eternal and unchanging.

From his observations of the comet's motion, he concluded that comets did not—as had previously been believed—belong to the sublunar world. He also determined that the orbit of the comet crossed the orbits of the planets, therefore crystalline spheres studded with planets could not exist either. These two celestial phenomena offered telling refutations of two key theses of the Aristotelian picture of the world. Tycho Brahe was an astronomer known for the reliability of his measurements and the keenness of his observations, but theory was not his strong suit. Nonetheless, he created a hybrid of the Ptolemaic and Copernican cosmologies, which for two generations served as a third alternative—a possible interpretation for the phenomena—besides the Ptolemaic and the Copernican systems (Quotation 3.10). Tycho Brahe left Earth at the center of his system, and in this sense, his system is geocen-

tric (or, more correctly, geostatic). However, he assumed that the Sun revolves about Earth while the other planets revolve about the Sun (Figures 3.25 and 3.61). Looking at it closely, we see that this is essentially a variant of the Ptolemaic system, since there the radius of the deferent, which has no physical significance, can be chosen arbitrarily, and therefore we may take all deferent circles to coincide with the orbit of the Sun about Earth. From the outset, Tycho Brahe went out of his way to avert all objections—whether religious, ideological, or physical—to the notion of a moving Earth, basing his system on the commonsense perception of a fixed and immovable Earth under our feet. At the same time, however, he could speak of distances and orbital periods exactly as in the Copernican system: in other words, he described a physical system that could be modeled. He expected his assistant Kepler to work out a detailed exposition of his theory.



Quotation 3.9, ***continued***

nor diminish, nor vary, either in number, magnitude, brightness, or any other dimension; but rather that it remains always the same, like to itself in every respect, without a single year wearing it down.

—Tycho Brahe, *De Nova Stella* [translated by Paul Franz]

I have taken numerous observations through appropriate instruments and have found from these, by means of spherical trigonometry, that this comet was so far distant from us that its greatest parallax at the horizon cannot have been greater than 15 minutes ... It thus follows through geometrical apportionment and reckoning, that this comet was at least 230 times the earth's radius above the earth ... it is easy to conclude and understand that this comet was high above the moon in the sphere of the planet Venus ...

The Aristotelian philosophy, [therefore], which we hear so often, cannot be valid in teaching that nothing new can originate in the heavens, and that all comets are located in the upper parts of the air ...

—Tycho Brahe [Christianson 1979]

◀ **Figure 3.25** The book *Almagestum Novum*, 1651, by the Jesuit priest Riccioli, containing a great deal of observational material, enjoyed great popularity among a wide readership. Riccioli presented Tycho Brahe's system in a simplified form, giving it preference over the Copernican system, as the title page shown here clearly illustrates [Zinner 1943].

However, by this time, "progressive intellectuals" had become Copernicans. Thus, for example, Cyrano de Bergerac sarcastically quotes a priest who admits that Earth indeed rotates on its axis, but only because the damned, trying to flee the fires of hell within, are turning it like a hamster spins the wheel in his cage.

Numerus. *Mensura*. *Pondus*. These words at the fingers of God's hand refer to the following sentence in the *Book of Wisdom*: "But thou hast ordered all things in measure and number and weight." On the ground sits Ptolemy; he looks appreciatively at Tycho's system with two centers, while his own system, the geocentric, lies in the corner.

The Hebrew text above God's hand represents the Tetragrammaton, the four consonants usually transliterated YHWH (Yahweh) and used as a biblical proper name for God. It is more correctly understood as the unspeakable name of God, for one reads it as *Adonai* (Lord, Kyrios, Dominus). This ancient symbol was crafted, perhaps by Moses himself, from three Hebrew words: *haya* (was), *hove* (is), and *yihye* (will be), that is, the eternal one. The significance of this interpretation is clarified in the caption to Figure 2.31 (in part after Rabbi Tamás Raj).

Quotation 3.10

The ponderable mass of the Earth—so unsuited to motion—is certainly unable to execute a threefold motion without violating the fundamental principles of physics; even Holy Writ speaks against it. ... I began a thoroughgoing investigation as to whether a hypothesis could be found that would be in complete agreement with observed phenomena and mathematical principles without coming into conflict with physics and eliciting the censure of the theologians. My efforts in this direction were successful beyond my expectation.

I believe with certainty and without doubt that the stationary Earth must be placed at the center of the universe, in accord with the opinions of physicists and astronomers of earlier times and the witness of the Word; however, I do not feel obliged to hold with the view of Ptolemy and the ancients that the Earth is at the center of the orbits of the planets: Namely, I assert that the celestial motions are arranged in such a way that that Earth is to be viewed as the center point of the motions of only the Moon, the Sun, and the eighth sphere—the one that is most distant. The other five planets orbit the Sun as though around their leader and king, and the Sun remains always at the center of their motion, while they follow it in their yearly motion. The Sun thus rules these motions, provides their rhythm and directs alone—like Apollo among the muses—the celestial harmony of the motions circling about it.

—Tycho Brahe, *Astronomiae instauratae progymnasmata*, 1602



▲ **Figure 3.26** Johannes Kepler (1571–1630): Born in Weil der Stadt (Württemberg). Studied theology in Tübingen, where he became acquainted with the Copernican system from Michael Maestlin. From 1594, taught in Graz, and

continued on next page

Tycho Brahe was a difficult person, and as handsomely as he was compensated for his investigations, he managed to spend it all on a princely lifestyle. Thus he fell into the disfavor of Frederick II's successor, and after a period of errantry, he arrived in Prague to serve as the court astronomer to Emperor Rudolf II. Here, Tycho Brahe became aware of the *Mysterium cosmographicum* of the young Kepler. This work was suffused with Platonic ideas, but it revealed Kepler's outstanding mathematical ability, which prompted Tycho Brahe to bring Kepler from Graz to Prague. At Tycho Brahe's death, in 1601, Kepler took over as court astronomer there.

As we shall see (Quotation 3.11), Kepler attached great importance to Tycho Brahe's measurements, but, as we see below, he felt that Tycho Brahe's cosmology was nothing more than an intellectual crutch for people lacking imagination:

In their place I cannot do other than substitute solely the opinion of Copernicus on the world, and, if it were possible, persuade everyone to believe it. Yet it is still a new idea to the common herd of scholars, and a doctrine which to many is quite absurd to hear, that the Earth is one of the planets, and is carried among the stars round an unmoving Sun. Therefore, those who are shocked at the novelty of the opinion should know that these speculations about harmonies also find a place in the hypotheses of Tycho Brahe, because that author has everything else which relates to the arrangements of the bodies and the combination of their motions in common with Copernicus. It is only the Copernican annual motion of the Earth which he transfers to the whole system of the planetary spheres, and to the Sun, which occupies the middle by the agreement of both authors. For by this transference of the notion it nevertheless comes about that the Earth, if not in that vast and immense space within the sphere of the fixed stars, yet at least in the system of the planetary world, takes the same position at any given time according to Tycho Brahe as Copernicus gives it. In fact, just as someone who draws a circle on paper moves the writing leg of his compasses round, whereas someone who fastens his paper or tablet to a revolving wheel describes the same circle, without moving the leg of his compasses or his pen, on the tablet as it moves round; in the same way in this case for Copernicus indeed the Earth traces out a circle by a real motion of its own body, passing in between the circles of Mars on the outside and Venus on the inside; but for Tycho Brahe the whole planetary system (in which among the others are also the circles of Mars and Venus) turns, like the tablet on the wheel, applying to the motionless Earth, as if to the pen of the man who turns the wheel, the blank space between the circles of Mars and Venus. The effect of this motion of the system is that the Earth marks on it the same circle round the Sun, intermediate between those of Mars and Venus, while it remains motionless itself, as according to Copernicus it marks out by a true motion of its own body, with the system at rest. Therefore, as harmonic study considers the motions of the planets as eccentric, as if viewed from the Sun, it may readily be understood that if an observer were on the Sun, even though it were in motion, to him the Earth, although it were at rest (to make a concession already to Tycho Brahe), would nevertheless appear to be going around an annual circuit, placed in between the planets, and also in an intermediate period of time. Hence if there is a man whose confidence is too weak for him to be able to accept the motion of the Earth among the stars, nevertheless he will be able to rejoice in the marvelous study of this absolutely divine mechanism, if he applies whatever he is told about the daily motions of the Earth on its eccentric to their appearance from the Sun, the same appearance as Tycho Brahe shows, with the Earth at rest.

—Johannes Kepler, *The Harmony of the World* [pp. 403–404]

3.2.4 *Celestial Harmony: Kepler*

Johannes Kepler (Figure 3.26) is one of the most fascinating personalities in the history of science: He was a religious man, enthralled by numerological mysticism, with an overflowing imagination, and yet he was prepared to look at facts with the utmost critical capacity. Reading his works is made difficult, but also fascinating—and in contrast to other greats of science—by the fact that he describes in detail the difficulties and false starts encountered in his research. If one pushes on to the results, one is captivated by Kepler's rapturous amazement, tinged with a degree of self-satisfaction, at the harmony he discovered. Descartes's *Discours* is also considered by many to be an autobiographical account in which the author describes his doubts and the gestation of the truths born of those doubts as they evolved. However, in the case of Descartes, we are dealing with a clear and concise train of thought, while with Kepler even the fundamental laws that bear his name can be teased out of his prose only with the greatest difficulty. The disciplined thoughts of Descartes appear in logical order, even *in statu nascendi*, whereas with Kepler they pour forth in an upheaval bearing all the joy and agony of their birth.

The basic idea in Kepler's early work *Mysterium cosmographicum*, which even in his later years he did not in any way repudiate, is that in the days of creation, God had the idea of arranging the planetary orbits in association with the five Platonic solids. The complicated spatial arrangements in Figure 3.27 are from Kepler's book.

The five regular solids are arranged in the following order about the Sun: the Sun is enclosed within an octahedron; then follow an icosahedron, a dodecahedron, a tetrahedron, and finally a hexahedron (cube). These nested Platonic solids determine the successive planetary orbits. Kepler even took care to note that the Platonic solids must be nested in such a way that enough space remained for the planetary epicycles. As a natural starting point, Kepler accepted the variant of the Copernican system in which the motion of the planets in their orbits about the Sun is corrected by epicycles. This model reflects the relationships between the average distances of the planets from the Sun to a surprisingly close degree of approximation. However, when one considers that the relationships among these distances do not in and of themselves characterize the solar system, the model loses all its significance: One may imagine a system with an arbitrary number of planets just as well as one with a fixed number, but having completely different distances from the Sun. It was only in 1619 that Kepler discovered a relationship between the distances from the Sun and the orbital periods of the planets, and it is this relationship that later allowed sweeping conclusions about the nature of the forces that operate between the Sun and the planets.

That Kepler discovered the three laws that bear his name, which are still valid and have become a standard part of the school curriculum, can be ascribed not only to the man's genius, but also to the fortunate concurrence of certain circumstances, which can be summarized as follows:

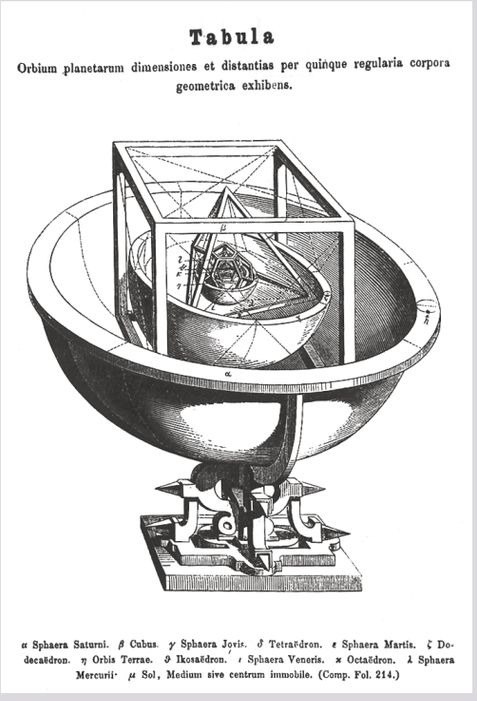
1. Kepler had at his disposal—after some wrangling with Tycho Brahe's estate—the entire collection of his predecessor's observations, which was the most precise and detailed set of measurements available at that time.
2. Of all the easily observed and well-studied planets, Mars possesses the orbit with the greatest eccentricity; therefore, the deviation of the orbit of Mars from a perfect circle can be easily measured.
3. The orbit of Earth is very close to circular, that is, it is an ellipse with very small eccentricity.

Figure 3.26 *continued*

composed astrological calendars with prognostications. The *Mysterium Cosmographicum* (1596) was also published here. In 1600, assistant to Tycho Brahe; from 1601, Brahe's successor in Prague as court astronomer to Rudolf II. Here in 1609, published the *Astronomia nova seu Physica coelestis* with a presentation of the first and second Keplerian laws, and the *Dioptrice* (1611), which contains the refraction law for small angles and the theory of the Galilean and Keplerian telescope. 1613–1626: Active in Linz; in this period, his treatise on how to determine the volumes of wine casks appeared: *Nova Stereometria doliorum vinariorum* (1615).

1619: Publication of the *Harmonices mundi*, in which we find Kepler's third law. The *Tabulae Rudolphianae* (1627) was long considered the most precise collection of tables of important astronomical data.

Kepler died in 1630, on a journey to the parliament in Regensburg, where he hoped to persuade the emperor to pay his salary, which was in arrears. Characteristic of Kepler is the following remark: *Ut oculus ad colores, auris ad sonos, ita mens hominis non ad quaevis sed ad quanta intelligenda condita est.* (As the eye was created for colors and the ear for sounds, so was human intelligence created for grasping not the qualitative but the quantitative.)



▲ **Figure 3.27** This is how Kepler imagined the relationship between the six planetary orbits and the five Platonic solids.

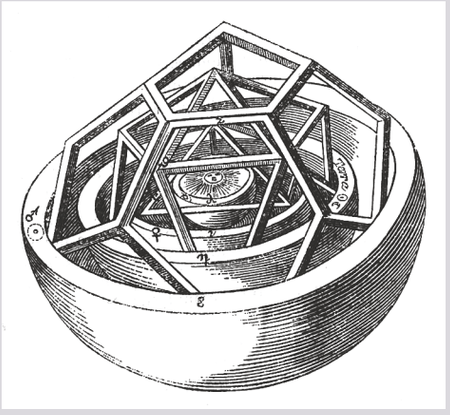
The model of the arrangement of the Platonic solids was constructed by Christophorus Leibfried in 1597. (University Library, Budapest.)

continued on next page

▼ Figure 3.27 *continued*

The Earth is the circle which is the measure of all. Construct a dodecahedron around it. The circle surrounding that will be Mars. Round Mars construct a tetrahedron. The circle surrounding that will be Jupiter. Round Jupiter construct a cube. The circle surrounding that will be Saturn. Now construct an icosahedron around the Earth. The circle inscribed within that will be Venus. Inside Venus inscribe an octahedron. The circle inscribed within that will be Mercury. There you have the explanation of the number of the planets.

—Johannes Kepler, *Mysterium Cosmographicum: The Secret of the Universe* [p. 69]





276 De Motibus Planetarum Harmonicis.

Fig. 45.

simi, in opposita proxime Solem veniant. In adjecto schemate pro singulis planetis facti sunt circuli terni, quorum nullus indicat viam ipsam planetae eccentricam, sed medius quidem, ut in Marte BE, est aequalis orbitae eccentricae, quoad longiorem ejus diametrum; orbita vero ipsa, ut AD, tangit AF superiorem ex tribus ab una plaga A, inferiorem CD ex opposita D. Circulus GH, punctis abumbratus et per centrum Solis descriptus, indicat viam Solis apud Tychonem Brahe; qui si hanc viam movetur, omnia omnino puncta totius systematis planetarii, hic depicti, viam incedunt aequalem, quilibet suam propriam. Et uno ejus puncto, scilicet centro Solis, in una sui circuli parte stante, ut hic in infima, omnia omnino puncta systematis stabunt in imis partibus suorum quodque circulorum. Veneris etiam tres circuli propter angustiam spatii coierunt in unum praeter voluntatem.

Tertio repetat lector ex Mysterio meo Cosmographico, quod edidi ante 22 annos: numerum planetarum seu curriculorum circa Solem desumtum esse a sapientissimo Conditore ex quinque figuris regularibus solidis, de quibus Euclides jam ante tot secula librum scripsit, qui Elementorum dicitur, eo quod ex propositionum serie sit constructus. Quod autem plura corpora regularia nequeant esse, hoc est, quod figurae planae regulares nequeant saepius quam quinquies in solidum congruere, id libro secundo hujus operis patuit.

Quarto, quod attinet proportionem orbitarum planetariarum, illa quidem inter binas orbitas vicinas semper est tanta, ut facile appareat, unamquamque illarum appropinquare proportioni unicae orbium unius ex quinque solidis figuris, scilicet orbis circumscripti ad orbem inscriptum figurae; at non tamen plane aequalis est, ut olim ausus fui de perfecta denique astronomia polliceri. Nam post absolutam demonstrationem intervallorum ex observationibus Brahei deprehendi hoc: cubi quidem angulis

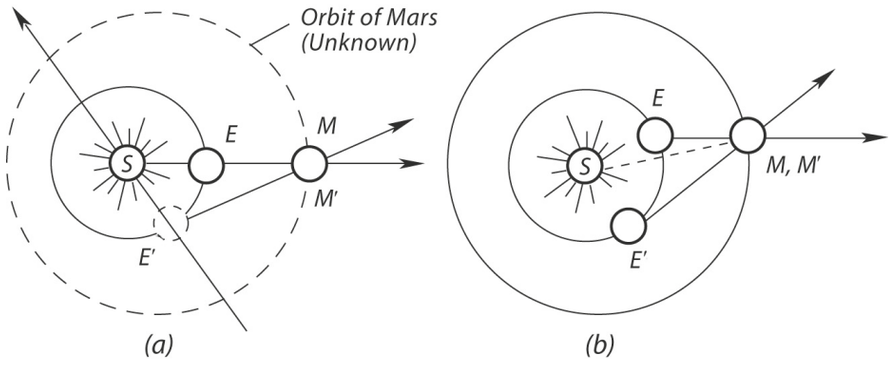
From the point of view of the history of science, the death of Tycho Brahe can be considered a positive event. He wanted his assistant Kepler to confirm his theory using his observational data, which, as we can see today, would have been a thoroughly fruitless undertaking. After Tycho Brahe's death, Kepler was free to follow his own ideas. Nonetheless, Kepler did hold Tycho Brahe's work in high esteem (Quotation 3.11).

At the beginning of his investigations, Kepler compared the observational data for the orbit of Mars with the theoretical results for both the Ptolemaic and Copernican systems. He did not share with Copernicus the idea of rejecting the use of equants, and he attempted to find ways of building them into the Copernican system when it turned out that the calculations in the original Copernican system led to results just as erroneous as those of the Ptolemaic system. He was hoping to force greater agreement between prediction and observation with the help of equants. It is indicative of the high value that Kepler placed on Tycho Brahe's data that the difference between theory and observation was a mere eight angular minutes. But Kepler knew that his predecessor's measurements were accurate to within two minutes, and he searched relentlessly for the cause of the discrepancy. It was this stubborn search that led to the discovery of elliptical orbits.

The train of thought that we are about to describe by which Kepler determined the shape of the planetary orbits belongs in its simplicity to the most remarkable achievements in the history of science. The challenge was to determine the orbit of a particular planet, in this case that of Mars, from the available observational data. We wish to emphasize that such a task is much more difficult than the basic problem in the Ptolemaic system, namely to describe the *apparent motion of a planet*, because there only the angular coordinates of the planet come into play in the calculations. Here, however, we are dealing with spatial planetary orbits for which not only the angular coordinates in which we observe the planet but also the distances from Earth and from the Sun have to be taken into account. Moreover, we must not forget that the measurements of the celestial bodies were performed from Earth, *whose orbit and orbital motion were as yet not known*.

Kepler first set himself the task of determining the orbit of Earth. We begin with the configuration *SEM* (Figure 3.28(a)), in which the Sun, Earth, and Mars all lie along a straight line. The Martian orbit is the dashed curve in the figure, so indicated to emphasize that we do not know this orbit at the outset, only a single position of the planet, and even this position only in terms of its direction. Kepler began with the observation that the orbital period of Mars around the Sun is 687 Earth days, and as a result, after that period of time, Mars will have returned to its initial position at the start of the observations. However, at that time, Earth will be located at a different point E' in its orbit. This point can be determined because we know the directional lines between Mars and Earth and between the Sun and Earth, and the intersection of those two lines yields Earth's location E'. After another 687 days, Mars will again be at its initial position, while Earth will have moved in its own orbit to a new position E''. From this position, the Mars–Sun angle can again be measured, and thus the position of Earth can again be determined, without having to know anything about the shape of the Martian orbit. It must, however, be noted here that the orbital measurements thus obtained are relative distances, determined in units of the mean distance between Earth and the Sun.

Once the orbit of Earth is known, a slight modification of the above procedure makes it possible to determine the orbit of Mars (Figure 3.28(b)). The next step is to select two positions of Earth in its orbit, separated by a period of 687 days.



◀ **Figure 3.28** **(a)** How Kepler determined the orbit of Earth. We begin with the configuration *SEM* and observe that after 687 days, Mars is again at its initial position (*M′*), whereas after this period of time, Earth finds itself at position *E′*. After measurement of the directions *E′M′* and *SE*, point *E′* is determined. After another 687 days, we determine an additional point of Earth's orbit.

(b) How Kepler determined the orbit of Mars when the orbit of Earth is known.

This is exactly the Martian orbital period, so the position of Mars will be identical for these two positions of Earth. If we now draw Earth's orbit, enter these two positions, and measure the direction of Mars from each position, then the intersection of the two direction lines yields a point on the orbit of Mars. The task is now to determine the directions in which Mars is to be observed for all pairs of Earth orbital positions that are 687 days apart. From the intersections of all such pairs of lines thus obtained, we can reconstruct the Martian orbit. Kepler had no need to carry out this entire series of observations: All he had to do was to consult Tycho Brahe's data for the corresponding dates. In addition, these data allow the determination of a planet's velocity along its orbit because the construction of the orbit automatically yields the times for the individual orbital points.

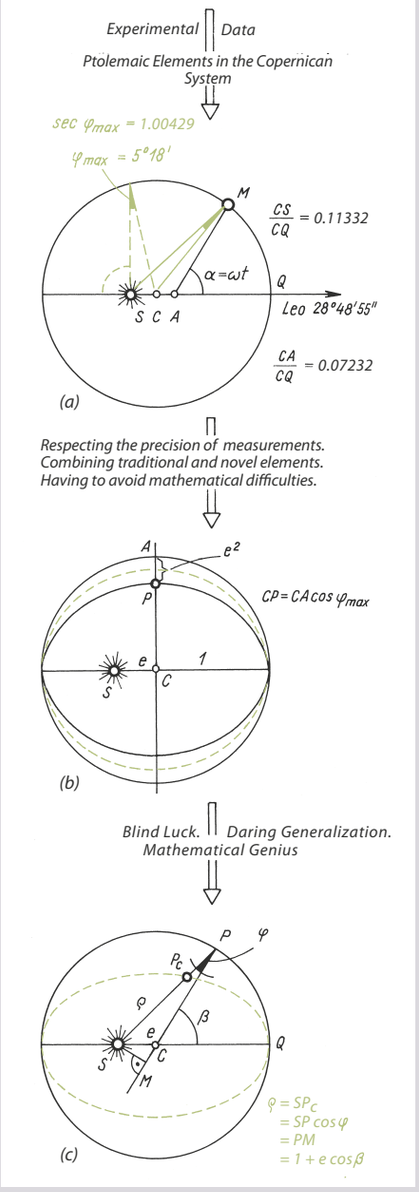
Thus, we have at our disposal the empirically determined orbital elements for the planet Mars. How, then, from these can Kepler's famous laws be derived?

As already mentioned, Kepler introduced daring and decisive innovations in two respects. First, he abandoned the requirement of circular orbits, thereby breaking once and for all with the Greek legacy. Second, he established a connection between motion and the force acting on the body, thereby taking the first steps in the direction of a physical astronomy, or more generally, a physical theory of motion. In the face of such decisive ideas, one would like to know how the scientist arrived at the result, and in getting the result what role was played by imagination, by audacity, by the knowledge of related sciences, or by sheer luck. We have already mentioned that, to our good fortune, Kepler described in his writings not only his results, but also the path to those results, with all the detours and blind alleys. We cannot hope to lay out Kepler's complete thought process, for such would require an entire book. Dijksterhuis [1956] has attempted to reconstruct Kepler's argumentation in abbreviated form. We shall attempt to simplify that presentation even further.

Following Copernicus, Kepler first tried to describe the orbit of Mars as a combination of circular orbits. In contrast to Copernicus, however, he did not mind using equants, as Ptolemy did, and thus he finally arrived at the Martian orbit depicted in Figure 3.29(a). The orbit thus obtained was in relatively good agreement with observation; however, there were points at which the discrepancy was greater than what was allowed by the accuracy of Tycho Brahe's measurements, namely, those eight minutes of arc referred to in Quotation 3.11. Furthermore, based on the measurements pertaining to Earth's orbit, Kepler believed that equants had to be used there also.

The Martian orbit thus obtained played a significant role in Kepler's further deliberations, although he viewed it as merely a provisional hypothesis (*hypothesis vicaria*). Kepler thus was able to make statements that are very similar to his first two laws, namely, that the Sun is located at a point removed from the center of the planetary orbit and that the velocity of a planet is inversely proportional to its distance from the Sun. The latter of these, as mentioned above, is only approximately correct. However, Kepler is thereby—without even yet mentioning elliptical orbits—already in possession of his second (correct) law: the law of areas.

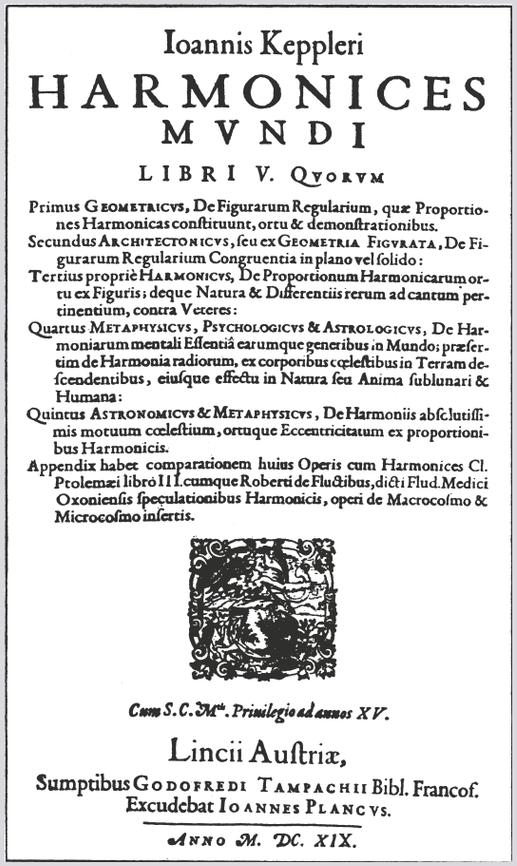
With the data on Mars's distance from the Sun and (based on his hypothesis related to the orbit) the positional data for Mars with respect to the Sun, Kepler could use three pairs of these values to determine the location of the orbit's aphelion (greatest distance from the Sun) and the eccentricity. Kepler had a sufficient amount of data and the corresponding pairs of data triples, each of which led to further orbital data. From these he concluded that the Martian orbit cannot be properly described with Copernican circles, even if these were supplemented with equants. He discovered further that the deviations became greater the farther the planet was located from its aphelion or perihelion. Therefore, the orbit must have an oval shape, as pictured in Figure 3.29(b). Such an oval orbit can be obtained also in the Ptolemaic



▲ **Figure 3.29** **(a)** Kepler's first attempt at a mathematical description of the experimentally determined orbit of Mars. He supplemented the simple Copernican system with equants.

(b) The appearance of the idea of an elliptical orbit: Kepler introduced it to make the mathematical computations for the application of the law of equal areas easier.

(c) The correct mathematical description of the planetary orbits.



Ioannis Keppleri

HARMONICES MVNDI

LIBRI V. QVORVM

Primus GEOMETRICVS, De Figurarum Regularium, quæ Proportiones Harmonicas conſtituunt, ortu & demonſtrationibus.
Secundus ARCHITECTONICVS, ſeu ex GEOMETRIA FIGVRATA, De Figurarum Regularium Congruentia in plano vel ſolido:
Tertius propriè HARMONICVS, De Proportionum Harmonicarum ortu ex Figuris; deque Natura & Differentiis rerum ad cantum pertinentium, contra Veteres:
Quartus METAPHYSICVS, PSYCHOLOGICVS & ASTROLOGICVS, De Harmoniarum mentali Eſſentiâ earumque generibus in Mundo; præſertim de Harmonia radiorum, ex corporibus cœleſtibus in Terram deſcendentibus, eiuſque effectu in Natura ſeu Anima ſublunari & Humana:
Quintus ASTRONOMICVS & METAPHYSICVS, De Harmoniis abſolutiſſimis motuum cœleſtium, ortuque Eccentricitatum ex proportionibus Harmonicis.
Appendix habet comparationem huius Operis cum Harmonices Cl. Ptolemæi libro III. cumque Roberti de Fluctibus, dicti Flud. Medici Oxonienſis ſpeculationibus Harmonicis, operi de Macrocoſmo & Microcoſmo inſertis.

Cum S. C. M^tis. Priuilegio ad annos XV.

Lincii Auſtriæ,
Sumptibus GODOFREDI TAMPACHII Bibl. Francof.
Excudebat IOANNES PLANCVS.

ANNO M. DC. XIX.

▲ **Figure 3.30** Title page of the book *The Harmony of the World*. A precise translation of the complete title would be, "Five books on the harmony of the world." (University Library, Budapest.)

Quotation 3.11

Since the divine benevolence vouchsafed us by TYCHO BRAHE, a most diligent observer, from whose observations the 8′ error in this Ptolemaic computation is shown, it is fitting that we with thankful mind both acknowledge and honor this benefit of God. For it is in this that we shall carry on, to find at length the true form of the celestial motions, supported as we are by these arguments showing [the Ptolemaic] suppositions to be fallacious. ... Now, because they could not have been ignored, these eight minutes alone will have led the way to the reformation of all astronomy, and have constituted the material for a great part of the present work.

—JOHANNES KEPLER, *New Astronomy* [p. 286]

system if an additional epicycle is introduced. In order to calculate an area such as was necessary for the law of areas, KEPLER introduced an ellipse as an approximation to the oval orbit with major axis equal to the diameter of the previously introduced circular orbit. As minor axis, he chose the width of the orbit as measured in the middle. Thus it was that in order to avoid mathematical difficulties and under the influence of Ptolemaic ideas, KEPLER arrived at an elliptical orbit. The width of the crescent-shaped area between the circle and the oval turned out to be $e^2 = 0.00858$. In comparing his elliptical orbit with the observational data, KEPLER now found that an excellent approximation would result if the width of the crescent were to have half the given value, that is, 0.00429. And here is where chance enters the picture: In his calculations of the position of Mars, KEPLER had very often made use of the secant of the maximal value of the angle marked in red in Figure 3.29(a), namely 5°18′, which has the value 1.00429. This value 0.00429 increased by 1 represents the "optimal" width of the crescent. Therefore, according to Figure 3.29(b), we must have the relationship

$$\frac{CA}{CP} = \frac{1}{1-e^2/2} \approx 1 + \frac{e^2}{2} = 1.00429,$$

which can be rewritten as

$$CP = CA\cos\varphi_{\max}.$$

Suppose—thus KEPLER's daring generalization—this relationship were valid for every point of the orbit; then the following relationship would hold for the distance from the Sun (Figure 3.29(c)):

$$\rho = SP_C = SP\cos\varphi = PM = 1 + e\cos\beta.$$

It only remained for KEPLER to prove that this relationship in fact describes the desired elliptical orbit, for which the agreement with the measurement data would be perfect. It would lead us too far afield to present KEPLER's proof, even in the simplified form presented by DIJKSTERHUIS. Consider, however, that we are still half a century from the Cartesian coordinate geometry, and therefore KEPLER had to overcome some truly daunting mathematical difficulties.

The first and second of KEPLER's laws appeared in his 1609 work *Astronomia nova*. The year 1609 is thus frequently cited as the starting point of the development of modern astronomy. Indeed, KEPLER was the first to reject the perfect circular motion of Platonic and Aristotelian physics, and with the laws that he discovered, it became possible to develop descriptive astronomy into a physical astronomy. That is how KEPLER himself must have sensed it, for in the complete title of his book, there appears the word *aitiologetos*, that is, logical explanation, but also the term *physica coelestis*. This choice of words suggests that KEPLER had also searched for a physical cause for planetary motion, although he still worked in the Aristotelian style where one relates the cause of motion to the velocity, not to the change in velocity. Yet, he did manage to observe that in some way, the Sun has an effect on the planets and that this effect falls off with increasing distance. KEPLER quantified this decrease in the form $1/r$, rejecting the correct inverse quadratic relationship on the assumption that the effect spreads only in the orbital plane in which case it will decrease as the first power of the orbital radius. If we employ the Aristotelian law of motion, with velocity proportional to force, then in today's notation we would obtain the relationship

$$v \propto \frac{1}{r},$$

from which it immediately follows that the product vr is equal to a constant. This relationship very roughly means that a greater distance from the Sun implies a smaller velocity, and a smaller distance implies a greater velocity. Today we know that the force is acting in the direction of the line between the Sun and the planet, *and not, as KEPLER still assumed, along the tangent to the orbit*. The correct interpretation of the action of the force was not expressed qualitatively until a generation after KEPLER, and it took two additional generations for a quantitative formulation.

Kepler's third law is to be found in his 1619 book *Harmonices mundi* (Figure 3.30), in which he was finally able to describe the harmony for which he had long searched: *The square of the time traveled by a planet in its orbit is proportional to the cube of the planet's mean distance from the Sun.*

Quotation 3.12 demonstrates the enthusiasm with which Kepler informed the world of this wonderful discovery. From the point of view of the history of science, this theorem is of great significance because Newton used it as an aid in formulating the universal theory of gravitation. Figure 3.31 depicts Kepler's three laws as we represent them today but described with Kepler's own lofty words.

Kepler's importance goes beyond astronomy. His investigations in optics and mathematics will be discussed in Sections 3.5 and 4.1. Here we would like to return to our earlier observation of the degree to which Kepler's capacity for precise observation and formulation of exact laws depended on his deep religiousness and inclination to numerology. As an example of the latter, he explained the reason for the number of planets by arguing that using five Platonic solids one can determine the positions of exactly six planets. Later, he gave the number of moons associated with each planet simply by assuming that these numbers should form a geometric progression. Therefore, since Earth has a single moon, Mars should have two, Jupiter four, and Saturn eight. Kepler understood celestial harmony as more than a metaphor; he linked the numerical relationships in the solar system with the harmonies of the musical scale. Thus the quotients of the angular velocities of the planets at perihelion and at aphelion should be for Earth, 16/15 (major second); for Mars, 3/2 (perfect fifth); and for Saturn, 5/4 (major third).

Kepler took the notion "music of the spheres" seriously, even assigning musical motifs to the planets (Figure 3.32).

It is characteristic of his religious mysticism that Kepler associated the fixed Cosmos with the Trinity, with the Sun representing the creative, life-giving Father; the fixed stars, the Son; and interstellar space, the Holy Ghost. This structuring of the immovable objects strengthened Kepler in his belief that the movable objects, among them the planets, must be parts of a divine harmony.

One of Kepler's tasks as court astronomer was the casting of horoscopes. Despite all his mysticism, he seems nonetheless not to have believed in astrology, for he observed, somewhat ironically, that God has given all of His creations a livelihood, and to the astronomer, He has given astrology.

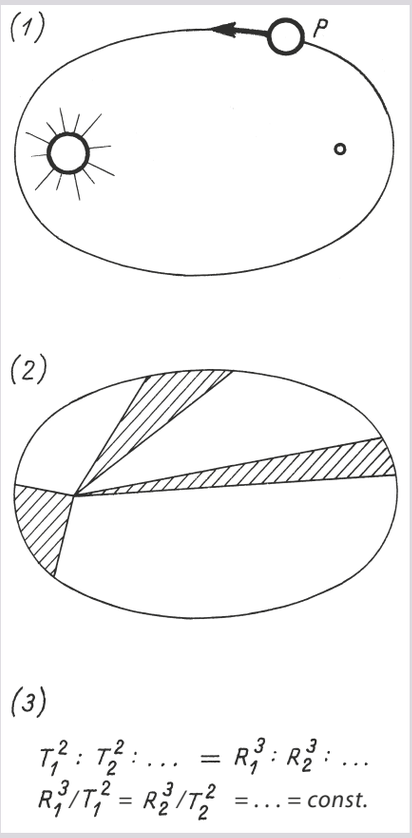
Many artists have been inspired by Kepler's life and his enthusiasm for the beauty of the laws underlying natural phenomena. In the twentieth century, the life of Kepler served the composer Hindemith as inspiration for an opera and for the symphony *Harmonia mundi*.

Kepler wrote his own epitaph:

> Mensus eram caelos, nunc terrae metior umbras;
> Mens coelestis erat, corporis umbra jacet.

(In life I measured the heavens, now I measure the darkness of Earth; Mind came down from the heavens, earth covers only my body.)

> Our admiration for this splendid man is accompanied by another feeling of admiration and reverence, the object of which is no man but the mysterious harmony of nature into which we are born. The ancients already devised the lines exhibiting the simplest conceivable form of regularity. Among these, next to the straight line and the circle, the most important were the ellipse



▲ **Figure 3.31** Kepler's three laws:

> O Thou Who dost by the light of nature promote in us the desire for the light of grace, that by its means Thou mayest transport us into the light of glory, I give thanks to Thee, O Lord Creator, Who has delighted me with Thy makings, and in the works of Thy hands have I exulted. Behold! Now, I have completed the work of my profession, having employed as much power mind as Thou didst give to me; to the men who are going to read these demonstrations I have made manifest the glory of Thy works, as much as of its infinity as the narrows of my intellect could apprehend.
>
> —Johannes Kepler, *The Harmony of the World* [trans. Wallis 1995, p. 240]

> I have demonstrated from the sure observations of Brahe that daily arcs, which are equal in one and the same eccentric circle, are not traversed with equal speed; but that these differing *delays in equal parts of the eccentric observe the ratio of their distances from the sun,* the source of movement; and conversely, that if equal times are assumed, namely, one natural day in both cases, the corresponding *true diurnal arcs of one eccentric orbit have to one another the ratio which is the inverse of the ratio of the two distances from the sun.* Moreover, I demonstrated at the same time that *the planetary orbit is elliptical and the sun, the source of movement, is at one of the foci of the ellipse.* …

continued on next page

Figure 3.31 *continued*

[A]t last the right ratio of the periodic times to the spheres … was conceived mentally on the 8th of March in this year One Thousand Six Hundred and Eighteen but unfelicitously submitted to calculation and rejected as false, finally, summoned back on the 15th of May, with a fresh assault undertaken, outfought the darkness of my mind by the great proof afforded by my labor of seventeen years on Brahe's observations and meditation upon it uniting in one concord, in such fashion that I first believed I was dreaming and was presupposing the object of my search among the principles. But it is absolutely certain and exact that *the ratio which exists between the periodic times of any two planets is precisely the ratio of the 3/2th power of the mean distances* … provided, however that the arithmetic mean between the two diameters of the elliptical orbit be slightly less than the longer diameter.

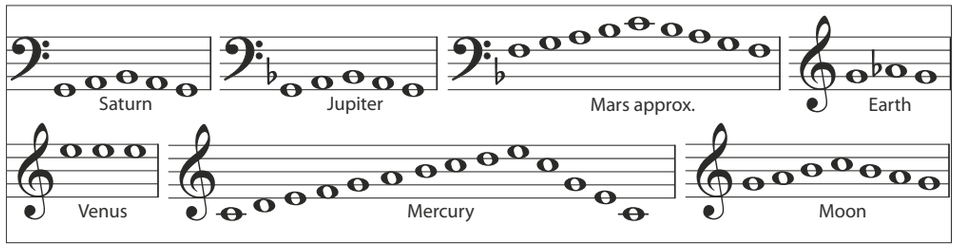
—Johannes Kepler, *Epitome of Copernican Astronomy* [pp. 178, 180]

Quotation 3.12

As regards that which I prophesied two and twenty years ago (especially that the five regular solids are found between celestial spheres), as regards that of which I was firmly persuaded in my own mind before I had seen Ptolemy's *Harmonies*, as regards that which I promised my friends in the title of this fifth book before I was sure of the thing itself, that which, sixteen years ago, in a published statement, I insisted must be investigated, for the sake of which I spent the best part of my life in astronomical speculations, visited Tycho Brahe, and took up residence in Prague: finally, as God the Best and Greatest, Who had inspired my mind and aroused my great desire, prolonged my life and strength of mind and furnished the other means through the liberality of the two Emperors and the nobles of this province in Austria-on-the-Anisana: after I had discharged my astronomical duties as much as sufficed, finally, I say, I brought it to light and found it to be truer than I had even hoped, and I discovered among the celestial movements the full nature of harmony, in its due measure, together with all its parts unfolded … not in that mode wherein I had conceived it in my mind (this is not last in my joy) but in a very different mode which is also very excellent and very perfect …

But now since the first light eight months ago, since broad day three months ago, and since the sun of my wonderful speculation has shone fully a very few days ago: nothing hold me back. I am free to give myself up to the sacred madness, I am free to taunt mortals with the frank confession that I am stealing the golden vessels of the Egyptians, in order to build of them a temple for my God, far

continued on next page



▲ **Figure 3.32** The musical motifs of the planets. (From Johannes Kepler, *The Harmony of the World,* after [Wallis, p 199].)

and the hyperbola. We see the last two embodied—at least very nearly so—in the orbits of the heavenly bodies.

It seems that the human mind has first to construct forms independently before we can find them in things. Kepler's marvelous achievement is a particularly fine example of the truth that knowledge cannot spring from experience alone but only from the comparison of the inventions of the intellect with observed fact.

—Albert Einstein, "Johannes Kepler" [pp. 265–266]

3.3 Galileo and Those Who Stood in His Shadow

3.3.1 *The Unity of the Celestial and Terrestrial Spheres*

For the general public as well as for scientific circles, Galileo (Figure 3.33) occupies a prominent place among the initiators and champions of the scientific revolution of the seventeenth century. This estimation can be attributed to his actual contributions to the natural sciences but also to his tragic fate. He was the first to come into conflict with powerful men and institutions because of a concrete physical theory and concrete facts and not because of general ideas about this world or the next.

A contributing factor to the emphasis on Galileo's central role is that it fits a common template whereby we simply associate a particular name with a particular idea, regardless of whether there were other important contributors. Just consider the quick associations—Greek philosophy: Aristotle and Plato; classical mechanics: Newton; and the theory of relativity: Einstein.

Another motive for overvaluation can be found in pedagogy, which seeks role models—perfect in their scientific as well as human conduct—and where such are lacking, they are fabricated.

As a result, in place of objective evaluation, a myth arises, which is more believable than myths of heroic or saintly deeds, for here, beneath the myth, lies an indisputable accomplishment (Quotation 3.13).

The danger of overvaluation becomes most acute in later, more sober, times, when the attempt is made to assign things to their proper places, for then we can see similar exaggeration, but in the opposite direction. In the case of Galileo, his human foibles, his nettlesome personality, his vanity, and his tendency to self-promotion have come to light (Quotation 3.14). In evaluating his scientific contributions, the roles of his predecessors, who employed similar methods or came up with similar findings, have been exaggerated. Thus, it was possible to call into question the originality of much of Galileo's work, particularly with the rediscovery of the results obtained in the Middle Ages, and even to diminish Galileo's heroic fight for truth, which, after all, ended in compromise.

First heroized and then demoted, Galileo is recognized by today's historians of science as having played a central role in the development of modern science. The debate now centers on determining the nature of his greatness more precisely, especially by looking at the work of his predecessors and contemporaries.

In what follows, we will first discuss Galileo's work in astronomy and mechanics; then we will look at his experimental methods and discoveries as well as his influence on the further development of science. Finally, we will attempt to present our current understanding of Galileo's importance based on the general assessment given here.

Although Galileo—as can be seen from his correspondence—was an adherent of the heliocentric system already in 1597, it was not until he became deeply involved with astronomy that he was truly convinced of its correctness. The impetus came from the news that a telescope had been constructed in the Netherlands. Galileo interrupted his research on motion, built a telescope for himself, and subsequently observed some remarkable celestial phenomena that had never been seen before. His observations were immediately published in the pamphlet *Sidereus nuncius* (Starry Messenger); see Figure 3.34. All of these phenomena would serve not only to strengthen Galileo in his belief in the Copernican system but also to demonstrate the insupportability of the entire edifice of Aristotelian physics.

As we have frequently seen in our discussions thus far, one of Aristotle's cardinal claims was that celestial and terrestrial phenomena were different in all respects: While on Earth all is transient, coming into being and passing away, the celestial sphere is unchanging, and there can be no talk of creation and dissolution. Motion on Earth can be disorderly and chaotic, whereas, in contrast, the celestial spheres course through their regular circular paths according to eternal and immutable laws. Bodies on Earth can assume the most varied forms, whereas those in the heavens are invariably perfect spheres. As objects of a higher order, the celestial spheres beam forth light; Earth alone is dark and gray.

But what do we now see in this ideal celestial realm? Let us begin with the Moon. With his telescope, Galileo could clearly see mountains, and from that time onward, the assertion that the surface of the Moon is similar to that of Earth could never be written off as the product of an overactive imagination. More important was his conclusion that the observable dark portion of the Moon, the part not lit up by the Sun, is also illuminated—admittedly so weakly that it is scarcely perceptible—by some other light source. Galileo correctly pointed out that this light actually comes from Earth, which reflects some of the light it receives from the Sun onto the Moon. He was the first to see that the wandering stars, that is, the planets, appear in the telescope as small round disks, while the fixed stars continue to be seen as bright points of light. He also interpreted this observation correctly, and his discovery of the phases of Venus served only to strengthen his argument that the planets, unlike the fixed stars, do not produce their own light, but acquire their illumination from the Sun (Figure 3.35(a)).

Galileo observed sunspots as well. However, he was not the only one to observe this phenomenon, and it is even probable that others—such as the Jesuit priest Scheiner, who appeared in the case against Galileo—had made this observation earlier. Galileo was able to determine that the sunspots move, and correctly inferred that this motion is due to the rotation of the Sun about its axis.

All of these observations convincingly established that terrestrial and celestial phenomena do not differ essentially in their nature and that it is only logical to include Earth among the other planets. But for the opponents of the Copernican teachings—and for its supporters as well—it was somewhat difficult to accept

Quotation 3.12, ***continued***

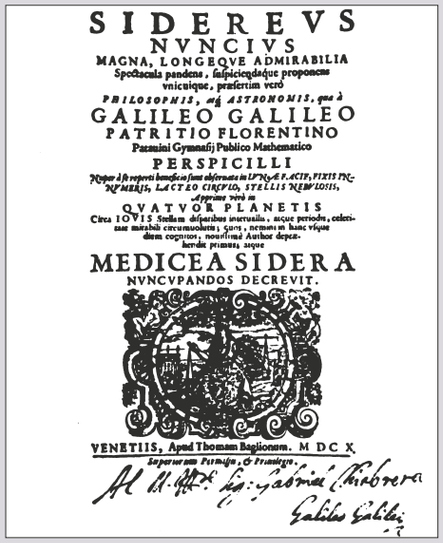
from the territory of Egypt. If you pardon me, I shall rejoice; if you are enraged, I shall bear up. The die is cast, and I am writing the book – whether to be read by my contemporaries or by posterity matters not. Let it await its reader for a hundred years, if God Himself has been ready for His contemplator for six thousand years.

—Johannes Kepler, *The Harmonies of the World* [trans. Wallis, pp. 169–170]



▲ **Figure 3.33** Galileo Galilei (1564–1642): Born in Pisa; his father, Vincenzo Galilei, was a mathematician and musician. Studied medicine at the University of Pisa; 1585: In Florence, learned of the work of Archimedes. 1589: Professor of mathematics at Pisa, where he wrote his treatise *De motu*, still entirely in the Aristotelian spirit. From 1592: Taught at the University of Padua. 1610: Left Padua to enter the service of the Medici in Florence; from this time on, principal mathematician of the grand duke of Tuscany. 1610: Publication of *Sidereus nuncius*, in which are presented the observations that he made with the telescope that he constructed. 1613: Publication of *Letters on Sunspots*. In 1616, he was officially warned against publicizing the Copernican theory. 1623: Publication of *Saggiatore*. In 1632, he presented his best-known work, *Dialogo sopra i due massimi sistemi del mondo*, to the public. This book led to the famous trial in 1633, in the course of which Galileo retracted his teachings. After the trial until his death, under house arrest in Arcetri, near Florence. In 1636, he finished work on the *Discorsi delle due nuove scienze*, which was published by Elsevier, in the Netherlands, in 1638.

Figure 3.34 Title page of *Sidereus nuncius*.


SIDEREVS
NVNCIVS
MAGNA, LONGEQVE ADMIRABILIA
Spectacula pandens, suspiciendaque proponens
vnicuique, præsertim verò
PHILOSOPHIS, nec non ASTRONOMIS, quæ à
GALILEO GALILEO
PATRITIO FLORENTINO
Patauini Gymnasij Publico Mathematico
PERSPICILLI
Nuper à se reperti beneficio sunt obseruata in LVNA FACIE, FIXIS PLVRIBVS, LACTEO CIRCVLO, STELLIS NEBVLOSIS,
Apprimè verò in
QVATVOR PLANETIS
Circa IOVIS Stellam disparibus interuallis, atque periodis, celeritate mirabili circumuolutis; quos, nemini in hanc vsque diem cognitos, nouissimè Author deprehendit primus, atque
MEDICEA SIDERA
NVNCVPANDOS DECREVIT.
VENETIIS, Apud Thomam Baglionum. M DC X.
Superiorum Permissu, & Priuilegio.
Galileo Galilei

Quotation 3.13

The matter is complicated even further by the fact that besides the genuine, though often one-sided, portraits of Galileo based on the study of his works, there exists a spurious picture presented by what may be called the Galileo-myth, the current popular representation. This latter is the picture that has been produced and is perpetuated by writers on modern physical science who feel the need of an historical introduction, but omit to perform the simple duty of exactness, which consists in verifying their statements from original sources. This picture is totally false, but it is much more vivid than any of the genuine ones, and consequently many readers will be inclined to be content with it. Moreover, it tends to simplify matters greatly: its splendor eclipses all secondary figures. It also furnishes a simple terminology because, whenever a term is needed to characterize the special features of classical science, the adjective Galilean at once suggests itself. It is therefore comprehensible that any criticism of this idealized picture tends to cause annoyance, while some Italian authors appear to feel hurt by it in their national pride. ...

Some people are inclined to call Galileo a martyr to science. Those who do so either are ignorant of the course of the trial of Galileo or do not know what a martyr is. Throughout the trial Galileo's behaviour was precisely contrary of that of a martyr; he did all he could to exculpate himself from the guilt which in a formal sense he had undoubtedly incurred, by the defense of the Copernican system, and in doing so he did not shrink from the most abject

continued on next page

that, even in a heliocentric system where the Sun is granted the central role, such an important celestial body as the Moon should revolve around Earth and not, like all others, around the Sun. For this very reason, the discovery of the moons of Jupiter was also of great significance. Here was another planet that had not just a single moon, but four of them (the number observed by Galileo), and that carried these four moons along in its orbit around the Sun. At the same time, Jupiter, with its four moons, also served as a graphic model for the solar system: As the moons revolve around Jupiter, so do the planets revolve about the Sun.

With the aid of his telescope, Galileo also settled the old argument regarding the nature of the Milky Way, for he was able to demonstrate that the continuous, cloudlike Milky Way decomposes into a multitude of stars.

Galileo's observations, described in *Sidereus nuncius*, created a public sensation. His findings became the subject of conversation not only among the educated class and those interested in science, but also among ordinary people. Before the pamphlet had even gone on sale, the English ambassador to Venice had already acquired a copy for his king. In an accompanying letter, he briefly describes its contents and observes that it is being discussed even on street corners. Many considered Galileo's discoveries as important as or even more important than the discovery of the New World (Quotation 3.15). Poets praised the glory of the discovery. Even the cardinal Barberini, who later as Pope Urban VIII would hand Galileo over to the Inquisition, was at this time one of his admirers.

It is characteristic of Galileo's genius that, whenever possible, he drew quantitative conclusions from his careful qualitative observations, and he even thought about the practical implications of his discoveries.

Galileo not only claimed that there are mountains on the Moon but also measured their height. His figures are in good agreement with those calculated today. The principle he used for his measurement was very simple. He observed that at the time of the half Moon, when the illuminated and dark portions of the Moon's surface are divided by one lunar diameter, bright points can be observed on the side lying in shadow as well. He concluded, correctly, that these bright spots correspond to mountain peaks illuminated by the Sun—just as on Earth, after the Sun has set on the plain, the high peaks still glow in sunlight (Figure 3.35(b,c)).

Galileo estimated the distance a in the figure in units of lunar diameters and thereby obtained the height of the mountains. He applied the Pythagorean theorem to the sides of the right triangle depicted in the figure to obtain

$$(R+x)^2 = R^2 + a^2.$$

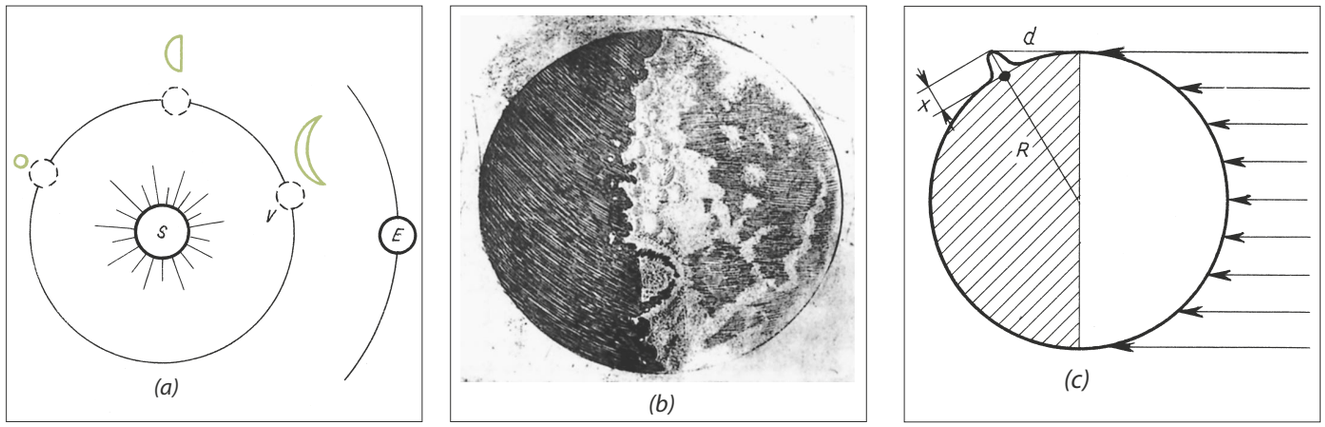
A simple rearrangement yields

$$x^2 + 2Rx - a^2 = 0,$$

from which follows the height of the mountains:

$$x = \frac{-2R + \sqrt{4R^2 + 4a^2}}{2} \approx \frac{a^2}{2R}.$$

For navigation on the high seas, the precise determination of time is of vital importance. Galileo proposed that the orbital periods of Jupiter's moons should be tabulated. Then, with the help of those tables, an observation of the position of the moons of Jupiter would allow time to be determined with great precision.



▲ **Figure 3.35** **(a)** With the discovery of the phases of Venus, Galileo could show, first of all, that the planets do not emit light but reflect the light from the Sun, and he could also support the Copernican system because the change in phases is easily explained in that system.

(b) Drawing of the Moon.

(c) Galileo's method for determining the height of mountains on the Moon.

This proposal turned out to be impractical, but it demonstrated Galileo's great originality and it represents, in addition, an attempt to transfer the results of pure science to real-world applications.

In *Sidereus nuncius*, Galileo already mentions his intention to write a book about the two cosmologies. What we have just described is a recapitulation of the dialogues of the first day in that book, *Dialogo* (Figure 3.36), where Galileo threw down the gauntlet in support of the unity of terrestrial and celestial physics, supplemented with proofs derived from observations of comets and novas (Plate XV). In the later portions of the book, Galileo refutes all the arguments that could be marshaled against the rotation of the Earth, using observations of phenomena here on Earth. We have already encountered these arguments in our discussion of Aristotle's medieval critics (Nicole Oresme, Jean Buridan). The most characteristic argument of this kind, and the one that seems to have the strongest scientific basis, is that a rock falling from a high tower would have to land behind the tower, since while the stone was falling, Earth would have turned from under it. Using numerous examples, Galileo demonstrated that such phenomena would proceed identically on a fixed Earth and on an Earth that rotates because the falling body already participates in any movement of Earth. Analogously, a person below deck on a ship moving at constant velocity in the middle of calm seas cannot tell whether the ship is moving or at rest. Today, this principle is enshrined in physics as the *Galilean principle of relativity*.

Galileo considered the most important part of his *Dialogo* to be the dialogue occurring on the fourth day. Here he maintains that there is a single phenomenon observable on Earth that can be explained only by Earth's revolution about the Sun: the occurrence of ocean tides. Galileo explains the tides incorrectly and chastises Kepler, who correctly suspected that the Moon has an important role to play in this phenomenon.

The *Dialogo*, even though Galileo was almost 70 when he wrote it, still exhibits many Aristotelian traits. Thus in the dialogue of the first day, the motion of the celestial spheres is already described in a completely Aristotelian sense: For the movement of the planets, among them Earth, uniform motion along circular paths is the only possibility. It is also worthy of note that the interlocutors in the

Quotation 3.13, ***continued***

self-humiliation. No one has the right to blame him for this; indeed, a scientific conviction is less apt to inspire heroism than a religious one. But one must not confuse the categories and honour him for an attitude over which it is best to draw a veil. On the other hand, all the parties concerned had good reason to be grateful to him for his attitude: the Roman Catholic Church, because she was not thus obliged to go to the extreme to which the constancy of the accused would have forced her; and natural science, because in the years vouchsafed to him after the trial Galileo presented it with the finest fruit of his mind, the *Discorsi*.

—E. J. Dijksterhuis, *The Mechanization of the World Picture* [pp. 334, 385]

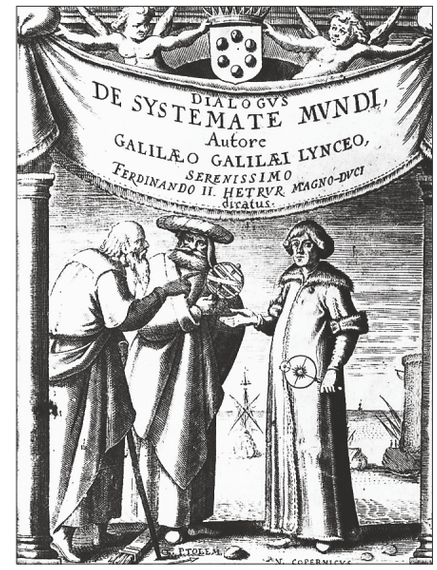
Quotation 3.14

With his keen sense of publicity and of the material value of his discoveries, which he found in no way incompatible with the pure joy of discovery, Galileo immediately tried to sell the titles of these stars in succession to the Duke of Florence (a Medici), to the King of France, and to the Pope, but the celestial honors seemed too expensive to all of them. ... These attempts, however, were to Galileo mere sideshows. He sensed at once the really revolutionary character of the new observations. Here he had for everyone to see the very model of Copernicus' system in the sky. This was knowledge not to keep but to broadcast. Within a month, in 1610, he had published what was clearly a scientific best seller, *Siderius Nuntius*, i.e. "Messenger from the Stars," in which his observations were set out briefly and plainly.

—J. D. Bernal, *Science in History* [pp. 292–293]


DIALOGO
DI
GALILEO GALILEI LINCEO
MATEMATICO SOPRAORDINARIO
DELLO STVDIO DI PISA.
E Filosofo, e Matematico primario del
SERENISSIMO
GR.DVCA DI TOSCANA.
Doue ne i congressi di quattro giornate si discorre sopra i due
MASSIMI SISTEMI DEL MONDO
TOLEMAICO, E COPERNICANO;
Proponendo indeterminatamente le ragioni Filosofiche, e Naturali tanto per l'vna, quanto per l'altra parte.
CON PRI VILEGI.
IN FIORENZA, Per Gio: Batista Landini MDCXXXII.
CON LICENZA DE' SVPERIORI.





▶ Figure 3.36
The *Dialogo* is the work of Galileo that has achieved the greatest popularity. It was published in 1632, in Italian, and in 1641, it appeared in Latin translation. In this book, three interlocutors—Salviati (speaking for Galileo), Sagredo (intelligent, sober thinker, and impartial discussion partner), and Simplicio (convinced adherent of Aristotle, named by Galileo after the Aristotelian commentator Simplicius)—converse over four days about the arguments and counterarguments for and against the two world systems. The book seeks to give the impression that the conclusion is completely entrusted to the reader, to be achieved through reading the arguments and counterarguments. In fact, however, the entire book consistently takes the side of the Copernicans.

I, Galileo, son of the late Vincenzo Galilei, Florentine, aged seventy years, arraigned personally before this tribunal and kneeling before you, Most Eminent and Reverend Lord Cardinals Inquisitors-General against heretical pravity throughout the entire Christian commonwealth, having before my eyes and touching with my hands the Holy Gospels, swear that I have always believed, do believe, and by God's help will in the future believe all that is held, preached, and taught by the Holy Catholic and Apostolic Church. But, whereas—after an injunction had been judicially intimated to me by this Holy Office to the effect that I must altogether abandon the false opinion that the Sun is the center of the world and immovable and that the Earth is not the center of the world and moves and that I must not hold, defend, or teach in any way whatsoever, verbally or in writing, the said false doctrine, and after it had been notified to me that the said doctrine was contrary to Holy Scripture—I wrote and printed a book in which I discuss this new doctrine already condemned and adduce arguments of great cogency in its favor without presenting any solution of these, I have been pronounced by the Holy Office to be vehemently suspected of heresy, that is to say, of having held and believed that the Sun is the center of the world and immovable and that the Earth is not the center and moves:

Therefore, desiring to remove from the minds of your Eminences, and of all faithful Christians, this vehement suspicion justly conceived against me, with sincere heart and unfeigned faith I abjure, curse, and detest the aforesaid errors and heresies and generally every other error, heresy, and sect whatsoever contrary to the Holy Church, and I swear that in future

continued on next page

book do not criticize Aristotle himself, but only his followers, who philosophize according to the letter but not the spirit of the master's teachings.

It is difficult to understand why, even in 1632, Galileo still held so rigidly to circular orbits considering that almost a quarter of a century had passed since Kepler had worked out the laws of motion for elliptical orbits. This is even more surprising because direct astronomical measurements had demonstrated that the assumption of circular motion for planets is untenable and that the further advancement of a physical astronomy founded on Copernicus's theory was impossible. It is likely that in this case—just as in his study of perturbations during free fall (such as air resistance)—Galileo thought that these deviations were due to secondary effects, which is to say that they did not change the fundamental nature of the phenomenon.

3.3.1.1 From the Dialogues

Simplicio: On earth I continually see herbs, plants, animals generating and decaying; winds, rains, tempests, storms arising; in a word, the appearance of the earth undergoing perpetual change. None of these changes are to be discerned in celestial bodies, whose positions and configurations correspond exactly with everything men remember, without the generation of anything new there or the corruption of anything old.

Salviati: But if you have to content yourself with these visible, or rather these seen experiences, you must consider China and America celestial bodies, since you surely have never seen in them these alterations which you see in Italy. Therefore in your sense, they must be inalterable.

Simplicio: Even if I had never seen such alterations in those places with my own senses, there are reliable accounts of them; besides which, cum eadem sit ratio totius et partium, those countries being a part of the earth like ours, they must be alterable like this.

Salviati: But why have you not observed this, instead of reducing yourself to having to believe the tales of others? Why not see it with your own eyes?

Simplicio: Because those countries are far from being exposed to view; they are so distant that our sight could not discover such alterations in them.

Salviati: Now see for yourself how you have inadvertently revealed the fallacy of your argument. You say that alterations which may be seen near at hand on earth cannot be seen in America because of the great distance. Well, so much the less could they be seen in the moon, which is many hundreds of times more distant. And if you believe in alterations in Mexico on the basis of news from there, what reports do you have from the moon to convince you that there are no alterations there? From your not seeing alterations in heaven (where if any occurred you would not be able to see them by reason of the distance, and from whence no news is to be had), you cannot deduce that there are none, in the same way as from seeing and recognizing them on earth you correctly deduce that they do exist here.

Simplicio: Among the changes that have taken place on earth I can find some so great that if they had occurred on the moon they could very well have been observed here below. From the oldest records we have it that formerly, at the Straits of Gibraltar, Abila and Calpe were joined together with some lesser mountains which held the ocean in check; but these mountains being separated by some cause, the opening admitted the sea, which flooded in so as to form the Mediterranean. When we consider the immensity of this, and the difference in appearance which must have been made in the water and land seen from afar, there is no doubt that such a change could easily have been seen by anyone then on the moon. Just so would the inhabitants of earth have discovered any such alteration in the moon; yet there is no history of such a thing being seen. Hence there remains no basis for saying that anything in the heavenly bodies is alterable, etc.

Salviati: I do not make bold to say that such great changes have taken place in the moon, but neither am I sure that they could not have happened. Such a mutation could be represented to us only by some variation between the lighter and the darker parts of the moon, and I doubt whether we have had observant selenographers on earth who have for any considerable number of years provided us with such exact selenography as would make us reasonably conclude that no such change has come about in the face of the moon. Of the moon's appearance, I find no more exact description than that some say it represents a human face; others, that it is like the muzzle of a lion; still others, that it is Cain with a bundle of thorns on his back. So to say "Heaven is inalterable, because neither in the moon nor in other celestial bodies are such alterations seen as are discovered upon the earth" has no power to prove anything.

Sagredo: This first argument of Simplicio's leaves me with another haunting doubt which I should like to have removed. Accordingly I ask him whether the earth was generable and corruptible before the Mediterranean inundation, or whether it began to be so then?

Simplicio: It was without doubt generable and corruptible before that flood, but that was so vast a mutation that it might have been observed as far as the moon.

Sagredo. Well, now; if the earth was generable and corruptible before that flood, why may not the moon be equally so without any such change? Why is something necessary in the moon which means nothing on the earth?

Salviati: A very penetrating remark. But I am afraid that Simplicio is altering the meaning a bit in this text of Aristotle and the other Peripatetics. They say that they hold the heavens to be inalterable because not one star there has

Figure 3.36 *continued*

I will never again say or assert, verbally or in writing, anything that might furnish occasion for a similar suspicion regarding me; but, should I know any heretic or person suspected of heresy, I will denounce him to this Holy Office or to the Inquisitor or Ordinary of the place where I may be. Further, I swear and promise to fulfill and observe in their integrity all penances that have been, or that shall be, imposed upon me by this Holy Office. And, in the event of my contravening (which God forbid!) any of these my promises and oaths, I submit myself to all the pains and penalties imposed and promulgated in the sacred canons and other constitutions, general and particular, against such delinquents. So help me God and these His Holy Gospels, which I touch with my hands.

[Having recited, he signed the attestation:]

I, the said Galileo Galilei, have abjured, sworn, promised, and bound myself as above; and in witness of the truth thereof I have with my own hand subscribed the present document of my abjuration and recited it word for word at Rome, in the convent of the Minerva, this twenty-second day of June, 1633.

I, Galileo Galilei, have abjured as above with my own hand.

—Galileo [quoted in Santillana 1955, pp. 312–313]

Quotation 3.15

I had this morning occasion for friendly dispute with Kepler when we were both lunching with the Ambassador of Saxony. ... He said concerning your book [the *Sidereus nuncius*] that truly it has revealed the divinity of your talent, but that you had given cause of complaint not only to the German nation but also to your own, since you make no mention of those writers who gave the signal and the occasion for your discovery, naming among them Giordano Bruno as an Italian, Copernicus and himself.

—Martin Hasdale, Letter to Galileo from Prague, April 15, 1610 [Singer 1950, p. 189]

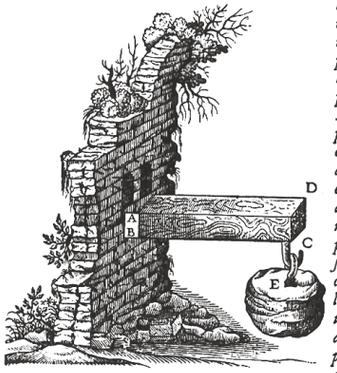
DISCORSI
E
DIMOSTRAZIONI
MATEMATICHE,
intorno à due nuoue scienze
Attenenti alla
MECANICA & I MOVIMENTI LOCALI,
del Signor
GALILEO GALILEI LINCEO,
Filosofo e Matematico primario del Sereniſſimo
Grand Duca di Toſcana.
Con vna Appendice del centro di grauità d'alcuni Solidi.



IN LEIDA,
Appreſſo gli Elſeuirii. M. D. C. XXXVIII.

114 DIALOGO SECONDO

ſin quì dichiarate, non ſarà difficile l'intender la ragione, onde auuenga, che vn Priſma, ò Cilindro ſolido di vetro, acciaio, legno, ò altra materia frangibile, che ſoſpeſo per lungo ſoſterrà graviſſimo peſo, che gli ſia attaccato, mà in trauerſo (come poco fa diceuamo) da minor peſo aſſai potrà tal'volta eſſere ſpezzato, ſecondo che la ſua lunghezza eccederà la ſua groſſezza. Imperò che figuriamoci il Priſma ſolido A B, C D *fitto in vn muro dalla parte* A B, *e nell'altra eſtremità s'intenda la forza del Peſo* E, *(intendendo ſempre il muro eſſer eretto all'Orizonte, & il Priſma, ò Cilindro fitto nel muro ad angoli retti) è manifeſto che douendoſi ſpezzare ſi romperà nel luogo* B, *doue il taglio del muro ſerue per ſoſtegno, e la* B C *per la parte della Leua, doue ſi pone la forza, e la groſſezza del ſolido* B A *e l'altra parte della Leua, nella quale è poſta la reſiſtenza, che conſiſte nello ſtaccamento, che s'hà da fare della parte del ſolido* B D, *che è fuor del muro, da quella che è dentro; e per le coſe dichiarate il momento della forza poſta in* C *al momento della reſiſtenza che ſtà nella*



166 DIALOGO TERZO

Motum æquabiliter, ſeu uniformiter acceleratum dicimus eum, qui à quiete recedens temporibus æqualibus æqualia celeritatis momenta ſibi ſuperaddit.

Salu. *Fermata cotal definizione vn ſolo principio domanda, e ſuppone per vero l'Autore, cioè;*

Accipio, gradus velocitatis ejuſdem mobilis ſuper diuerſas planorum inclinationes acquiſitos tunc eſſe æquales, cum eorumdem planorum elevationes æquales ſint.

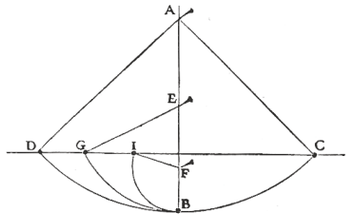
Chiama la eleuazione di vn piano inclinato la perpendicolare, che dal termine ſublime di eſſo piano caſca ſopra la linea orizontale prodotta per l'infimo termine di eſſo piano inclinato; come per intelligenza: eſſendo la linea A B *parallela all'orizonte, ſopra 'l quale ſiano inclinati li due piani* C A, C D; *la perpendicolare* C B *cadente ſopra l'orizontale* B A, *chiama l'Autore la eleuazione de i Piani* C A, C D, *e ſuppone, che i gradi di velocità del medeſimo mobili ſcendente per li piani inclinati* C A, C D, *acquiſtati ne i termini* A D, *ſiano eguali, per eſſer la loro eleuazione l'iſteſſa* C B. *E tanto anco ſi deue intendere il grado di velocità, che il medeſimo cadente dal punto* C *harebbe nel termine* B.



Sagr. *Veramente mi par, che tal ſuppoſto habbia tanto del probabile, che meriti di eſſer ſenza controuerſia conceduto: intendendo ſempre, che ſi rimouano tutti gl'impedimenti accidentarii, & eſterni; e che i piani ſiano ben ſolidi, e terſi, & il mobile di figura perfettiſſimamente rotonda, ſi che & il piano, & il mobile non habbiano ſcabroſità. Rimoſſi tutti i contraſti, & impedimenti, il lume naturale mi detta ſenza difficoltà, che vna Palla graue, e perfettamente rotunda ſcendendo per le linee* C A, C D, C B, *giungerebbe ne i termine* A D B, *con impeti eguali.* Salu.

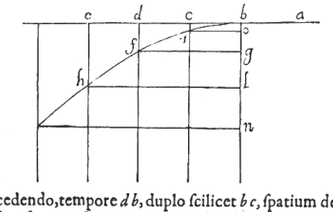
DEL GALILEO. 167

Salu. *Voi molto probabilmente diſcorrete: mà oltre al veriſimile voglio con vna eſperienza accreſcer tanto la probabilità, che poco gli manchi all'aggagliarſi ad vna ben neceſſaria dimoſtrazione. Figuratevi queſto foglio eſſere vna parete eretta all'orizonte, e da vn chiodo fitto in eſſa pendere vna palla di piombo d'un oncia, ò due ſoſpeſa dal ſottil filo* A B *lungo due, ò tre braccia perpendicolare all'orizonte; e nella parete ſegnate vna linea orizontale* D C *ſegante à ſquadra il perpendicolo* A B, *il quale ſia lontano dalla parete due dita in circa: trasferendo poi il filo* A B *con la palla in* A C, *laſciate eſſa palla in libertà, la quale primieramente vedreteſcendere deſcriuendo l'arco* C B D, *e di tanto trapaſſare il termine* B, *che ſcorrendo per l'arco* B D *ſormonterà ſino quaſi alla ſegnata parallela* C D, *reſtando di peruenirui per piccoliſſimo interuallo, toltogli il preciſamente arriuarui dall'impedimento dell'aria, e del filo. Dal che poſſiamo veracemente concludere, che l'impeto acquiſtato nel punto* B *dalla palla nello ſcendere per l'arco* C B, *fù tanto, che baſtò à riſoſpingerſi per vn ſimile arco* B D *alla medeſima altezza: fatta, e più volte reiterata cotale eſperienza, voglio che ficchiamo nella parete raſente al perpendicolo* A B *vn chiodo, come in* E,



242 DIALOGO QVARTO

tionem quadratorum ipſarum, *c b, d b, e b*, ſeu dicamus, in ratione earundem linearum duplicata. Quod ſi mobili ultra *b* verſus *c* æquabili latione lato deſcenſum perpendicularem ſecundum quantitatem *c i* ſuperadditum intelligamus, reperietur tempore *b c* in termino *i* conſtitutum. Vlterius autem



procedendo, tempore *d b*, duplo ſcilicet *b c*, ſpatium deſcenſus deorſum, erit ſpatii primi *c i* quadruplum: demonſtratum enim eſt in primo tractatu, ſpatia peracta à gravi motu naturaliter accelerato eſſe in duplicata ratione temporum. Pariterque conſequenter ſpatium *e h*, peractum tempore *b e*, erit ut 9. adeo ut manifeſtè conſtet, ſpatia *e h, d f, c i*, eſſe inter ſe ut quadrata linearum *e b, d b, c b*. Ducantur modò à punctis *i, f, h*, rectæ *i o, f g, h l*, ipſi *e b* æquidiſtantes; erunt *h l, f g, i o*, lineæ lineis *e b, d b, c b*, ſingulæ ſingulis æquales; nec non ipſæ *b o, b g, b l*, ipſis *c i, d f, e h* æquales. Eritque quadratum *h l* ad quadratum *f g*, ut linea *l b* ad *b g*: & quadratum *f g* ad quadratum *i o*, ut *g b* ad *b o*. Ergo puncta *i, f, h*, ſunt in una eademque linea Parabolica. Similiterque demonſtrabitur, aſſumptis quibuſcunque temporis particulis æqualibus cujuſlibet magnitudinis, loca mobilis, ſimili motu compoſito lati, iiſdem temporibus in eadem linea parabolica reperiri, ergo patet propoſitum.

Salu.

DEL GALILEO. 243

Salu. *Queſta concluſione ſi raccoglie dal conuerſo della prima delle due propoſizioni poſte di ſopra, imperò che deſcritta per eſempio la Parabola per li punti* b h, *ſe alcuno delli* 2 f, i, *non fuſſe nella deſcritta linea parabolica, ſarebbe dentro, ò fuori; e per conſeguenza la linea* f g *ſarebbe ò minore, ò maggiore, di quella che andaſſe a terminare nella linea Parabolica: onde il quadrato della* h l *non al quadrato della* f g, *ma ad altro maggiore, ò minore harebbe la medeſima proporzione che hà la linea* l b *alla* b g, *ma la hà al quadrato della* f g. *adunque il punto* f *è nella Parabolica; e così tutti gl'altri, &c.*

Sagr. *Non ſi può negare che il diſcorſo ſia nuouo, ingegnoſo e concludente, argomentando ex ſuppoſitione, ſupponendo cioè, che il moto trauerſale ſi mantenga ſempre equabile, e che il naturale deorſum parimente mantenga il ſuo tenore d'andarſi ſempre accelerando ſecondo la proporzion duplicata de i tempi: e che tali moti, e loro velocità nel meſcolarſi non ſi alterino, perturbino, & impediſchino, ſi che finalmente la linea del Proietto non vadia nella continuazion del moto à degenerare in vn'altra ſpezie; coſa che mi ſi rappreſenta come impoſſibile. Imperò che, ſtante che l'aſſe della Parabola noſtra, ſecondo 'l quale noi ſupponghiamo farſi il moto naturale de i graui, eſſendo perpendicolare all'Orizonte, va à terminar nel centro della Terra, & eſſendo che la linea Parabolica ſi và ſempre ſlargando dal ſuo aſſe, niun Proietto andrebbe già mai à terminar nel centro, ò ſi vi andrebbe, come par neceſſario, la linea del Proietto tralignerebbe in altra diuerſiſſima dalla Parabolica.*

Simpl. *Io a queſte difficoltà ne aggiungo dell'altre: vna delle quali è che noi ſupponghiamo, che il piano orizontale il quale non ſia nè accliue, nè decliue, ſia vna linea retta; quaſi che vna ſimil linea ſia in tutte le ſue parti egualmente diſtante dal centro, il che non è vero; perche partendoſi dal ſuo mezo và verſo le eſtremità ſempre più, e più allontanandoſi dal centro, e però aſcendendo ſempre; il che ſi tira in conſeguenza eſſere impoſſibile, che il moto ſi perpetui, anzi che nè pur per qualche ſpazio ſi mantenga equabile, mà*

▲ **Figure 3.37** The *Discorsi* (known as *Dialogues Concerning Two New Sciences*) is GALILEO's most mature work and the one of greatest importance for the further development of science. It was written after his trial, when GALILEO was under house arrest, and it includes his earlier investigations on mechanics. In the *Discorsi*, we meet again the interlocutors from the *Dialogo (Dialogue Concerning the Two Chief World Systems)*. On the third and fourth days of the four-day conversation, the discussion turns to free fall, experiments with an inclined plane, and the parabolic path of a projectile. The book was written in Latin with insertions in Italian. The fifth and sixth days were reconstructed from GALILEO's posthumous papers.

continued on next page

ever been seen to be generated or corrupted, such being probably a lesser part of heaven than a city is of the earth; yet innumerable of the latter have been destroyed so that not a trace of them remains.

Sagredo: Really, I thought otherwise, believing that Simplicio distorted this exposition of the text so that he might not burden the Master and his disciples with a notion even more fantastic than the other. What folly it is to say, "The heavens are inalterable because stars are not generated or corrupted in them." Is there perhaps someone who has seen one terrestrial globe decay and another regenerated in its place? Is it not accepted by all philosophers that very few stars in the heavens are smaller than the earth, while a great many are much bigger? So the decay of a star in the heaven would be no less

momentous than for the whole terrestrial globe to be destroyed! Now if, in order to be able to introduce generation and corruption into the universe with certainty, it is necessary that as vast a body as a star must be corrupted and regenerated, then you had better give up the whole matter; for I assure you that you will never see the terrestrial globe or any other integral body in the universe so corrupted that, after having been seen for many ages past, it dissolves without leaving a trace behind.

Salviati: But to give Simplicio more than satisfaction, and to reclaim him if possible from his error, I declare that we do have in our age new events and observations such that if Aristotle were now alive, I have no doubt he would change his opinion. This is easily inferred from his own manner of philosophizing, for when he writes of considering the heavens inalterable, etc., because no new thing is seen to be generated there or any old one dissolves, he seems implicitly to let us understand that if he had seen any such event he would have reversed his opinion, and properly preferred the sensible experience to natural reason. Unless he had taken the senses into account, he would not have argued immutability from sensible perceptions not being seen.

—Galileo, *Dialogue Concerning the Two Chief World Systems* [pp. 47–51]

3.3.2 *Inclined Planes, Pendulums, and Projectile Motion*

While the *Dialogo* is of decisive importance both in overcoming theoretical obstacles to the further development of science and for the popularization of science, the *Discorsi*, in contrast (Figure 3.37), with its findings in the field of descriptive kinematics, served as a starting point for the mechanics of the seventeenth century.

In treating free fall, Galileo began with the assumption that motion occurring in nature with variable velocity proceeds according to the simplest possible laws, and that nothing could be simpler than equal increase in velocity in equal periods of time. Today we would say that Galileo took as his starting point the idea of motion with constant acceleration (Quotation 3.16). Here the question arises as to why, with the same argument, the simplest motion would not be the equal increase in velocity after traversing equal *distances*. Galileo's reply was that such motion is logically untenable. We will return to this point later.

The correctness of the assumption of motion with constant acceleration could not be shown by direct experiment because instantaneous velocity could not be measured. Therefore, it was necessary to derive logical consequences of the assumption that could then be actually measured.

First, Galileo wrote down the distance traveled by a free-falling body as a function of time; that is, he multiplied the average velocity by time. From this, he obtained the following rule: The distances traveled by a body in free fall at various times are related to one another as the squares of the times that it took the body to traverse the given distances. (See Figure 3.38.)

In today's notation, but sticking to Galileo's conception, we might summarize his train of thought as follows: Let velocity be proportional to time, $v = at$, for some constant a. If a body starts with velocity zero, then the mean velocity is given by

$$v_{\mathrm{m}} = \frac{v}{2} = \frac{at}{2}.$$

Figure 3.37 *continued*

My purpose is to set forth a very new science dealing with a very ancient subject. There is, in nature, perhaps nothing older than motion, concerning which the books written by philosophers are neither few nor small; nevertheless I have discovered by experiment some properties of it which are worth knowing and which have not hitherto been either observed or demonstrated. Some superficial observations have been made, as, for instance, that the free motion [*naturalem motum*] of a heavy falling body is continuously accelerated; but to just what extent this acceleration occurs has not yet been announced; for so far as I know, no one has yet pointed out that the distances traversed, during equal intervals of time, by a body falling from rest, stand to one another in the same ratio as the odd numbers beginning with unity.

It has been observed that missiles and projectiles describe a curved path of some sort; however, no one has pointed out the fact that this path is a parabola. But this and other facts, not few in number or less worth knowing, I have succeeded in proving; and what I consider more important, there have been opened up to this vast and most excellent science, of which my work is merely the beginning, ways and means by which other minds more acute than mine will explore its remote corners.

This discussion is divided into three parts; the first part deals with motion which is steady or uniform; the second treats of motion as we find it accelerated in nature; the third deals with the so-called violent motions and with projectiles. [pp. 153–154]

The parts of the *Discorsi* that deal with motion (third and fourth days, example above) contain material appropriate for a high-school curriculum. In contrast, the discussions of the first and second days, contain interesting ideas on the hardness of objects and on the effect of a vacuum on that hardness. Here we present extracts that contain the germ of modern model theory.

Sagredo: Yes, that is what I mean; and I refer especially to his last assertion, which I have always regarded as false, though current, opinion; namely, that in speaking of these and other similar machines one cannot argue from the small to the large, because many devices which succeed on a small scale do not work on a large scale. Now, since mechanics has its foundation in geometry, where mere size cuts no figure, I do not see that the properties of circles, triangles, cylinders, cones and other solid figures will change with their size. If, therefore, a large machine be constructed in such a way that its parts bear to one another the same ratio as in a smaller one, and if the smaller is sufficiently strong for the purpose for which it is designed, I do not see why the larger also should not be able to withstand any severe and destructive tests to which it may be subjected.

Salviati: … Sagredo, you would do well to change the opinion which you, and perhaps also many other students of mechanics, have entertained concerning the ability of machines and structures to resist external disturbances, thinking that when they are built of the same material and maintain the same ratio between

continued on next page

Figure 3.37 *continued*

parts, they are able equally, or rather proportionally, to resist or yield to such external disturbances and blows. For we can demonstrate by geometry that the large machine is not proportionally stronger than the small. Finally, we may say that, for every machine and structure, whether artificial or natural, there is set a necessary limit beyond which neither art nor nature can pass; it is here understood, of course, that the material is the same and the proportion preserved. … Please observe, gentlemen, how facts which at first seem improbable will, even on scant explanation, drop the cloak which has hidden them and stand forth in naked and simple beauty. Who does not know that a horse falling from a height of three or four cubits will break his bones, while a dog falling from the same height or a cat from a height of eight or ten cubits will suffer no injury? Equally harmless would be the fall of a grasshopper from a tower or from the distance of the moon. … [pp. 2–3]

What follows is from the second day.

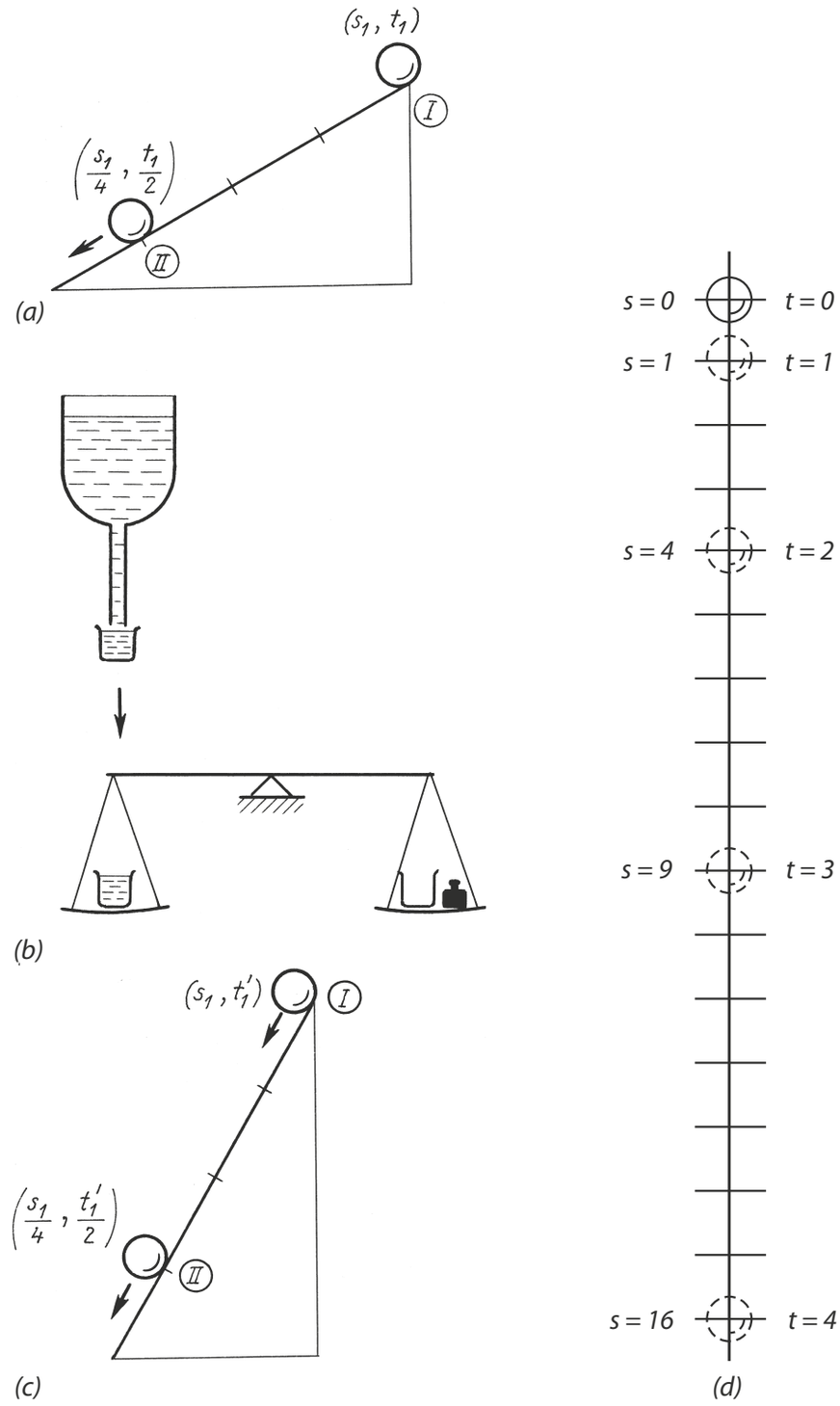
From what has been already demonstrated, you can plainly see the impossibility of increasing the size of structures to vast dimensions either in art or in nature; likewise, the impossibility of building ships, palaces, or temples of enormous size in such a way that their oars, yards, beams, iron-bolts, and, in short, all their other parts will hold together; nor can nature produce trees of extraordinary size because the branches would break down under their own weight; so also it would be impossible to build up the bony structures of men, horses, or other animals so as to hold together and perform their normal functions if these animals were to be increased enormously in height; for this increase in height can be accomplished only by employing a material which is harder and stronger than usual. [p. 130]

From this, it follows that the distance traveled is

$$s = v_m t = \frac{at}{2} t = \frac{1}{2} a t^2,$$

from which it immediately follows that

$$\frac{s}{t^2} = \frac{a}{2} = \text{constant},$$



▶ **Figure 3.38 (a)** The experimental principle: The ball starts at the highest point of the inclined plane of length s_1 and reaches the bottom in time t_1. If it starts at one quarter of the length, it will take one half of that time.

(b) Galileo determined the time by weighing the amount of water that flowed out of a vessel.

(c) As the steepness of the plane is increased, we approach free fall.

(d) For motion with constant acceleration, the distance traveled is proportional to the square of the time. The successive distances traveled in equal intervals of time are proportional to the successive odd integers: 1, 3, 5, 7, etc.

or

$$\frac{s_1}{t_1^2} = \frac{s_2}{t_2^2} = \cdots.$$

This result can be experimentally verified, since both the distance and time can easily be measured, and then it can be determined whether the above relation is satisfied for associated distance–time pairs. In trying to carry out direct measurements during free fall, one is confronted with the particular difficulty of having to measure very small time intervals in order to verify the law. Galileo overcame this difficulty by employing an inclined plane with a small angle of inclination. Without changing the temporal nature of the phenomenon, he was able, so to speak, to slow down free fall to the point where he could measure the requisite time intervals with sufficient accuracy using the instruments available to him. Galileo argued that by increasing the plane's angle of inclination, that is, with increasing steepness of the inclined plane, one gradually approaches the case of free fall.

In these descriptions of kinematic processes, for the first time in the history of mechanics we encounter a detailed presentation of an experiment as well as the experimental conditions in a form that we would expect in a scientific publication today (Quotation 3.17). It is noteworthy that Galileo used a variation of the water clock to measure time (Figure 3.38(b)). He let water run through a thin tube from a container during the time that the ball rolled down the inclined plane. He weighed the water on an accurate scale and from this he determined the time of the roll. In this way, he was actually able to prove that for a given inclined plane, the ratio of the distance traveled to the square of the elapsed time is constant (Figure 3.38(d)).

Galileo was also able to verify a simple relationship between the distance traveled in successive time intervals. He found that these successive lengths are in proportion to the successive odd integers.

Let us denote the length of a selected time interval by t_0. Then the distance traveled after n time intervals from the start of the experiment is given by

$$s_n = \frac{1}{2} a \left(n t_0\right)^2,$$

and the distance after $n-1$ such intervals is given by

$$s_{n-1} = \frac{1}{2} a \left(n-1\right)^2 t_0^2.$$

From this, we may compute the difference of the two distances, obtaining

$$\Delta s_n = s_n - s_{n-1} = \frac{1}{2} a t_0^2 \left[n^2 - (n-1)^2 \right] = \frac{1}{2} a t_0^2 (2n-1),$$

and indeed, for the successive path lengths, we have

$$\Delta s_1 = \frac{1}{2} a t_0^2, \quad \Delta s_2 = \frac{1}{2} a t_0^2 \times 3, \quad \Delta s_3 = \frac{1}{2} a t_0^2 \times 5, \quad \ldots,$$

or, in general,

$$\Delta s_1 : \Delta s_2 : \Delta s_3 : \ldots = 1 : 3 : 5 : \ldots.$$

Let us summarize Galileo's method, now considering the roles of experiment versus theoretical considerations.

Quotation 3.16

When, therefore, I observe a stone initially at rest falling from an elevated position and continually acquiring new increments of speed, why should I not believe that such increases take place in a manner which is exceedingly simple and rather obvious to everybody? If now we examine the matter carefully we find no addition or increment more simple than that which repeats itself always in the same manner. This we readily understand when we consider the intimate relationship between time and motion; for just as uniformity of motion is defined by and conceived through equal times and equal spaces (thus we call a motion uniform when equal distances are traversed during equal time-intervals), so also we may, in a similar manner, through equal time-intervals, conceive additions of speed as taking place without complication; thus we may picture to our mind a motion as uniformly and continuously accelerated when, during any equal intervals of time whatever, equal increments of speed are given to it.

—Galileo, *Dialogues Concerning Two New Sciences* [pp. 154–155]

Quotation 3.17

A piece of wooden [plank], about 12 cubits long, half a cubit wide, and three finger-breadths thick, was taken; on its edge was cut a channel a little more than one finger in breadth; having made this groove very straight, smooth, and polished, and having lined it with parchment, also as smooth and polished as possible, we rolled along it a hard smooth, and very round bronze ball. Having placed this board in a sloping position, by lifting one end some one or two cubits above the other, we rolled the ball, as I was just saying, along the channel, noting, in a manner presently to be described, the time required to make the descent. We repeated this experiment more than once in order to measure the time with an accuracy such that the deviation between two observations never exceeded one-tenth of a pulse-beat. Having performed this operation and having assured ourselves of its reliability, we now rolled the ball only one-quarter the length of the channel; and having measured the time of its descent, we found it precisely one-half of the former. Next we tried other distances, comparing the time for the whole length with that for the half, or with that for two-thirds, or three-fourths, or indeed for any fraction; in such experiments, repeated a full hundred times, we always found that the spaces traversed were to each other as the squares of the times, and this was true for all inclinations of the plane, i.e., of the

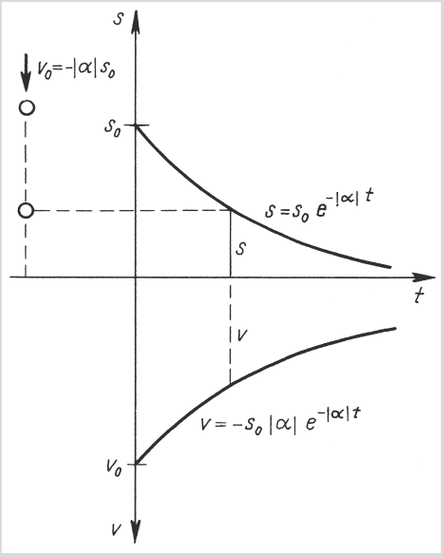
continued on next page

Quotation 3.17, ***continued***

channel, along which we rolled the ball. We also observed that the times of descent, for various inclinations of the plane, bore to one another precisely that ratio which, as we shall see later, the Author had predicted and demonstrated for them.

For the measurement of time, we employed a large vessel of water placed in an elevated position; to the bottom of this vessel was soldered a pipe of small diameter giving a thin jet of water, which we collected in a small glass during the time of each descent, whether for the whole length of the channel or for a part of its length; the water thus collected was weighed, after each descent, on a very accurate balance; the differences and ratios of these weights gave us the differences and ratios of the times, and this with such accuracy that although the operation was repeated many, many times, there was no appreciable discrepancy in the results.

—Galileo, *Dialogues Concerning Two New Sciences* [pp. 171–172]



▲ **Figure 3.39** Motion with velocity proportional to the path length is also possible, contrary to Galileo's assertion.

1. Galileo begins with an explanation of the terms that he uses; in the case at hand, he defines what is meant by motion with uniform acceleration.
2. Then he formulates a hypothesis concerning the expected distances.
3. From this hypothesis he then mathematically derives relationships that can be proven experimentally.
4. Finally, Galileo performs experiments to test the theoretical predictions.

If we investigate the individual steps in greater detail, we can see right away that *the experimental results were not the immediate starting point*. One could start, of course, with everyday experience, which gives us only an imprecise notion of the process of free fall: we perceive that the velocity of a free-falling object increases. The crucial point is that Galileo began by focusing on a concrete natural phenomenon. This first step in Galileo's scientific approach is *the formulation of concepts and hypotheses*. However, recognizing that Galileo did not give precedence to the experimental results should in no way be taken as a negative judgment. Today it is generally believed that the intimate interplay between theory and experiment, precisely as practiced by Galileo, promises the greatest chance of success.

Let us now consider to what extent Galileo's reasoning was correct when he took as the simplest possible starting point a proportionality between velocity and time. As we have already mentioned, there was also the possibility of choosing the change in velocity as being proportional to the distance traveled, which was rejected by Galileo as *logically untenable* (Quotation 3.18). A historical curiosity is—as can be seen from his letters and the quoted selection from the *Discorsi*—that Galileo had initially seen this second approach as the simpler and therefore chose it as his starting point. It is also remarkable that using a false assumption, namely, that the velocity in free fall grows in proportion to the distance traveled, Galileo still arrived at the correct relationship between distance traversed and time taken. The underlying logical train of thought was, of course, incorrect because from false premises one can arrive at a correct result only with faulty logic.

It is also worth noting that there are real physical situations—naturally not in free fall—where velocity is, in fact, proportional to the distance traveled, which proves that such assumptions do not always lead to a logical contradiction.

The relationship $v = \alpha s$ can be written as

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \alpha s, \tag{1}$$

and the solution to this differential equation is well known to be

$$s = s_0 e^{\alpha t}.$$

Today, we can say under what circumstances such motion can arise. Namely, if we differentiate equation (1) with respect to time, then we obtain the relationship

$$\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = \alpha \frac{\mathrm{d}s}{\mathrm{d}t} = \alpha v.$$

This relationship can also be interpreted as an equation of motion. On the left-hand side is the acceleration (which we think of as multiplied by the unit mass), and on the right-hand side is the force. A force proportional to the velocity—in the case of bodies moving in air—can be identified with the force of friction to within a good approximation. In this case, the constant α must be negative, that is, $\alpha = -|\alpha|$. In the absence of other forces, motion is possible only if the object has been imparted some initial velocity. In this case, the velocity decreases to zero, according to the formula

$$v = \frac{\mathrm{d}s}{\mathrm{d}t} = -s_0 |\alpha| e^{-|\alpha| t},$$

where the distance traveled is obtained as $s = s_0 e^{-|\alpha| t}$. We see here yet again that velocity and path length are indeed proportional (Figure 3.39).

One may rightly introduce the counterargument here that s is not the distance traveled from the start of the motion. However, if we count time from $t = -\infty$ and employ a force that increases proportional to the velocity, then the objection is no longer valid.

One may also question the legitimacy of Galileo's bold extrapolation from an inclined plane with an inclination of 10–15° to vertical free fall, which we may think of as motion on an inclined plane with inclination 90°. On a plane with a small inclination, a ball will roll; with increasing steepness, the ball at some point begins to glide; and it is clear that with a vertical plane, there is no longer any rolling at all. We cannot be sure that these facts do not have a fundamental influence on the character of the motion.

Based on our current knowledge, we know that the type of motion—rolling or sliding—does indeed influence the numerical values; however, it leaves the *character* of the law, and in particular the uniform increase in velocity, unchanged. Therefore, the extrapolation to free fall used by Galileo is justified even by today's standards. However, he would have obtained the wrong numerical value for the acceleration; indeed, in his single surviving note, the numerical value for the terrestrial acceleration is quite far (5 m/s^2) from the correct value, which is difficult to reconcile with his painstaking process of measurement. Another possible explanation is that Galileo considered the numerical value itself, which today we see as fundamental, an insignificant detail.

Let us apply to the case of a rolling object (ball, cylinder, ring) the law of conservation of energy

$$mgh = \frac{1}{2} m v^2 + \frac{1}{2} I \omega^2,$$

which is, of course, an anachronism for Galileo's time.

Here h denotes the height from the which the object falls, ω the angular velocity ($v = r\omega$), and I the moment of inertia. Since we have $I = (1/2)mr^2$ for a (homogeneous) cylinder, $I = mr^2$ for a thin-walled cylinder, and $I = (2/3)mr^2$ for a sphere, we obtain the energy equation

$$v^2 = \frac{2gh}{1+\kappa}$$

for some constant κ; that is, we are dealing with an acceleration of $g/(1 + \kappa) = 2g/3$ for the cylinder, $g/2$ for the ring, and $5g/7$ for the sphere. In passing from rolling to gliding, the relationships become, of course, more complicated.

Galileo also proved some very elegant theorems regarding motion in a plane, which are often used today as examples in college physics courses. One of these is that a ball rolling on an inclined plane representing a chord of a given circle, as indicated in Figure 3.40, will traverse any such incline in the same period of time.

This theorem can be understood by using standard college physics: Between the linear path with angle of inclination α and time t, we have the relationship

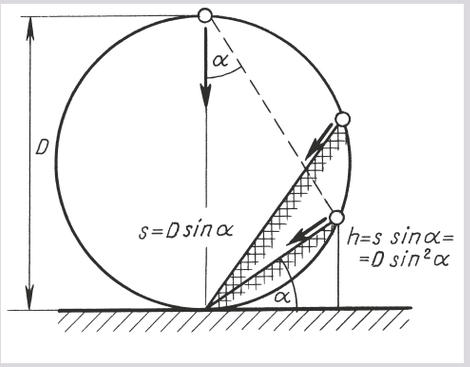
$$s = \frac{1}{2} g \sin\alpha t^2,$$

where g is the acceleration due to gravity (in the case of frictionless gliding; for a rolling object, we must use $g/(1 + \kappa)$). From this, we obtain the time

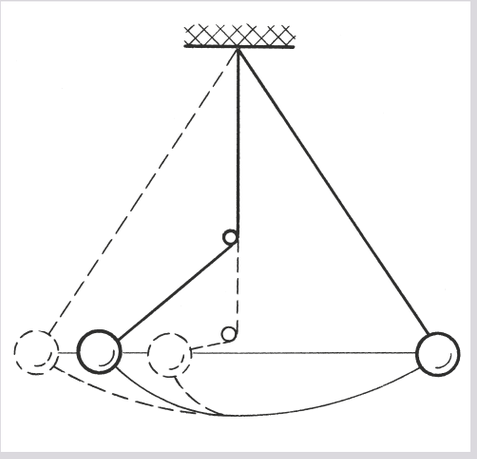
$$t = \sqrt{\frac{2s}{g \sin\alpha}}.$$

From the figure, we see that the length s of the path is given by $s = D\sin\alpha$, and we thus have

$$t = \sqrt{\frac{2D\sin\alpha}{g\sin\alpha}} = \sqrt{\frac{2D}{g}},$$



▲ **Figure 3.40** If the endpoints of inclined planes with the same base point lie on a circle, then the time for a ball to reach the bottom is the same in each case.



▲ **Figure 3.41** With such an experiment, Galileo was able to support the hypothesis that the velocity of a ball rolling down a plane depends only on the height of the start.

Quotation 3.18

Salviati: The present does not seem to be the proper time to investigate the cause of the acceleration of natural motion concerning which various opinions have been expressed by various philosophers, some explaining it by attraction to the center, others to repulsion between the very small parts of the body, while still others attribute it to a certain stress in the surrounding medium which closes in behind the falling body and drives it from one of its positions to another. Now, all these fantasies, and others too, ought to be examined; but it is not really worth while. At present it is the purpose of our Author merely to investigate and to demonstrate some of the properties of accelerated motion (whatever the cause of this acceleration may be)—meaning thereby a motion, such that the momentum of its velocity [*i momenti della sua velocità*] goes on increasing after departure from rest, in simple

continued on next page

Quotation 3.18, ***continued***

proportionality to the time, which is the same as saying that in equal time-intervals the body receives equal increments of velocity. ...

Sagredo: So far as I can see at present, the definition might have been put a little more clearly perhaps without changing the fundamental idea, namely, uniformly accelerated motion is such that its speed increases in proportion to the space traversed; ...

Salviati: It is very comforting to me to have had such a companion in error; and moreover let me tell you that your proposition seems so highly probable that our Author himself admitted, when I advanced this opinion to him, that he had for some time shared the same fallacy. ...

Simplicio: I am one of those who accept the proposition, and I believe that ... the [velocity] of the falling body is doubled when it falls from a doubled height; these propositions, it appears to me, ought to be conceded without hesitation or controversy.

Salviati: And yet they are as false and impossible as that motion should be completed instantaneously; and here is a very clear demonstration of it. If the velocities are in proportion to the spaces traversed, or to be traversed, then these spaces are traversed in equal intervals of time; if, therefore, the velocity with which the falling body traverses a space of eight feet were double that with which it covered the first four feet (just as the one distance is double the other) then the time-intervals required for these passages would be equal. But for one and the same body to fall eight feet and four feet in the same time is possible only in the case of instantaneous [discontinuous] motion; but observation shows us that the motion of a falling body occupies time, and less of it in covering a distance of four feet than of eight feet; therefore it is not true that its velocity increases in proportion to the space. ...

Sagredo: ... But now, continuing the thread of our talk, it would seem that up to the present we have established the definition of uniformly accelerated motion which is expressed as follows:

> A motion is said to be equally or uniformly accelerated when, starting from rest, its momentum (celeritatis momenta) receives equal increments in equal times.

Salviati: This definition established, the Author makes a single assumption, namely,

> The speeds acquired by one and the same body moving down planes of different inclinations are equal when the heights of these planes are equal. ...

Sagredo: Your assumption appears to me so reasonable that it ought to be conceded without question, provided of course there are no chance

continued on next page

which means that the time required for a ball to roll from the highest point to the base of the circle is independent of the angle α, and in this time period, under free fall, the ball will traverse a length equal to the diameter of the circle.

In his proof, Galileo used another theorem, one that was to play an important role later in the history of mechanics. Namely, he assumed as his initial hypothesis that the velocity of the object at the bottom of the incline depends only on the height of the starting point. He demonstrated this theorem with the help of a pendulum (Figure 3.41). He displaced the pendulum through a certain angle, clamping the thread at different points (as shown in the figure) while the pendulum was swinging so that the pendulum swung on a shorter and shorter radius. He found that the height to which the pendulum rose in each case was the same as the height at which it started. We see in this observation a germ of the law of conservation of energy or more specifically the transformation of kinetic and potential energy one into the other. For free fall, it is easy to derive a relationship between the height from which the object is dropped and its velocity: The height is proportional to the square of the time, while the velocity is proportional to the time, from which it follows that the height is proportional to the square of the velocity.

Galileo's method of proof is based on the idea that the longer the inclined plane, the average velocity of the object rolling down will be greater in the same proportion, because the *height* of the plane is greater in just the necessary proportion.

In today's notation, Galileo's idea can be represented in the following way: The time the object travels is given by

$$t = \frac{s}{v_m},$$

where the average (mean) velocity satisfies

$$v_m \propto \sqrt{h}.$$

Furthermore, we have

$$s = D\sin\alpha, \quad h = s\sin\alpha = D\sin^2\alpha,$$

yielding

$$t = \frac{s}{v_m} \propto \frac{D\sin\alpha}{\sqrt{D\sin^2\alpha}} = \sqrt{D}.$$

From this it follows that the time of the motion is the same for all such paths. Furthermore, we see that this time is proportional to the square root of the diameter of the circle, or equivalently, to the square root of the radius.

Another valid conclusion reached by Galileo is that the time of motion of an object along two chords, as shown in Figure 3.42, is less than that along a single chord joining the same two points. Clearly, the basis for this assertion is that the increase in average velocity on the twofold path more than makes up for the increase in the length of the total path. This greater average velocity is the result of a greater acceleration on the first, steeper chord, with the result that the body achieves a greater velocity in a shorter period of time. The time can be decreased even further by replacing a twofold path with a polygonal one with a larger number of sides. In this way, Galileo arrived at the valid conclusion that an object reaches the base of the circle fastest by traveling along the circle itself.

On the basis of the previous considerations, Galileo constructed two additional theorems, which, although incorrect, stimulated further research because of the way in which the problem was posed. The first of these theorems represents an *invalid*

generalization of the theorem presented above: if we join two points not at the same height by an arbitrary curve, then the circular arc has the property that an object moving along it passes from the higher point to the lower point in the least amount of time. The correct theorem discussed above had as its content the minimality property of the circular arc in relation to arbitrary inscribed polygonal paths, and in comparison to such polygonal paths, the circular path is indeed minimal with respect to time. However, if we allow other types of curves, then some may be discovered along which an object will take less time to travel than along a circle. However, with his false theorem, GALILEO alerted the scientific community to the problem of the *brachistochrone*, that is, the curve with the sought-after minimality property.

Another of GALILEO's false conclusions was that objects starting at various points on the circle would all take the same time to reach the base of the circle by moving along the circle. The discovery of the correct curve satisfying this property had to wait for the next generation. If we apply this theorem to the motion of a pendulum, then we may conclude that GALILEO believed erroneously that the period of oscillation of the pendulum is independent of the amplitude.

GALILEO's lasting contribution is that he established the independence of the period of a pendulum from the weight of the pendulum bob and from its material composition. In fact, for small amplitudes, the period is essentially independent of the amplitude (Figure 3.43). However, this is no longer true for large amplitudes.

Today, the motion of the mathematical pendulum would be described thus: Its equation of motion is

$$ml\frac{\mathrm{d}^2\alpha}{\mathrm{d}t^2} = -mg\sin\alpha,$$

where, for a small angle α, we may use the approximation

$$\sin\alpha \approx \alpha.$$

We then obtain

$$\frac{\mathrm{d}^2\alpha}{\mathrm{d}t^2} = -\frac{g}{l}\alpha,$$

with the solution

$$\alpha = \alpha_0 \sin\sqrt{\frac{g}{l}}t, \quad T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{l}{g}}.$$

We see that the period of oscillation is constant and proportional to the square root of the length of the pendulum. For arbitrary amplitudes, the solution of the equation of oscillation takes the form of an elliptic integral.

Indeed, if instead of α, we introduce a new variable ψ defined by

$$\sin\frac{\alpha}{2} = \sin\frac{\alpha_0}{2}\sin\psi = k\sin\psi,$$

where α_0 denotes an arbitrarily large initial angle (the velocity at this angle is zero), then we obtain for the period T the following expression, which depends on the parameter k, a characteristic of the initial angle:

$$T = 4\sqrt{\frac{l}{g}}\int_0^{\pi/2}\frac{\mathrm{d}\psi}{\sqrt{1-k^2\sin^2\psi}} = 2\pi\sqrt{\frac{l}{g}}\left[1+\left(\frac{1}{2}\right)^2k^2+\left(\frac{1\times3}{2\times4}\right)^2k^4+\cdots\right].$$

The integral appearing here is called a *complete elliptic integral of the first kind*.

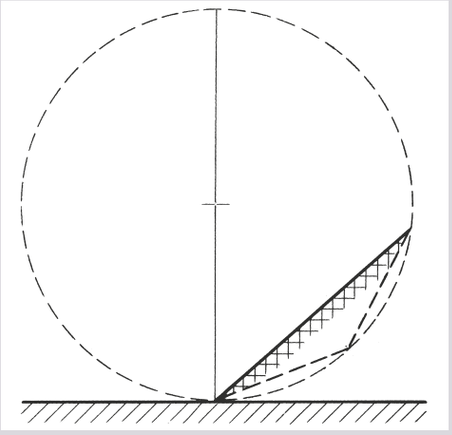
Another significant accomplishment by GALILEO is the recognition that projectiles travel along a parabolic path. To prove this assertion, GALILEO employed the prin-

Quotation 3.18, ***continued***

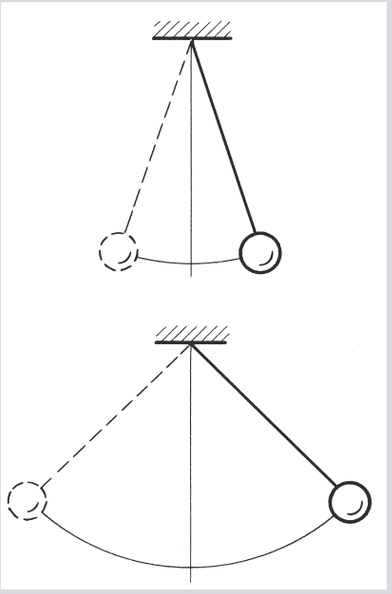
or outside resistances, and that the planes are hard and smooth, and that the figure of the moving body is perfectly round, so that neither plane nor moving body is rough.

Salviati: Your words are very plausible; but I hope by experiment to increase the probability to an extent which shall be little short of a rigid demonstration. Imagine this page to represent a vertical wall, with a nail driven into it; and from the nail let there be suspended a lead bullet of one or two ounces by means of a fine vertical thread, say from four to six feet long ... [Figure 3.41]

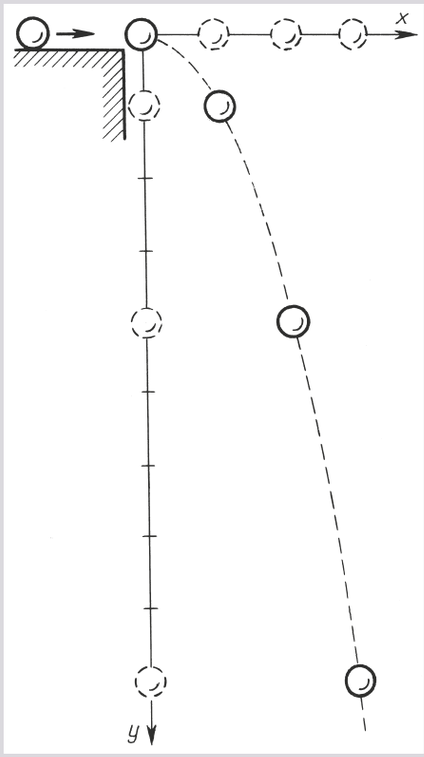
—GALILEO, *Dialogues Concerning Two New Sciences* [pp. 160–163]



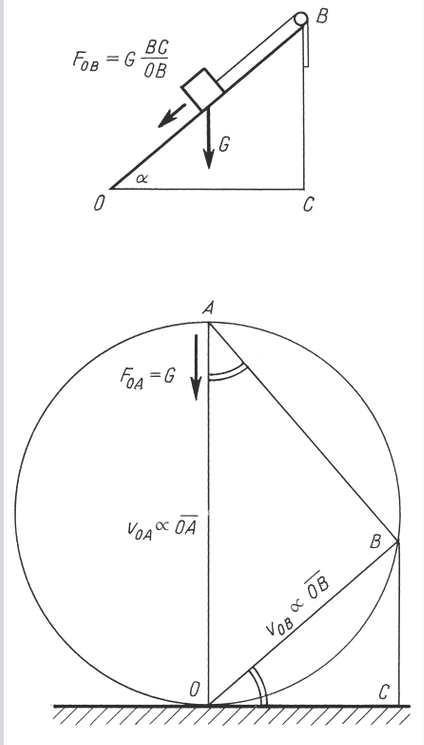
▲ **Figure 3.42** GALILEO determined correctly that a ball rolling along a circular arc will reach the base point sooner than one traversing a straight line, or any polygonal path with vertices lying along the arc.



◀ **Figure 3.43** The period of a mathematical pendulum consisting of a mass suspended from a massless thread is independent of the mass, but it *does depend* on the amplitude of deflection. GALILEO did not recognize this dependence, and he incorrectly generalized the rule that is valid for small deflections.



▲ **Figure 3.44** A projectile shot horizontally executes two independent motions: It moves at constant velocity in the horizontal direction and falls with constant acceleration in the vertical direction. The result is a parabolic path.



▲ **Figure 3.45** Galileo's (incorrect) dynamical explanation for why planes with different inclinations but whose endpoints lie on the same circle yield the same traverse times.

ciple of independence of motion. As depicted in Figure 3.44, an object leaves a horizontal surface with a certain velocity. The object's subsequent motion can be interpreted as the resultant of two independent motions: The object continues to move in the horizontal direction with constant velocity; at the same time, it commences free fall in the vertical direction. These two motions combine to create a parabolic path.

As we know today, this path may be described as follows: An object begins its motion in the horizontal direction with velocity v_0, and after time t, it has traversed the distance

$$x = v_0 t$$

in this direction. Meanwhile, in the vertical direction, during this same time interval, the object traverses the distance

$$y = \frac{1}{2} g t^2.$$

Using the first equation, we may represent time as a function of the x-coordinate, namely, $t = x/v_0$, and thereby obtain, after substituting this expression into the second equation,

$$y = \frac{1}{2}\frac{g}{v_0^2} x^2.$$

This equation represents a parabola in the xy-coordinate system. Of course, Galileo did not have such a coordinate representation at his disposal and therefore required a more complicated method of proof to establish his conclusion.

3.3.3 *Galileo's Greatness*

If we wish to assess Galileo's accomplishments, it must first be made clear that he investigated the *kinematics* of free fall and so described the motion of objects without associating this motion with a cause—the effective force. Thus in no way can Galileo be considered to be a pioneer of *dynamics*, in which case he would have explained motion in terms of the effect of force. Indeed, he consciously restricted his investigations to a mathematical description of observed phenomena and had no intention of searching for the effective cause (Quotation 3.18). Furthermore, when Galileo did attempt to establish a connection between motion and force, he proceeded, in his later years, in the usual Aristotelian manner. Thus, for example, he also gave a dynamical derivation of the theorem discussed earlier in which objects require the same amount of time to traverse various paths in a circle with the same base point. His initial premise was that the driving force down the incline stands in the same relation to the force acting vertically as that of the height of the path to its length, a relationship that was already known to Jordanus Nemorarius. If we now invoke the (completely false) Aristotelian thesis that the velocity is proportional to the effective force—where velocity is understood to mean the average velocity—then we find that the longer the inclined plane, the greater the effective force, and hence the average velocity increases as well. Therefore, the ratio between length and velocity remains constant. In other words, the running time is independent of the length of the inclined plane.

Galileo's reasoning is illustrated in a little more detail in Figure 3.45.

The force acting along the path (inclined plane) OB is given by

$$F_{OB} = G\frac{BC}{OB} \quad (= G\sin\alpha).$$

Since triangles OAB and OBC are similar, we have

$$\frac{BC}{OB} = \frac{OB}{OA} \quad (= \sin\alpha),$$

whence

$$F_{OB} = G\frac{OB}{OA},$$

or

$$\frac{F_{OB}}{G} = \frac{OB}{OA}, \quad \frac{F_{OB}}{F_{OA}} = \frac{OB}{OA}.$$

If we now set the velocity to be proportional to the effective force, we have additionally that

$$\frac{F_{OB}}{F_{OA}} = \frac{OB}{OA} = \frac{v_{OB}}{v_{OA}},$$

or

$$\frac{OB}{v_{OB}} = \frac{OA}{v_{OA}}.$$

This relationship allows us to conclude that objects traveling along paths *OB* with arbitrary angle of inclination require the same amount of time as an object in free fall along the diameter *OA*.

This leads us to the question of how Galileo could employ the false Aristotelian thesis to arrive at a doubtlessly correct result. We find that this was simply a happenstance due to the particular features of the experiment. To see this, let us use today's methods and write down the equation of motion, taking into account the fact that we have constant acceleration,

$$F = m\frac{\mathrm{d}v}{\mathrm{d}t} = m\frac{v}{t} = m\frac{2v_{\mathrm{m}}}{\mathrm{t}},$$

and hence

$$Ft = 2mv_{\mathrm{m}}, \quad \text{or} \quad Ft \propto v_{\mathrm{m}}.$$

We find that the *product of force and time* is proportional to the average velocity. In the present problem, the traversal times for all the paths are the *same*, and consequently, quite by chance, the average velocity happens to be proportional to the force acting along the incline, which is why Galileo was able to arrive at a correct result from an incorrect hypothesis.

It is, of course, not at all surprising that the influence of Aristotle is strong in Galileo's early works. The physics textbook written by Galileo's teacher is well known. It is evident from this work that Galileo did not study the most up-to-date physics even of that time. The extent of the influence on him of the Aristotelian theory of motion can best be understood from the fact that even such a great mind as Galileo could not, even by the end of his life, escape from under Aristotle's spell. We have already spoken, in fact, of Galileo's belief in circular motion as the natural motion for all bodies.

In his investigations into motion on an inclined plane, Galileo came quite close to the laws of inertia of Newtonian mechanics, according to which force is required to *change* the state of motion, not to *maintain* it. Thus Galileo asserted that an object on a smooth horizontal surface would maintain its state of motion indefinitely if no frictional forces were present (Quotation 3.19). However, by state of motion Galileo understood not motion in a straight line, but motion in a circular orbit (Figure 3.46). Motion of indefinite duration along a straight path was unthinkable for Galileo, since he imagined the cosmos as closed and finite.

This notion of circular motion crops up again in Galileo's work when he attempted to answer the question of how an extraterrestrial observer would observe the motion of a stone falling from a high tower. Galileo assumed that in such a case, the path of the stone would appear to describe a part of the circular path depicted in Figure 3.47, which passes through the center of Earth.

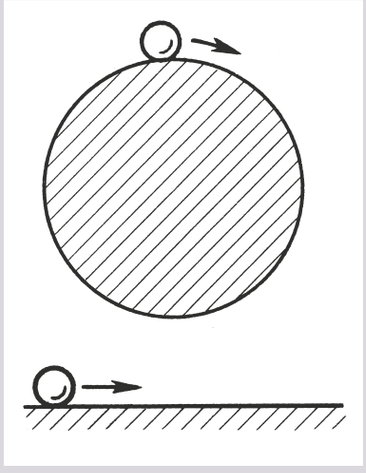
We may summarize Galileo's lasting contributions as follows:

In astronomy he was the first to observe and interpret a host of phenomena, among them the moons of Jupiter, lunar mountains, sunspots, the Milky Way

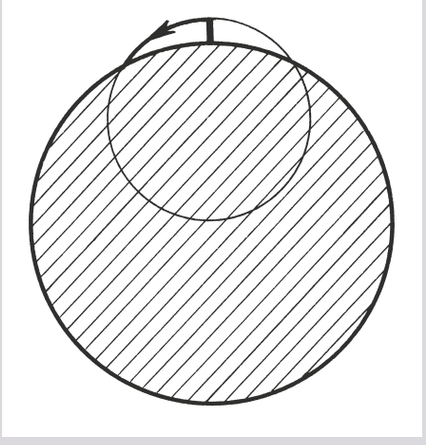
Quotation 3.19

Furthermore we may remark that any velocity once imparted to a moving body will be rigidly maintained as long as the external causes of acceleration or retardation are removed, a condition which is found only on horizontal planes; for in the case of planes which slope downwards there is already present a cause of acceleration, while on planes sloping upward there is retardation; from this it follows that motion along a horizontal plane is perpetual ... .

—Galileo, *Dialogues Concerning Two New Sciences* [p. 215]



▲ **Figure 3.46** For Galileo, motion without application of force—inertial motion—is motion along a circle.



▲ **Figure 3.47** Galileo imagined that a stone falling from a tower would traverse a circular path from the point of view of an extraterrestrial observer.

▼ **Figure 3.48** Simon Stevin (1548–1620): Bookkeeper for an Antwerp merchant, then tax official in Bruges; at 35, he began studies at the University of Leyden; from 1604, senior supervisor of agriculture and hydraulic engineering. His most important works are *De Thiende* (1585), the first systematic treatment of decimal fractions; *De Beghinselen der Weeghconst* (1586), introduction to the force parallelogram, the decomposition of forces on an inclined plane, experiments on the falling times of lead spheres of various masses; and *De Beghinselen des Waterwichts* (1586), equilibrium of fluids, pressure distribution, flotation properties of ships. Stevin also worked on the problem of magnetic declination of the compass needle.



understood as a grouping of stars, and the phases of Venus. Through these observations, he both gave an indirect proof for the correctness of the heliocentric system and brought the physics of the heavens and of Earth closer together.

Not only did Galileo introduce a method for the cultivation of the natural sciences, but he also successfully demonstrated how to apply it in concrete instances. We would like to stress this here especially because Bacon also offered a method for scientific investigation but he was unable to apply it to any concrete problem.

We emphasize two decisive steps in Galileo's method: (1) the definition of concepts as the proper starting point, and (2) the selection of a simple form of motion from among the complicated motions that appear in nature and the selection of idealized experimental conditions (completely smooth surface, completely smooth ball), which are never fully realized in nature. This way and *only* this way was it possible to treat the phenomena quantitatively. In the history of physics, Galileo was the first to speak of the necessity of ignoring insignificant influences, and he was bold enough to extrapolate from the actually measured values to the values to be expected in the case of an idealized object.

In the introduction to this book, when we discussed the various viewpoints of evaluating scientific progress, we singled out as a crucial factor the construction of a proper abstract model that does not exclude any significant aspects of the phenomena but at the same time makes them accessible to mathematical treatment. From this point of view, Galileo's contribution represents a milestone in the history of science.

Galileo wrote in Italian, the vernacular language of the people, in a clear and simple style. He was thus able to inspire an interest in science among the educated classes, but this did not filter down to the broader masses. It cannot, therefore, be said that science had suddenly become a matter of general public attention, but it was beginning to come into fashion. Among the educated people, humanists, and scientists, Galileo's experiments and discoveries had become a topic of conversation.

3.3.4 *In the Background: Stevin and Beeckman*

We will have to come back once more to the appraisal of Galileo after we look into what was accomplished in the fields of mechanics and astronomy during Galileo's lifetime outside of Italy.

Galileo had noteworthy predecessors who dealt with methodological questions and recognized the importance of experimentation and who while still paying homage to the great figures of the past, were willing to abandon the findings of classical science where necessary (Quotation 3.20). Perhaps the best expression of the doubts that arose in the appraisal of Galileo can be found in Einstein's foreword to the English translation of one of Galileo's books (see Quotation 3.21).

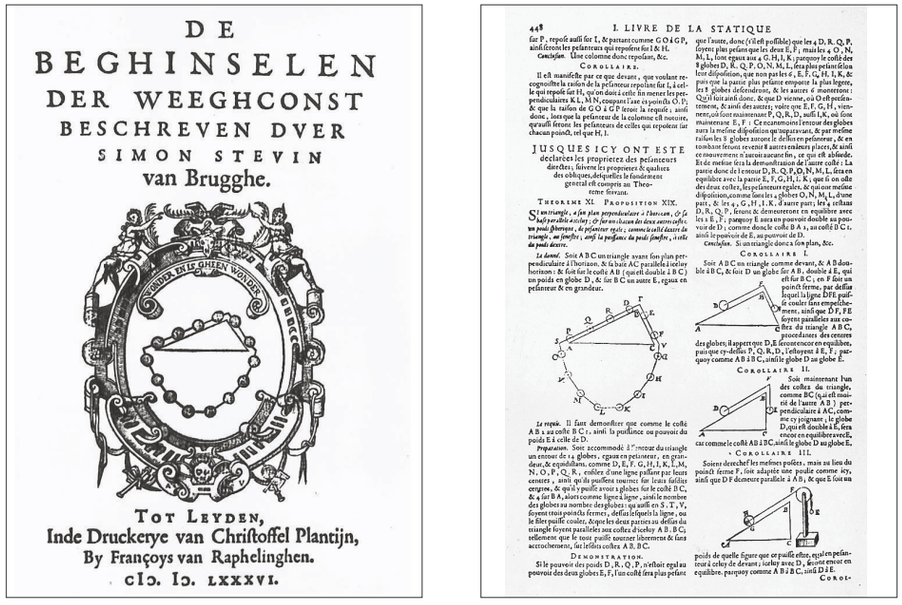
Above all, we should mention here a contemporary of Galileo, Simon Stevin (Figure 3.48), about whom we shall have much to say in our discussion of statics. In his 1586 book (Figure 3.49), we find the following description of an experiment with falling bodies, which Stevin conducted along with Jan de Groot. (Note that de Groot was the father of the famed jurist known as Hugo Grotius.)

> Let us take (as the very learned Mr. Jan Cornets de Groot, most industrious investigator of the secrets of Nature, and myself have done) two spheres of lead, the one ten times larger and heavier than the other, and drop them together from a height of 30 feet onto a board or something on which they give a perceptible sound. Then it will be found that the lighter will not be ten times longer on its way than the heavier, but that they fall together on to the

board so simultaneously that their two sounds seem to be one and the same rap. The same is found also to happen in practice with two equally large bodies whose gravities are in the ratio of one to ten.

We did not describe in any detail Galileo's famous free-fall experiment from the Leaning Tower of Pisa because, in all probability, Galileo never performed such an experiment. Nowhere did he write of such an experiment, although he was not the least bit hesitant to write about his successes. Some suspect that in his youth, Galileo did indeed undertake such an investigation but drew from it the conclusion supporting Aristotelian physics, to wit, that a heavy body falls faster than a lighter one. It is a curious circumstance that posterity has ascribed more importance to an experiment that in all likelihood was *never performed by* Galileo than to actual experiments performed by others.

If we wish to give Galileo credit for having given in his 1638 book *Discorsi* a precise mathematical description of a concrete motion—the uniformly accelerating


DE
BEGHINSELEN
DER WEEGHCONST
BESCHREVEN DVER
SIMON STEVIN
van Brugghe.

TOT LEYDEN,
Inde Druckerye van Christoffel Plantijn,
By Françoys van Raphelinghen.
CIↃ. IↃ. LXXXVI.

448 I. LIVRE DE LA STATIQUE

JUSQUES ICY ONT ESTE declarées les proprietez des pesanteurs directes; suivent les proprietez & qualitez des obliques, desquelles le fondement general est compris au Theoreme suivant.

THEOREME XI. PROPOSITION XIX.

Le donné. Soit ABC un triangle avant son plan perpendiculaire à l'horizon, & la baze AC parallele à iceluy horizon: & soit fait le coslé AB (qui est double à BC) un poids en globe D, & sur BC un autre E, egual en pesanteur & en grandeur.

Le requis. Il faut demonstrer que comme le coslé AB a au coslé BC; ainsi la puissance ou pouvoir du poids E a celle de D.

Preparation. Soit accommodé à l'entour du triangle un entour de 14 globes, egaux en pesanteur, en grandeur, & equedistans, comme D, E, F, G, H, I, K, L, M, N, O, P, Q, R, coulans d'une ligne pallante par leurs centres, ainsi qu'ils puissent tourner sur leurs fardilles cergeons, de sorte qu'il y puisse avoir 4 globes sur le coslé BC, & 7 sur BA, alors comme ligne à ligne, ainsi le nombre des globes au nombre des globes: qui aussi en S, T, V, soyent trois poinctels fermez, dessus lesquels la ligne, ou le fillet puisse couler, & que les deux parties au dessus du triangle soyent parallèles aux coslez d'iceluy AB, BC; tellement que le tout puisse sourner librement & sans accrochement, tel lesdits coslez AB, BC.

DEMONSTRATION.

Si le pouvoir des poids D, R, Q, P, n'estoit egal au pouvoir des deux globes E, F, l'un coslé sera plus pesant que l'autre, donc (s'il est possible) que les 4 D, R, Q, P, soyent plus pesans que les deux E, F; mais les 4 O, N, M, L, sont egaux aux 4 G, H, I, K; parquoy le coslé des 8 globes D, R, Q, P, O, N, M, L, sera plus pesant selon leur disposition, que non pas les 6, E, F, G, H, I, K, & puis que la partie plus pesante emporte la plus legere, les 8 globes defendront, & les autres 6 monteront: Qu'il soit ainsi donc, & que D vienne, où O est pesantement, & ainsi des autres; voilà que E, F, G, H, viennent où sont maintenant P, Q, R, D, aussi I, K, où sont maintenant E, F: Ce changement retourne des globes aura la mesme disposition qu'auparavant, & par mesme raison les 8 globes auront le dessus en pesanteur, & en tombant feront revenir 8 autres en leurs places, & ainsi ce mouvement n'auroit aucune fin, ce qui est absurde. Et de mesme sera la demonstration de l'autre coslé: la partie donc de l'entour D, R, Q, P, O, N, M, L, sera en equilibre avec la partie E, F, G, H, I, K; que si on oste des deux coslez, les pesanteurs egales, & qui ont mesme disposition, comme les 4 globes O, N, M, L, d'une part, & les 4, G, H, I, K, d'autre part; les 4 restans D, R, Q, P, seront de mesme entre en equilibre avec les 4 E, F; parquoy E aura en pouvoir double au pouvoir de D; comme donc le coslé BA a au coslé BC; ainsi la puissance de E, au pouvoir de D.

Conclusion. Si un triangle donc a son plan, &c.

COROLLAIRE I.

Soit ABC un triangle comme devant, & A B double à B C, & soit D un globe sur AB, double à E, qui est sur BC, en 8 sur un poinct fermé, par dessus lequel la ligne DFE puisse couler sans empeschement, ainsi que DF, FE soyent parallèles aux costez du triangle ABC, procedantes des centres des globes; il appert que D, E, seront encore en equilibre, puis que cy-dessus P, Q, R, D, l'estoyent à E, F; parquoy comme AB à BC, ainsi le globe D au globe E.

COROLLAIRE II.

Soit maintenant l'un des costez du triangle, comme BC (qui est moitié de l'autre AB) perpendiculaire à AC, comme cy joignant; le globe D qui est double à E, sera encore en equilibre avec E, car comme le coslé AB à BC, ainsi le globe D au globe E.

COROLLAIRE III.

Soient derechef les mesmes posées, mais au lieu du poinct fermé F, soit adjoûté une poulie comme icy, ainsi que DF demeure parallele à AB, & que E soit un poids de quelle figure que ce puisse estre, qui est pesanteur à celuy de devant; iceluy avec D, seront encore en equilibre, parquoy comme A B à BC, ainsi D à E.

COROL-

▲ **Figure 3.49** Title page and excerpt from one of Stevin's works.

motion—then we should also note the fact that Kepler in his 1609 book *Astronomia nova* published his first two laws of planetary motion, including a precise description of a body in motion on an elliptical path under variable velocity. Indeed, Kepler goes even further and seeks the ultimate cause that could explain how such motion would come into being. And we should not conceal here the fact that Kepler in this search also allowed himself to be guided strictly by Aristotelian ideas.

The degree to which the time was ripe for the methods used by Galileo and for the results he obtained can be seen in the diary of Isaac Beeckman (1588–1637), in which one can find notes beginning in 1618 relating to free fall. The content of these notes can be summarized as follows:

Let us suppose that gravitation acts in the form of a large number of impulses in rapid succession. Assume that each impulse results in an equally large increase in velocity. Let us suppose further—and here we have the first statement of the principle of inertia in the modern sense of the word—that *the object retains the velocity that has been imparted to it*. If we assume a velocity v_0 imparted by a single impulse and a time interval t_0 between impulses, then the velocity increases stepwise, as shown in

Quotation 3.20

Clearer proofs, in the discovery of secrets, and in the investigation of the hidden causes of things, being afforded by trustworthy experiments and by demonstrated arguments, than by the probable guesses and opinions of the ordinary professors of philosophy: so, therefore, that the noble substance of that great magnet, our common mother (the earth), hitherto quite unknown, and the conspicuous and exalted powers of this our globe, may be the better understood, we have proposed to begin with the common magnetick, stony, and iron material, and with magnetical bodies, and with the nearer parts of the earth which we can reach with our hands and perceive with our senses; then to proceed with demonstrable magnetick experiments; and so penetrate, for the first time, into the innermost parts of the earth. ...

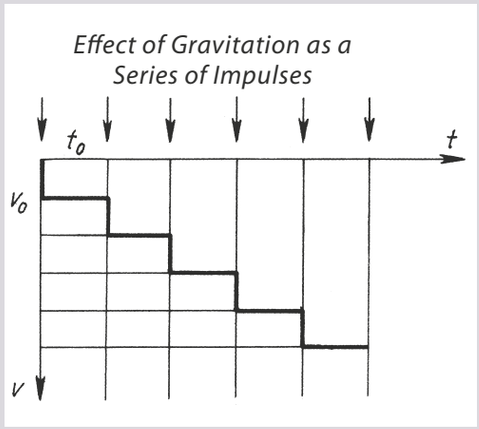
But to you alone, true philosophers, honest men, who seek knowledge not from books only but from things themselves, have I addressed these magnetical principles in this new sort of Philosophizing. But if any see not fit to assent to these self-same opinions and paradoxes, let them nevertheless mark the great array of experiments and discoveries (by which notably every philosophy flourisheth), which have been wrought out and demonstrated by us with many pains and vigils and expenses. In these rejoice, and employ them to better uses, if ye shall be able. ...

Therefore we but seldom quote ancient Greek authors in our support, because neither by using Greek arguments nor Greek words can the truth be demonstrated or elucidated more precisely or more significantly. For our doctrine magnetical is at variance with most of their principles and dogmas. ...

Sometimes we use new and unusual words, not that by means of foolish veils of vocabularies we should cover over the facts with shades and mists (as Alchemists are wont to do) but that hidden things which have no name, never having been hitherto perceived, may be plainly and correctly enunciated. ...

To those early forefathers of philosophy, Aristotle, Theophrastus, Ptolemy, Hippocrates, and Galen, let due honor be ever paid: for by them wisdom hath been diffused to posterity; but our age hath detected and brought to light very many facts which they, were they now alive, would gladly have accepted. Wherefore we also have not hesitated to expound in demonstrable hypotheses those things which we have discovered by long experience. Farewell.

—William Gilbert (Gilberd), *On the Magnet* (*De Magnete*), 1601



▲ **Figure 3.50** Beeckman's argument that gravitation acts in small impulses like hammer blows. The body does not lose the velocity that it has already gained.

Quotation 3.21

A man is here revealed who possesses the passionate will, the intelligence, and the courage to stand up as the representative of rational thinking against the host of those who, relying on the ignorance of the people and the indolence of teachers in priest's and scholar's garb, maintain and defend their positions of authority. His unusual literary gift enables him to address the educated men of his age in such clear and impressive language as to overcome the anthropocentric and mythical thinking of his contemporaries and to lead them back to an objective and causal attitude toward the cosmos, an attitude which had become lost to humanity with the decline of Greek culture.

In speaking this way I notice that I, too, am falling in with the general weakness of those who, intoxicated with devotion, exaggerate the stature of their heroes. It may well be that during the seventeenth century the paralysis of mind brought about by the rigid authoritarian tradition of the Dark Ages had already so far abated that the fetters of an obsolete intellectual tradition could not have held much longer—with or without Galileo.

—Albert Einstein, Foreword to Stillman Drake's translation of the *Dialogue Concerning the Two Chief World Systems*

Figure 3.50. As the frequency of the impulses increases, the step function passes in taking the limit to a straight line, and we may reach—*as a logical inference*—the conclusion that the velocity is proportional to the time. From this conclusion, then, follow all the relationships that Galileo had derived from *hypotheses*.

The path followed during the time interval $t = nt_0$ can be calculated thus: At velocity v_0 attained from the first imparted impulse, the object traverses the distance v_0t_0. Then with velocity $2v_0$, attained after the second impulse, the distance $2v_0t_0$ is traversed. And finally, with velocity nv_0 after the nth impulse, the distance nv_0t_0 is traversed. This yields a total distance traveled of

$$s = v_0t_0 + 2v_0t_0 + \cdots + nv_0t_0 = v_0t_0(1 + 2 + \cdots + n).$$

Upon performing the summation in parentheses, we obtain

$$s = v_0t_0\frac{n(n+1)}{2} = \frac{v_0t_0}{2}\left(n^2 + n\right).$$

If we now assume that the number of impulses is very large, then we may write, in the limit,

$$s \approx \frac{v_0t_0}{2}n^2 = \frac{1}{2}\frac{v_0}{t_0}(nt_0)^2 = \frac{1}{2}\frac{v_0}{t_0}t^2.$$

There is a particularly interesting entry in Beeckman's diary, and it is worth quoting the original Latin: *Haec ita demonstravit Mr. Peron*. Concealed behind that name is the young Descartes, at the time 22 years of age and completing his military service.

Beeckman had many truly revolutionary ideas, and there is the opinion that Descartes has Beeckman to thank for many ideas. Beeckman theorized that heat results from the motion of atoms and that water in a pump is lifted by the external air pressure and not by a *horror vacui*.

Coming back now to Galileo, it almost seems as if he, as well as the educated public, took no notice of the advances that had already been made in physics. This was indeed true to some extent, especially with respect to Kepler's laws, but also with Beeckman's work. However, it was not possible for further progress to have been made at the time by proceeding directly from these works. In both cases, further development would have required a full knowledge of the equations of motion. It was not until Newton that something new could be fashioned from Kepler's laws, so for the nearly 100 years until Newton, these laws would have no direct impact on the development of physics. In contrast, Galileo's laws regarding the motion of falling bodies could be utilized at once, and the two succeeding generations were able to work out all the ideas contained—or hidden—in Galileo's work. At the same time we should emphasize the difficulty of Galileo's achievement in comparison with Kepler's, insofar as the abstract conditions, that is the motions of planets, were readily given by nature, while Galileo had to create the abstract conditions himself in the setups of his experiments.

3.3.5 *The Possibility of Connection*

The questions in dynamics that remained open after Galileo's death are summarized in what follows.

Today, of course, we know what the goal was that had to be reached. It was the fundamental equation of dynamics, namely

$$\text{force} = \text{mass} \times \text{acceleration}.$$

Said another way, force is equal to the change in momentum per unit time. Among the basic notions of force, mass, velocity, and acceleration, Galileo clarified and

provided a mathematical description for the last two: velocity and acceleration. In this respect, the work with the problem of free fall was a fortunate step. However, it is also specifically in the case of free fall that we encounter a startling fact that was to play a very important role in the future. In contrast to Aristotelian dynamics, in which a large force produces a large velocity and therefore a heavy body should fall faster than a lighter one, the new dynamics will assert that a large force results not in a large velocity, but a *large change in velocity*. But how is it then, that a heavy body, upon which clearly a larger force is acting due to its greater weight, should be accelerated in free fall at the same rate as a lighter one? One of Galileo's correspondents, Giovanni Battista Baliani (1582–1666), grasped the essential point of the phenomenon when he suggested that the mass of an object has a double function: The mass is manifested, on the one hand, as an *agens*, that is, as gravitational mass in the modern sense of the term, and, on the other hand, as *passum*, or inertial mass. Today, we know that the acceleration of an object in free fall is independent of its weight because the mass appearing in the formula for gravitational force and the inertial mass, which characterizes the resistance of the object to a change in its state of motion, are proportional.

The difference between inertial and gravitational mass and the circumstances under which the one or the other property of mass appears can be easily demonstrated. If we hold a heavy metal ball in our hand, then the force that our muscles have to produce is given by the *gravitational mass* of the ball. If we now lay the ball on a completely smooth horizontal surface, the gravitational force can no longer play a role in its motion. In order for us to move the ball along the surface back and forth in a straight line at a particular frequency and amplitude, we need to apply a force that is related to the *inertial mass* of the ball. The course followed by Galileo was indeed suitable for arriving at the law of transformation between kinetic and potential energy and to solve complicated problems, but it was not suitable to lead the way to the fundamental laws of dynamics. In the given historical situation, making a connection between the two fundamental quantities of dynamics, force and mass, was rendered impossible by this double function of mass.

Even today, the role of momentum and the laws of change in momentum are most easily recognized and investigated in the context of a collision. The behavior of elastic balls moving on a horizontal surface can be studied independently of external influences and the phenomenon of collision is a commonplace experience for everyone. One of the crucial studies carried out in the period between Galileo and Newton—encompassing one or two generations—in addition to the study of complicated kinetic processes in a gravitational field was the investigation of the laws of collision. With Descartes, whose grandiose cosmology can be compared to an enormous game of billiards, the laws of collision assume a particular importance.

Before turning to our next topic, we offer one final observation on Galileo. Whatever our opinion may be with regard to his importance as an individual or his place in the history of science, we can still, while leafing through his works, find ourselves left spellbound three and a half centuries later by the depth of his thought and the magic of his style (Quotations 3.22 and 3.23).

Quotation 3.22

Philosophy is written in the grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is impossible to understand a single word of it [without these, one wanders about in a dark labyrinth].

—Galileo, *The Assayer* (*Il Saggiatore*) [pp. 237–238]

Quotation 3.23

But Nature, on the other hand, is inexorable and immutable; she never transgresses the laws imposed upon her, or cares a whit whether her abstruse reasons and methods of operation are understandable to men. ... For the Bible is not chained in every expression to conditions as strict as those which govern all physical effects; nor is God any less excellently revealed in Nature's actions than in the sacred statements of the Bible.

—Galileo, Letter to the Grand Duchess Christina [pp. 182–183]

3.4 The New Philosophy: Doubt Becomes Method

3.4.1 Francis Bacon and the Inductive Method

From the beginning of the seventeenth century, scientists in different fields—physicians, physicists, and astronomers—became increasingly aware of the fact that the

Quotation 3.24

Sagredo: I can put myself in Simplicio's place and see that he is deeply moved by the overwhelming force of these conclusive arguments. But seeing on the other hand the great authority that Aristotle has gained universally; considering the number of famous interpreters who have toiled to explain his meanings; and observing that the other sciences, so useful and necessary to mankind, base a large part of their value and reputation on Aristotle's credit; Simplicio is confused and perplexed, and I seem to hear him say, "Who would there be to settle our controversies if Aristotle were to be deposed? What other author should we follow in the schools, the academies, the universities? What philosopher has written the whole of natural philosophy, so well arranged, without omitting a single conclusion? Ought we to desert that structure under which so many travelers have recuperated? Should we destroy that haven, that Prytaneum where so many scholars have taken refuge so comfortably; where, without exposing themselves to the inclemencies of the air, they can acquire a complete knowledge of the universe by merely turning over a few pages? Should that fort be leveled where one may abide in safety against all enemy assaults?"

I pity him no less than I should some fine gentleman who, having built a magnificent palace at great trouble and expense, employing hundreds and hundreds of artisans, and then beholding it threatened with ruin because of poor foundations, should attempt, in order to avoid the grief of seeing the walls destroyed, adorned as they are with so many lovely murals; or the columns fall, which sustain the superb galleries, or the gilded beams; or the doors spoiled, or the pediments and the marble cornices, brought in at so much cost—should attempt, I say, to prevent the collapse with chains, props, iron bars, buttresses, and shores.

Salviati: Well, Simplicio need not yet fear any such collapse; I undertake to insure him against damage at a much smaller cost. There is no danger that such a multitude of great, subtle, and wise philosophers will allow themselves to be overcome by one or two who bluster a bit. Rather, without even directing their pens against them, by means of silence alone, they place them in universal scorn and derision. It is vanity to imagine that one can introduce a new philosophy by refuting this or that author. It is necessary first to teach the reform of the human mind and to render it capable of distinguishing truth from falsehood, which only God can do.

—Galileo, *Dialogues Concerning Two New Sciences* [pp. 56–57]

Aristotelian ideas that had long held sway in their various specialties were no longer tenable. However, in the universities, in the teaching methods as well as in the sciences, the emphasis was on the humanities (philosophy other than natural philosophy, that is, logic, metaphysics, rhetoric, and ethics), where the failure of Aristotelianism was not at all evident. Thus, we cannot yet speak about a complete repeal of Aristotle's philosophy. Nevertheless, little by little, the lack of a unified philosophical system began to make itself felt. Already in the *Dialogo*, the confirmed believer of the Aristotelian worldview, Simplicio, despairs that without a closed system, what could be the secure basis for the discovery of truth or for proving an assertion either true or false. At the same time, he deplored the decline of the Aristotelian edifice of ideas, in whose construction and embellishment so many eminent thinkers had participated (Quotation 3.24). Thus, in educated circles the need for a new philosophical basis was seen as a necessity. In part, a new method for the establishment of truth was awaited, while, in addition, a complete system to be built on this new method was envisioned in which everything would again assume its proper place. With such a new system, it should be possible to decide whether an *arbitrary* assertion is true or not. Ultimately, what was hoped for was the birth of a new philosophical system that would replace the Aristotelian worldview.

Both Francis Bacon (Figure 3.51) and Descartes consciously assumed not only the role of destructive critic, but also that of creator. Each offered a method for the discovery of scientific truth. Descartes, in addition, laid out a complete philosophical system. The goal that they pursued is clearly seen in the formulations that they employ in their works. Bacon deliberately titles his best-known work *Novum Organon*, to contrast it with Aristotle's *Organon*. Bacon set out in his book to provide scientists with as powerful a tool as the deductive logic of Aristotle had been. Descartes made it clear in the foreword to his *Principia Philosophiae* that he had created his system in order to do away with the system of Aristotle.

Bacon is frequently referred to as the father of English empiricism; in natural philosophy, he is credited as the founder of the inductive method. In agreement with Descartes, he reaches the conclusion that in the quest for scientific truth, we must first cleanse our minds in a critical way of all preconceptions; in other words, any scientific investigation should begin with the methodical doubting of all previously accepted knowledge. Except for this common starting point, the ideas of these two philosophers are as far from each other as is possible within philosophy.

Bacon begins by taking into consideration all factors that might cloud our reason. He calls such factors idols, illusions, or mirages. He then goes on to distinguish four types of illusion:

The *idola tribus*, or illusions of the tribe, are characteristic of all humanity. We may compare the human mind to a fun-house mirror, which reflects the outside world not in its true nature, but in a distorted fashion. Thus we are inclined to see order and regularity even where none exist. One result of this might be the belief, for example, that the planets have perfect circular orbits. Once we have convinced ourselves of such a regularity, we are then predisposed to bend the facts to match our theory and turn a blind eye to facts that would prove it false. We believe in superstitions, such as the prophetic power of dreams and omens, and in astrology, because we are willing to retain in our memories only those events that confirm the prediction. After analyzing all these shortcomings, Bacon offers some advice: We should be suspicious of every observation and every theory that our minds find particularly attractive.

The *idola specus*, or illusions of the cave, arise from the characteristics of the individual, reflecting personal character, predispositions, and education. We find, for

example, those who are theoretical and those who are practical; those predisposed to an analytic approach, others to a synthetic one; those fired up with enthusiasm for everything old and sanctified by tradition, and those who are only satisfied with whatever is new.

The *idola fori*, or illusions of the marketplace, result from interpersonal contact. A great deal of confusion is caused by the fact that the language for everyday communication is different from that used in scientific discourse. Many a philosophical problem is merely the result of improper use of language.

Finally, the *idola theatri*, or illusions of the theater, are the most dangerous of all because it is so difficult to free oneself of them. Whereas the other three forms of illusion are seen to be products of ordinary human existence, of our nature, and of our daily activities, we acquire for ourselves this fourth type of illusion only with a great expenditure of effort. These illusions are all the false doctrines and dogmas that we learn through the painstaking study of various philosophical systems and theories. BACON calls these "illusions of the theater" because, in his view, all recognized philosophical systems can be likened to theatrical pieces written by the great philosophers, in which "you may observe the same thing which is found in the theater of the poets, that stories invented for the stage are more compact and elegant, and more as one would wish them to be, than true stories out of history."

Let us now proceed to the examination of nature with our mind cleared of these illusions and attempt to describe simple phenomena or to discover the laws underlying them. We can approach this goal only if we make inquiries of nature itself. The best method of inquiry is the inductive method—not the inductive method of the ancient logicians, in which a bold generalization is made from a few examples, but a refined procedure of tabulation. Let us examine the nature of some physical characteristic, for example, the nature of heat. Whether warm or cold, heat can be looked upon as a simple property because it is accessible to our direct perception. We now construct a table (Table 3.2) in which we record the situations in which heat is encountered, noting down any accompanying phenomena. Then we prepare another table corresponding to the situations in which objects are cold, that is, in which heat is lacking. In this table, as well, we enter all accompanying phenomena. Finally, in a third table we note the situations in between. These are the famous Baconian "Tables of First Presentation," the tables for presence, absence, and degrees. We can now investigate which accompanying phenomena are always present in association with heat but are lacking when heat is absent. In this way, we may discover a connection between heat and a particular phenomenon and thereby find the property that we may hypothesize to be the cause of heat or that characterizes its basic nature. It was precisely in connection with heat that, after lengthy analysis, BACON discovered that heat and motion could be considered related phenomena. Finally, he arrived at the single correct result obtained with his method: that heat is by nature a form of motion.

BACON called this finding one of the first vintage, and if we follow BACON's ideas further, we may arrive at laws of the second and later vintages from a number of results of the first vintage with the application of new experimental findings. BACON's minutely described prescription for discovering natural laws is not usable in practice, and no one has ever discovered any laws by this method. BACON himself was exclusively a philosopher; he was not competent in science and, what is worse, either did not know or did not understand the scientific work of his contemporaries; in any case, these things were of no interest to him.



▲ **Figure 3.51** FRANCIS BACON (1561–1626): English statesman and philosopher. After studies in Cambridge, opened a legal firm; member of Parliament; in 1618, named lord chancellor under JAMES I. In 1621, imprisoned on charges of bribery, pardoned by the king, and then withdrew from public life.

He established a reputation as a writer with his *Essays*, 1597, complete (third) edition, 1625.

BACON intended to present his philosophical system in full in a planned multivolume work, *Instauratio Magna scientiarum* (great renewal of the sciences). The second volume of this work is the *Novum Organon* (1620). In his utopian novel *The New Atlantis* (1626), he describes a social organization in which learned societies play a prominent role. Later, the Royal Society was founded, based on this model. In his book *De dignitate et augmentis scientiarum*, he classified the sciences, distinguishing the sciences of memory, imagination, and reason, that is, history, poetry, and philosophy. In the *Encyclopédie*, DIDEROT started out from BACON's classification.

> We think some ground of hope is afforded by our own example, which is not mentioned for the sake of boasting, but as a useful remark. Let those who distrust their own powers observe myself, one who has amongst my contemporaries been the most engaged in public business, who am not very strong in health (which causes a great loss of time), and am the first explorer of this course, following the guidance of none, nor even communicating my thoughts to a single individual; yet having once firmly entered in the right way, and submitting the powers of my mind to things, I have somewhat advanced (as I make bold to think) the matter I now treat of. Then let others consider what may be hoped from men who enjoy abundant leisure, from united labours, and the succession of ages, after these suggestions on our part, especially in a course which is not confined, like theories, to individuals, but admits of the

continued on next page

Figure 3.51 *continued*

best distribution and union of labour and effect, particularly in collecting experiments. For men will then only begin to know their own power, when each performs a separate part, instead of undertaking in crowds the same work.

—Francis Bacon, *Novum Organum*, Book I.113

These three faculties form at the outset the three general divisions of our system of human knowledge: History, which is related to memory; Philosophy, which is the fruit of reason; and the Fine Arts, which are born of imagination. Placing reason ahead of imagination appears to us to be a well-founded arrangement and one which is in conformity with the natural progress of the operations of the mind. Imagination is a creative faculty, and the mind, before it considers creating, begins by reasoning upon what it sees and knows. Another motive which should decide us to place reason ahead of imagination is that in the latter faculty the other two are to some extent brought together. The mind creates and imagines objects only insofar as they are similar to those which it has known by direct ideas and by sensations. The more it departs from these objects, the more bizarre and unpleasant are the beings which it forms. Thus, in the imitation of Nature, invention itself is subjected to certain rules. It is principally these rules which form the philosophical part of the Fine Arts, which is still rather imperfect because it can only be the work of genius, and genius prefers creation to discussion.

—d'Alembert, *Preliminary Discourse to the Encyclopedia of Diderot* [p. 51]

▼ **Table 3.2** Bacon's tables for explaining the nature of heat. From the given compilation of facts, Bacon concluded that heat is a form of motion. His starting point is correct, and in the case considered, it led—by chance—to a correct conclusion, although in general, his procedure leads nowhere.

Tabula Essentiae et Praesentiae	*Tabula Graduum*	*Tabula Declinationis sive Absentiae*
Sun	Animal heat increases with work, fever, and the consumption of alcohol.	Moonlight
Fire		Water
Thermal spring		Cold wind
Friction		A cellar in summer
Slacking of lime		
	The amount of heat from the Sun depends on its position in the sky.	

Thus he held the Copernican model of the solar system—one hundred years after its formulation—to be an error, and although he was aware of Gilbert's pioneering work on magnetism, he could not appreciate it. He had no knowledge of Harvey's theory of the circulation of the blood, in spite of the fact that he was a patient of Harvey. It can be inferred from a number of places in Bacon's writings that, although he understood the importance of the interplay of theory and experiment for successful scientific work, his emphasis was rather on the collection of empirical data. He failed to appreciate the role of mathematics, and he condemned intuition entirely. No wonder that some remarked that if Bacon's method had been followed, neither Newtonian mechanics nor especially the theories of relativity and quantum mechanics would have ever come into being.

In defense of Bacon, we might add that in his case—as in the case of many other scholars—we are dealing with a phenomenon that in cybernetics is known as *overregulation*. At a time when truth was being sought in philosophical works of bygone eras or in abstract speculation, an effective reaction had to take the form of an overemphasis on the theme of "back to nature." That Bacon maintained his clear-sightedness is demonstrated in an oft-cited and very graphic image in which he compares purely empirical work to the activity of an ant dragging various objects into an unordered heap, and purely theoretical work to "spiders who make cobwebs out of their own substance." However, one arrives at truth only through activity that can be compared with that of the honeybee, who collects nectar from flowers and processes it into honey. In another place, Bacon writes, *neither the hand alone nor the spirit left to itself can create anything of significance*.

Bacon's other great accomplishment was that he recognized that the purpose of the natural sciences is the discovery of natural laws so that the forces of nature can be made to work in the service of mankind (Quotation 3.25). He consciously promulgated this view. Thus, the goal is no longer knowledge for its own sake, nor to try to understand the will of God, but rather human progress and the well-being of humanity.

Bacon was, in fact, religious, but he advocated the point of view that theological considerations have no place in scientific investigation. He was the first to stress the independent character of science, free from religious ideology.

In summary, we may say that Bacon correctly identified the goal of scientific research and that he correctly pointed out that we are able to understand natural laws only by consulting nature itself; moreover, he *correctly understood the autonomous character of science*.

Bacon's inductive method, however, placed far too much emphasis on the importance of cataloguing and is unusable in practice. Bacon *ignored the scientific discoveries of his time and the methods employed in achieving them*.

3.4.2 *A Method for Discovering Certain Truth: Descartes*

Bacon's doubt in the correctness of all existing theories led him to collect the concrete facts of nature. Descartes (Figures 3.52–3.54, Plate XVI), having a fundamentally different cast of mind, proposed a completely different course for arriving at certain truth. He had obtained the best possible scientific education available in his time at the Jesuit school in the city of La Flèche. It was there that he became convinced that everything that he had learned needed to be reconsidered and that there was no truth that was beyond criticism, with one exception: the truth of mathematics. He even wrote down why he believed in the truth of mathematical theorems: The mathematical *method* is itself the guarantee for the discovery of certain truth.

Namely, mathematics begins with truths that are clear and evident to the intellect. These are the axioms. From the axioms one can arrive at the most complex theorems by way of a chain of ideas, each link of which is clearly and evidently derived from the previous one. Descartes was dreaming of a natural science in which every assertion could be the last link in a chain of ideas that begins with an indubitable truth and consists of clear and logical steps. Therefore, Descartes set out to find this clear and evident initial truth, and for solving this problem, he himself prescribed the four rules that he recommended to others in their investigations:

1. Only those conclusions are to be accepted as true that we recognize as being clearly and evidently true, which means that they appear so clearly and evidently to our minds that all possibility of doubt is excluded.
2. When the problem to be solved is complicated it should be broken down into the greatest number of steps possible, so that it will be easier to solve.
3. We should organize our thoughts and proceed step by step from simple and easily confirmed statements to more complicated relationships.
4. Finally, at each step, the entire chain forged thus far should be examined so that one is certain that nothing essential has been overlooked.

Let us now attempt to discover a fundamental truth. We first cast away all preconceived opinions and we proceed with doubt about ourselves, about others, and about all books, even the most venerable (Quotation 3.26). But even our senses could deceive us, and therefore these are not to be trusted either. The same holds for our reasoning, for how many false theories and theses have thus been brought forth! The reality of the outside world is also to be doubted, for how often have we believed in the existence of particular external circumstances that then turned out to be illusions. Among all these doubts we find but a single certainty—the fact that we doubt; and in this way we arrive at our first certain knowledge, which is so sure and so evident that it can be attacked by no skeptic. This fundamental truth is the following: *Cogito, ergo sum*; I think, therefore I am (Quotation 3.27).

Doubt is an act of thought, for the existence of my body could be an illusion. My thinking self, which I can perceive clearly and distinctly, is thus distinguished from my body. But what guarantees the actual existence of the outside world? What gives us the certainty that a statement about the external world is in fact true and not an illusion? Descartes continues his proof as follows: My mind is clearly incomplete, since up until now, only doubt is certain, while knowledge is uncertain. However, knowledge overcomes doubt. Moreover, I have a clear and distinct idea about a perfect being. This idea of a perfect being, however, cannot arise from an imperfect spirit because it is a clear and distinct truth that an effect cannot be greater than the effective cause. In this way one may establish the existence of a perfect being.

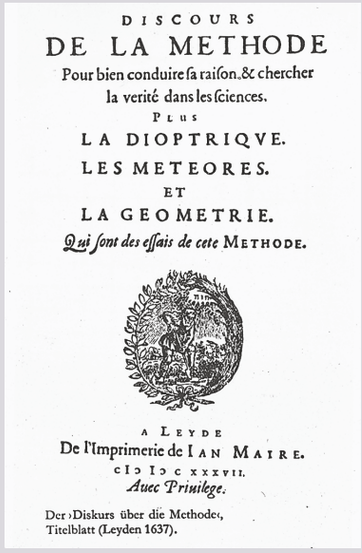
After the proof of the existence of the soul and of God, it was easy for Descartes to prove the actual existence of the outside world and to demonstrate that we are capable of understanding it correctly: from the perfection of God it follows that He would not deceive us.

Descartes's starting point, as well as the proof of the existence of God, is nothing new in the history of philosophy. Saint Augustine presents a similar approach for proving the existence of his own person (Section 2.5). Moreover, Saint Anselm of Canterbury formulated a proof that is similar to the one given by Descartes. The originality of the Cartesian circle of ideas lies in its emphasis on his fundamental truth—*cogito, ergo sum*—as well as in the structure of his argumentation. Earlier, the actual existence of the outside world and the corporeality of the individual were



▲ **Figure 3.52** René Descartes (1596–1650): Born in Le Haye, in Touraine, the son of an impoverished noble who had served as a jurist. From his eighth to his sixteenth year, Descartes attended the Jesuit college in La Flèche. In 1617, he served as a volunteer at the military academy in Breda (Holland), then traveled through Europe, and in 1619, he joined the army of the duke of Bavaria. During this time, on a certain cold November day while sitting by a warm stove, Descartes arrived at the fundamental idea of a new philosophy. In 1623, he traveled to Italy, but there is no indication that he had the intention of visiting Galileo. In 1629, he moved to Holland, seeking a peaceful environment for his work. He worked out his world system in the years 1629 to 1633. With an eye on Galileo's fate, Descartes did not dare publish his book *Le Monde ou traité de la lumière*, fearing that things might go for him as they had for Galileo. Thus, this book did not appear in the form originally intended, although many of the ideas were taken up in other of Descartes's works. The book *Regulae ad directionem ingenii*, written in 1628, was published only after Descartes's death. In 1637, the *Discours de la Méthode* was published; in 1641, *Meditationes de Prima Philosophia*; and in 1644, *Principia Philosophiae*. Descartes's last work was the *Traité des passions de l'Âme* (1649). In 1649, he moved to Stockholm at the invitation of Queen Christina, and there, unable to withstand the harsh climate, succumbed to pneumonia the following year.

> However, in addition to the rest, it must be firmly fixed in our memory as the supreme infallible rule that those things which have been revealed to us by God must be believed to be the most certain of all. And that, although perhaps some spark of the light of reason might seem to very clearly and evidently suggest to us something else; nonetheless, trust must be placed solely in divine authority rather than in our own judgment. But in those matters about which divine faith teaches us nothing, it by no means becomes a philosopher to accept as true something which he has never perceived to be true; and to trust in his senses, that is, the unconsidered judgments of his childhood, more than in mature reason.
>
> —René Descartes, *Principles of Philosophy,* I.75



DISCOURS
DE LA METHODE
Pour bien conduire sa raison, & chercher
la verité dans les sciences.
Plus
LA DIOPTRIQVE.
LES METEORES.
ET
LA GEOMETRIE.
Qui sont des essais de cette METHODE.

A LEYDE
De l'Imprimerie de IAN MAIRE.
cIↃ IↃ C XXXVII.
Avec Privilege.

Der ›Diskurs über die Methode‹, Titelblatt (Leyden 1637).

◀ Figure 3.53 The *Discours de la Méthode* appeared in 1637. This book was in fact planned as an introduction to three treatises that appeared at the same time: *La Géométrie*, *La Dioptrique*, and *Les Météores*. DESCARTES wished in these books to give proof of the practicability of his method.

The book *Les Météores* deals with cloud formation, rain, ice, heat, thunder, and lightning. The most significant scientific achievement here is the explanation of the formation of rainbows. *La Dioptrique* has the more practical goal of showing how to construct good lenses. This book is divided into two parts: vision, including the structure of the eye, and the law of refraction.

The book *La Géométrie* presents a foundation for analytic geometry. It is clear from DESCARTES's correspondence that he had worked out the most important results in this book 20 years earlier. In contrast to the other two books, this one is very difficult to understand, as if DESCARTES had set out deliberately not to care about giving a clear presentation. In the *Géométrie*, notation that is still used today was introduced and a theory of equations is given (Quotations 3.26 and 3.27).

> In the *Dioptrique* and in *Météores* I attempted to prove that my method is better than the customary, and I am convinced that with my *Géométrie* I succeeded in the proof. At the very beginning of the book I was able to solve a problem that, according to PAPPUS, none of the ancients were able to solve; and we may add that none of the moderns were able to solve either … and with regard to the second part, about the nature and properties of curves it seems to me that my geometry is to ordinary geometry as CICERO's *Rhetoric* is to a child's ABCs …
>
> People will have to realize that here I have accomplished something that is more than what was done by my predecessors; moreover, everybody also has to realize that our successors will never find anything in this area that I could not have found myself, just as well as they, if I had taken the trouble to seek it.
>
> —RENÉ DESCARTES, letter to MERSENNE, December 1637

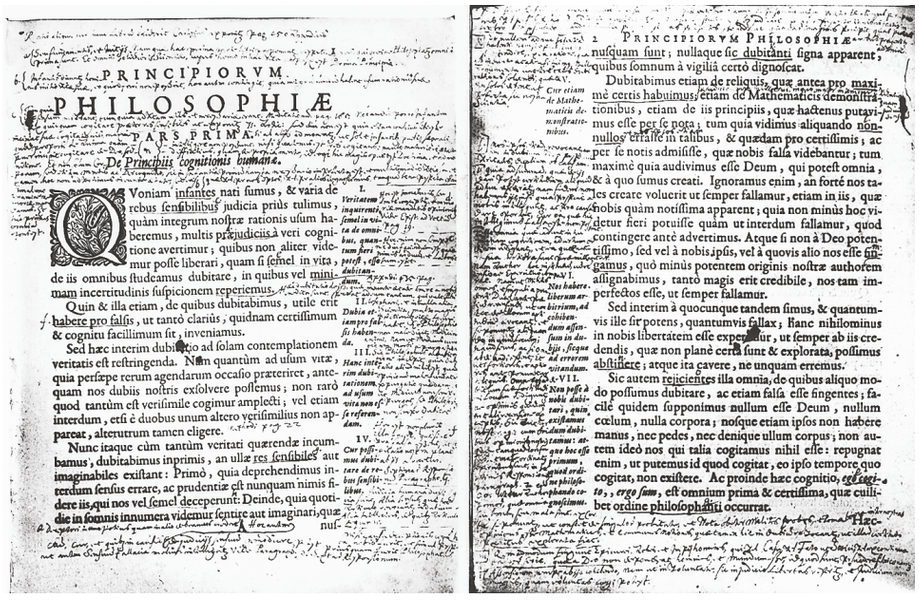
Quotation 3.25

Knowledge and human power are synonymous, since the ignorance of the cause frustrates the effect; for nature is only subdued by submission, and that which in contemplative philosophy corresponds with the cause in practical science becomes the rule.

—FRANCIS BACON, *Novum Organum*, Book I.3

taken as given, and from those axioms one sought to comprehend and prove the existence of God and the human soul. For DESCARTES, the idea of the soul and a supreme being were the clear and distinct givens, and from these he arrived at corporeal existence as well as the reality of the outside world (Figure 3.55).

At this point in DESCARTES's argument, we are ensured that the objects of the world are not illusions. So, we can trust in our clear and distinct perception. Many properties of objects—their color, smell, warmth, or coldness—may still be deceptive: by some they may be judged as being one way, by others another way. What is clear and evident for every observer, however, is the awareness of the three-dimensional extent of the objects; the extent is therefore the fundamental corporeal property.



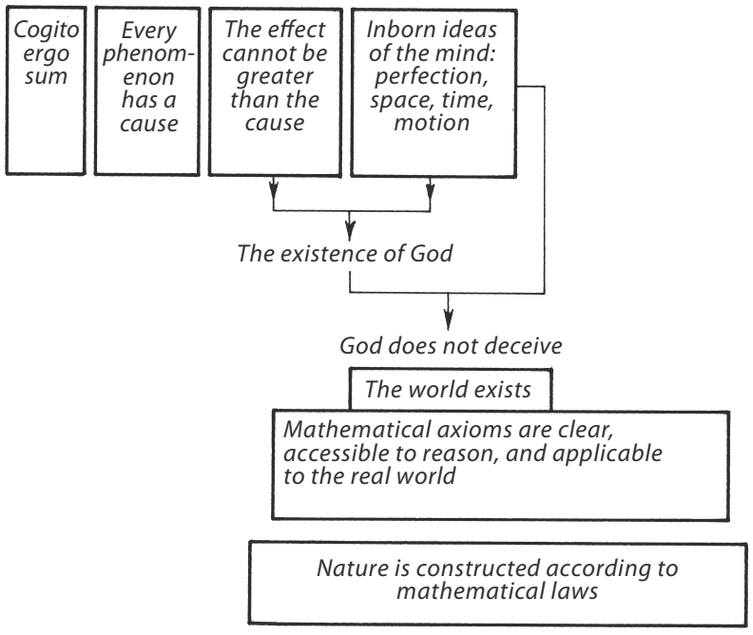
PRINCIPIORVM
PHILOSOPHIÆ
PARS PRIMA.
De Principiis cognitionis humanæ.

PRINCIPIORVM PHILOSOPHIÆ

▲ Figure 3.54 The four-part *Principia Philosophiae* appeared in 1644. In the first part, which contains the basic philosophical theses, the ideas that appeared in the *Discours* are organized and sharpened. In the second part, DESCARTES deals with the properties and laws of matter and motion. In the third part, DESCARTES's world model is explained on the basis of these laws, and in the fourth part, the properties of Earth are considered. Here appears the Cartesian explanation of the gravitational force as well as an extensive discussion of the tides.

It is here that the notion of duality enters into the Cartesian system, an issue which will cause many problems in the future. In addition to the thinking substance, the *res cogitans*, or more simply the soul, there appears the *res extensa*, the corporeal substance that is the physical matter. Already in DESCARTES's lifetime, the question was discussed regarding how these two substances can relate to each other, that is, how the soul can interact with the material world (Figure 3.56).

For the development of physics there are two significant aspects of Cartesian philosophy: The emphasis on the mathematical or deductive method was important for the theoretical development, whereas for concrete investigations, it was crucial that DESCARTES saw physical extent, and the closely related idea of motion, as primary characteristics of objects. With this, DESCARTES required that all properties of matter, including the effect of gravity, be derived from the extent and movement of objects. With this requirement he gave physics a program. Characteristic of him is the following statement, which is reminiscent of ARCHIMEDES: *Give me extent and motion, and from that I will construct the entire world*. With such a point of view, it is understandable that interaction between two objects is possible only with the impact of one upon the other, so that collision processes and laws of



▲ **Figure 3.55** Logical connections among Descartes's fundamental ideas.

collision attained particular significance in the Cartesian natural philosophy. We have already mentioned that the simplest path to the notion of change of momentum, and so ultimately to Newton's laws of motion, leads through collision processes. We do not know whether it was only coincidence or necessity that in applying mathematical methods to philosophy, Descartes worked out a conception of matter that was just what physics needed at that time for its further development.

3.4.3 *The Cartesian Laws of Motion*

Let us turn to Descartes's laws of motion, which are important from the standpoint of physics (Figure 3.57):

1. A body at rest remains at rest as long as no force acts on it; a body in motion continues its motion with unchanging velocity until it interacts with something that changes its motion.
2. Every moving body strives to maintain *linear* motion (Quotation 3.28).

Descartes recognized that a force is necessary to move an object in a circular orbit. We meet this Cartesian thesis already in his unpublished work *Le monde*. It is worth mentioning here that Galileo believed until a much later date that circular motion was a natural state in which no force is required to maintain the motion.

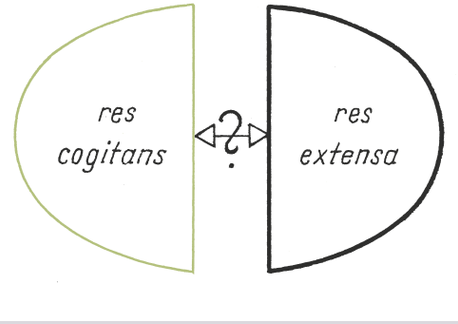
3. If a moving body meets another body and if it has less force to continue its motion than the other to resist the motion, then it will change its direction without a loss of its motion; but if it has greater force, then it takes the other body along with it, thereby losing as much of its motion as it imparts to the second body. For example, if a body collides with a larger body at rest, it will rebound back in the direction whence it came without any loss of motion.

From the general principles of collisions, Descartes derived eight concrete rules. The first of these is that if two identical moving bodies of equal hardness approach each other at the same velocity, then on collision, each body returns in the direction from which it came without a loss of speed. This law is simply the law of

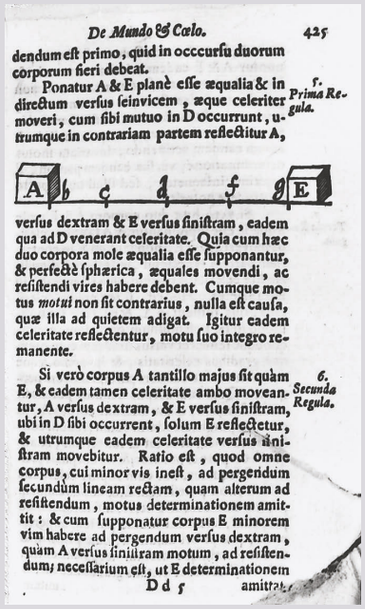
Quotation 3.26

And so, in order to philosophize seriously, and to discover the truth about all things which can be known: first, all prejudices must be abandoned or else we must carefully avoid trusting any of the opinions accepted by us in the past unless we first ascertain that they are true by submitting them to a new examination. Next, in order to proceed correctly, we must pay attention to the notions which we ourselves have in us; and all those, and only those, which we clearly and distinctly know while we thus attend to them, are to be judged true. In doing so, we shall first notice that we exist, insofar as our nature is that of a thinking thing; and at the same time we shall also notice both that God is, and that we depend upon Him, and that from a consideration of His attributes we can investigate the truth of the remaining things, since He is their cause. Finally, we must note that, in addition to the notions of God and of our mind, there is also in us the knowledge of many statements of eternal truth, for example that no thing can be produced from nothingness, etc.; and similarly, there is the knowledge of a certain corporeal nature, or one extended, divisible, mobile, etc.; and also the knowledge of certain sensations which affect us, for example, pain, colors, flavors, etc. (although we do not yet know what causes us to be thus affected). And comparing these things with those which we formerly more confusedly thought, we shall acquire the practice of forming clear and distinct conceptions of all things which can be known. And the chief principles of human knowledge seem to me to be contained in these few.

—René Descartes, *Principles of Philosophy*, I.75



▲ **Figure 3.56** According to Descartes, there are two independent substances. The question of the Cartesian duality of the world, the interaction between the two substances, became the focus of philosophical discussions in the following century.



De Mundo & Cœlo. 425

dendum eſt primo, quid in occurſu duorum corporum fieri debeat.

Ponatur A & E planè eſſe æqualia & in directum verſus ſeinvicem, æque celeriter moveri, cum ſibi mutuo in D occurrunt, utrumque in contrariam partem reflectitur A, 1. Prima Regula.

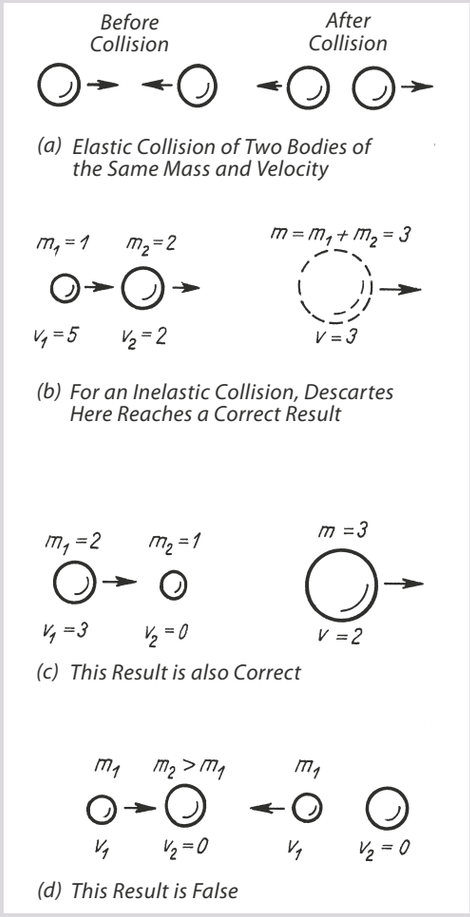
A b c d f g E

verſus dextram & E verſus ſiniſtram, eadem qua ad D venerant celeritate. Quia cum hæc duo corpora mole æqualia eſſe ſupponantur, & perfectè ſpharica, æquales movendi, ac reſiſtendi vires habere debent. Cumque motus *motui* non ſit contrarius, nulla eſt cauſa, quæ illa ad quietem adigat. Igitur eadem celeritate reflectentur, motu ſuo integro remanente.

Si verò corpus A tantillo majus ſit quàm E, & eadem tamen celeritate ambo moveantur, A verſus dextram, & E verſus ſiniſtram, ubi in D ſibi occurrent, ſolum E reflectetur, & utrumque eadem celeritate verſus ſiniſtram movebitur. Ratio eſt, quod omne corpus, cui minor vis ineſt, ad pergendum ſecundùm lineam rectam, quam alterum ad reſiſtendum, motus determinationem amittit: & cum ſupponatur corpus E minorem vim habere ad pergendum verſus dextram, quàm A verſus ſiniſtram motum, ad reſiſtendum; neceſſarium eſt, ut E determinationem 6. Secunda Regula.

Dd 5 amittat

◀ **Figure 3.57** This is how the Cartesian laws of motion were described in a 1679 university textbook. *(Antonii Le Grand Institutio Philosophiae secundum Principia D. Renati Descartes. In usum juventutis academicae*. Noribergae. 1679.)



▲ **Figure 3.58** Special cases of collision, one with an invalid description. Descartes erroneously assumed that $m|\mathbf{v}|$ is constant.

collisions applied to completely elastic bodies of equal mass, which is valid also today. However, most of the other rules are erroneous. Figure 3.58 shows one invalid case and some correct ones. The errors were already pointed out in Descartes's lifetime. Even Descartes himself realized that his results did not always agree with everyday experience. Nevertheless, he provided a living example of the kind of human weaknesses suggested by Bacon, seeking the cause in the experimental conditions rather than in an error of his theory (Quotation 3.29). We know today, of course—and this was even pointed out by the young Huygens—that the errors in the laws follow from the fact that Descartes failed to distinguish between elastic and inelastic collisions, and furthermore, that he did not take into account the vectorial nature of momentum, which in the simple cases he investigated meant that he did not treat velocity as a signed quantity.

After his formulation and detailed elaboration of the basic laws of motion, Descartes set out to explain, with the help of these laws, all phenomena of the visible world. Such a project was, of course, grandiose, and it can be immediately seen that it was over-hasty and therefore impracticable. One cannot hope on the basis of such simple laws, not even expressed in quantitative form, to understand the complex processes that govern the world. However, it is worth going somewhat more deeply into the system proposed by Descartes, not only because of the boldness of his ideas, but also because their exposition had so many kernels of truth in them that would later be successfully taken up in the development of individual branches of science.

3.4.4 *The First Cosmogony*

Descartes wished to provide not only an explanation for the current state of the world, but also to give an answer to the question of how this state could have arisen. In posing such a question, he was obliged to assuage both his own conscience and those of his readers that his theory did not stand in contradiction to the Bible (Quotation 3.30). Here he did not turn to the usual formula that his theory was merely of a hypothetical nature, its truth content uncertain, and that it provided at most—as with celestial mechanics—an improved computational method. Rather, he stated explicitly that his explanation for the origin of the world was surely false because it was well known that the world, together with everything on it, was created by God in six days in all its perfection. Then what was the point of such a theory of origins? Descartes's answer was that we can still ask the question of whether the current state of the world might also be understood in terms of its development in accordance with natural laws from some initial state. Human reason would be just as reassured from such a reconstruction and understanding as if the assumptions with regard to the initial state were actually true.

Let us assume, following Descartes, that at the beginning the world consisted of a completely homogeneous material endowed with a certain amount of motion. The originally homogeneous world broke into parts, and through collisional processes, three primordial materials were formed (Figure 3.59). The first of these materials is fine and flexible; it corresponds essentially to Aristotelian fire. (In the first presentation of his theory in *Le Monde*, Descartes actually used this designation.) The second material, which in Aristotelian nomenclature corresponds to air, is characterized by mobility and sphericity. Finally, there are the larger and coarser pieces, corresponding to Aristotelian earth. A mixture of these three materials fills space continuously, without gaps. That is, in the Cartesian theory there are no atoms and neither

is there vacuum, so that all of space is somehow filled with a mixture of three fluids. Due to the initial motion, gigantic whirlpools formed in the fluid, which eventually led to a separation of the different sorts of materials. At the center of these vortices, the fine, glowing matter collected, out of which the Sun and fixed stars were formed. The heavier types of matter worked their way toward the outside because a greater force acts on them due to the rotational motion of the vortex. Associated with each fixed star, there is a vortex-filled region of space in which matter of the third type accumulates, eventually forming the planets. The smooth, elastic spheres of the second type of matter fill all of space, acting as transmitters of light.

Figure 3.60 depicts some of these vortices. They border on one another in such a way that they do not impede each other's motion. It is frequently the case that the poles of some vortices will touch the stream of matter along the equator of other vortices. This matter strives to move away from the axis or rotation, with the various types of matter behaving according to their various degrees of mobility, but this will be successful only where there is no opposing effect along the axis of an adjacent vortex. Thus, along their axis from both sides there is a constant flow of matter into the interior of the vortices that collects at their center. This is how a constantly turbulent flow of fine matter forms around the fixed stars, which is the source of their light.

Now we must imagine the motion taking place within the solar system: the matter whirling about the Sun carries Earth with it, while the matter whirling around Earth carries the Moon along. To see how this might work, consider a vortex on the surface of a river that moves along with the current. Descartes could claim that the planets *do not move* because motion consists in a change of position relative to the immediate environment, and this position does not change (Quotation 3.31, Figure 3.61).

Descartes viewed attractive force acting at a distance as an occult phenomenon that could not exist. With the movement of vortices, he not only explained the peculiarities of the planetary orbits, but also solved the problem of gravitation, the question of why bodies are heavy, or, more precisely, why a heavy body attempts to fall toward the center of Earth. Descartes illustrated this as follows (Figure 3.62): Suppose we shake a mixture of lead shot and wood shavings in a glass vessel until it is homogeneous and then rotate the mixture about a vertical axis. We would observe that the wood shavings move toward the axis of rotation, whereas the lead shot is pressed against the glass wall of the vessel. (Note that this is the principle of centrifugal separation that is widely used in industry and in science today.)

Of course, for this to work, we would have to believe that it is the second type of matter, which is fine and mobile, that corresponds to the lead shot. By the way, Descartes could conclude from his theory even the fact that a body thrown with sufficiently great force would not return to Earth.

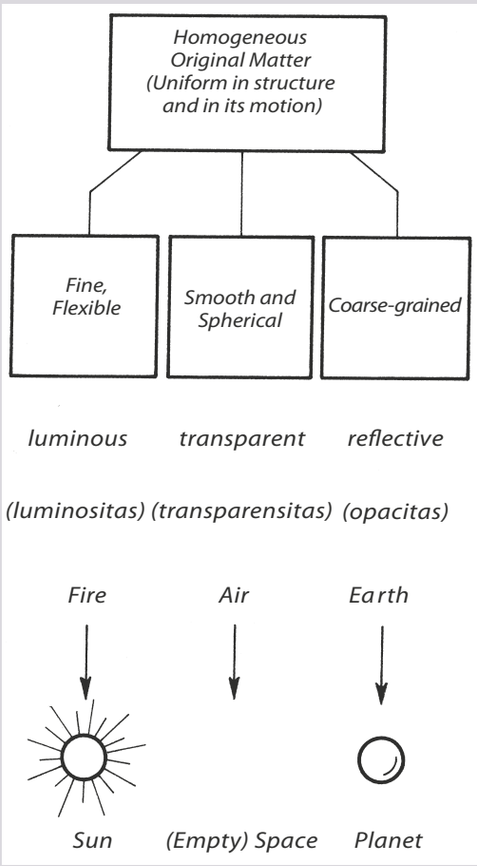
The positive aspects of the Cartesian system of ideas can be summarized as follows:

Despite its metaphysical background, this interpretation is decidedly and unequivocally materialistic.

Descartes emphasizes the material unity of the entire world and does not content himself with the conclusion that celestial and terrestrial laws are alike in principle, but asserts that the laws of motion for the planets—or more generally for all celestial bodies—are identical to those for terrestrial objects.

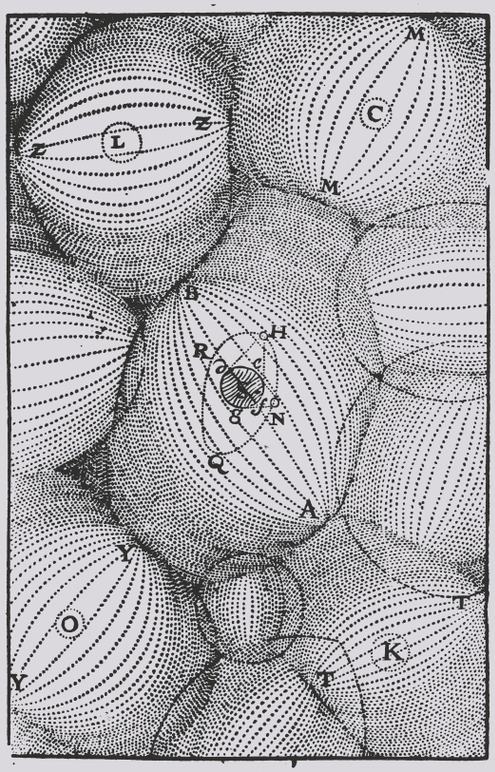
Descartes strives for clarity, and he emphasizes that all observable phenomena are amenable to rational understanding and demonstration.

This striving for clarity is closely connected with Descartes's attempt to explain the universe on the basis of the laws of mechanics (Quotation 3.32). This effort for a mechanistic explanation of the world will be dominant for the next three

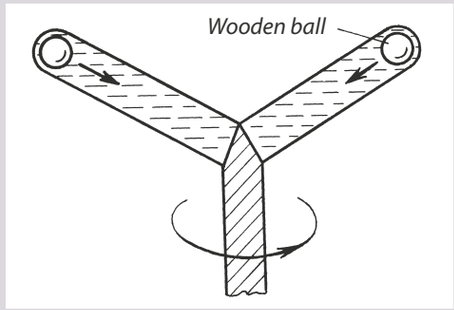


▲ **Figure 3.59** The three "basic" materials of the Cartesian world.

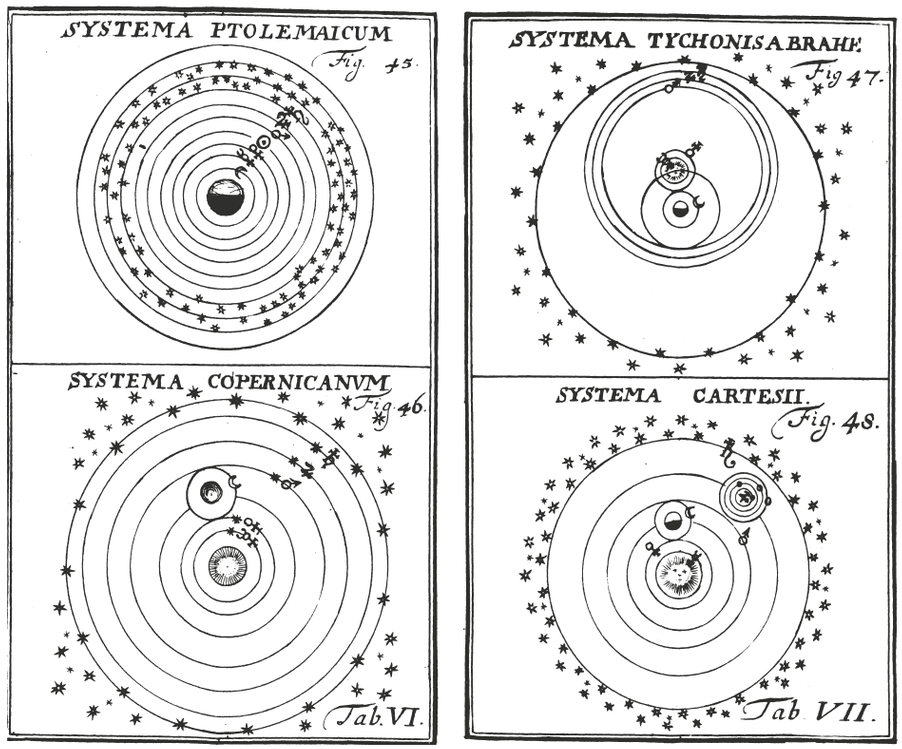
> And so now we have two very different kinds of matter, which can be called the two first elements of this visible world. The first is that of the matter which has so much force of agitation that, by colliding with other bodies, it is divided into particles of indefinite smallness, and which adapts its shapes to fill all the narrow parts of the little angles left by the others. The second is that of the matter which is divided into spherical particles, admittedly very small if compared with those bodies which our eyes can discern; yet of a certain and determined quantity and divisible into others much smaller. And, in a short while, we shall discover the third, which is composed of parts which are either much bulkier or have shapes less suited to movement. And we shall show that all the bodies of this visible world are composed of these three elements: the Sun and the fixed Stars of the first, the Heavens of the second, and the Earth, the Planets, and the Comets of the third. For since the Sun and the fixed Stars emit light, the Heavens transmit it, and the Earth, the Planets, and the Comets reflect it: we shall, without being inaccurate, ascribe this threefold difference in visual appearance to three elements.
>
> —René Descartes, *Principles of Philosophy*, III.52



▲ **Figure 3.60** The axes of the represented vortices are denoted by *AB*, *TT*, *YY*, *ZZ*, and *MM*, while the associated centers (fixed stars) are denoted by *S*, *K*, *O*, *L*, and *C*. Along the axis *AB*, fine matter enters the vortex from vortices *L* and *K*, rotates about the point *S*, and then exits along a spiral path, so that it eventually flows into vortices *O* and *C* along the corresponding vortical axes. The characteristic quantities of the second material—size and speed—eventually exhibit certain regularities. Within the orbit *HNQR*, the particles become finer and move faster as they approach the center of the vortex, whereas outside the orbit their sizes remain constant but their speed increases with increasing distance from the orbit. Therefore, this orbit is where their speed is the least. Matter rotating about the star can accumulate and form spots on the surface of the star, which can eventually lead to the star coming apart and another vortex incorporating its material. It is thus that planets and comets are formed.



▲ **Figure 3.62** This is how Descartes explained the origin of gravitation. Huygens (1690, *Discours de la cause de la pesanteur*) raised the objection that under this interpretation, the weight of a body would be greater if it contained less material.



▲ **Figure 3.61** Even at the beginning of the eighteenth century, the Cartesian system was considered one of the possible—and most progressive—systems. (H. Winckler, *Institutiones Philosophiae Wolfianae*. Lipsiae 1735.)

centuries to come. Even up to the end of the nineteenth century, no less a figure than Lord Kelvin maintained that natural phenomena can be truly explained only if we work with material forms that satisfy the laws of hydrodynamics. Even if the modern hydrodynamic models of nuclear physics and the hydrodynamic interpretation of quantum mechanics cannot be linked directly to Descartes, still their origin is to be sought with him.

Descartes was the first to give a role to a conservation law, even if not to its full significance (Quotation 3.33).

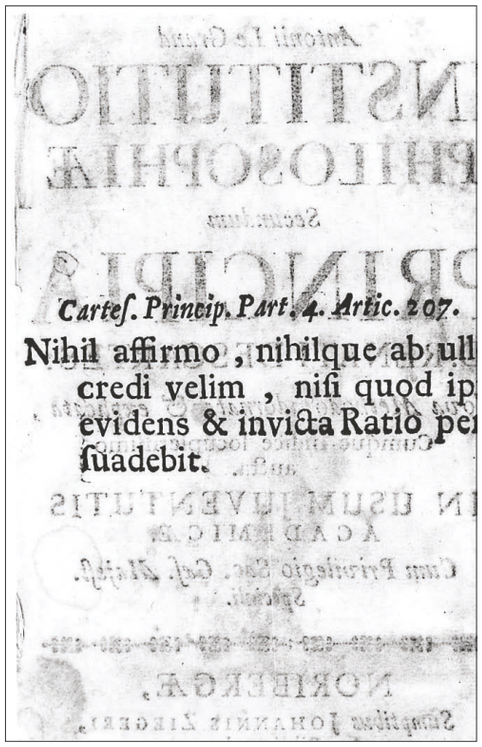
For the first time in the history of science, an explanation for the origin and development of the existing cosmos is given.

Despite his religious leanings, Descartes made a decisive break with the theological interpretation of the material world (Quotation 3.34).

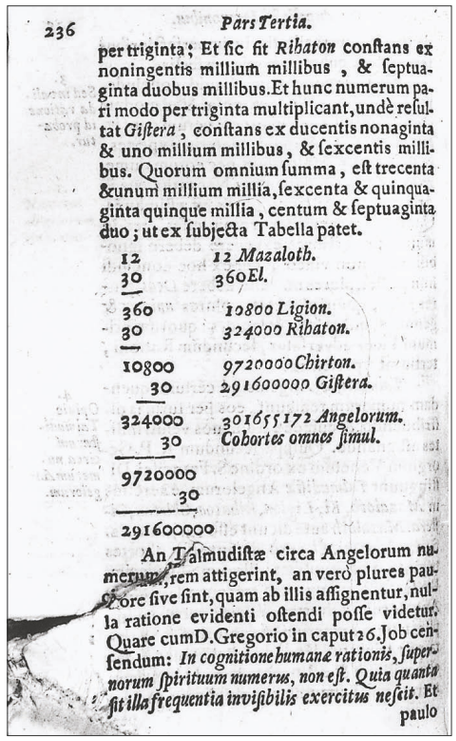
Descartes made a name for himself in science not only for the work discussed above. Whenever we refer to the orthogonal coordinate system as "Cartesian," we honor above all his accomplishments in mathematics, and his contributions to optics are recalled in the Snell-Descartes law of refraction. We shall have more to say about these in the following sections.

Descartes's greatest mistake lies in his overemphasis on reasoning to the detriment of experimentation. He relied so heavily on the validity of clearly perceptible truths in the real world that he assigned to experience only the subordinate role of supporting these truths and providing simple-minded explanations for them. He did not grant experience the role of arbitrator.

The acceptance of clear and distinct ideas is up to the individual for whom their truth is indeed evident. But what kind of truths this can lead to is illustrated in Figures 3.63 and 3.64.



▲ **Figure 3.63** Descartes's influence is demonstrated by the number of books that take his basic idea as their motto. He is quoted like the Bible. Here we can read: *I affirm nothing, nor would I wish anyone to believe anything except when he is convinced by clear and irrefutable reason*.



▲ **Figure 3.64** However, quoting Descartes has not prevented the author of this work from calculating the number of angels to nine significant figures.

Toward the end of the century, Huygens recognized that reason and experience are of equal importance (Figure 3.65); however, the Cartesian ideas continued to remain in force, although their progressive character was gradually diminishing, so that in the end, the designation "Cartesian" became synonymous with "Aristotelian" and was considered reactionary.

Today, we consider Descartes—even taking into account his shortcomings—an outstanding personality both in the history of science and the history of philosophy.

3.4.5 *On the Periphery of Western Culture*

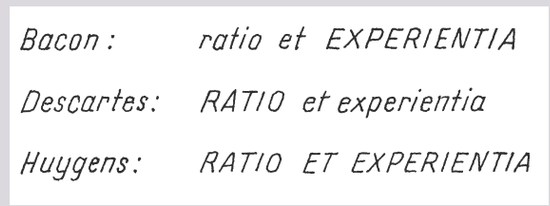
As we saw in Section 2.1.2, the Latin culture emanating from Rome and the Greek Orthodox culture spreading from Byzantium met in the eastern border areas of Poland and Hungary. Thus the typically Western institutions (universities, knightly orders) as well as spiritual, intellectual, and artistic currents (Gothic, Renaissance, Reformation) arrived in the far eastern part of Europe either late or not at all. We mention the remarkable fact that the easternmost Gothic cathedral was built in Braşov. This city (at the time in Hungary, today in Romania) was founded by the German knightly order that had been summoned there by the Hungarian king to defend the eastern border.

If we observe the immediate vicinity along this border, we see some institutions of higher education being created and then rapidly disappearing, which is an indication that in this area, some of the more powerful rulers were making attempts at transplanting Western culture, which, however, for a variety of reasons was never able to take root. Thus, for example, as early as the first half of the

Quotation 3.27

Thus, since our senses sometimes deceive us, I decided to suppose that nothing was exactly as our senses would have us imagine. And since there are men who err in reasoning, even in the simplest matters in geometry, and commit paralogisms, judging that I was just as prone to err as the next man, I rejected as false all the reasonings that I had previously taken for demonstrations. And finally, taking into account the fact that the same thoughts we have when we are awake can also come to us when we are asleep, without any of the latter thoughts being true, I resolved to pretend that everything that had ever entered my mind was no more true than the illusions of my dreams. But immediately afterward I noticed that, during the time I wanted thus to think that everything was false, it was necessary that I, who thought thus, be something. And noticing that this truth—*I think, therefore I am*—was so firm and so certain that the most extravagant suppositions of the skeptics were unable to shake it, I judged that I could accept it without scruple as the first principle of the philosophy I was seeking.

—René Descartes, *Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences*, Part IV [p. 17]



▲ **Figure 3.65** By the end of the century, reasoning and experimentation reached a healthy balance.

Quotation 3.28

Furthermore, from the same immutability of God, we can obtain knowledge of the rules of laws of nature, which are the secondary and particular causes of the diverse movements which we notice in individual bodies. The first of these laws is that each thing, provided that it is simple and undivided, always remains in the same state as far as is in its power, and never changes except by external causes. Thus, if some part of matter is square, we are easily convinced that it will always remain square unless some external intervention changes its shape. Similarly, if it is at rest, we do not believe that it will ever begin to move unless driven to do so by some external cause. Nor, if it is moving, is there any significant reason to think that it will ever

continued on next page

Quotation 3.28, ***continued***

cease to move of its own accord and without some other thing which impedes it. We must therefore conclude that whatever is moving always continues to move as far as it is in its own power. ...

The second law of nature [which I observe] is: that each part of matter, considered individually, tends to continue its movement only along straight lines, and never along curved ones.

—René Descartes, *Principles of Philosophy*, II.37, 39 [pp. 59–60]

Quotation 3.29

We have discovered certain principles concerning material things, and there can be no doubt about the truth of these principles, since we sought them by the light of reason and not through the prejudices of the senses.

—René Descartes, *Principles of Philosophy*, III.1 [p. 79]

Quotation 3.30

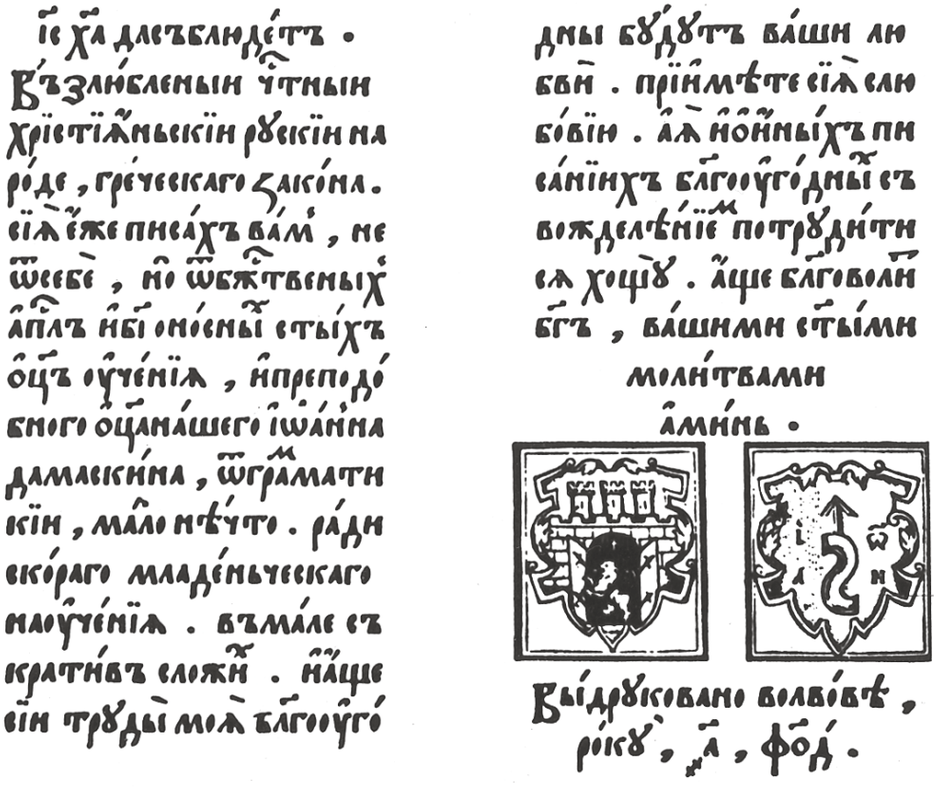
In addition to these rather general things, I could also include here, among the phenomena, several other specific things, concerning not only the Sun, the Planets, the Comets, and the fixed Stars, but also the Earth: that is, everything which we see [around the Earth, or which occurs] on its surface. For indeed, in order to know the true nature of this visible world, it is not sufficient to find some causes by which one can explain what appears in the heaven very far away from us; it is necessary also to be able to deduce from them the things we see very close to us [and which affect us more noticeably]. But even so, [I think] it is unnecessary for us to consider them all [immediately, and deem it preferable for us to attempt first] to discover the causes of these more general things [which I have enumerated, so that we may see subsequently whether we can also deduce from these same causes all the more specific things which we shall have ignored while searching for these causes]. And we shall know that we have correctly determined these causes when we observe that we can explain, by their means, not only those phenomena which we have considered up to now, but also everything else about which we have not previously thought. ...

And certainly, if the principles which I use are very obvious, if I deduce nothing from them except by means of a Mathematical sequence, and if what I thus deduce is in exact agreement with all natural phenomena; it seems [to me] that it would be an

continued on next page

eleventh century, the court of Yaroslav I, the Wise (978–1054), prince of Kiev, was the scene of an intensive cultural life. Anna Yaroslavna, the prince's daughter and later wife of the French king Henry I, was one of the few women in Europe who could read and write. Kievan Rus eventually collapsed after the Mongol invasion in 1240. Hungary also suffered from this invasion, but was able to regain its power and independence within a generation, and by the end of the Middle Ages, it had become a great power in Central Europe. Under the rule of Louis I, "the Great" (1326–1382), a personal union with Poland was established, and an attempt was made to enter the mainstream of European culture—for example, with the establishment in 1367 of a short-lived university in Pécs. In the second half of the fifteenth century, the court of the Hungarian king Matthias Corvinus (1443–1490) became a nexus of Renaissance culture. The reader can see from the reproductions in the color plates of codices of the Bibliotheca Corviniana, the library founded by Matthias, just how great the enthusiasm and receptiveness were for knowledge in the natural sciences. However, further developments were impeded by the Turkish conquest of 1526 and the century and a half of Ottoman rule that followed. The university founded in 1367 ceased to exist, and the famous library of Matthias Corvinus was scattered to the four corners of the globe.

The worldview of Cardinal Péter Pázmány (1570–1637), founder of an institution of higher learning from which the University of Budapest later developed, is still that of Aristotle, a medieval vision of a world created especially for mankind; yet the Cardinal's fiery sermons bear witness to an enthusiasm for knowledge and science and a feeling for nature. The *Magyar Encyklopédia* of János Apáczai Csere, published in 1653—only three years after the death of Descartes—already takes as its starting point the Cartesian dictum *cogito*.

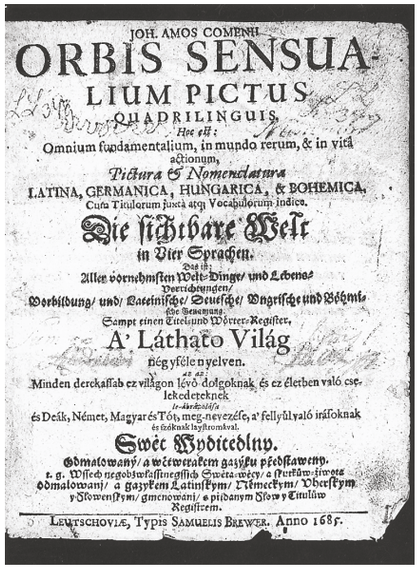


▲ **Figure 3.66** Afterword from a book printed by Ivan Fedorov (Lvov 1574).

The revival of culture in Poland is also linked with the name of a talented dynasty. JADWIGA, daughter of LOUIS I, heir to the Polish throne, married the Lithuanian grand duke JOGAILA. Under the rule of this dynasty (1386–1572), Kraków developed into a European cultural center. COPERNICUS studied for a time at its university, founded in 1364.

Russia was a special case in a number of respects. It was outside the embrace of the Roman Catholic Church, and was therefore essentially excluded from the Latin cultural realm. By way of the Orthodox Church, it had closer contact with Byzantine culture. Political circumstances also hindered cultural development. For almost 300 years (until 1480), Russia suffered under the Tatar yoke, in spite of a number of military victories over the invaders (Kulikovo, 1380). The result, as has been mentioned, was a dearth of some of the institutions characteristic of the Middle Ages (universities, knightly orders) in the West. Intellectual and cultural life was confined to religious activities, having only loose connections with science, and then only for the purpose of adjusting the calendar and fixing the dates of religious festivals. The first printing press was established by IVAN FEDOROV in Moscow under the auspices of IVAN IV, "THE TERRIBLE"; see Figure 3.66. The first book to be printed was *Apostolos* (1564). It was not until the seventeenth century that schools of higher learning were established for priests, and it was to be yet another century before access to Western science was achieved. The groundwork for this and its implementation are largely the work of LOMONOSOV.

We shall not speak here of the Czech lands because they did not belong to the peripheral states. Bohemia was a part of the Habsburg Empire, and indeed, Prague, the one-time Habsburg residence, was one of the centers of European intellectual life since the late Middle Ages and the source of newer intellectual currents (such as the Hussite movement). Especially worthy of mention is the famous Czech pedagogue COMENIUS (JAN AMOS KOMENSKÝ, 1592–1670), who was active also in Hungary, traveled to England, and was invited to Sweden by QUEEN CHRISTINA. His pedagogical views were strikingly modern. He was an advocate of general education, and he emphasized the important role of national languages. Figure 3.67 shows an extract from one of his more important works, the


JOH. AMOS COMENII.

ORBIS SENSUALIUM PICTUS QUADRILINGUIS,

Hoc eſt:

Omnium fundamentalium, in mundo rerum, & in vitâ actionum,

Pictura & Nomenclatura

LATINA, GERMANICA, HUNGARICA, & BOHEMICA,

Cum Titulorum junctâ atq; Vocabulorum indice.

Die ſichtbare Welt

in Vier Sprachen.

Das iſt:

Aller vornehmſten Welt-Dinge/ und Lebens-Verrichtungen/

Vorbildung/ und/ Lateiniſche/ Deutſche/ Ungriſche und Böhmiſche Benennung.

Sampt einen Titel- und Wörter-Regiſter,

A' Láthato Világ

négyféle nyelven.

Minden derekaſſab ez világon levő dolgoknak, és ez életben való cselekedeteknek

és Deák, Német, Magyar és Tót, meg-nevezése, a' fellyűl-való irásoknak

Swět Wydietedlny.

Odmalowaný/ a wěcmi wſſem gazyku předſtawený.

t. g. Wſſecky negobzwlaſſtnegſſich Směta-wěcy/ a ſkutkůw-zywota

Odmalowaní/ a gazykem Latinſkym/ Německym/ Uherſkym y Slowenſkym/ gmenowaní/ a pidanym Slow y Titulůw Regiſtrem.

LEUTSCHOVIÆ, Typis SAMUELIS BREWER. Anno 168 5.


(208)

CIII.

Sphæra Cœleſtis. Die Himels-Kugel.

Az Égi Golyobis. Kule Nebeſka.

Aſtro-

▲ **Figure 3.67** From the *Orbis sensualium pictus* of COMENIUS. Here COMENIUS promotes the compromise worldview proposed by TYCHO BRAHE. (Library of the Tiszántuli Reformed Parish, Sárospatak; photograph by István Sipos.)

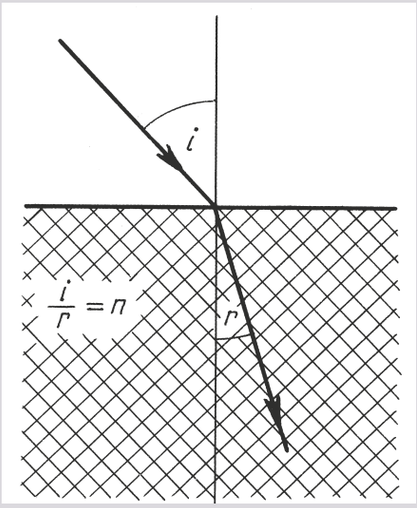
Quotation 3.30, ***continued***

injustice to God to believe that the causes of the effects which are in nature and which we have thus discovered are false. For we would then be accusing Him of having made us so imperfect as to be liable to make mistakes, even when correctly using our reason [which He has given us]. ...

However, [because the matters I am treating here are of no little importance] and because I should perhaps be thought too presumptuous if I stated that I had discovered truths [which others have failed to discover], I prefer to make no decision about it; and, [in order that each reader may be free to form his opinion], I wish what I shall write later to be taken only as an hypothesis [which is perhaps very far from the truth]. But, even though these things may be thought to be false, I shall consider that I have achieved a great deal if all the things which are deduced from them are entirely in conformity with the phenomena: for, if this comes about, my hypothesis will be as useful to life as if it were true, [because we will be able to use it in the same way to dispose natural causes to produce the effects which we desire]. ...

Indeed, in order to better explain natural things, I may even retrace their causes here to a stage earlier than any I think they ever passed through. [For example], I do not doubt that the world was created in the beginning with all the perfection which it now possesses; so that the Sun, the Earth, the Moon, and the Stars existed in it, and so that the earth did not only contain the seeds of plants but was covered by actual plants; and that Adam and Eve were not born as children but created as adults. The Christian faith teaches us this, and natural reason convinces us that this is true; because, taking into account the omnipotence of God, we must believe that everything He created was perfect in every way. But, nevertheless, just as for an understanding of the nature of plants or men it is better by far to consider how they can gradually grow from seeds than how they were created [entire] by God in the very beginning of the world; so, if we can devise some principles which are very simple and easy to know and by which we can demonstrate that the stars and the Earth, and indeed everything which we perceive in this visible world, could have sprung forth as if from certain seeds [even though we know that things did not happen that way]; we shall in that way explain their nature much better than if we were merely to describe them as they are now [or as we believe them to have been created]. And because I think I have discovered some principles of this kind, I shall here briefly expound them.

—RENÉ DESCARTES, *Principles of Philosophy*, III.42–45 [pp. 104–106]



▲ **Figure 3.68** Kepler used a simple proportion, $i = nr$, between i and r, even though he knew that it was only approximately correct.

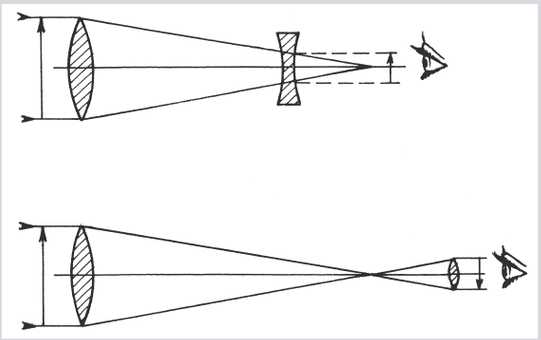
▼ **Figure 3.69** The Dutch, or Galilean, telescope consists of a converging lens as objective and a dispersing lens as ocular and produces an upright image.

In 1610/1611, Kepler formulated the optical theory of the telescope. Based on this theory, in 1611, Scheiner built the astronomical, or Keplerian, telescope. This consists of two converging lenses and produces an inverted and reversed image.

Today's familiar image formation and lens formulas, or more generally, the fundamentals of geometric optics, were worked out, after the correct formulation of the refraction law in 1637, by Cavalieri (1647), Barrow, Huygens, Newton, and Halley (1693). For thin lenses and paraxial rays, the formulas are

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{f},\quad \frac{1}{f}=(n-1)\left(\frac{1}{r_1}+\frac{1}{r_2}\right),$$

where a is the image distance, b the objective distance, f the focal length, n the index of refraction, and r_1, r_2 the radii of curvature of the bounding surfaces.



Orbis sensualium pictus (the world in pictures), which explains a number of phenomena and dispenses scientific knowledge in several languages: originally written in Latin and German, the edition shown includes Hungarian and Czech as well.

If we are looking for the periphery not only to the east, but to the west as well, then we should mention the Americas, and North America in particular. At the beginning of the seventeenth century, the first English, Dutch, and French settlers arrived in the territory of what are now the United States of America and Canada. The Pilgrims, the first settlers of the New England states, landed in 1620 aboard the *Mayflower*, that is, at the time when Galileo, Kepler, and Bacon lived. The first North American printing press was established in 1638, and in 1636 the first university, Harvard, opened its doors. Intellectual life in the colonies was limited at first to literary activity in both spiritual and secular areas (travel literature, poetry). It would be another hundred years before American science would begin not only to absorb knowledge from Europe, but to extend it as well (Benjamin Franklin).

In closing, we wish to emphasize it was not only in Europe that culture, in the larger sense of the word, flourished. One needs to think only of the achievements in the Byzantine world (icon painting, church music, architecture) or the active cultural life at the court of the Moldavian prince Stephen III, the Great (ca. 1433–1504), or at the court of Stefan Dušan (1308–1355), king of the Serbian Empire. The reason that we do not devote more attention to these other regions is only that they do not lie along the main path of our central interest: the development of the new physical worldview.

3.5 Light, Vacuum, and Matter through the Middle of the Seventeenth Century

3.5.1 *The Snell–Descartes Law*

Our knowledge of the nature of light at the beginning of the seventeenth century was essentially the same as that summarized by Ptolemy and expanded and further developed by Ibn al-Hasan (Alhazen). The invention of the telescope, however, gave new impetus to both geometric and physiological optics, that is, the study of the eye. The first significant progress in this area was made by Kepler. He carried out extensive experiments on refraction, and although he realized that the relationship between the angle of incidence and the angle of refraction was not a proportional one and that for large angles the deviation from proportionality was significant, still he failed to express his results in a satisfactory mathematical form. Having determined that proportionality represents a good approximation up to angles of about 30°, in his calculations regarding image formation with lenses, and above all, telescopes, he used the then usual relationship

$$\frac{i}{r}=n$$

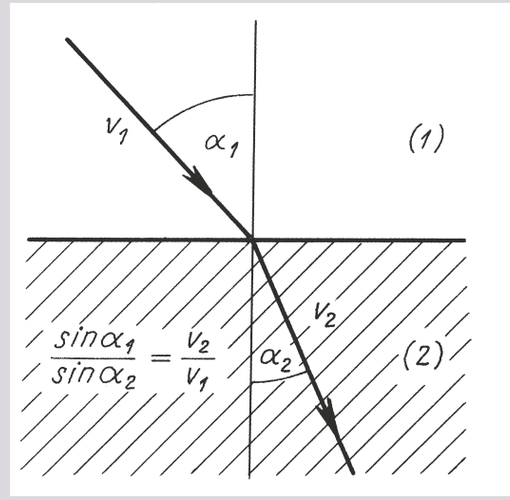
(Figures 3.68 and 3.69). It was also Kepler who discovered and described the phenomenon of total reflection and introduced the concept and terminology of the focal point. He correctly interpreted the role of the lens of the eye and also discovered the law that the degree of illumination of an object falls off with the square of the distance from the light source. He assumed, however, that the index of refraction is proportional to the density of the medium. This assertion was refuted by Thomas Harriot.

The Jesuit Christof Scheiner conducted a number of important and valuable investigations into vision and the structure of the human eye. We encountered his name earlier, in connection with the trial against Galileo, and he was also involved in a priority dispute over the discovery of sunspots and so as well that of the rotation of the Sun. The title of his work *Oculus, hoc est: fundamentum opticum* (the eye as the basis of optics) indicates the importance that he attached to the role of the eye in the study of optics. He began by studying the surgically removed eyes of oxen; then he experimented with human eyes and determined that the optical image in the eye is formed on the retina. Descartes, who carried out similar experiments, considered Scheiner's work important and made reference to it. In connection with Scheiner's investigations, it is worth returning once again to the point made in the introduction to this book: It is a mistake to denigrate out of hand the opponents of science's pioneers and to brand them as narrow-minded men clinging to an outmoded ideology. It is widely believed that the opponents of Galileo—where one thinks primarily of the Jesuits—were hidebound men who clang to Peripatetic ideas that had ossified into dogma and who were disinclined to let themselves be convinced by observing the real facts of nature. Here we can see that the experiments of Scheiner, particularly those on the human eye, prove the opposite with their audacity and concrete findings.

As we have already mentioned in our discussion of the work of Descartes, his investigations in the domain of optics were included with his work on geometry and the description of meteorological phenomena in the illustrated appendix to the *Discours.* It is here that one finds the statement that space is completely filled with three types of matter, one of which is responsible for the propagation of light. The statement of the refraction law in the form

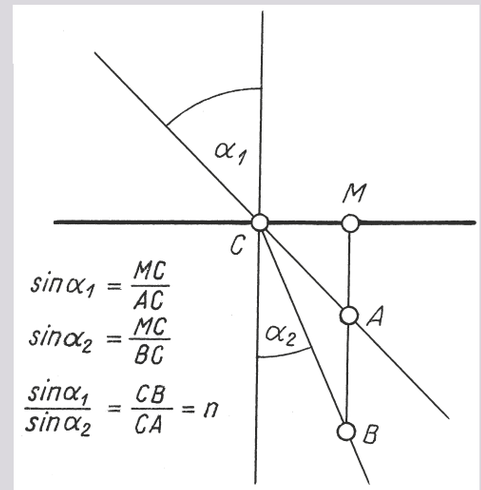
$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{v_2}{v_1}$$

is attributed to Descartes. This law states that the quotient of the *sine* of the angle of incidence and the *sine* of the angle of refraction is constant for a given medium, and this constant is called the index of refraction (Figure 3.70).



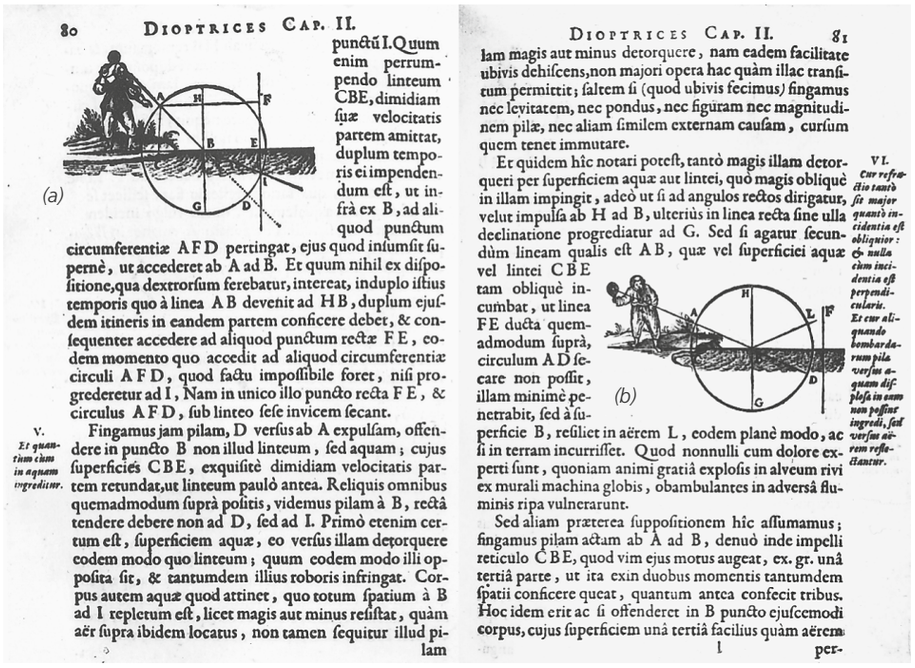
▲ **Figure 3.70** Descartes related the law of refraction to the velocity of light.

Descartes's law of refraction should more properly be called the Snell–Descartes law (and indeed, the law frequently is known by the abbreviated name of Snell's law) because Willebrord Snell (1580–1626), a professor at the University of Leyden, had certainly discovered this law by 1620 and promulgated it in his teaching (Figure 3.71). It is also certain that Descartes presented the law in printed form only in 1637, and he did not state that he had learned of the law from someone else. However, Descartes was always very parsimonious in acknowledging the contributions of others, and so the lack of a source is no proof that he did not know of Snell's result. In 1662, Vossius stressed in an article that in writing his book, Descartes must have known of Snell's work (Quotation 3.35). From today's perspective, we might find it a bit surprising that Descartes was not accused of plagiarism in his lifetime, not even in the heat of the arguments arising from his writings. Such suspicions were raised only after his death. Today, we know from his correspondence that Descartes was working on a law of refraction almost a decade before the publication of *Discours*, and it is therefore plausible that at the time of writing his book, he knew of Snell's result and therefore should have cited it, but that he had nevertheless discovered it independently. One aspect of Descartes's contribution is undisputed: He attempted to give a clear interpretation of the law of refraction and to that end adopted an approach still used today to derive the "law of refraction" for the paths of electrons in an electric field.



▲ **Figure 3.71** Snell formulated the law as $nCA = CB$.

▶ **Figure 3.72** **(a)** DESCARTES's explanation of the law of refraction. The horizontal component of velocity of the ball does not change, while the vertical component decreases. **(b)** DESCARTES's explanation of the reflection of light.



80 DIOPTRICES CAP. II.

punctũ I. Quum enim perrumpendo linteum CBE, dimidiam ſuæ velocitatis partem amittat, duplum temporis ei impendendum eſt, ut infrà ex B, ad aliquod punctum circumferentiæ AFD pertingat, ejus quod inſumſit ſupernè, ut accederet ab A ad B. Et quum nihil ex diſpoſitione, qua dextrorſum ferebatur, intereat, induplo iſtius temporis quo à linea AB devenit ad HB, duplum ejuſdem itineris in eandem partem conficere debet, & conſequenter accedere ad aliquod punctum rectæ FE, eodem momento quo accedit ad aliquod circumferentiæ circuli AFD, quod factu impoſſibile foret, niſi progrederetur ad I, Nam in unico illo puncto recta FE, & circulus AFD, ſub linteo ſeſe invicem ſecant.

V. *Et quantum cum in aquam ingreditur.*

Fingamus jam pilam, D verſus ab A expulſam, offendere in puncto B non illud linteum, ſed aquam; cujus ſuperficiés CBE, exquiſitè dimidiam velocitatis partem retundat, ut linteum pauló antea. Reliquis omnibus quemadmodum ſuprà poſitis, videmus pilam à B, rectâ tendere debere non ad D, ſed ad I. Primò etenim certum eſt, ſuperficiem aquæ, eo verſus illam detorquere eodem modo quo linteum; quum eodem modo illi oppoſita ſit, & tantumdem illius roboris infringat. Corpus autem aquæ quod attinet, quo totum ſpatium à B ad I repletum eſt, licet magis aut minus reſiſtat, quàm aër ſupra ibidem locatus, non tamen ſequitur illud pilam

DIOPTRICES CAP. II. 81

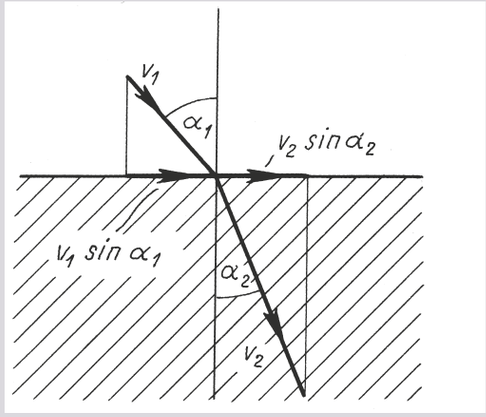
lam magis aut minus detorquere, nam eadem facilitate ubivis dehiſcens, non majori opera hac quàm illac tranſitum permittit; ſaltem ſi (quod ubivis fecimus) fingamus nec levitatem, nec pondus, nec figuram nec magnitudinem pilæ, nec aliam ſimilem externam cauſam, curſum quem tenet immutare.

Et quidem hìc notari poteſt, tantò magis illam detorqueri per ſuperficiem aquæ aut lintei, quò magis obliquè in illam impingit, adeò ut ſi ad angulos rectos dirigatur, velut impulſa ab H ad B, ulteriùs in linea recta ſine ulla declinatione progrediatur ad G. Sed ſi agatur ſecundùm lineam qualis eſt AB, quæ vel ſuperficiei aquæ vel lintei CBE tam obliquè incumbat, ut linea FE ducta quemadmodum ſuprà, circulum ADſecare non poſſit, illam minimè penetrabit, ſed à ſuperficie B, reſiliet in aërem L, eodem planè modo, ac ſi in terram incurriſſet. Quod nonnulli cum dolore experti ſunt, quoniam animi gratiâ explosis in alveum rivi ex murali machina globis, obambulantes in adversâ fluminis ripa vulnerarunt.

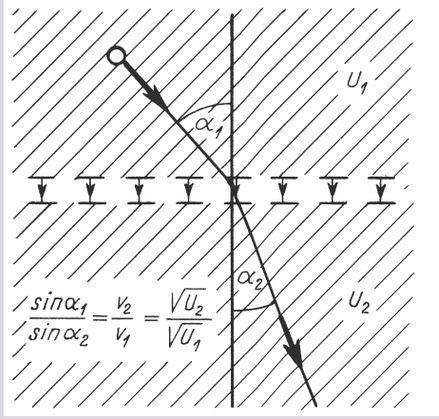
VI. *Cur refractio tantò ſit major quantò incidentia eſt obliquior: & nulla cùm incidentia eſt perpendicularis. Et cur aliquando bombardarum pila verſus aquam diſplosa in eam non poſſint ingredi, ſed verſus aërem reflectantur.*

Sed aliam præterea ſuppoſitionem hìc aſſumamus; fingamus pilam actam ab A ad B, denuò inde impelli reticulo CBE, quod vim ejus motus augeat, ex. gr. unâ tertiâ parte, ut ita exin duobus momentis tantumdem ſpatii conficere queat, quantum antea confecit tribus. Hoc idem erit ac ſi offenderet in B puncto ejuſcemodi corpus, cujus ſuperficiem unâ tertiâ facilius quàm aëream

l per-



▲ **Figure 3.73** To derive the law of refraction, it suffices to assume that the component of velocity parallel to the surface does not change.



▲ **Figure 3.74** The path of an electron is broken in crossing the double layer. From the law of diffraction, we may conclude a connection between the optical index of refraction and the square root of the potential.

Following DESCARTES's interpretation, let us consider first the reflection of light. According to Figure 3.72(b), which we have taken from DESCARTES's book, we throw a ball at an oblique angle onto a flat surface and observe that the angle at which the ball bounces off the surface is equal to the angle of incidence. This law can easily be understood in terms of the laws of mechanics, in that the component of the ball's velocity parallel to the surface does not change, while the vertical component keeps its magnitude but changes sign.

Now imagine that instead of a flat elastic surface we had stretched a piece of canvas and hurled the ball at an angle toward the canvas with enough force to pierce the canvas. According to DESCARTES, we may again assume that the components of velocity parallel to the surface remain unchanged. In breaking through the surface, the magnitude of the vertical component is reduced, and the path of the ball is altered, as shown in Figure 3.72(a). We can even see from the figure that in analogy with the experiment described above, as a ray of light transitions from air into glass, the *speed of light in glass should be greater than that in air*, since the light ray makes a smaller angle with the axis of incidence in the glass than in the air. From Figure 3.73, we easily derive the relation

$$v_1 \sin\alpha_1 = v_2 \sin\alpha_2,$$

from which follows the refraction law

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{v_2}{v_1}.$$

The incorrect conclusion derived from the Cartesian model that light travels faster in a transparent medium than it does in air hung on stubbornly for more than two centuries; even NEWTON—starting with a different model—came to the same conclusion.

We mentioned above that even today, using a line of reasoning similar to that of DESCARTES, we can solve a problem of quite a different nature, namely, the fundamental problem of electron optics. Let us consider, using Figure 3.74, the path of an electron that moves at the outset along a straight line in a region of space characterized by a constant potential and then meets a planar boundary surface separating the

first region of space from a second region also characterized by a constant potential, but with a different value. After crossing the boundary, the electron, of course, continues to travel in a straight line because no force acts in a region of constant potential. The question is this: Through what angle, with respect to the normal to the boundary, is the electron's path bent with respect to the electron's angle of incidence? Let us consider what is happening physically at the border between the two regions. We see that an accelerating or decelerating force acts, which, like the stretched canvas in the Cartesian model, changes the magnitude of only the component of the velocity perpendicular to the boundary plane. Thus in electron optics, one arrives at the *refraction law for electron paths*, and from this law, the fundamental principles for the construction of electron microscopes.

Descartes used the law of refraction to interpret a wide variety of phenomena and to solve a number of different problems. His optics is consciously oriented toward practical application, his essay on the subject having been written with the express purpose of making possible the construction of devices that would either make life more comfortable for mankind or facilitate the study of nature.

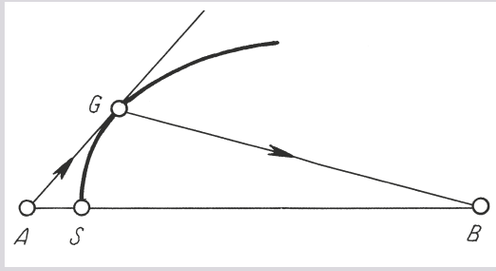
Descartes also posed the question of the optimal form of rotationally symmetric lenses. He searched for the surface in which different rays emerging from a point on the rotational axis could refract and again cross the axis at another single point (Figure 3.75–3.77). The curve that Descartes derived, and which bears his name (Cartesian oval), when revolved about the axis of symmetry, defines lenses that have this property.

Descartes also studied vision following Scheiner (Figure 3.78). He analyzed in detail the effect of different types of lenses and how they alter the optical path. In essence, he specified the optical data for lenses that correct for near- and farsightedness.

Descartes's greatest contribution to the study of optics is considered to be his explanation of the origin of rainbows. Although the Cartesian theory of color—which cannot be sketched here because of its complexity—was not suitable for explaining how the rainbow's colors come into being, it correctly gave both the rainbow angle and the conditions for the appearance of a rainbow.

Descartes is often criticized for having attempted to explain all phenomena from fundamental principles, supporting his argument exclusively on logical reasoning, and even if he didn't entirely reject experimentation, he was at least indifferent to it. We have already seen that with respect to his laws of motion, this critique is by and large valid. However, in formulating his theory of the rainbow, he performed endless experiments with a water-filled glass sphere. It had long been known that the rainbow had something to do with the refraction of light by raindrops hovering in the air. It is for this reason that Descartes made use of a spherical glass container, so as to have a model of the raindrop. Figure 3.79 depicts the path of light in the formation of the primary rainbow. The light ray is refracted on the outer surface of the water droplet, then reflected on the inner surface (yet there is *no total reflection*!). Finally, it exits the drop with yet another refraction. The secondary rainbow arises from a second reflection on the inner surface of the water droplet followed by the light ray's exit from the droplet. What had to be explained was why the rays responsible for the rainbow were all within a certain angle. Descartes made extensive calculations, collected his results in a table, and concluded that a bundle of light rays traveling in parallel lines and entering a water droplet will exit more or less in parallel only if the Sun, raindrop, and observer are suitably oriented. In all other cases, the bundle of rays diverges on exiting the raindrop, and our eyes will not perceive the light (Figure 3.80).

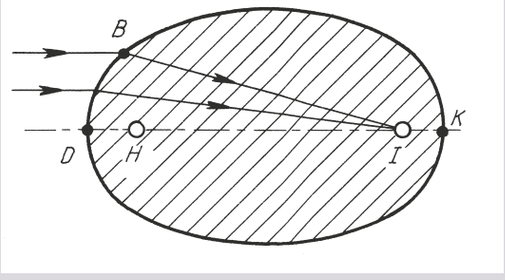
Another piece of evidence for attributing an experimental streak to Descartes is that he constructed a simple device for determining the angle of refraction. This apparatus is depicted in Figure 3.81, taken from Descartes's book. The direction of



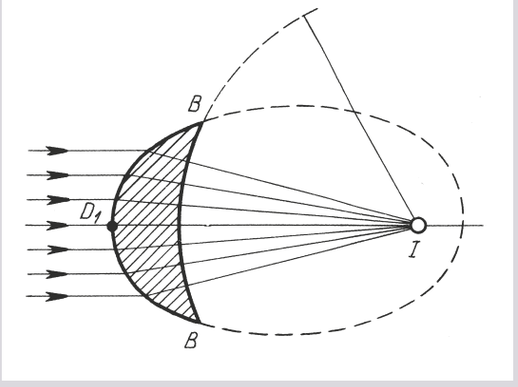
▲ **Figure 3.75** Descartes raises the question of the perfect lens form. What form must the curve *SG* assume so that all rays emitted from point *A* are focused at point *B*? The answer is given by

$$\frac{AG - AS}{BS - BG} = n,$$

which is the equation of a Cartesian oval.



▲ **Figure 3.76** Ovals can degenerate into ellipses or hyperbolas. In the case shown, the rays are focused at the point *I* when $n = DK/HI$.



▲ **Figure 3.77** If we draw a circle around the focal point *I*, then the path of the rays in the resulting lens is the same as that in Figure 3.76 because the rays striking perpendicular to the sphere's surface are not refracted.

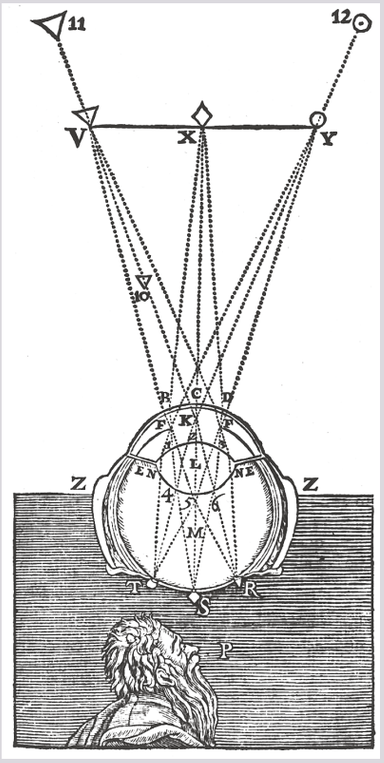
However, a truly perfect lens should unite rays emitted from an arbitrary point in space to a single image point. Furthermore, if the image points lie in a plane perpendicular to the lens axis, the image points should also lie in a plane, and moreover, the image enlargement should be constant in the entire image space. A simple lens is far from being capable of satisfying all these requirements. L. P. Seidel (1821–1896) systemized the image imperfections and identified seven such errors. Five of them occur even

continued on next page

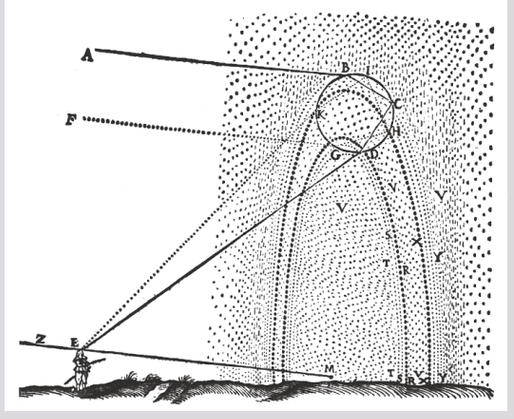
Figure 3.77 *continued*

with monochromatic light: image field curvature, distortion, spherical aberration, coma (asymmetry), and astigmatism. The errors caused by dispersion are the chromatic aberrations, or color errors: point locus errors and lateral chromatic aberration.

Individual errors can be overcome (achromatic and anastigmatic lenses). However, all image errors cannot be corrected simultaneously. Also, the wave nature of light leads to diffraction, which can have a significant effect on the sharpness of the image.



◀ **Figure 3.78** The relationship between the image on the retina and the position of the object imaged.



▲ **Figure 3.79** Descartes's explanation of the origin of the rainbow. The principal arc arises from single internal reflection, while the secondary arc is the result of double internal reflection.

a light bundle is determined by two holes bored in the framework. The light then travels through a prism constructed of the material under investigation. The light is refracted, and it creates a spot of light on the base of the apparatus. In the figure, Descartes indicates the calculations necessary for determining the index of refraction.

Although Descartes makes definite statements about the speed of light in various media while discussing the law of refraction, he still claims that the propagation of light does not require any time at all. He uses the following vivid metaphor to explain: We perceive light like a blind man probing for objects around him with the help of his stick, namely, at the exact moment that the stick touches the object. What is more, Descartes was so convinced of this conclusion, based on general philosophical principles, that in a letter to Mersenne he stated that if someone could prove the falsity of this formulation, he would gladly admit that he knows nothing about philosophy! Just imagine the weight of such a declaration directly from Descartes, who was well aware of his own abilities and the importance of his ideas!

Perhaps Descartes's assertion that no perceptible time is required for the propagation of light is to be interpreted in the sense that the time in question is very small because the speed of light is very large in comparison to all other speeds of propagation—for example, to the speed of sound. Descartes was aware of Galileo's experiment designed to measure the speed of light by using a flashing signal lamp and noting how long it took for the returned signal to arrive. This technique had proved successful in experiments to measure the speed of sound; however, it had to fail for the speed of light because the reaction time of the person observing the signal is much greater than the time it takes light to cover any distance on Earth. A further indication of Descartes's genius is that he suspected that the most suitable technique for measuring the speed of light would involve some phenomenon involving a celestial body at a time that could be precisely calculated well in advance. He carefully compared the calculated and observed times of lunar eclipses and found that the results supported his assumption that light propagates in an immeasurably brief period of time. These ideas of Descartes are significant because it was only one generation later that the Danish astronomer Ole Rømer used yearly repeating irregularities in the transit times of the moons of Jupiter in front of or behind the planet to determine the velocity of light for the first time, obtaining a result that is in good agreement with today's value.

3.5.2 *Fermat's Principle*

The Cartesian hypothesis regarding the velocity of light was not generally accepted. Pierre de Fermat (1601–1665) sharply rejected Descartes's assumption, asserting that, on the contrary, a denser material would clearly offer greater resistance to the movement of light, and consequently the speed of light in such media must be smaller (Quotation 3.36). At the same time, starting from the conviction that all processes in nature proceed in the most economical way, Fermat managed to derive his famous principle (1662), which today has lost none of its validity. A key factor was Fermat's great skill and the enjoyment he took in solving extreme value problems (Quotation 3.37). Fermat's principle yields the following results with regard to the laws of refraction (Figure 3.82): Let us choose two points in two homogeneous media of different densities separated by a planar boundary, and naturally also assume that in each of the media, light propagates in a straight line. In principle, there are infinitely many paths by which light could travel in a straight line from one point to the boundary, and then, refracted at the appropriate angle, proceed to the selected

point inside the second medium. Out of all these geometrically conceivable paths, light will actually follow the one for which the time of travel is minimal. With the notation from the figure, this condition may be expressed mathematically as follows:

$$\frac{l_1}{v_1} + \frac{l_2}{v_2} = \text{minimum}.$$

It is worthwhile at this point to reconstruct the proof given by FERMAT in detail in order to show the wealth of original ideas that had to be marshaled in order to solve a relatively simple extreme value problem before the invention of calculus. It is also interesting to compare the method that we are about to describe with the one that we shall look at when we discuss Maupertuis's principle (Section 4.2), which serves the same ends.

We begin by drawing in Figure 3.83 the actual path for which

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{v_1}{v_2}$$

holds and then drawing a circle of arbitrary radius with center at the point of refraction, so that the above equation can be written in the form

$$\frac{DN\,/\,MN}{NS\,/\,NH} = \frac{DN}{NS} = \frac{v_1}{v_2} > 1,$$

where we have used that $MN = NH$.

The travel times are then related by

$$\frac{t_1}{t_2} = \frac{l_1}{l_2}\frac{v_2}{v_1},$$

and from this general relationship, we obtain

$$\frac{t_{MN}}{t_{NH}} = \frac{MN}{NH}\frac{IN}{MN} = \frac{IN}{NH}$$

and

$$\frac{t_{MR}}{t_{RH}} = \frac{MR}{RH}\frac{PR}{MR} = \frac{PR}{RH},$$

as well as, finally,

$$\frac{t_{MNH}}{t_{MRH}} = \frac{IN + NH}{PR + RH}.$$

One must now show that

$$PR + RH > IN + NH,$$

where

$$\frac{MN}{IN} = \frac{MR}{PR} = \frac{DN}{NS}.$$

To this end, we specify points O and V such that

$$\frac{MN}{DN} = \frac{RN}{NO} \quad \text{and} \quad \frac{DN}{NS} = \frac{NO}{NV}. \tag{1}$$

From $DN < MN$, it follows that $NO < RN$, and from $NS < DN$, we obtain $NV < NO$. Furthermore, from

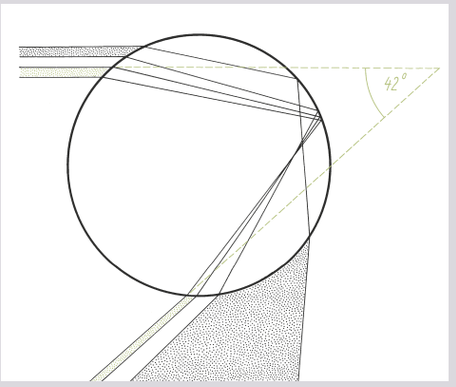
$$MR^2 = MN^2 + NR^2 - 2MN \times NR\cos(MNR)$$

and

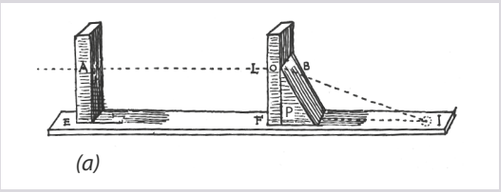
$$-MN\cos(MNR) = DN,$$

we obtain

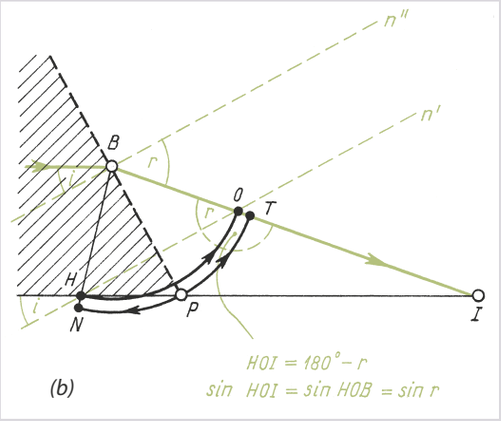
$$MR^2 = MN^2 + NR^2 + 2DN \times NR.$$



▲ **Figure 3.80** The only rays that remain parallel after exiting the water drop are those forming an angle of 42° with the incident ray.



(a)



(b)

▲ **Figure 3.81** **(a)** The experimental determination of the index of refraction. The ray L exits the prism P at point B.

(b) DESCARTES's construction for determining the index of refraction: We first draw the given points B, P, and I, and then the arc PT with radius BP. We then draw an arc PN such that the length of PT is equal to that of PN. We obtain the point H as the intersection of the lines BN and IP. We then draw a circle about point B with radius BH, thereby obtaining the point O. The line joining O with the point H is the normal n' to the surface of the prism. From the law of sines, we have, in the triangle HOI,

$$\frac{OI}{HI} = \frac{\sin OHI}{\sin HOI} = \frac{\sin OHI}{\sin HOB} = \frac{\sin i}{\sin r} = \frac{1}{n}.$$

Quotation 3.31

And it is important to remember here what was said earlier concerning the nature of movement; i.e., that (if we are speaking properly and in accordance with the truth of the matter) it is only the transference of a body from the vicinity of those bodies which are immediately contiguous to it, and considered to be at rest, into the vicinity of the others. However, in common usage, all action by which any body travels from one place to another is often also called movement; and in this sense of the term it can be said that the same thing is simultaneously moved and not moved, according to the way we diversely determine its location. From this it follows that no movement, in the strict sense, is found in the Earth or even in the other Planets; because they are not transported from the vicinity of the parts of the heaven immediately contiguous to them, inasmuch as we consider these parts of the heaven to be at rest. For, to be thus transported, they would have to be simultaneously separated from all [the contiguous parts of the heaven], which does not happen. However, because the matter of the heaven is fluid, sometimes some of its particles, and sometimes others, move away from the Planet to which they are contiguous, and this by a movement which must be attributed solely to them and not to the Planet: in the same way as the partial transferences of water and air which occur on the earth's surface are usually attributed, not to the earth, but to those portions of water and air which are transported. ...

Now that we have, by this reasoning, removed any possible doubts about the motion of the Earth, let us assume that the matter of the heaven, in which the Planets are situated, unceasingly revolves, like a vortex having the Sun as its center, and that those of its parts which are close to the Sun move more quickly than those further away; and that all the Planets (among which we [shall from now on] include the Earth) always remain suspended among the same parts of this heavenly matter. For by that alone, and without any other devices, all their phenomena are very easily understood. Thus, if some straws [or other light bodies] are floating in the eddy of a river, where the water doubles back on itself and forms a vortex as it swirls; we can see that it carries them along and makes them move in circles with it. Further, we can often see that some of these straws rotate about their own centers, and that those which are closer to the center of the vortex which contains them complete their circle more rapidly than those which are further away from it. Finally, we see that, although these whirlpools always attempt a circular motion, they practically never describe perfect circles, but

continued on next page

From the defining equation for the points O and V, we have

$$DN \times NR = MN \times NO,$$

so that from

$$NO^2 < NR^2,$$

we obtain

$$MR^2 = MN^2 + NR^2 + 2MN \times NO > (MN + NO)^2$$

and

$$MR > MN + NO.$$

With the help of the relations

$$\frac{DN}{NS} = \frac{MN}{IN} = \frac{NO}{NV} = \frac{NO + MN}{NV + IN} = \frac{MR}{RP}$$

and

$$MR > MN + NO,$$

we obtain

$$RP > NV + IN.$$

If we are able to prove $RH > HV$, then we will surely have

$$PR + RH > IN + NV + VH \equiv IN + NH.$$

To prove the inequality, we use

$$RH^2 = NH^2 + NR^2 - 2NH \times NR\cos(HNR) = NH^2 + NR^2 - 2NR \times SN \qquad (2)$$

as well as the relations

$$\frac{HN}{DN} = \frac{MN}{DN} = \frac{NR}{NO} \quad \text{and} \quad \frac{DN}{NS} = \frac{NO}{NV},$$

from which we obtain

$$\frac{HN}{NS} = \frac{NR}{NV} \Rightarrow SN \times NR = HN \times NV.$$

But then from equation (2), taking into account

$$NR > NV,$$

we have

$$RH^2 = NH^2 + NR^2 - 2HN \times NV > NH^2 + NV^2 - 2HN \times NV,$$

or

$$RH > NH - NV = HV,$$

which was to be proved.

Descartes's authority was so great that Fermat was warned that publication of his theory would bring vigorous opposition. Moreover, every experiment seemed to verify the Cartesian refraction law. What astonished Fermat most was that from his theory, he exactly derived the Cartesian law, except that a different relationship was

obtained between the index of refraction and the velocities of light in the two media (Quotation 3.38). Physically, only the ratio between the two velocities, namely, the index of refraction, could be determined, so there appeared to be no significant difference in practice. However, it was the principle itself against which the Cartesians protested, for in their view, Fermat wanted to smuggle something into the theory that was inconsistent with a mechanistic interpretation (Quotation 3.39). Aside from the assumption of the simplicity of natural processes, Fermat had not invoked any metaphysical basis, in contrast to Maupertuis, who two generations later would declare that behind the principle of minimality could be seen the rational and purposeful action of a higher being. Fermat was thus able to respond to the attacks in a relaxed and mildly ironic manner (Quotation 3.40).

The characteristics of Aristotelian physics have been summarized in Table 1.4 (see page 75). We can see that the Aristotelian theses regarding motion as well as the hierarchy of celestial and terrestrial physics had all been rejected by the middle of the seventeenth century, and in some special cases the correct answer had already been born. But what had most stubbornly survived were precisely those ancient theories that were the first in human history to be advanced as science: the speculations on the structure of matter. In this regard, the seventeenth century made no progress beyond new doubts and misgivings.

3.5.3 *Vacuum and Air Pressure*

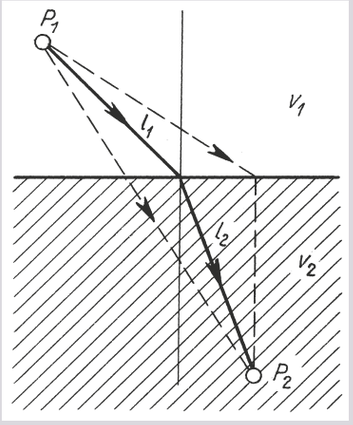
It is surprising that the Cartesian worldview, which was formulated with the purpose of replacing the Aristotelian worldview in its entirety, still had two characteristic Peripatetic traits: that a force can be transmitted only between two bodies in contact with each other, and that a vacuum cannot exist. To be sure, these ideas within the Cartesian system had positive aspects as well: The effective force arising from contact is no longer necessary for *maintaining* a state of motion, but only for *changing* it; furthermore, from the standpoint of natural philosophy, the ancient occult qualities were reduced to such geometric concepts as spatial extent and motion, which made it possible to formulate quantitative laws. Even the thesis of the impossibility of the vacuum should be understood within the framework of the Cartesian program: If there is no vacuum, then action at a distance between two bodies could be reduced to immediate interactions, so that off-key words such as "sympathy" and "affinity" could be eliminated from physics.

The Aristotelian assumption of *horror vacui* once again became the center of interest when it was observed that pumps are able to raise water only up to a certain height. Even Galileo dealt with this problem, from both practical and philosophical points of view. In the very beginning of his 1638 book *Discorsi*, he discussed the question of whether the solidity of an object can be explained by considering that in the moment of a fission a vacuum would be created between the surfaces of any two parts, so it is nature's horror of the vacuum that will resist this fracture. The interlocutors in the *Discorsi* realize that in the breaking up of solid objects, other effects are also significant, but in the case of water, only the *horror vacui* can play a role. Galileo also provided a protocol (Figure 3.84) by which this *horror* could be measured. It uses a cylinder filled with water that is closed from below with a piston. When weights are hung from the piston, at a certain weight the piston is pulled from the cylinder; this weight is then a measure of the *horror vacui*. At first sight, this notion seems hopelessly naive, but in this way, Galileo made the *horror vacui* measurable and assigned a numerical value to it, so that it ceased to be

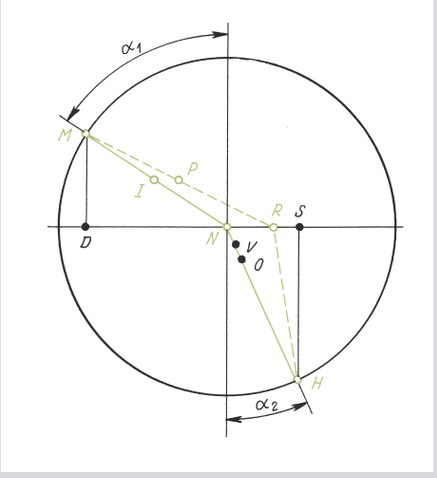
Quotation 3.31, ***continued***

sometimes become too great in width or in length, [so that all the parts of the circumference which they describe are not equidistant from the center]. Thus we can easily imagine that all the same things happen to the Planets; and this is all we need to explain all their remaining phenomena.

—René Descartes, *Principles of Philosophy*, III.28, 30 [pp. 94–95, 96]



▲ **Figure 3.82** Determination of the path traversed in minimal time. The shortest path would be the line P_1P_2, which, however, results in a relatively longer time spent in the denser medium, in which light travels more slowly.



▲ **Figure 3.83** Fermat's method for deriving the law of refraction.

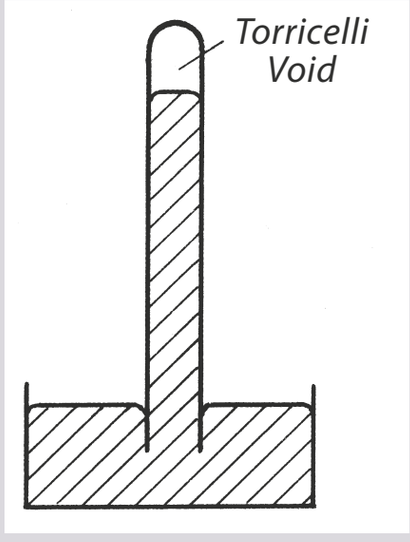


DEL GALILEO. 15

appartarla piglieremo vna materia continua, le cui parti manchino di ogni altra resistenza alla separazione fuor che di quella del Vacuo, quale à lungo è stato dimostrato in certo Trattato del nostro Acc^co. esser l'Acqua. Talche qualunque volta si disponesse vn Cilindro d' Acqua, e che attratto si sentisse resistenza allo staccamento delle sue parti, questo da altra cagione, che dalla repugnanza al Vacuo, non potrebbe riconoscersi. Per far poi vna tale esperienza mi son' immaginato vn' artifizio, il quale con l'aiuto di vn poco di disegno meglio, che con semplici parole, potrò dichiarare. Figuro questo C A B D essere il profilo di vn Cilindro di metallo, ò di vetro, che sarebbe meglio voto dentro, mà giustissimamente tornito, nel cui concauo entri con esquisitissimo contatto vn Cilindro di legno, il cui profilo noto E G H F, il qual Cilindro si possa spignere in sù, e'n giù: e questo voglio, che sia bucato nel mezzo, si che vi passi vn filo di ferro oncinato nell' estremità K, e l'altro capo I vadia ingrossandosi in forma di Cono, ò turbine, facendo che il foro fatto nel legno sia nella parte di sopra esso ancora incauato in forma di Conica superficie aggiustata puntualmente per riceuere la Conica estremità I del Ferro I K, qualunque volta si tiri in giù dalla parte K. Inserto il legno, ò vogliamolo chiamar Zaffo E H nel cauo Cilindro A D non voglio, ch' arriui fino alla superior superficie di esso Cilindro, mà che ne resti lontano due, ò tre dita: e tale spazio deue esser ripieno di Acqua, la quale vi si metterà tenendo il vaso con la bocca C D all' in sù, e calcandoui sopra il Zaffo E H col tenere il turbine I remoto alquanto dal cauo del legno, per lasciar l'esito all' aria, che nel calcare il Zaffo se n'uscirà per il foro del legno, che perciò si fà alquanto più largo della grossezza dell' astricciuola di ferro I K. Dato l'esito all' aria, e ritirato il ferro, che ben

A B I E F G H K C D

▲ **Figure 3.84** GALILEO's proposal for measuring the *h orror vacui*.



▲ **Figure 3.85** TORRICELLI's experiment.

an occult quality and became a physical quantity, with only its name to remind us of its murky origins. Therefore, not only the phenomenon itself but also its cause could be investigated more easily. Yet GALILEO's explanation for why the water in a pump's suction tube cannot rise higher than a certain value was still completely false: by analogy with the previous experiment, he imagined that a taller column of water would break apart under its own weight.

In connection with GALILEO's kinematics, we have confronted several times the surprising fact that GALILEO could with persuasive force and great rhetorical skill make claims or defend ideas that the scientific vanguard had already left behind. Thus in 1618, BEECKMAN had already found that the water in a suction tube can only climb to a height of eighteen cubits and correctly identified the reason: that it was pushed to that height by air pressure. The first proof of this assertion capable of erasing all doubt, and which then became part of general scientific knowledge, was carried out a generation later in the work of PASCAL. Previously (1643), however, GALILEO's pupil EVANGELISTA TORRICELLI (1608–1647), together with VIVIANI, carried out the famous Torricelli experiment (Figure 3.85), which today is still one of the standard experiments performed in physics classrooms.

If we immerse a mercury-filled glass tube, closed at one end and open at the other, into a tub filled with mercury and then hold it upright, we find that the column of mercury frees itself from the upper end of the tube and stabilizes at a height of seventy-six centimeters above the surface of the mercury in the tub. This arrangement is of great significance because working with the shorter column of mercury is much simpler than trying to perform the analogous experiment with a ten-meter-high column of water. Both GALILEO and TORRICELLI had already conjectured that the density of the fluid would play a decisive role. Even from the—completely incorrect—Galilean theory that the water in the suction tube climbs only so high because the water column would break apart under its own weight, one may still conclude that columns of mercury are much more prone to detachment because they are heavier due to their greater density.

The other great significance of using mercury is the seemingly irrelevant fact that due to the smaller size the *tubes can be made of glass and one can see* what happens inside. In contrast, in the suction tube of a water pump it is impossible to tell whether the water is sticking to the piston or whether some sort of vacuum has arisen. The most exciting question of the following decade and the subject of most passionate disputes was the nature of the *Torricelli void*, that is, the space between the upper end of the glass tube and the mercury. It was necessary to explain what was in the empty space and why the mercury always rose to a specific height. To each of these questions there were, in principle, two answers: One may assume or reject the existence of a vacuum in the empty space, and the cause of the limited height of the column of mercury may be sought either in air pressure or in *horror vacui* (Quotation 3.41).

The possible answers may even be combined, and in fact, each combination had its prominent advocates. The supporters of the Aristotelian worldview, of course, maintained that because of the impossibility of a vacuum, the Torricelli void must be filled with matter of some kind and that the cause of the limited height to which the mercury rises is to be sought in the *horror vacui* now tamed as a measurable quality. DESCARTES and his followers correctly saw the cause of the limited height of the column in the external air pressure, but they insisted that a vacuum could not exist in the Torricelli void. It was finally PASCAL who, after careful and extensive experiments, came to the correct conclusion that the external air pressure is responsible for the limited height of the column *and* that a vacuum exists in the Torricelli void.

We have to add immediately, however, that this interpretation is not completely correct either, since the Torricelli void is in fact filled with *mercury vapor*. The amount of such vapor is considerable because the vapor pressure of mercury at room temperature is 10^{-3} torr. In the mercury diffusion pumps used in most vacuum equipment, the mercury vapor is precipitated by means of vapor traps in order to obtain a better vacuum. Fortunately, the presence of saturated mercury vapor *does not affect* the phenomena observed in these experiments.

We mention as a curiosity the theory of the Jesuit priest Étienne Noël, who was the teacher of Descartes at the school at La Flèche and who later become a follower and disseminator of the Cartesian views. In Noël's opinion, the sinking column of mercury in the glass tube takes the "fire particles" from the glass along with it—particles with which the glass was saturated when it was manufactured at high temperatures—leaving pores behind through which "pure air" can stream into the void from outside. "Pure air" is one of the four Aristotelian elements, a component of the real air of the atmosphere.

One of Pascal's (Figure 3.86) famous experiments, the experiment of a void within a void (*vide dans le vide*) (Figure 3.87), already provides a satisfactory explanation that the rising column of mercury in the tube is caused by external air pressure. This experiment is also of interest because at that time there were no air pumps, and Pascal had to invent another way to vary the air pressure inside a container. He solved the problem as follows: He placed a Torricelli device in a large glass container whose lower opening was closed at the start of the experiment by a membrane. He filled the container with mercury through the upper opening and closed it with a lid so that it was airtight. He then immersed the entire construction with the end closed by the membrane in a mercury-filled tub and finally punctured the membrane. In this way, Pascal was able to investigate the conditions in the Torricelli void, where—if we use today's language—the inner Torricelli tube served as a barometer. As the mercury flowed out of the broad glass container, the height of the column of mercury in the tube diminished. First of all, Pascal saw that in the case of a Torricelli tube placed in a Torricelli void, the level of liquid in the tube and in the basin were the same. If one now allows air to bubble into the Torricelli void of the large glass container from the outside, then the height of the mercury column in the inner barometer increases. One cannot but conclude that the column of mercury is pushed to a certain height by the pressure of the air.

To investigate the role of the density of the liquid employed, Pascal performed an interesting experiment with two 14-meter tubes attached to ship's masts that could be tilted, filling one tube with water and the other with red wine. The experiment received considerable advance publicity, and Pascal posed the question to five hundred curious onlookers as to which liquid, water or wine, would sink lower in the tube. Most decided in favor of the wine, reasoning that red wine contained more "spirit" and therefore must generate greater pressure in the Torricelli void. Pascal, however, already knew that only the ratio of the densities mattered, and alcohol being lighter than water, he could predict that the wine, because of its significant alcohol content, would sink less than the water. The experiment, of course, bore out Pascal's prediction.

As we have already mentioned, Pascal was very cautious in drawing conclusions from the results of his experiments, and for a long time, he considered both *horror vacui* and air pressure as equally possible causes. He wanted to remain true to the fundamental philosophical principles that he had outlined in a reply to a critical letter from the above-mentioned Jesuit Noël (Quotation 3.42).



▲ **Figure 3.86** Blaise Pascal (1623–1662): Son of a highly educated jurist and well-known mathematician, who educated his children himself. Pascal's first work, which appeared in 1640, was *Essais pour les coniques*, which contains Pascal's theorem on hexagons inscribed in conics. From 1642 to 1644, he constructed an adding machine, which may be seen as a predecessor of today's calculating machines. The 1640s were the most fruitful years for Pascal in mathematics and physics: laws of hydrostatics; study of the vacuum; the binomial theorem; methods for determining area, which contain elements of the calculus; fundamentals of probability theory. In the 1650s, the increasing influence of Jansenism led to Pascal's joining the Jansenists in 1655 and living in Port Royal. (Jansenism is based on the teachings of Saint Augustine. The Jansenists raised a strong polemic against the Jesuits, which Pascal endorsed in his literary works *Les Provinciales* and *Pensées*.)

The unit of pressure in the international system is the pascal.

There is a Pascal's law in hydrostatics, a Pascal's theorem in geometry, and a Pascal's triangle in mathematics.

As a writer, Pascal created masterworks of the French language.

As a philosopher, he was the first to make significant inroads into the problem of the limits of what can be discerned and set the "arguments of the heart" against those of reason. His intuitionism had a direct effect on Rousseau, but also on Bergson and even on the modern existentialists.

> Let us divide our credulity and suspicion with more justice, and limit this respect we have for the ancients. As reason gives it birth, she ought also to measure it; and let us consider that if they had continued in this restraint of not daring to add

continued on next page

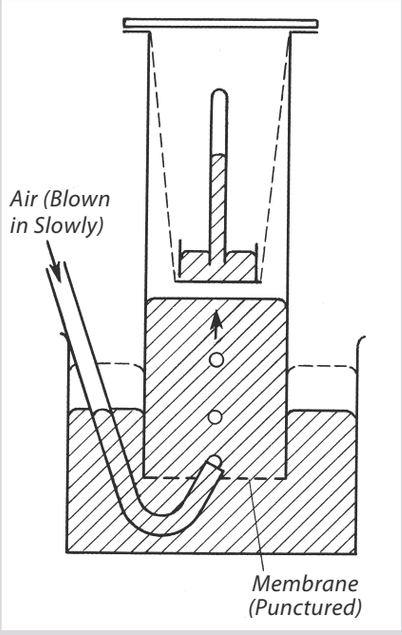
Figure 3.86 *continued*

any thing to the knowledge which they had received, or if those of their times had made the like difficulty in receiving the novelties which they offered them, they would have deprived themselves and their posterity of the fruit of their inventions.

As they only made use of that which had been bequeathed to them as a means whereby to gain more, and as this happy daring opened to them the way to great things, we should take that which they acquired in the same manner, and by their example, make of it the means and not the end of our study, and thus strive while imitating to surpass them.

For what is more unjust than to treat our ancestors with more deference than they showed to those who preceded them, and to have for them that inviolable respect which they have only merited from us because they had not the like for those who possessed the same advantage over them?

—Blaise Pascal, "Preface to the Treatise on Vacuum" [pp. 547–548]



▲ **Figure 3.87** Schematic representation of the experiment "void within a void" (after Hund).

The decisive experiment was designed by Pascal in 1647 and was carried out by his brother-in-law Florin Perier on the mountain Puy-de-Dôme. Pascal had correctly argued that if the height of the mercury column depended on the pressure caused by the weight of air, then the height of the column should decrease as the altitude above sea level increases. The very carefully executed experiments indeed showed an easily measurable difference of eight centimeters between the reading of the barometer set at the foot of the mountain and of the one installed 800 meters higher. This experiment, performed on September 19, 1648, belongs among the "grand" experiments in the history of physics. It was, of course, not described in Pascal's book that appeared in 1647; the world first learned about it in October 1648 from the article *Récit de la grande expérience de l'équilibre des liqueurs*. Pascal's findings were received with rapt interest by scholars; yet after his death, his relatives, the executors of his estate, all but apologized for Pascal's experiments on air pressure by remarking that the readers might get a false impression if they knew only this work, which was really unbecoming of such a wonderful genius.

Following in the wake of Pascal's investigations was a clever experiment by Edme Mariotte (ca. 1620–1684)—who, independently of Robert Boyle, discovered the Boyle-Mariotte gas law—which complemented the experiment carried out on Puy-de-Dôme. Replacing altitude with depth, Mariotte placed barometers underwater at various depths (Figure 3.88) and found that the height of the column of mercury increased with depth, exactly according to the expected relationship, namely the ratio of the densities of mercury and water.

Concurrently with the Italian and French investigations, in the city of Magdeburg that was still suffering from the turmoil following the Thirty Years' War, Otto von Guericke (1602–1686, Figure 3.89), the mayor, carried out several interesting experiments. He considered it evident that if we remove matter, for example water or air, from an enclosed space, then a vacuum will result and the external air pressure will weigh on this vacuum. He first tried to create a vacuum by pumping the water out of a water-filled barrel. This was unsuccessful because air immediately rushed in between the staves of the barrel. Guericke's greatest contribution to science was the construction of the air pump. Using it, he was able to pump the air out of a metal vessel, thereby creating a vacuum within. He had many problems at the beginning, for the pressure of the outside air flattened the vessel. Eventually he was able to construct vessels of sufficient strength, and with these, Guericke performed some notable experiments. In 1654, he conducted the demonstration for which he is famous, in which he joined together two hemispherical shells and pumped the air out of the resulting sphere, at the parliament building in Regensburg in the presence of the emperor. Two teams of eight horses were unable to overcome the air pressure and pull the hemispheres apart. After air was allowed to enter, the two hemispheres immediately fell apart. Figure 3.90 depicts two similar demonstrative experiments.

Guericke's experiments received wide publicity through the 1658 work *Mechanica hydraulica-pneumatica* by the Jesuit Gaspar Schott.

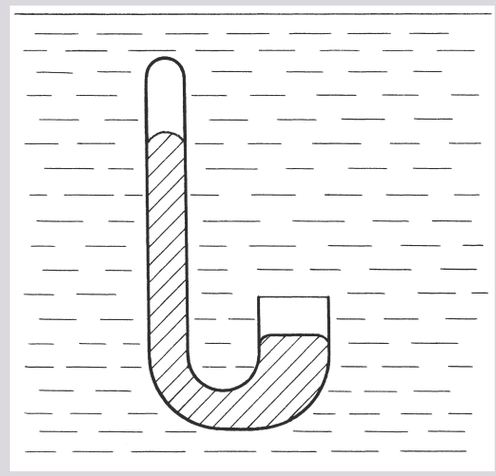
This work was continued primarily by Robert Boyle (1627–1691; Figure 3.91) with his assistant Hooke, who later became famous. They, too, built an air pump, but instead of simply constructing a replica of Guericke's invention, they developed an improved variant (Figure 3.92). As a result of their extensive experiments with vacuums, they determined that light can travel through a vacuum, but sound cannot. They also discovered that combustion cannot take place in a vacuum, and that animals placed in a vacuum die.

Unlike Guericke, Boyle investigated the structure of matter extensively, but cautiously avoided getting involved in the dispute over the nature of the vacuum. He declared that by "vacuum," he simply meant a region of space in which there is no air.

Figure 3.93 shows an experimental setup that still plays an important role in physics instruction as a clear presentation of the gas laws. If this U-shaped glass tube, which is closed on one side and open on the longer side, is filled with ambient air at the beginning of the experiment, we find that, as we pour mercury in the longer side, the volume of the trapped air decreases. Boyle used this experiment to refute the skeptical notion that the pressure of air would not be able to raise a column of mercury to a height of 76 centimeters. He showed that in the open side, an arbitrarily high mercury column can be created—just by adding sufficient amounts of mercury—and that the air in the closed side, now in a smaller volume, will keep the equilibrium. This experiment can be used to verify the Boyle-Mariotte law, according to which the product of the volume and pressure of a gas is constant, assuming constant temperature.

While Boyle was a committed proponent of the corpuscular theory of matter, he made no attempt to interpret the relationship between the volume and pressure of a gas on the basis of this theory. In fact, the inference was there to be made that the pressure exerted by the gas should be a consequence of the movement of atoms, that is, that the force exerted by a gas on the walls of a container is the result of particles striking the walls at high speeds and being reflected from the surface. The detailed mathematical treatment of this idea was for a long time hindered by the circumstance that according to Newton, such a force could also be exerted by atoms at rest. Newton's reputation was such that it was only in 1738 that Daniel Bernoulli was finally able to make significant progress in this area.

The practical significance of the work of Torricelli, Pascal, Guericke, and Boyle was quickly recognized. For example, the relationship between air pressure and weather was soon determined and utilized for forecasting.



▲ **Figure 3.88** Mariotte's experiment to explain the role of external pressure: he placed barometers at various depths of water.



▲ **Figure 3.89** Title page of Guericke's famous book: *Otto von Guericke's New (So-Called) Magdeburg Experiments on Empty Space*. (Library of the Technical University, Budapest.)

3.5.4 *Uncertain Steps on the Path to Modern Chemistry*

Unlike the other challenges to the Aristotelian conception in the seventeenth century, the investigations into the structure of matter led to only minor differences. That is, of course, not surprising. Rather what is astounding is that the pre-Socratic natural philosophers focused their scientific interest precisely on the most difficult of problems. Namely, as we have frequently stressed, in studying the nature of matter, it is extremely difficult to separate out the elementary phenomena that in their clear form would reveal the fundamental relationships. By the seventeenth century, chemistry and metallurgy could boast a number of practical results, but none of these were of a nature to offer insight into how to investigate the structure of matter. Not only the solution, but even the correct formulation of the question, had to wait for the eighteenth and nineteenth centuries. It is not as if there had been no relevant attempts in the seventeenth century. Every one of the scholars mentioned above did some work on this problem. Galileo, for example, although proceeding with great caution, made attempts to answer some related questions using a renewed version of Democritus's atomic theory. The main difficulty here was, in fact, an ideological one, for ever since antiquity, atheism was tightly linked to atomism as its theoretical basis. We should recall Lucretius's poem, in which he links the atomic idea into an organic whole with the categorical denial of all supernatural powers. The existence of eternal and unalterable atoms is irreconcilable both with creation and with divine influence.



106 LIBER TERTIUS.

ICONISMUS XII.

terim autem, ut aliquid fiat, ideoque clavicula (*videatur Iconismi* XI. *Figura* V) noſtræ phialæ, infra ſimul ventilio aliquo inſtructa eſt, per quod ſyringe aër inflari nec tamen regredi poteſt; referatâ itaque prius clave ut aër intret & globum repleat, ac deinde, ſyringis ope tribus vel quatuor vicibus inflato aëre, diſrumpuntur hæ phialæ ſponte ſuâ, à ſe ipſis. Unde reverâ poſſumus dicere, illud quod à 24 equis Imò (quando Hemiſpheria ſunt majora) à 100 & ultrà Equis, non queat divelli, ſaltem hoc modo ab uno viro poſſe diſſlari, vel flatu diſjungi.

CAPUT XXV.

Aliud Experimentum, quo oſtenditur Lances ſeu Phialas præ-dictas, pondere divelli.

PRæter ea, quæ hactenus acta ſunt, de commemorando cogitavi: Utpote has phialas *A B*, ſi pondere quodam decidenti divellerentur, tantum fragorem quantum bombarda, dum pyrii pulveris operâ glo-



DE PROPRIIS EXPERIMENTIS. 109

ICONISMUS XIV.

CAPUT XXVII.

De Vaſe Vitreo, plures quàm viginti vel 50 & plures homines robuſtos attrahere valente.

CUm in Comitiis Ratiſponæ anno 1654 habitis, Reipubl. cauſâ degerem, & *Cæſar. Majeſtati* FERDINANDO TERTIO, ut & *Rom. Regiæ Majeſtati* FERDINANDO QUARTO, utriuſque memoriæ beatiſſimæ, quorundam Experimentorum meorum, inter quæ ceu nova antecedens numerabatur, ſpecimen, juſſus præberem, atque de illis prius diſſererem, quod ſi quis hominum halitum oris ſui in hoc vitrum *Recipiens nempe dictum* efflaret, illeſimul animam eſſet expiaturus; Princeps *de Auersberg, dictæ Regiæ Majeſtatis Aulæ Præfectus*, antè adducì non potuit, quò hiſce verbis meis fidem haberet, quam ipſe, quæ in lagenâ vitreâ facta eſſent, oculis ſuis vidiſſet. Sed cum etiam tunc temporis alia

O 3

▲ **Figure 3.90** GUERICKE's experiments with the hemispheres (top) and with a cylinder and a piston (bottom).

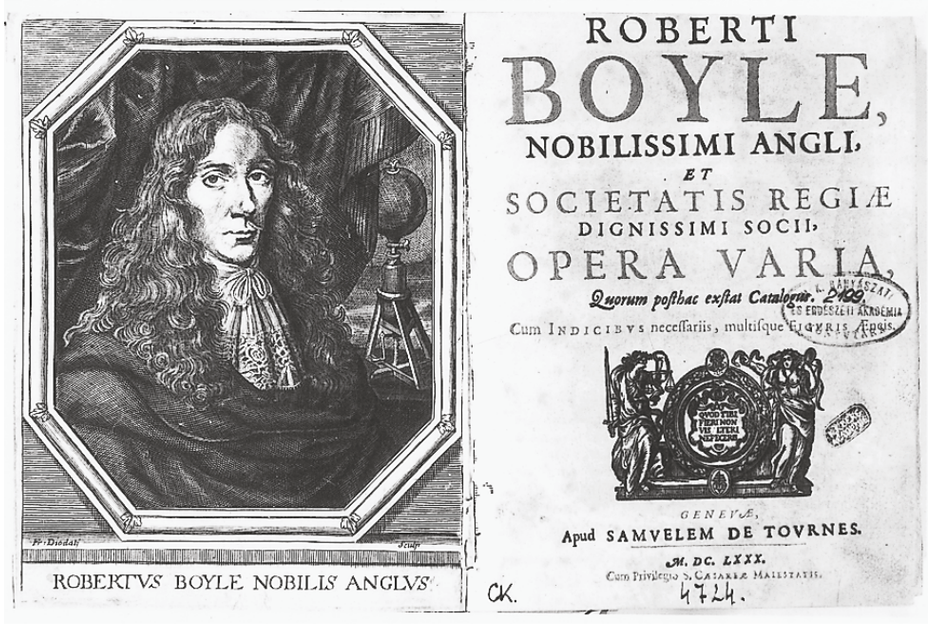
It was Father PIERRE GASSENDI (1592–1655) who finally brought atomic theory into a form that was acceptable to Christian ideology. According to him, atoms are not eternal, but were created by God, and on Judgment Day, God could destroy them should he so choose. Their motion is not random, but is subordinate to the will of God. The fundamental properties of all atoms are their hardness and mutual impenetrability. They differ in regard to size (*moles*), form (*figura*), and motion (*pondus*). GASSENDI thought of this last quantity as some kind of *impetus*. The atoms can combine in various ways. From here on the treatment followed the path outlined already by DEMOCRITUS and ARISTOTLE. Groups of atoms differ with respect to arrangement (*ordo*) and position (*situs*). He gives examples: those that differ with respect to arrangement could be compared to anagrams such as *Roma* and *Amor* or *Laurus* and *Ursula*, where the only difference is the arrangement of the letters, for example, *R* and *A* are at different positions. GASSENDI expanded this graphic metaphor and concluded that just as an immeasurable number of books can be written by using the letters of the alphabet, so can the different groupings of atoms account for all the infinite variety of forms in nature. According to GASSENDI, each atom at the moment of its creation received an impulse that remains with it, unlike the Cartesian theory, in which only the total impulse of the universe is what remains unchanged. GASSENDI tried to bring atomic theory into harmony with the Aristotelian four elements. To that end, he postulated that atoms prefer particular forms (*concretiunculae*) that correspond to the four elements.

We have already discussed DESCARTES's notions about the structure of matter. His ideas did have some chance of being developed into quantitative laws for the behavior of matter in a macroscopic scale. However, the Cartesian assertion that an empty region of space is an absurdity represents an extraordinary limitation because it then follows that in the entire universe, the only possible motions would be those similar to the flow of incompressible fluids. Therefore, even DESCARTES's most loyal followers, in their attempts to explain the structure of macroscopic bodies, chose to allow for the existence of a vacuum, even though they continued to accept the fundamental Cartesian assumptions of the three types of matter.

As a result, until the middle of the seventeenth century, the following theoretical models existed, to some extent independently and to some extent intertwined:

- The four Aristotelian elements, earth, water, air, and fire. However, even the followers of ARISTOTLE were forced to give up the idea of an infinite divisibility of matter and to speak of a "least part" of each element (*minima naturalia*).
- The idea going back to PARACELSUS that everything in the world is composed of three fundamental principles (*tria prima*): mercury, salt, and sulfur. This hypothesis remained extremely popular for some time.
- The Cartesian theory, according to which there are three types of matter with particles of various degrees of fineness that filled space completely. Each of these types of matter was responsible for a different group of phenomena.
- Finally, the atomic theory of DEMOCRITUS in the form propounded by GASSENDI, which we discussed above.

In 1661, BOYLE published *The Sceptical Chymist*. Like GALILEO's books, it is cast in dialogue form, and the title suggests that BOYLE was less interested in constructing new theories than in criticizing the old ones. In the book, one of the three interlocutors declares his allegiance to the teachings of ARISTOTLE, another is a follower of PARACELSUS, and the third is the skeptical chemist BOYLE himself. His main line of attack is a sharp criticism of the Aristotelian ideas in their

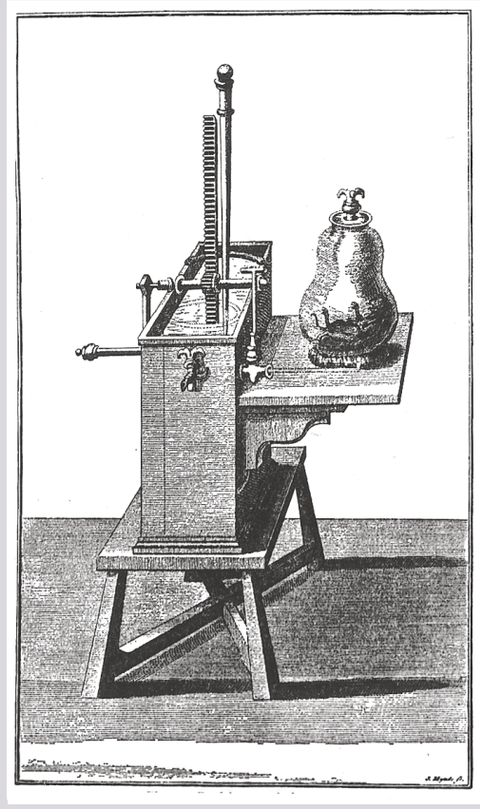


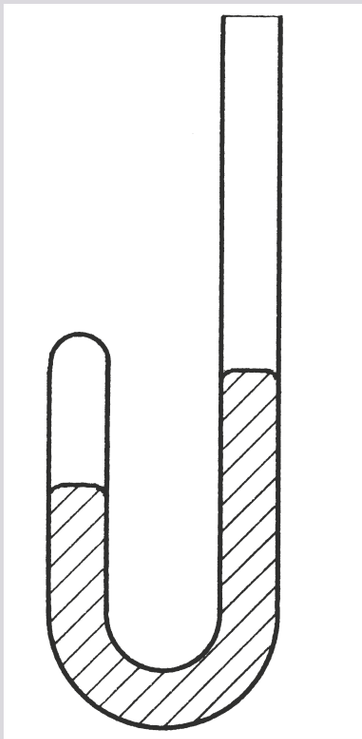
◀ **Figure 3.91** Boyle's likeness and the title page of his book. (Library of the University for Heavy Industry, Miskolc, Hungary.)

original form, suggesting that the Aristotelian four elements are clearly incapable of fulfilling the task originally conceived for them, namely, that of constructing the varied manifestations of the universe, because they themselves can be decomposed into further components. Moreover, other materials disintegrate when subjected to various treatments—primarily when heated or burnt—not just into four, but into even greater number of constituents. Boyle's book is frequently cited as representing the beginning of modern chemistry. In this work—more fully developed in the second edition of 1680—we meet the notion of chemical element in an embryonic form. In the 1666 article, *The Origin of Forms and Qualities according to the Corpuscular Philosophy*, Boyle compellingly unites the theory of Gassendi and the Cartesian system. Not only in this work, but also in another work written for this express purpose, he made a great effort to show that the corpuscular theory was compatible with religious sensibilities. He went so far as to maintain that involvement in experimental science helps to make a person a better Christian. He carefully circumvents the term "atomic theory" and speaks instead of "corpuscles" to avoid the association between atomic theory and atheism mentioned earlier.

Boyle, by the way, believed in the dream of all alchemists to make gold from a base metal. He even wrote a letter to the Royal Society on this subject. It should not surprise us that the author of a book called *The Sceptical Chemyst* should have been involved in such activities, which at first glance seem to belong to the realm of the occult, for we must not forget that the supporters of the corpuscular theory believed that the various forms of matter differ only in the arrangement or position of their otherwise identical components. Thus there should be no reason—at least theoretically—why some technique or other could not succeed in altering the arrangement or position of the components in such a way as to result in the structure of a different material. When in contemporary nuclear physics we bombard a nucleus with high-energy particles until it loses a proton and turns into a different chemical element, we are doing just that. Today, of course, we know that the mystical dream of the alchemists and Boyle's ideas based on atomic theory could not have been realized only because the equipment to produce the necessary high energies for such a transmutation was then unavailable. With his investigations,

▼ **Figure 3.92** Boyle's air pump.





▲ **Figure 3.93** The U-tube, often used in physics instruction, open at one end and sealed at the other. Boyle carried out experiments with his assistant Hooke using instruments of this type.

Quotation 3.32

But I attribute determined figures, and sizes, and movements to the imperceptible particles of bodies, as if I had seen them; and yet I acknowledge that they are imperceptible. And on that account, some readers may perhaps ask how I therefore know what they are like. To which I reply: that I first generally considered, from the simplest and best known principles (the knowledge of which is imparted to our minds by nature), what the principal differences in the sizes, figures, and situations of bodies which are imperceptible solely on account of their smallness could be, and what perceptible effects would follow from their various encounters. And next, when I noticed some similar effects in perceptible things, I judged that these things had been created by similar encounters of such imperceptible bodies; especially when it seemed that no other way of explaining these things could be devised. And, to this end, things made by human skill helped me not a little: for I know of no distinction between these things and natural bodies, except that the operations of things made by skill are, for the most part, performed by apparatus large enough to be easily perceived by the senses: for this is necessary so that they can be made by

continued on page 242

Boyle attracted the attention of Newton, who had an intense interest in similar questions. However, Newton's critical sense about unsupported speculation saved him at least to the extent that he never published anything on the topic.

We do not wish to discuss in detail the ideas of Huygens, who advanced the furthest in understanding the structure of matter, but still failed to draw from his theory quantitative inferences of either chemical or mechanical character.

We must now reach a bit into the future. The basic ideas of phlogiston theory appeared in the seventeenth century (Joachim Becher, 1635–1682), and even the first detailed exposition of the theory was given in 1697, by Georg Ernst Stahl (1660–1734). However, its general acceptance and status as prevailing theory only reached its full flowering in the eighteenth century. The most ordinary and obvious chemical reaction, namely combustion, was explained by the hypothesis that combustible materials (coal, oil, sulfur) contain *phlogiston* (φλογιστός = burnt), which escapes during combustion. The more phlogiston a material contains, the more easily it burns; thus coal, for example, consists almost entirely of phlogiston. The rusting of metals was also imagined as such a process. The nature and properties of phlogiston remained unexplained: Many considered it more a principle than a form of matter; others saw it as a weightless (perhaps even characterized by negative weight) fluid of a special kind. Its role in combustion was, one might say, the opposite of that assigned today to oxygen. The reduction of metallic ore could be represented as follows:

$$\text{ore} + \text{phlogiston} \rightarrow \text{metal};$$

today, we would write

$$\text{ore} - \text{oxygen} \rightarrow \text{metal}.$$

What we today call pure metal was considered by the supporters of the phlogiston theory to be a compound. The phlogiston required for reduction is provided by the burning of coal. There were experimental results available—generally accurate—on weight changes in chemical reactions. However, any discrepancies from theory were generally explained away by additional assumptions.

Despite its fundamental falsity, the phlogiston theory played a progressive role; it was able to provide useful directions for practical application (the large-scale production of sulfuric acid was achieved through the guidance of Stahl). Thus it comes as no surprise that such great physicists as Cavendish and Priestly supported this theory. It was not until the end of the eighteenth century that Lavoisier finally succeeded in providing an alternative corresponding to the modern conception of chemistry.

3.6 After Descartes and before Newton: Huygens

3.6.1 Huygens's Axioms on Dynamics

When we speak of Huygens's talent, his accomplishments, and his significance, we ought to do so in superlatives, although we should choose our words with care because the seventeenth century was a century of geniuses (Table 3.3). Many of those listed in the table have earned great renown from particular points of view:

Kepler had the liveliest imagination but was also the most precise and disciplined observer.

Galileo was a sober researcher with well-defined objectives who worked with the most modern methods of his time.

Descartes was a prisoner of his own theory, yet his work had enormously far-reaching and significant influence.

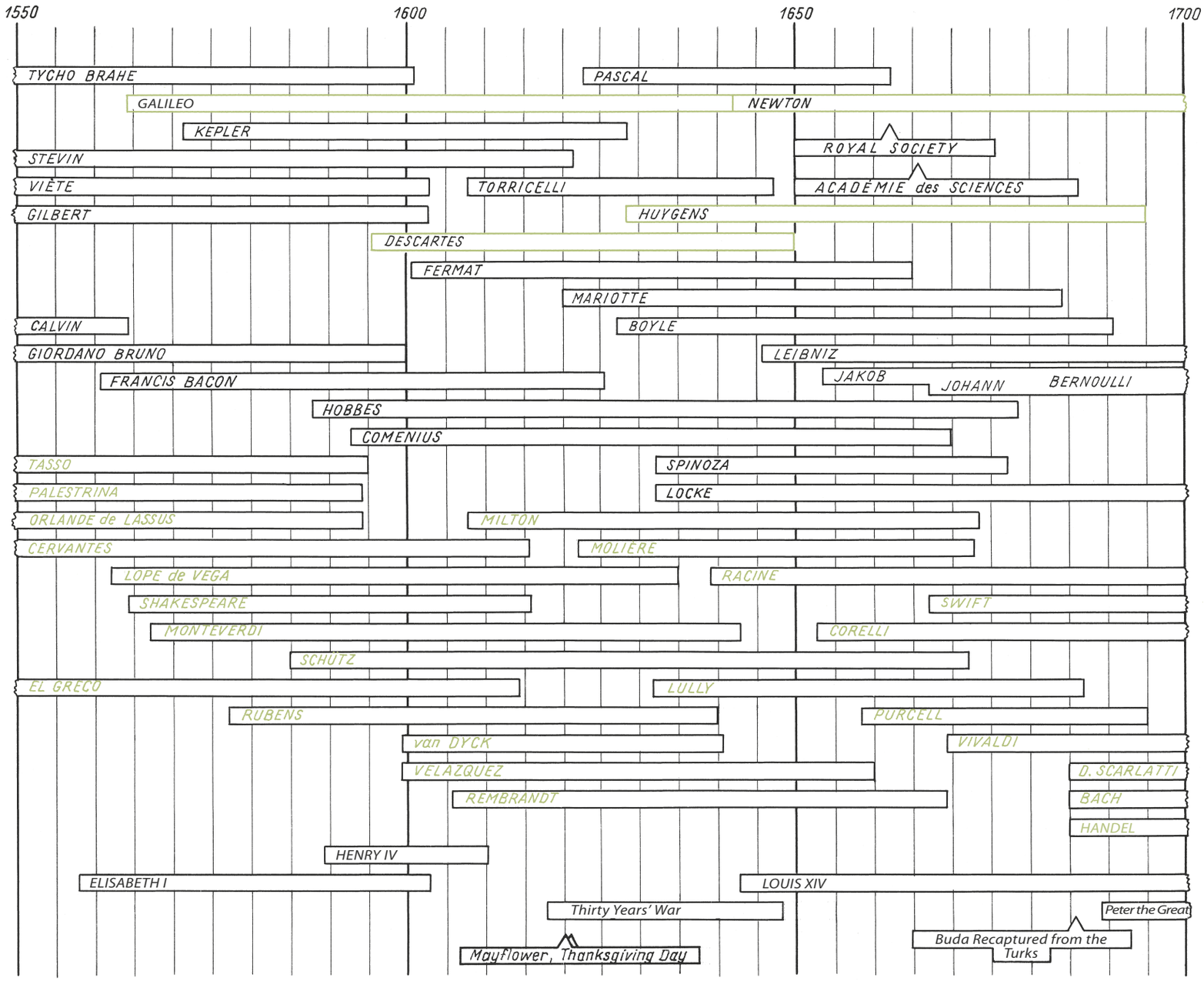


Table 3.3 Chronology of the century of genius. The "century of change" has a compelling narrative. It began with the burning at the stake of Giordano Bruno, and then reached such an intellectual high point with Galileo and Descartes that the achievement of further heights would appear impossible, although it was indeed achieved with Huygens and Newton. From this summit, things proceeded at a calmer pace into the following century. The goal was to understand the world as it appeared to the inquisitive human mind, that is, how it was formed by the Creator. The worldview of the century, in this sense, remained static with only a few exceptions.

Among the experimental tools used in the seventeenth century, we mention here the microscope (likely invented in 1590 by the Dutchman Zacharias Janssen), which played an important role in the study of plants and animals. R. Hooke, *Micrographia* (1665): plant cells; Antonie van Leeuwenhoek (1632–1723): bacteria, protozoa; Marcello Malpighi (1628–1694): tissues, red blood cells, study of embryos, capillaries; William Harvey (1578–1657): the 1628 work *De motu cordis et sanguinis*, in which Harvey describes the circulation of the blood and the action of the heart, marks a turning point. Comenius (1592–1670): founder of scientific pedagogy; *Didacta Magna*, published in Czech in 1632 and in Latin in 1638; *Orbis sensualium pictus*, published in Latin and German in 1658, in Latin, German, and Hungarian in 1669, and in Latin and Hungarian in 1675; *Prodromus pansophiae* (harbinger of universal knowledge) in 1637.

A number of significant scientific academies were also founded in this epoch, to which we shall return in Section 4.1. In this connection we must mention here the name Marin Mersenne (1588–1648), who took on himself the role of scientific journal and a learned academy. Through his travels and correspondence, he maintained personal or written contact with the most important thinkers of his time: Galileo, Descartes, Desargues, Pascal, the Huygens family, and so on. To report a discovery to Mersenne was tantamount to having it published throughout Europe. Himself a scientist, he extended, in his *Harmonie Universelle* (1636–1637), based on his own experiments with long stretched strings, the work of Benedetti, Beeckman, and Galileo on the relationship between pitch and frequency and between the physical properties of the string and frequency. Joseph Sauveur (1653–1716) continued work in this direction: From him we have the word "acoustics"; he discovered overtones and the phenomenon of beats.

continued on next page

Quotation 3.32, ***continued***

men. On the other hand, however, natural effects almost always depend on some devices so minute that they escape all senses. And there are absolutely no judgments [or rules] in Mechanics which do not also pertain to Physics, of which Mechanics is a part or type: and it is as natural for a clock, composed of wheels of a certain kind, to indicate the hours, as for a tree, grown from a certain kind of seed, to produce the corresponding fruit. Accordingly, just as when those who are accustomed to considering automata know the use of some machine and see some of its parts, they easily conjecture from this how the other parts which they do not see are made: so, from the perceptible effects and parts of natural bodies, I have attempted to investigate the nature of their causes and of their imperceptible parts. ...

And although perhaps in this way it may be understood how all natural things could have been created, it should not therefore be concluded that they were in fact so created. For just as the same artisan can make two clocks which indicate the hours equally well and are exactly similar externally, but are internally composed of an entirely dissimilar combination of small wheels: so there is no doubt that the greatest Artificer of things could have made all those things which we see in many diverse ways. And indeed I most willingly concede this to be true, and will think that I have achieved enough if those things which I have written are only such that they correspond accurately to all the phenomena of nature, [whether these effects are produced by the causes I have explained or by others]. And indeed this will also suffice for the needs of everyday life, because Medicine and Mechanics, and all the other arts which can be perfected with the help of Physics, have as their goal only those effects which are perceptible and which accordingly ought to be numbered among the phenomena of nature.

—René Descartes, *Principles of Philosophy*, IV.203, 204 [pp. 285–286]

Table 3.3 *continued*

Interesting synchronies

1619: Kepler—*Harmonices mundi*; Velázquez—"Adoration of the Kings"; Schütz—*Psalms of David*; founding of the clearing bank in Hamburg; beginning of the large-scale transportation of African slaves to the Americas; study of the optics of the eye by Scheiner.

1632: Galileo—*Dialogo*; Rembrandt—"The Anatomy Lesson of Dr. Nicolaes Tulp"; Gustavus Adolphus killed at the Battle of Lützen against Wallenstein; in this year, Locke, Spinoza, Vermeer, Lully, and van Leeuwenhoek were born.

1664–1666: Newton's "anni mirabiles"; Frans Hals—"Women Regents of the Haarlem Almshouse"; Poussin—"Apollo and Daphne"; death of Poussin; Schütz—*Christmas Oratorio*; Hooke—*Micrographia*; death of Fermat; Grimaldi—diffraction of light at an opaque edge; Spinoza—*Renati Descartes Principia Philosophiae more geometrico demonstrata*; Molière—*Tartuffe*, *Don Juan*; Racine—*Alexander the Great*.

1687: Newton—*Principia*; explosion of gunpowder at the Parthenon; death of Lully; beginning of the parliament at Bratislava, at which the succession of the male descendants of the Habsburgs were recognized by the Hungarian Estates. In the previous year, Buda was recaptured from the Turks.

In Pascal, theoretical and practical gifts were linked most harmoniously to a literary predisposition.

Newton was indisputably the giant of the century without peer.

How does Huygens fit into this pantheon? First of all, we have to say that Huygens was neither a philosopher nor a writer. He was a professional scholar, physicist, and mathematician, like Archimedes, with whom he may be compared with respect to his method and his role in science.

With Huygens (Figure 3.94), physics reached a level comparable to that of an elementary course in college physics today. His investigations—aside from the qualitative aspects of his theory of light—go beyond high-school material and are an organic component of mechanics as taught at universities today. With them, physics became an independent field of science, one that to learn and develop further would require a person's complete devotion. Previously, physicists had been marginal figures in the history of philosophy—perhaps with the exception of Huygens's contemporary Leibniz—and conversely, only a few ideas of the great philosophers found any use in physics.

In his writings, Huygens did not start out with general philosophical principles; instead, what underlies his thought are simple, vivid, and quantitatively formulated *physical* principles. All of his results are sound even to this day; they have practical applications, and indeed, Huygens himself put many of them into practice. He built on the work of his predecessors, first of all Galileo, but Descartes also had a great influence on him, even though Huygens protested to the contrary (Quotation 3.43). Huygens brought the work of his predecessors along a straight path a step closer to the closure eventually realized with Newtonian mechanics. Where his investigations were not complete, such as, for example, his recognition of the transformations between potential and kinetic energy, he anticipated the results that were later derived from Newtonian mechanics.

With that said, we may now invoke the appropriate superlatives of Huygens: He was a consummately critical and disciplined physicist, surpassed in his century only by Newton in the number of concrete results achieved.

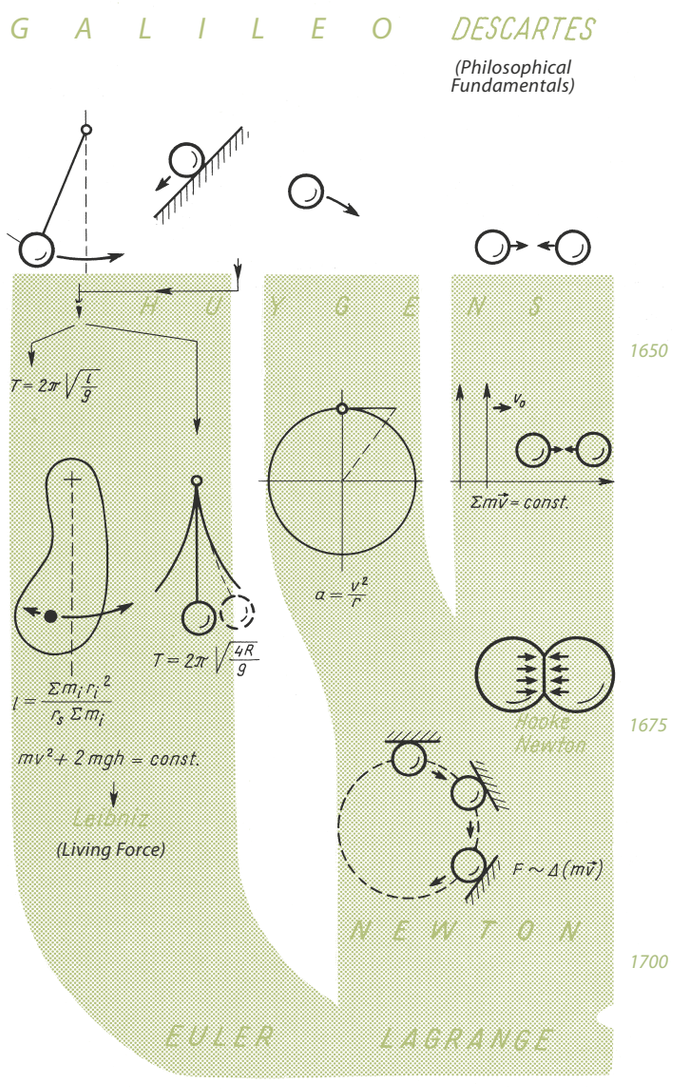
We have attempted with Figure 3.95 to present the state of knowledge in mechanics at the time of Galileo's death. There are two paths of development to be seen, one of which leads to quantitative solutions to complex problems, and even to a particular formulation of the law of conservation of energy. This path builds on the experiments and results of Galileo on the inclined plane, the pendulum, and projectile motion.

Another path, with two branches, leads directly to Newton's fundamental laws. It begins with collision phenomena, in which the significance of the quantities *momentum* and *change in momentum* are the most obvious. The other branch involves the

investigation of the motion of bodies along a curved path, where at first, of course, consideration was limited to simple circular motion. In considering such motion, the vectorial nature of the quantities velocity and change in velocity becomes clear.

Let us proceed now with Huygens along the path laid out by Galileo. We accept the laws that Galileo discovered concerning motion along an inclined plane: velocity grows in proportion to time, and distance traveled is proportional to the square of the time if the body begins with zero velocity. Huygens re-derived these results with a somewhat different reasoning. In his *Horologium oscillatorium*, Huygens formulated two hypotheses:

1. Principle of inertia: A body left to itself moves uniformly along a straight line.
2. Principle of superposition of motions.



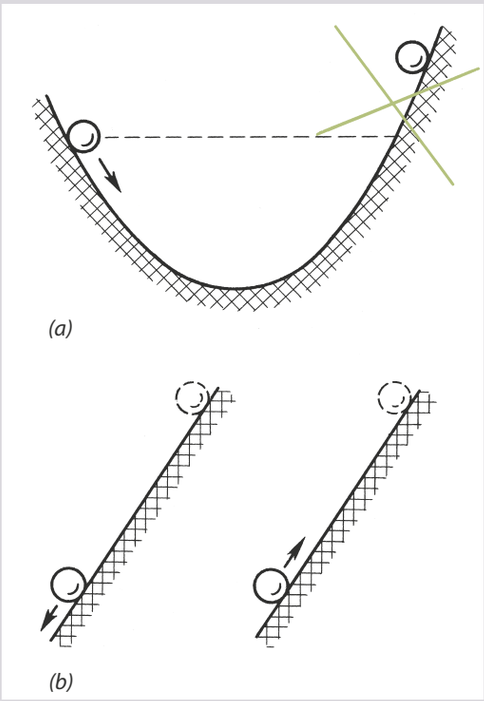
▲ **Figure 3.95** The road from Galileo to Newton: Galileo gave a quantitative description of motion in a gravitational field; Descartes laid emphasis on collision processes, and Huygens continued both of these directions. He found the precise relationships of the period of a mathematical pendulum and derived the law of collision from the general principle that all coordinate systems that move at a constant velocity with respect to one another are equivalent. His results, leapfrogging the laws of motion, could later be understood as beautiful examples of the principle of conservation of energy, for which reason we have let this part of Huygens's work rejoin the main flow *after* Newton. The laws of motion themselves were derived from detailed investigation of motion along a curved—circular—path and of the collision process.



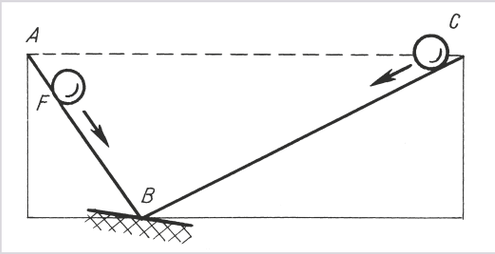
▲ **Figure 3.94** Christiaan Huygens (1629–1695): Born in The Hague; first instruction from his father, a well-known politician, linguist, musician, and mathematician. Finished legal studies, also very early involvement with mathematical problems, obtaining his first results at age 17 by solving problems of the catenary curve. At 22, treatise on determining the areas of conic sections (*Cyclometriae*, 1651); at 25, gave the best known value for *p* (*De circuli magnitudine inventa*, 1654); also contributions to probability theory. Huygens began his work on problems of physics with astronomical investigations; with his brother, he built his own telescope. In these investigations, which led to the discovery of a ring about Saturn, one of Saturn's moons (*Systema Saturnium*), and the Orion nebula, Huygens was faced with the task of determining times precisely. In 1657, he constructed a pendulum clock and then was intensively involved with the problem of finding a pendulum with period independent of maximal deflection (cycloidal pendulum). His results in this area—as was typical for him—were published only after great delay (*Horologium oscillatorium*, 1673). Another very significant work, *De motu corporum ex percussione*, appeared only after his death, in 1703, although he had presented a summary to the Royal Society as early as in 1669. His best-known treatise is the *Traité de la lumière* (1690). In the foreword, Huygens gave this reason for the delay in publication: he needed time to translate the book into Latin, the language of science.

In the period just prior to Newton, that is, in the third quarter of the seventeenth century, Huygens enjoyed a great reputation among physicists. In 1663, he was named a member of the Royal Society, and in 1665, was nominated by Colbert to be president of the Académie Française. In 1681, he left France, on account of his poor health, and was unable to return following the 1685 revocation of the Edict of Nantes. Huygens died in the city of his birth in 1695.

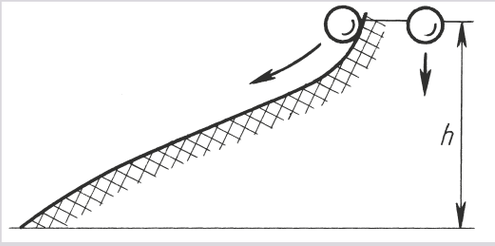
Huygens was also exemplary as a man, for he recognized the achievements of others despite his critical attitude. Although he devoted all his energies to science, he did not live the life of a recluse; this is clear from the number of poems addressed by Huygens to beautiful women of his time. Characteristic of the broad spectrum of his interests is the manuscript of a science-fiction novel found among his posthumous papers, as well as a design for an internal combustion engine.



▲ **Figure 3.96** HUYGENS's two hypotheses: **(a)** The center of gravity of an object moving due to its weight can never rise to a height greater than the height at which it began. **(b)** Motion is reversible.



▲ **Figure 3.97** HUYGENS's proof that spheres rolling on inclined planes with differing angles with reach the base with the same speed if they are started from the same height.



▲ **Figure 3.98** If a body is released along an arbitrary inclined surface, then its speed is equal to that of a body under free fall dropped from the same height.

HUYGENS considered vertical free fall itself an instance of compound motion: A body falls with the velocity attained at the end of the first moment, as a constant velocity, to which the velocities attained in the succeeding moments are superposed. In the *Discorsi*, GALILEO hypothesized that under ideal conditions, the velocity at the foot of an inclined plane depends only on the height of the starting point (Quotation 3.18). HUYGENS proved this theorem with the help of another postulate, which later turned out to be enormously fruitful. He generalized a theorem due to TORRICELLI from the area of statics to state the following principle:

> *If an arbitrary system of bodies moves under the exclusive influence of forces due to weight, then the common center of gravity of the bodies will at no time be higher than it was at the beginning of the motion.*

To simplify the principle for the case of a single body: a body moving under the influence of its own weight cannot reach a height higher than its initial position (Figure 3.96(a)). In this form, the principle is self-evident.

Let us now add to this another principle of HUYGENS that can be proved already with the help of the initial hypotheses: *Under ideal conditions, motion is reversible*. In the simplest case, this theorem states that if a body moves down an inclined plane with initial velocity zero at the highest point and arrives at the foot with a final velocity v_f, then it can reach the highest point with velocity zero if it starts at the foot with the same velocity v_f (Figure 3.96(b)).

Now let us see how with the help of these two principles, HUYGENS derived the result that two bodies starting at the same height on inclined planes of differing angles will arrive at the feet of their respective planes with the same velocity (see Figure 3.97). Suppose that a ball rolls from point C to point B, and its velocity is less than that of the ball that rolls from point A to point B, and suppose further that its velocity is equal to that of a ball that began its motion on the inclined plane AB at point F. Let us then let a ball roll from point F. According to these assumptions, it must attain the velocity that will allow it to reach point C if we let it roll up plane BC. But the point C is higher than point F, which means that the center of gravity of a body moving only under the influence of the forces due to weight has moved higher than its initial position, which is impossible according to the first principle. We note here that according to HUYGENS, the change in direction from the motion on inclined plane FB to motion on BC can be accomplished by the ball ricocheting off a flat plate arranged at a suitable angle at the base of both inclined planes.

In an analogous fashion, one can prove that the velocity of a ball rolling from point A to point B cannot be less at point B than the velocity of the ball rolling from C to B at that point. Combining these two results allows us to conclude that the two balls must have the same velocity.

Utilizing a similar method, HUYGENS also showed that the final velocity depends only on the initial height even when the inclined surface is not flat but follows an arbitrary curve (Figure 3.98). Furthermore, he proved that the time of the fall (for a given initial height) is proportional to the length of the inclined plane (Figure 3.99). These results actually go back to GALILEO, but it was HUYGENS who gave them a logical structure.

From the equation $s = \frac{1}{2} g \sin\alpha t^2$ follows

$$t = \sqrt{\frac{2s}{g \sin\alpha}},$$

and since

$$\sin\alpha = \frac{h}{s},$$

we have

$$t = \sqrt{\frac{2s^2}{gh}} = s\sqrt{\frac{2}{gh}}.$$

Since g and h are constant, we in fact have that $t \propto s$.

3.6.2 *The Mathematical Pendulum*

The velocity at any given point of a body moving along a curvilinear path as described in the above experiments can be easily determined from the vertical distance traversed. However, determining the time that it takes the body to traverse the given distance presents serious difficulties because we must consider that the angle of inclination of the path can change from moment to moment. With integral calculus the problem would be relatively easy to solve. However, we must not forget that Huygens did not have this mathematical tool at his disposal, for at the time, Leibniz and Newton were just beginning to conceive of it. For this very reason, the work of Huygens on the mathematical, physical, and cycloidal pendulums must be viewed as truly remarkable.

In discussing the work of Galileo, we mentioned that he mistakenly assumed that the period of a pendulum was independent of the distance traveled by the pendulum bob, that is, of the amplitude, even though this assumption is valid only for small amplitudes. Galileo's other mistake, which is closely related to this one, was to assume that bodies rolling from various points along the interior of a circle require the same amount of time to reach the base of the circle.

We shall now sketch—essentially following Dugas [1950]—the train of thought by which Huygens succeeded in determining the time required for a body to reach the base of a circle if the starting point is not too far distant. This same line of reasoning gives the period of a mathematical pendulum for small amplitudes. The condition "not too far distant" means here that the angle of inclination of the tangent line to the circular path at the starting point should be sufficiently small so that the sine of this angle can be approximated by its tangent.

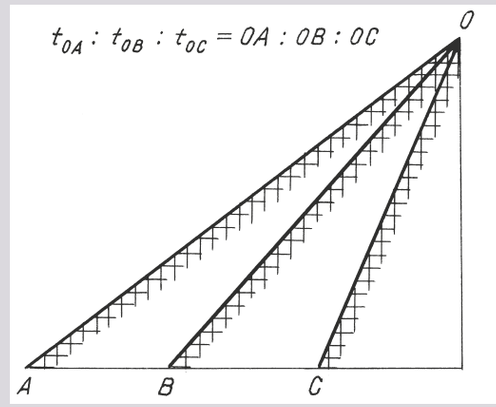
Since calculations with parabolas are easier to work out than those with circles, let us first approximate the circle by a parabola, as shown in Figure 3.100, whose curvature at the vertex is equal to that of the given circle. Let us divide the parabola into a series of small line segments (Figure 3.101) and attempt to determine the time necessary for the traversal of one of these segments. These small time intervals are then to be summed so that we may obtain the time it takes the body to reach the lowest point of its path. Such a time represents one-fourth the period of the pendulum. In this way, Huygens obtains the value

$$T = 2\pi\sqrt{\frac{l}{g}}$$

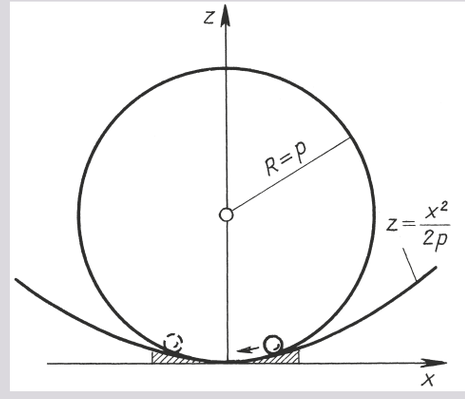
for the period, which can be found today in every physics reference book.

Let us now follow Huygens's line of thought further, remaining as faithful as possible to his ideas while making use of methods of higher mathematics that were available only later. In the coordinate system of Figures 3.100 and 3.101, the equation of the parabola is given by

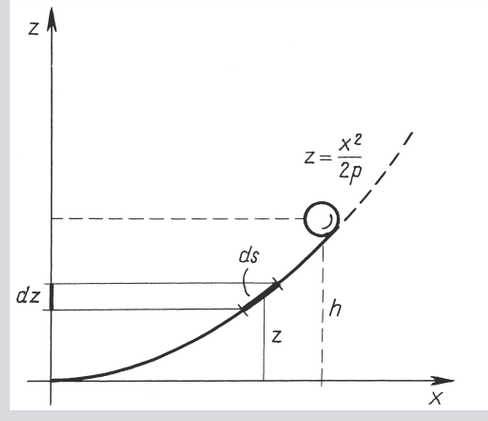
$$z = \frac{x^2}{2p}.$$



▲ **Figure 3.99** The time needed by spheres to roll down the inclined planes is proportional to the length of the plane in the case of identical starting heights.



▲ **Figure 3.100** The circle serving as the incline can be approximated in a neighborhood of its base by a parabola.

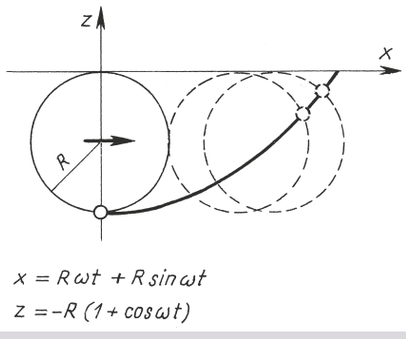


▲ **Figure 3.101** The travel time of a sphere on a parabolic path is the sum of the times taken to traverse the elementary distances ds.

Quotation 3.33

After having examined the nature of movement, we must consider its cause, which is twofold: [we shall begin with] the universal and primary one, which is the general cause of all the movements in the world; and then [we shall consider] the particular ones, by which individual parts of matter acquire movements which they did not previously have. As far as the general [and first] cause is concerned, it seems obvious to me that this is none other than God Himself, who, [being all-powerful] in the beginning created matter with both movement and rest; and now maintains in the sum total of matter, by His normal participation, the same quantity of motion and rest as He placed in it at that time. For although motion is only a mode of the matter which is moved, nevertheless there is a fixed and determined quantity of it; which, as we can easily understand, can be always the same in the universe as a whole even though there may at times be more or less motion in certain of its individual parts. That is why we must think that when one part of matter moves twice as fast as another twice as large, there is as much motion in the smaller as in the larger; and that whenever the movement of one part decreases, that of another increases exactly in proportion. We also understand that it is one of God's perfections to be not only immutable in His nature, but also immutable and completely constant in the way he acts. Thus, with the exception of those changes which either manifest experience or divine revelation renders certain, and which we either perceive or believe to occur without any change on the part of the Creator; we must not suppose that there are any others in His works, for fear of accusing Him of inconstancy. From this it follows that it is completely consistent with reason for us to think that, solely because God moved the parts of matter in diverse ways when He first created them, and still maintains all this matter exactly as it was at its first creation, and subject to the same law as at that time; He also always maintains in it an equal quantity of motion.

—René Descartes, *Principles of Philosophy*, II.36 [pp. 57–58]



▲ **Figure 3.102** A point on a circle rolling on the *x*-axis (touching it from below) describes a cycloidal path whose equation is given in the figure.

It can easily be proved that the parameter p appearing here is equal to the radius of curvature, since this follows in general from

$$\frac{1}{R} = \frac{z''}{\left(1+z'^2\right)^{3/2}},$$

and with

$$z'' = \frac{1}{p}, \qquad z'\big|_{x=0} = \frac{x}{p}\bigg|_{x=0} = 0,$$

we in fact obtain

$$\frac{1}{R} = \frac{1}{p}.$$

The velocity of the body beginning at height h that is at height z at time t is given by

$$v^2 = 2g(h-z),$$

or, equivalently,

$$\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = 2g(h-z).$$

Between the differential arc length $\mathrm{d}s$ and the vertical differential $\mathrm{d}z$, we have the relationship

$$\mathrm{d}s = \frac{\mathrm{d}z}{\sin\alpha}.$$

If we use the reasonable small-angle approximation $\sin\alpha \approx \tan\alpha$ and note the relationship

$$\tan\alpha = z' = \frac{x}{p} = \sqrt{\frac{2z}{p}},$$

which follows from the equation of the parabola, we then obtain

$$\frac{p}{2z}\left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2 = 2g(h-z).$$

From this, we obtain that the time required to transit the vertical differential $\mathrm{d}z$ is

$$\mathrm{d}t = \frac{1}{2}\sqrt{\frac{R}{g}}\frac{\mathrm{d}z}{\sqrt{(h-z)z}}, \tag{1}$$

and so the total transit time is given by

$$t_0 = \frac{T}{4} = \frac{1}{2}\sqrt{\frac{R}{g}}\int_0^h \frac{\mathrm{d}z}{\sqrt{(h-z)z}} = \frac{\pi}{2}\sqrt{\frac{R}{g}}.$$

Thus was Huygens able to derive the formula—still in use today—for the period of the mathematical pendulum.

3.6.3 *The Cycloidal Pendulum*

Huygens also posed the question whether there exists a curve for which the transit time of a body does not depend on the starting point (that is, on the distance from the base point). One of Huygens's most brilliant accomplishments was finding an answer to this question and showing that the *cycloid* (Figure 3.102) has this property. Not only did he solve this problem theoretically, but he even constructed

a pendulum whose period is constant regardless of the amplitude. We cannot present the original demonstration of this property of the cycloid, since Huygens required twelve theorems for the complete derivation, none of which is easy to grasp. Therefore, we shall once again attempt to present an abbreviated form of the derivation, making use of modern mathematical methods.

Given an arbitrary angle of inclination α, the time required to transit the vertical differential dz is given by

$$\mathrm{d}t = \frac{\mathrm{d}z}{\sqrt{2g(h-z)}} \frac{1}{\sin\alpha}.$$

As we have seen above, the height h can be derived from the transit time if the equation for dt can be expressed in the form of equation (1) above. We indeed arrive at relation (1) if the curve is so constructed that at every point, we have

$$k \sin\alpha = \sqrt{z},$$

where k is some constant. In this case, the value of the integral is independent of h. Therefore, Huygens formulated the problem as follows (Figure 3.103): we seek a curve for which the points F given by the equation

$$BF = \frac{BE}{TE} k$$

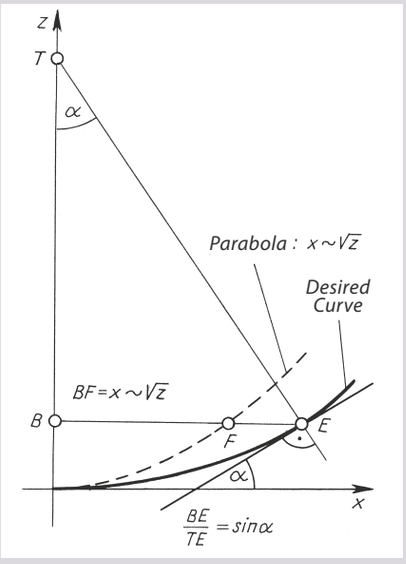
lie on a parabola. Thus he arrived at the cycloid, for which he was able to demonstrate the required property.

Huygens was still not satisfied with the rigor of the above derivation; here we must bear in mind that his infinitesimal methods suffered from problems characteristic of the labor pains preceding the birth of calculus. However, now with knowledge of the result, he was able to proceed in his work *Horologium oscillatorium* with Archimedean rigor. Thus he arrived first at a cycloidal curve as the final formula for the constant period, and from that, he derived the formula for the thread pendulum, valid for small amplitudes.

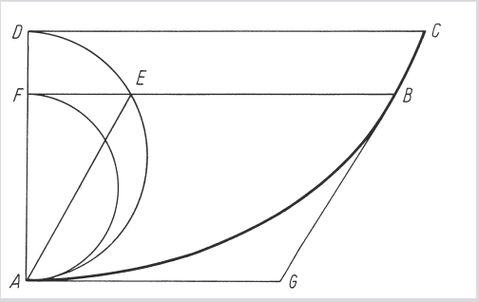
In all this, we have to make use of a property of the cycloid (Figure 3.104), namely, that the tangent BG at an arbitrary point B of this curve is parallel to the line AE, which connects E, the point of intersection of the generating circle and the line BF parallel to the directrix DC, to the base point A of the cycloid: $BG \parallel AE$. In his Theorem XXIV, part of a series of logically linked theorems, Huygens proves (by dividing the cycloid into elementary inclined planes) that the time T_1 required for transiting the cycloidal curve BA is related to the time T_2 during which the body traverses the line BG with velocity $v_0/2$ (where v_0 is the velocity that it would achieve from sliding along the line BG) as the circumference of the semicircle of diameter FA is related to the line segment FA. However, T_2 is nothing other than the time required to transit the segment BG (and simultaneously to transit the inclined plane AE). But Galileo had already demonstrated (Figure 3.40) that the inclined plane AE is transited in the same time as that required for free fall along the diameter AD. Thus Huygens arrives at his Theorem XXV:

> *If a body slides along an upwardly concave cycloid with vertical axis, then it always takes the same amount of time to reach the lowest point, regardless of the point at which the motion begins. This time is related to the time of freefall along the cycloidal axis as half the circumference of a circle to its diameter.*

In today's notation, we obtain from the formula $t = \sqrt{2s/g}$ for the time of free fall from height $2R$ the value $T_2 = \sqrt{4R/g}$. Thus, the above theorem yields $T_1 / T_2 = \pi/2$; hence, $T_1 = \frac{\pi}{2}\sqrt{4R/g}$. From this, we obtain $T_0 = 4T_1 = 2\pi\sqrt{4R/g}$ for the period of the cycloidal pendulum. If we now consider that the radius of curvature of the cycloid at the base point is equal to $4R$, we obtain, to good approximation, for the period of the thread pendulum of length $l = 4R$, the formula $T = 2\pi\sqrt{l/g}$.



▲ **Figure 3.103** If the point F determined by the equation $BF = (BE/TE)k$ lies on a parabola, then we have $BF \approx \sqrt{z}$, and from $BE/TE = \sin\alpha$, we see that indeed $\sqrt{z} = k \sin\alpha$. for points on the desired curve.

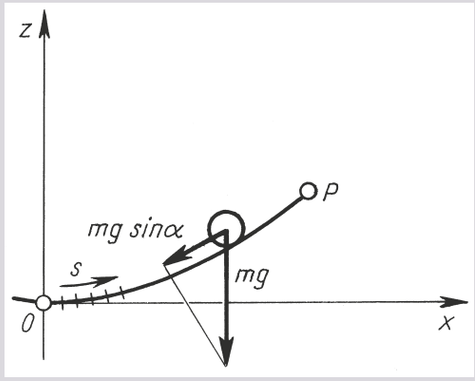


▲ **Figure 3.104** Drawing accompanying Theorem XXV in the *Horologium Oscillatorium*. This is the theorem in which Huygens arrives at the formula for the period of the cycloidal pendulum.

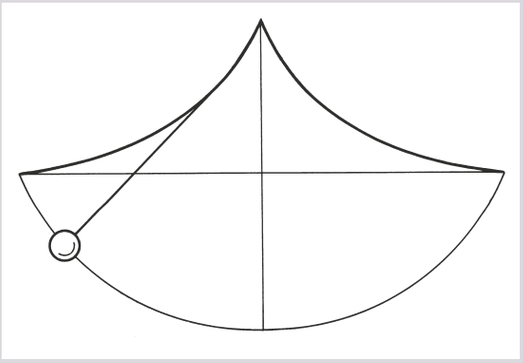
Quotation 3.34

For although, from a moral point of view, it may be a [good and] pious thought to believe that God created all things for us, since this may move us all the more to love Him and to give thanks to Him for so many blessings; and it is true in a sense, since there is no created thing which we cannot put to some use, even if it is only a matter of exercising our minds by contemplating it and, by means of it, being moved to praise God: it is, however, in no way likely that all things were made for us in the sense that God had no other purpose in creating them. And [it seems to me that] it would be clearly ridiculous to attempt to use such an opinion to support reasonings about Physics; for we cannot doubt that there are many things which are currently in the world, or which were formerly here and have already ceased to exist, which no man has even seen or known or used.

—Descartes, *Principles of Philosophy*, III.3 [p. 85]



▲ **Figure 3.105** To describe the motion along a cycloidal path, we choose as variable the (signed) arc length measured from the point 0.



▲ **Figure 3.106** A body suspended from a thread will describe a cycloidal arc if the thread is forced to follow guide plates of cycloidal shape.

Quotation 3.35

... He must therefore have borrowed his principle of refraction from Snell ... but does not mention, as is generally his way, this name.

—Isaac Vossius, *De Lucis Natura et Proprietate*, 1662

Quotation 3.36

M. Descartes has never demonstrated his principle. For apart from the fact that comparisons are not of much use as foundations for demonstrations, he uses his own in the wrong way and even supposes that the passage of light is easier in dense bodies than in rare ones, which is clearly false.

—Pierre de Fermat, Letter to de la Chambre, January 1, 1662 [Dugas 1957, p. 254]

In today's textbooks of theoretical physics, this problem is handled as follows (Figure 3.105): To write down the equation of motion, the component of force parallel to the tangent to the curve of motion must be known. As we may read off from the figure, this is equal to $-mg\sin\alpha$, so that we obtain the equation of motion

$$m\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -mg\sin\alpha.$$

From the parametric representation of the cycloid

$$x = R\omega t + R\sin\omega t,$$
$$z = -R(1+\cos\omega t),$$

we use

$$s = \int_0^P \mathrm{d}s = \int_0^t \frac{\mathrm{d}s}{\mathrm{d}t}\mathrm{d}t$$

and

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2} = \sqrt{(R\omega)^2(1+\cos\omega t)^2 + (R\omega)^2\sin^2\omega t} = R\omega\sqrt{2(1+\cos\omega t)} = 2R\omega\cos\frac{\omega t}{2}$$

to obtain its arc length in the form

$$s = \int_0^t 2R\omega\cos\frac{\omega t}{2}\mathrm{d}t = 4R\sin\frac{\omega t}{2}.$$

If we now set

$$\sin\alpha = \frac{\tan\alpha}{\sqrt{1+\tan^2\alpha}} = \sqrt{\frac{z'^2}{1+z'^2}} = \sqrt{\frac{\sin^2\omega t}{2(1+\cos\omega t)}} = \sqrt{\frac{1-\cos^2\omega t}{2(1+\cos\omega t)}} = \sqrt{\frac{1-\cos\omega t}{2}} = \sin\frac{\omega t}{2},$$

it then follows that

$$s = 4R\sin\frac{\omega t}{2}, \quad s = 4R\sin\alpha, \quad \sin\alpha = \frac{s}{4R}, \quad \sin\alpha = \sin\frac{\omega t}{2}.$$

The equation of motion can then be brought into the form

$$\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -g\frac{s}{4R}, \quad \text{or} \quad \frac{\mathrm{d}^2 s}{\mathrm{d}t^2} + \frac{g}{4R}s = 0,$$

which has the solution

$$s = s_0\sin\sqrt{\frac{g}{4R}}t, \quad \omega = \frac{2\pi}{T} = \sqrt{\frac{g}{4R}}, \quad T = 2\pi\sqrt{\frac{4R}{g}}.$$

The body thus executes, for arbitrary maximal amplitude s_0, a harmonic oscillation with constant angular frequency, that is, with constant period.

If the amplitude-independent period is actually to be realized with a thread pendulum, then the question arises as to how the pendulum bob attached to a thread can be made to execute a cycloidal path. Huygens solved this problem by arranging, as shown in Figure 3.106, curved guide plates in such a way that the thread presses against them during the pendular motion. At first, he attempted to realize the pendulum experimentally with an amplitude-independent period by changing the form of the guide plates. From the ideas described above, Huygens then realized that the pendulum bob must be guided along a cycloidal path, and we have him to thank for the further significant observation that the pendulum bob describes a cycloidal path

precisely when the string presses against a cycloid (Figure 3.107). This fact follows from the property of the cycloid that its *involute* is again a cycloid. (An involute is a curve traced by a point on a taut string as it unwinds from another curve; the original curve is called the *evolute* of its involute.) The cycloid along which the pendulum bob moves is the *involute* of the cycloid of the guide curves.

Huygens did not have to discover all the properties of the cycloid mentioned here on his own because at the time this curve was already the focus of mathematical interest. However, the discovery of the connection between the cycloid's evolute and involute is Huygens's own.

3.6.4 *The Physical Pendulum*

Another of Huygens's outstanding accomplishments is the determination of the period of the physical pendulum, or, more precisely, the determination of the length of a mathematical pendulum with the same period as the physical pendulum. Let us construct a physical pendulum with three bobs, as shown in Figure 3.108, joined by a rigid, massless rod. If the pendulum is moved from its rest position, then it oscillates about the point O. For our present deliberations, let us start with Huygens's two principles and investigate the state of the pendulum at an arbitrary point in time. We imagine freeing the bobs from their rigid linkage and redirecting their motion vertically upward, which could be accomplished through a collision with a suitably oriented plane. If we then fix each bob at the maximum height that it reaches, the center of gravity of the system comprising the three free bobs cannot be higher than it was in its original position. But it also cannot be lower because we could then consider the new position as the starting position, and due to the reversibility of the motion, the bobs would arrive at a final position corresponding to the original starting position. In the new final position, the height of the center of gravity would be greater than in the original position, which is impossible. It follows then, that at any time, the center of gravity in such an imaginary arrangement must be at exactly its initial height. Later on, we shall investigate the connections between this result and the law of conservation of energy.

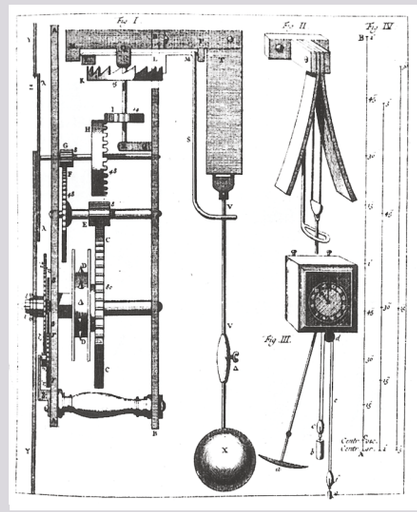
Now we shall determine the length of the mathematical pendulum that oscillates with the same frequency as a given physical pendulum following Huygens's original exposition. Here Huygens is working completely in the spirit of Archimedes, and so he begins with the final result. He specifies the length of the equivalent mathematical pendulum and proves that the assumption of a lack of agreement between the periods of the two pendulums leads to a contradiction. Figure 3.109 shows the physical pendulum and the mathematical pendulum to which it is equivalent with regard to its period. Huygens now states that an agreement of the periods can be achieved if the length of the mathematical pendulum satisfies the condition

$$l = \frac{m_1 r_1^2 + m_2 r_2^2 + m_3 r_3^2}{\left(m_1 + m_2 + m_3\right) r_{\mathrm{cg}}}. \qquad (2)$$

The meanings of the quantities in the formula can be read from the figure, where the position of the bob of an equivalent mathematical pendulum is also shown.

The two pendulums oscillate identically if, for an arbitrary angle, the velocity v_m of the mathematical pendulum agrees with the velocity v_{ph} of the corresponding point of the physical pendulum.

We now assume that the two velocities differ and first consider the case that the velocity of the bob on the mathematical pendulum is less than that of the corresponding point on the physical pendulum.



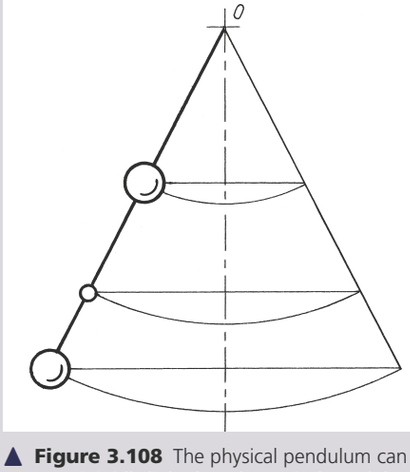
▲ **Figure 3.107** Huygens's cycloidal pendulum clock.

Quotation 3.37

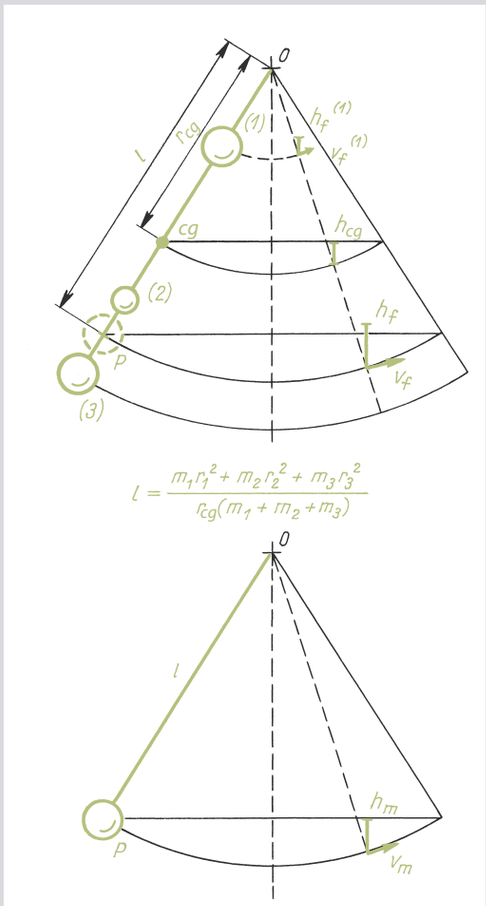
... the principle, so common and so well-established, that Nature always acts in the shortest ways. ...

... in order that my conscience might be satisfied, it was not sufficient to have found a point such as F through which the natural motion was accomplished more quickly, more easily and in less time than by the straight line P_1P_2. It was also necessary to find the point which allowed the passage from one side to the other in less time than any other there might be. In this connection, it was necessary to use my method of *maximis* and *minimis*, which is rather successful for expediting this kind of problem.

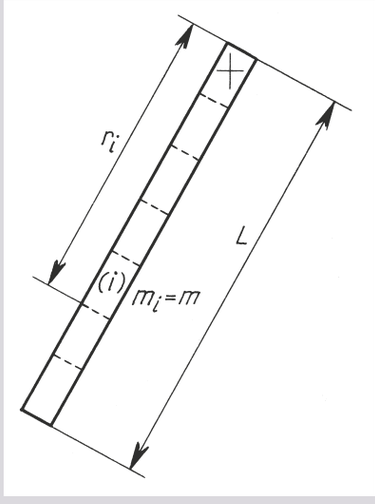
—Pierre de Fermat, Letter to de la Chambre, January 1, 1662 [Dugas 1957, pp. 254–255]



▲ **Figure 3.108** The physical pendulum can be viewed as a system of rigidly linked mathematical pendulums of different lengths with common suspension point. Huygens answered the question of the period using his two hypotheses.



▲ **Figure 3.109** To determine the period of the physical pendulum, HUYGENS sought the length of a mathematical pendulum with identical period.



▲ **Figure 3.110** Determination of the equivalent length of a homogeneous rod. The method used is a typical example of how HUYGENS proceeded without using integral calculus.

That is,

$$v_{ph} > v_m.$$

But from this, it follows that the fictive height of the center of gravity associated with the velocity of the physical pendulum is greater than that of the mathematical pendulum, that is,

$$h_{ph} > h_m.$$

We can also read from the figure that the velocity of the point mass labeled (1) stands in relation to the velocity v_{ph} as the distance of the mass (1) from the pivot point is to the distance of the point P:

$$\frac{v_{ph}^{(1)}}{v_{ph}} > \frac{r_1}{l}.$$

The associated fictive heights are proportional to the squares of the velocities:

$$\frac{h_{ph}^{(1)}}{h_{ph}} = \left(\frac{v_{ph}^{(1)}}{v_{ph}}\right)^2 = \left(\frac{r_1}{l}\right)^2.$$

From this, we obtain the relationship

$$h_{ph}^{(1)} = \frac{r_1^2}{l^2} h_{ph}.$$

Finally, from the assumption $h_{ph} > h_m$, we have

$$h_{ph}^{(1)} > \frac{r_1^2}{l^2} h_m,$$

and analogous reasoning yields, for masses (2) and (3), the following:

$$h_{ph}^{(2)} > \frac{r_2^2}{l^2} h_m, \quad h_{ph}^{(3)} > \frac{r_3^2}{l^2} h_m.$$

If we now multiply the last three inequalities successively by m_1, m_2, m_3, then we obtain, on adding the results,

$$m_1 h_{ph}^{(1)} + m_2 h_{ph}^{(2)} + m_3 h_{ph}^{(3)} > (m_1 + m_2 + m_3) h_{cg},$$

and finally,

$$\frac{m_1 h_{ph}^{(1)} + m_2 h_{ph}^{(2)} + m_3 h_{ph}^{(3)}}{m_1 + m_2 + m_3} > h_{cg}.$$

On the left-hand side of this relationship is the height of the center of gravity for the three masses for an arbitrary position of the pendulum. It follows from the inequality that this height is greater than the height of the center of gravity of the pendulum bob in its initial state, which is impossible because of the first principle.

Similarly one can prove—by taking into account the reversibility of motion—that the condition $v_{ph} > v_m$ also leads to an absurd result, so that, in fact, the length of the mathematical pendulum whose period agrees with that of the physical pendulum under investigation is correctly given by equation (2).

This holds as well, of course, for an arbitrary number of masses joined by a rigid rod.

In the general case, the length of the equivalent mathematical pendulum is given by

$$l = \frac{\sum_i m_i r_i^2}{\sum_i m_i r_i} = \frac{\sum_i m_i r_i^2}{r_{cg} \sum_i m_i}.$$

It is worth noting how HUYGENS determined the equivalent pendulum length for a homogeneous rod without the use of integral calculus. As shown in Figure 3.110, we divide the rod of length L into n pieces,

so that the distance of the ith piece from the axis of rotation is given by

$$r_i = \frac{i}{n} L.$$

In the case of a homogeneous rod, each section has the same mass, so that the mass can be factored out of the summation. Thus we have

$$l = \frac{m\sum_i \left(\frac{i}{n}L\right)^2}{m\sum_i \frac{i}{n}L} = \frac{\left(\frac{L}{n}\right)^2 \sum_{i=1}^{n} i^2}{\frac{L}{n}\sum_{i=1}^{n} i}.$$

If we now take into account that

$$\sum_{i=1}^{n} i^2 = \left(\frac{1}{3} + n\frac{2}{3}\right)(1 + 2 + \cdots + n)$$

and that for very large n the approximations

$$\frac{1}{n^3}\sum_{i=1}^{n} i^2 \to \frac{1}{3},$$
$$\frac{1}{n^2}\sum_{i=1}^{n} i \to \frac{1}{2}$$

hold, then for the equivalent pendulum length we get the following formula:

$$l = \frac{2}{3} L.$$

We should point out the fact that the moment of inertia and its mathematical representation appear here for the first time in the history of physics. However, Huygens did not yet give a particular name to the quantity $\Sigma_i m_i r_i^2$.

Huygens established all the theorems that we meet today in theoretical mechanics in connection with the physical pendulum; for example, he determined that the pivot point O and the point P are reciprocal, or, more simply put, interchangeable.

Of greater importance from the point of view of the history of science is the fact that with his theorem on the constancy of the height of the virtual center of gravity, Huygens in fact formulated the law of conservation of energy. In today's usual notation, Huygens's theorem can be formulated as follows: Let us denote by h_i the height, and by v_i the velocity, of the center of gravity of mass i at an arbitrary time. Then the virtual height associated with this velocity is given by

$$\frac{v_i^2}{2g},$$

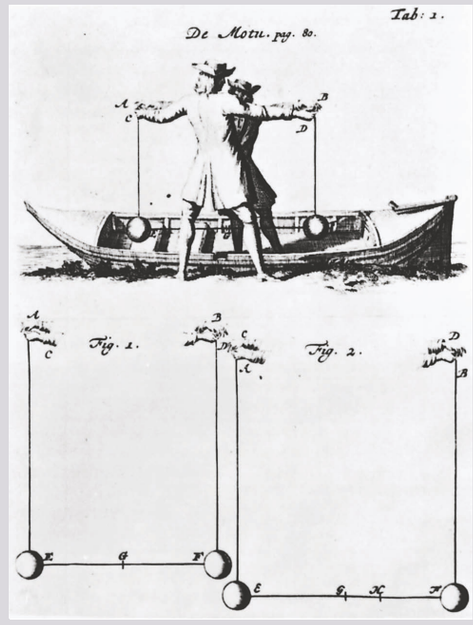
and the (total) height of the center of gravity at an arbitrary time is given by

$$\frac{\sum_i m_i g\left(h_i + \frac{v_i^2}{2g}\right)}{\sum_i m_i g} = \frac{\sum_i m_i g h_i + \sum_i \frac{1}{2} m_i v_i^2}{\sum_i m_i g}.$$

By Huygens's theorem, this quantity remains constant throughout the motion, and since the denominator is constant, so must be the numerator, that is,

$$\sum_i m_i g h_i + \sum_i \frac{1}{2} m_i v_i^2 = \text{constant}.$$

This is precisely the equation expressing the conservation of energy: *the sum of kinetic and potential energies is constant during the course of motion*. However, we stress here once again that Huygens did not write down the relation in this form.

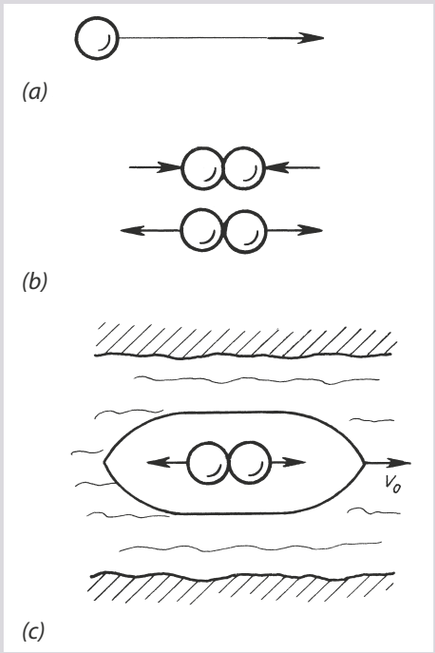


▲ **Figure 3.111** Huygens does not explain why the laws of mechanics should be the same on a boat as on the shore, but he postulates it and derives laws from the requirement of this invariance. In the theory of relativity, one can similarly derive the law of the dependence of mass on velocity from the equivalence of the laws of collision in two frames of reference moving relative to each other.

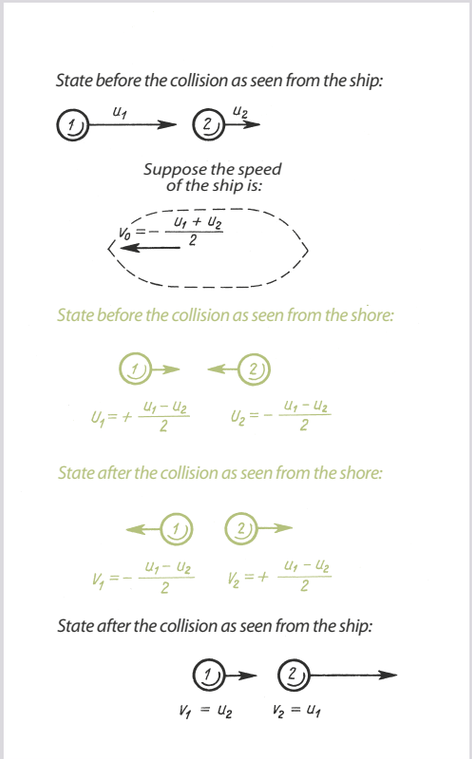
Quotation 3.38

The reward of my work has been [the] most extraordinary, most unexpected, and most fortunate that I have ever obtained. For after having gone through all the equations, multiplications, antitheses and other operations of my method, and finally having settled the problem ... I found that my principle gave exactly the same proposition for the refractions that M. Descartes has established. I was so surprised by a happening that was so little expected that I only recovered from my astonishment with difficulty. I repeated my algebraic operations several times and the result was always the same, though my demonstration supposes that the passage of light through dense bodies is more difficult than through rare ones—something I believe to be very true and necessary, and something which M. Descartes believes to be the contrary.

—Pierre de Fermat, Letter to de la Chambre, January 1, 1662 [Dugas 1957, p. 257]



▲ **Figure 3.112** The initial assumptions that led to the laws of (elastic) collision.



▲ **Figure 3.113** HUYGENS's derivation of the laws for elastic collisions by bodies of equal mass using the equivalence of frames of reference moving linearly with constant velocity relative to each other.

3.6.5 *The Collision Laws as Consequences of the Equivalence of Inertial Systems*

In considering laws of collision, HUYGENS attacked a problem that lay directly on the path to NEWTON's fundamental laws (Figure 3.111). HUYGENS began with the following three basic assumptions:

1. A body in motion strives to maintain its motion along a straight line with constant velocity until it encounters some obstacle (Figure 3.112(a)).
2. If two identical balls with equal but oppositely directed velocities collide, then after the impact, their directions of motion are reversed without the magnitude of their velocities being altered (Figure 3.112(b)).
3. The laws of collision are the same for an observer on a boat that is moving with constant velocity as for an observer on shore (Figure 3.112(c)).

> *Here we have the first explicit formulation of the relativity principle of classical mechanics for coordinate systems that are moving at constant velocity in relation to each other. Moreover, this principle can be used to derive quantitative laws (Figure 3.113).*

With the help of these three axioms, HUYGENS was now able to determine, for an elastic collision of balls of equal mass, the velocities after impact for *arbitrary* initial velocities. To do so, he simply chose the velocity of the boat such that the second axiom could be employed. For example, suppose that the velocities of the two balls before the collision are, with respect to the boat u_1 and u_2, and after the collision they are v_1 and v_2. From the shore, the corresponding velocities are observed to be U_1 and U_2, and V_1 and V_2. Let us now choose the boat's velocity to be

$$v_0 = -\frac{u_1 + u_2}{2},$$

so that an observer on shore observes the velocities before the collision to be

$$U_1 = u_1 - \frac{u_1 + u_2}{2} = \frac{u_1 - u_2}{2},$$
$$U_2 = u_2 - \frac{u_1 + u_2}{2} = \frac{u_2 - u_1}{2}.$$

Since we have $|U_1| = |U_2|$, the observer on shore can use the third axiom, and the velocities after impact are therefore

$$V_1 = \frac{u_2 - u_1}{2}, \quad V_2 = \frac{u_1 - u_2}{2}.$$

From these equations, we may conclude that the velocities after impact as measured on the boat are

$$v_1 = V_1 + \frac{u_1 + u_2}{2} = \frac{u_2 - u_1}{2} + \frac{u_1 + u_2}{2} = u_2$$

and

$$v_2 = V_2 + \frac{u_1 + u_2}{2} = \frac{u_1 - u_2}{2} + \frac{u_1 + u_2}{2} = u_1;$$

thus, after impact, the balls exchange velocities.

To deal with collisions involving bodies of unequal masses, Huygens introduced an additional axiom, by which he in fact defined the concept of elastic collision: If after a collision of two bodies, the absolute value of the velocity of one body remains the same, then the absolute value of the velocity of the second body is also unchanged.

This axiom can be formulated as follows. If

$$v_1 = -u_1,$$

then

$$(v_1 - v_2) = -(u_1 - u_2),$$

from which it follows immediately that if we have

$$v_1 = -u_1 + x,$$

then the velocity of the second body must be

$$v_2 = -u_2 + x.$$

Huygens combined the laws of collision with his successfully tested principle on the center of gravity. He regarded the collision as if the two balls colliding on a horizontal plane had acquired their initial velocities by virtue of having rolled down inclined planes of a given height corresponding to their velocities. These heights, for the two bodies, are given by

$$h_1^{\mathrm{b}} = \frac{u_1^2}{2g}, \quad h_2^{\mathrm{b}} = \frac{u_2^2}{2g},$$

and we now define the height h_s^{b} of the center of gravity of the two bodies before impact by

$$h_1^{\mathrm{b}} m_1 + h_2^{\mathrm{b}} m_2 = (m_1 + m_2) h_{\mathrm{cg}}^{\mathrm{b}}.$$

The superscript b indicates a state before the collision; let the superscript a indicate the state after impact. After the collision, we allow the two balls to run up inclined planes, with the heights reached given by

$$h_1^{\mathrm{a}} = \frac{v_1^2}{2g}, \quad h_2^{\mathrm{a}} = \frac{v_2^2}{2g}.$$

Then the height of the center of gravity of the two bodies after the collision will be given by

$$h_1^{\mathrm{a}} m_1 + h_2^{\mathrm{a}} m_2 = (m_1 + m_2) h_{\mathrm{cg}}^{\mathrm{a}},$$

and since this must be the same as in the initial state, we have

$$h_1^{\mathrm{b}} m_1 + h_2^{\mathrm{b}} m_2 = h_1^{\mathrm{a}} m_1 + h_2^{\mathrm{a}} m_2,$$

or

$$m_1 u_1^2 + m_2 u_2^2 = m_1 v_1^2 + m_2 v_2^2.$$

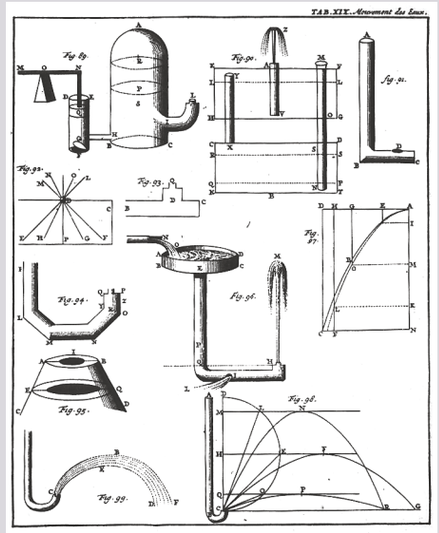
Today, we would interpret this relationship as the conservation law for kinetic energy in an elastic collision. If here we set $v_1 = -u_1 + x$ and $v_2 = -u_2 + x$, then we

Quotation 3.39

[Fermat's principle is] a principle which is moral and in no way physical; which is not, and which cannot be, the cause of any effect of Nature. ... That path, which you reckon the shortest because it is the quickest, is only a path of error and bewilderment, which Nature in no way follows and cannot intend to follow.

—Claude Clerselier, Letter to Fermat, May 6, 1662 [Dugas 1957, p. 258]


OEUVRES
DE
M^R. MARIOTTE,
de l'Académie Royale des Sciences;
DIVISÉES EN DEUX TOMES,
Comprenant tous les Traitez de cet Auteur, tant ceux qui avoient déja paru ſéparément, que ceux qui n'avoient pas encore été publiez;
Imprimées ſur les Exemplaires les plus exacts & les plus complets,
Revuës & corrigées de nouveau.
TOME PREMIER.
A LEIDE,
Chez PIERRE VANDER Aa,
Marchand Libraire, Imprimeur de l'Univerſité & de la Ville.
M D C C XVII.

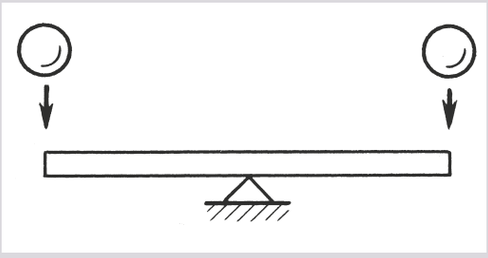


▲ **Figure 3.114** Title page and typical figures from Mariotte's book. (Library of the University for Heavy Industry, Miskolc.)

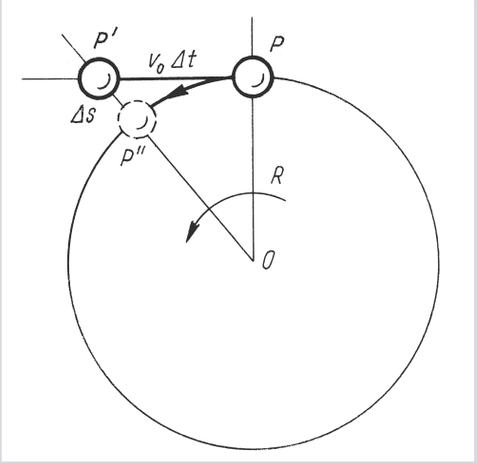
Quotation 3.40

I have often said to M. DE LA CHAMBRE and yourself that I do not claim and that I have never claimed, to be in private confidence of Nature. She has obscure and hidden ways that I have never had the initiative to penetrate; I have merely offered her a small geometrical assistance in the matter of refraction, supposing that she has need of it. But since you, Sir, assure me that she can conduct her affairs without this, and that she is satisfied with the order that M. DESCARTES has prescribed for her, I willingly relinquish my pretended conquest of physics and shall be content if you will leave me with a geometrical problem, quite pure and *in abstracto*, by means of which there can be found the path of a particle which travels through two different media and seeks to accomplish its motion as quickly as it can.

—PIERRE DE FERMAT, Letter to CLERSELIER, May 21, 1662 [Dugas 1957, p. 259]



▲ **Figure 3.115** MARIOTTE also investigated the circumstances under which a balance remains in equilibrium after the impact of bodies.



▲ **Figure 3.116** HUYGENS determined acceleration under uniform circular motion in this way.

obtain an equation that allows us to determine the unknown quantity x:

$$m_1u_1^2 + m_2u_2^2 = m_1u_1^2 + m_2u_2^2 - 2\left(m_1u_1 + m_2u_2\right)x + \left(m_1 + m_2\right)x^2,$$

or

$$x = \frac{2\left(m_1u_1 + m_2u_2\right)}{m_1 + m_2}.$$

Other researchers besides HUYGENS were at work on the problem of collision processes at the same time. Of these, MARIOTTE's work is worth mentioning. While his quantitative results did not go beyond the work of HUYGENS, he placed great emphasis on experimentation. Using a simple apparatus consisting of two balls hanging from strings, MARIOTTE (Figure 3.114) verified and corroborated the collision laws, including results for the heights that appeared in HUYGENS's work simply as virtual quantities. We point out in particular two of his achievements: He correctly recognized the role played by mass in collisions (Quotation 3.44), and he attempted to determine the force that results when a substance streaming at a given velocity meets an obstacle (such as a stream of water against the pan of a balance). Moreover, he investigated the conditions under which a balance remains in equilibrium when bodies with differing momenta strike the two pans of a balance (Figure 3.115).

3.6.6 *Circular Motion*

Finally, we offer a few words on HUYGENS's work on circular motion. Imagine that we are sitting on the rim of a horizontal wheel rotating on a vertical axis. If we hold an object of a certain mass by an attached string, then we can observe the effect of a force that attempts to move the object radially outward. HUYGENS gave this force the name *centrifugal force*. To determine its magnitude, he proposed the following thought experiment (Figure 3.116). Suppose that the object has horizontal velocity v_0 at the point P, and if it were not held by the string, it would maintain this velocity, arriving at the point P' in the small time interval Δt. The observer sitting on the wheel would observe the object moving radially outward with increasing velocity. The distance Δs of the object from the observer after time Δt can be approximately determined from the following formula:

$$R^2 + \left(v_0\Delta t\right)^2 = \left(R + \Delta s\right)^2,$$

$$\text{or} \quad \Delta s^2 + 2R\Delta s = \left(v_0\Delta t\right)^2.$$

The approximation is that the points P', P'', and O are all assumed to lie on a straight line. This is not exactly so, because in reality it is the length of the arc $P'P''$ that is equal to the length PP'. By further assuming that

$$\Delta s^2 \ll 2R\Delta s,$$

we get

$$\Delta s = \frac{1}{2}\frac{v_0^2}{R}\left(\Delta t\right)^2.$$

Comparing this formula with the equation $s = \frac{1}{2}at^2$ for motion with constant acceleration, we see that they are of the same general form, and that we also found a value for the acceleration:

$$a = \frac{v_0^2}{R}.$$

Newton also derived this relationship and was disappointed to learn that Huygens had already published the same results.

Huygens's train of thought is of great importance to the history of science because here it was established—contradicting the Peripatetics and even Galileo—not only that to maintain circular motion, a force is always required (which, by the way, Descartes already knew), but also that a numerical value could be calculated for this force. In this way, Huygens smoothed the way for a precise determination of acceleration in motion along curved paths.

As we have already mentioned, Huygens was no philosopher; his strength lay—as we have seen—in the establishment of simple, reasonable, but very productive fundamental physical principles. Nonetheless, Quotation 3.45, taken from the foreword to his book dedicated to problems of optics (*Traité de la lumière*), represents one of the most cogent formulations of the basic principles of natural philosophy.

3.7 Newton and the Principia: The Newtonian Worldview

3.7.1 *The Tasks Awaiting the Advent of Newton*

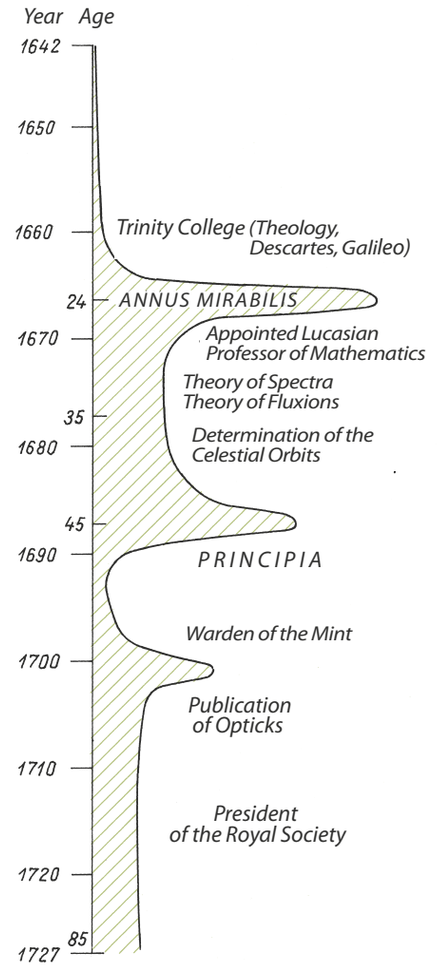
In the previous sections, we have sketched the path and followed the ideas leading to a new dynamics. Let us see—by summarizing the results of the first seven to eight decades of the seventeenth century—what was there for Newton to build on and what tasks were awaiting him.

We have spoken so far of three strands by which these ideas developed: free fall, collision, and circular motion.

- *Free fall.* The problem comprises the kinematics of bodies moving with constant acceleration, the proportionality of the distance traveled to the square of the time, and the surprising fact that every object—at least under ideal conditions—falls with the same acceleration. This fact greatly simplifies the kinematic description, but it also complicates its dynamic interpretation. Huygens's wide-ranging investigation into the problem of free fall did not bring us significantly closer to the goal, even though it was indirectly very useful because it showed that the right choice of an initial starting point—such as Huygens's principle on the center of gravity that was discussed in detail—can yield a broad multitude of concrete results.
- *Collision.* The momentum—that is, the product of the mass of the body and its velocity—as well as its change over time clearly play key roles.
- *Circular motion.* The important realization is that in order to maintain such motion, a force is necessary, contrary to the Peripatetic belief that circular motion can to some extent be seen as inertial, or naturally given, motion. More generally, it is precisely with circular motion, as the simplest form of curvilinear motion, that the vector nature of velocity and the change of velocity are the most evident and are also quantitatively accessible.

Behind all this lies a new law of inertia, recognized as final and irrevocable, according to which motion is a *state* and not a *process* and an effective cause is needed not to maintain it but to change it.

Finally, Descartes put his stamp on the worldview of physicists by requiring a unified explanation of celestial and terrestrial phenomena, and further requiring that this explanation be clearly formulated, meaning that interaction is possible only by immediately visible and perceptible contact.



▲ **Figure 3.117** Important events and creative periods in Newton's life. See also Plate XVII.

Isaac Newton (1642–1727): According to the Julian calendar then in use, born Christmas day in 1642; however, on the Continent, the new year had already begun. Newton's father had died several months before his son's birth. From 1661, Newton, with the support of his uncle, studied mathematics at Trinity College, Cambridge. During the plague epidemic in 1665, he withdrew to his estate in Woolsthorpe; this and the following year are known as Newton's "anni mirabiles" (Plate XVII). At only 24 years of age, Newton conceived the fundamental ideas for the binomial theorem, differential calculus, theory of color, centripetal force, laws of motion, and theory of gravitation. On his return to Cambridge, he dealt with problems of optics; in 1668, he completed a reflecting telescope. In 1669, he was appointed to the Lucasian Chair of Mathematics at the University of Cambridge, to succeed Isaac Barrow. In 1672, he presented his *Theory of Light and Colors* to the Royal Society; this book precipitated such dispute that he decided not to publish anything further. A summary of his work on optics was not published until

continued on next page

Figure 3.117 *continued*

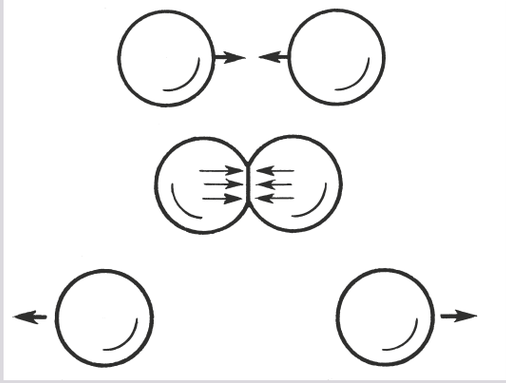
1704, in the book *Opticks*. In 1684, on the urging of Halley, he began to write the *Principia*; Halley then assumed the costs of publication. In the years 1692 and 1693, Newton suffered the consequences of a severe nervous breakdown; he recovered, retaining his full intellectual capacities, although he made no further significant scientific contributions in the remaining 35 years of his life. That he was capable of such is demonstrated by his solution in a single night of a problem posed by Bernoulli (1696), although the task was estimated to take six months, and the solution of a problem posed by Leibniz almost in the moment that he became aware of it (1716).

In 1699, Newton was named warden of the Royal Mint, and in 1705, he was knighted by the queen. From 1703 until his death in 1727, he was president of the Royal Society. He is buried in Westminster Abbey.

Quotation 3.41

It is often said in explanation of the ... force which prevents the quicksilver from falling down, as it would naturally do, [that it] is internal to the vessel, arising either from the vacuum or from some exceedingly rarefied substance; but I assert that it is external and that the force comes from without. On the surface of the liquid which is in the bowl there rests the weight of a height of fifty miles of air; then what wonder is it if into the vessel, in which the quicksilver has no inclination and no repugnance, not even the slightest, to being there, it should enter and should rise in a column high enough to make equilibrium with the weight of the external air which forces it up?

—Evangelista Torricelli, Letter to Michelangelo Ricci, June 11, 1644



▲ **Figure 3.118** The phases of a symmetric elastic collision.

Toward the end of the century, a fourth strand waiting to be spun together with the above three, the problem of planetary motion, moved to the forefront. In this manner Kepler's laws could finally get the recognition they merited.

What was awaiting Newton was the task of uniting these more-or-less independent problems into a single worldview (Figure 3.117, Plate XVII).

In the end, the unified development of mechanics and indeed of the entire physical worldview rest on two fundamental realizations made by Newton. The first is the law of motion, which established a quantitative relationship between the change of a state of motion and the underlying force, that is, the recognition of the relationship

$$\text{force} = \text{mass} \times \text{acceleration}.$$

The second is the universal law of gravitation, according to which the attractive gravitational force between any two bodies is proportional to the product of their masses and is inversely proportional to the square of the distance between them.

The first law can provide the force if we know the motion, or the motion if we know the force, reducing all the previous problems of this form in the history of science to special cases; Newton himself provided posterity with an almost inexhaustible supply of new applications. The second law guarantees the unity of the celestial and terrestrial worlds because the path of a stone falling from a tower and the path traced by the Moon or any planet could now be calculated according to the same law.

In the following sections, we attempt to present as simply as possible, using today's common terminology, the train of thought that led Newton to these laws. In his *Principia*, the bible of classical physics, Newton's ideas and results already appear in their final form and in great generality wrapped, as it were, in the ceremonial vestments inherited from Euclid and Archimedes of theorems followed by proofs. We may learn about the difficult process of the birth of his ideas partly from Newton himself and partly from his contemporaries. We can especially rejoice in the fact that Newton had the habit of noting his yet unripe thoughts and calculations in a journal. This journal, which Newton called his *waste book*, is, with its collection of fragments of ideas, perhaps the most important surviving material in Newton's own hand. In the modern reappraisal of these papers, many exciting pieces of information for the history of science have come to light; however, due to their large number, we can only occasionally touch on these.

3.7.2 *A Force Is Not Required to Maintain a State of Motion but to Change It*

Let us now consider the steps that led to the formulation of the law of motion. We have frequently spoken of the importance of collision processes for the discovery of this law. Let us consider the simplest of such processes, the head-on collision of two elastic balls of equal masses moving at equal speeds. The result of the collision—both balls ricocheting off each other with reversed directions and the same speeds—is so obvious that Huygens, as we have seen, used it as an axiom in his treatment of more general collisions.

Let us consider in a bit more detail what happens physically during the course of a collision, irrespectively of how quickly it takes place. As depicted in Figure 3.118, the two balls are elastically deformed, and therefore each pushes against the other. As a result of this force, each of the balls is slowed down, and there is a moment at which the velocities of the two balls are zero. Newton's first significant insight in

connection with this is that a force is required to bring a moving object to a standstill (that is, to reduce a momentum to zero), and this force is due to the pressure from the elastic deformation of the balls.

In the next phase of the collision, as a result of the forces exerted by the deformed balls, both balls recover their original velocities, but in the opposite directions. Newton's second observation is that the force needed to create a particular motion is the same as what is required to make that motion cease.

Newton's third observation is that during their interaction, each body affects the other with an equal but opposite force. A generalization of this fact later led to Newton's third axiom.

The collision, which lasts a very short time, is in fact a transition between two inertial states of the bodies, that is before and after the collision, they execute linear motion at constant velocity in the absence of any outside influence. The result of the collision can therefore be described quantitatively because uniform linear motion can be characterized by well-defined state parameters, namely, momentum and velocity (Figure 3.119).

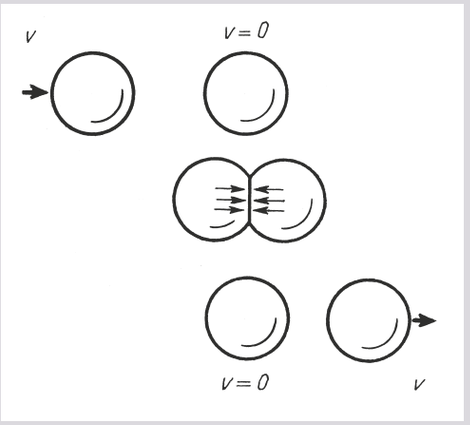
The force acting on a body in motion along a curved path can be easily understood if we reduce the effect to a series of collisions. Suppose a body is moving, as shown in Figure 3.120, along the linear path *AB* with constant velocity when it collides at point *P* with another body in such a way that, at impact, a momentum of magnitude mv_p is transmitted to the body in the direction *PC*. The body then continues its path in the direction *PP′*, and suppose that at point *P′* it collides with another body so that its direction changes once again. The result of this sequence of collisions is that the body travels along a polygonal path.

A very important step toward generalization is if we assume that an abrupt change in the state of motion does not have to be the consequence of a collision, in other words through direct contact, but that it could also be caused by any type of force. Even gravitational force can be regarded from this point of view as a consequence of brief impulses, as had already been done by Beeckman (see Section 3.3).

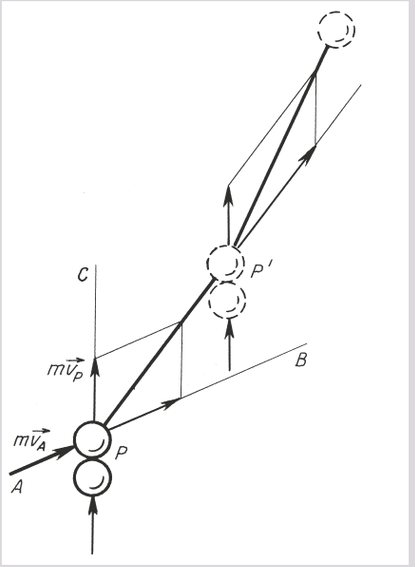
Here we call attention to a peculiarity in the terminology. In keeping with Newton, here we use the notion of "force" differently from the way it is used today. According to our current nomenclature, the quantity that results in a given change in momentum is not called a force; today, force means the *change* in momentum *per unit of time*. A given change in momentum is equal to the product of the force and the time over which the force acts, or more precisely, equal to the time integral over the force.

If we consider again what happens during the second phase of the elastic collision depicted in Figures 3.118 and 3.119. We see that, for example, the ball on the right is accelerated from its resting position due to the effective "force" of the pressure in the direction of that pressure, and as a result it acquires a certain momentum. This momentum is therefore proportional to the "force" and has the same direction. What happens now at point *P* of Figure 3.120? If no physical effects were acting here on the bodies, the ball would continue moving along the line *AB*. Keeping in mind the principle of "independence of motion" formulated by Galileo, we see that the motion in the direction *PC* combines with the original motion, and so we may formulate the following theorem on motion resulting from a collision: *The change in momentum is proportional to and has the same direction as the effective "force."*

We stress once more that here we have been using the term "force" in Newton's sense; in today's language we would substitute the product of force and time.



▲ **Figure 3.119** In this particular collision, the ball on the left will come to a stop, while the equal but opposite effect will accelerate the other ball to its original velocity.

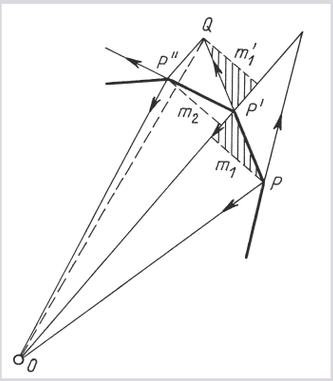


▲ **Figure 3.120** We can force a body to traverse a polygonal path through a series of collisions. In passing to the limit, we obtain motion along a curved path.

Quotation 3.42

Let all the disciples of Aristotle gather together all the strength in the writings of their master and his commentators in order, if they can, to make these things reasonable by means of the horror of the vacuum. Except that they know that experiments are the true masters that must be followed in physics. And that what has been accomplished in the mountains reverses the common belief of the world that Nature abhors a vacuum; it has also established the knowledge—which will never die—that Nature has no horror of a vacuum, and that the heaviness of the mass of air is the true cause of all the effects which have previously been attributed to this imaginary cause.

—Blaise Pascal [Dugas 1957, p.171]



◀ **Figure 3.121** For central forces, the area law holds: The surfaces of the triangles *OPP'* and *OP'P''* are equal. Since the time intervals are equal, we have *PP'* = *P'Q*, and therefore $m_1 = m_1'$ and *OP'* || *QP''*, so that $m_1 = m_2$ and $OP'm_1 = OP'm_2$.

[37]

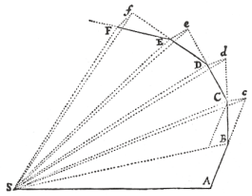
SECT. II.

De Inventione Virium Centripetarum.

Prop. I. Theorema. I.

Areas quas corpora in gyros acta radiis ad immobile centrum virium ductis describunt, & in planis immobilibus consistere, & esse temporibus proportionales.

Dividatur tempus in partes æquales, & prima temporis parte describat corpus vi insita rectam *AB*. Idem secunda temporis parte, si nil impediret, recta pergeret ad *c*, (per Leg. I) describens lineam *Bc* æqualem ipsi *AB*, adeo ut radiis *AS*, *BS*, *cS* ad centrum actis, confectæ forent æquales areæ *ASB*, *BSc*. Verum ubi corpus venit ad B, agat vis centripeta impulsu unico sed magno, faciatq; corpus a recta *Bc* deflectere & pergere in recta *BC*. Ipsi *BS* parallela agatur *cC* occurrens *BC* in *C*, & completa secunda temporis parte, corpus (per Legum Corol. 1) reperietur in *C*, in eodem plano cum triangulo *ASB*. Junge *SC*, & triangulum *SBC*, ob parallelas *SB*, *Cc*, æquale erit triangulo *SBc*, atq; adeo etiam triangulo *SAB*. Simili argumento si vis



▲ **Figure 3.122** The area law in Newton's *Principia*.

Quotation 3.43

Mr. Descartes had discovered how to make his conjectures and fictions be taken for truths. And something similar happened to those who read his *Principles of Philosophy* as to those who read the Romans who are pleasing and give the impression of being true histories. The novelty of the figures and little particles and vortices make it very charming. It seemed to me, when I read this book of *Principles* for the first time, that everything went swimmingly, and I thought, whenever I had some difficulty with it, that it was my fault for not clearly understanding his thought. I was only 15 or 16 years old. But having since then discovered things in it that are manifestly false, and others that are very unlikely, I have thoroughly returned from my former obsession, and at the present time I can find almost nothing that I can approve as true in his entire physics, or metaphysics, or meteorology.

—Christiaan Huygens, *Euvres Complètes*, vol. 10 [translated by Paul Franz]

As we have already explained, the forces acting on the vertices of the polygonal path can be of a quite general nature. For example, it can be a *central force*, directed toward one fixed point, the center, from any point in space. The importance of investigating this configuration is clear: Planetary motion involves precisely such a force. We assume now, following Figures 3.121 and 3.122, that the force directed toward a central point acts in the form of a rapid succession of impulses and that the path consequently has the form of a polygonal line. It is easy to see from the figure that the areas of the triangles that lie between the segments of the path and the center are equal; for example, the area of triangle *OPP'* is equal to the area of triangle *OP'P''*, since the two triangles share the common side *OP'*, and the corresponding altitudes m_1 and m_2 are equal. This observation leads to Kepler's second law—the law of areas—where we see that this law is valid for an arbitrary central force and that no additional assumptions about the dependence of the force on distance are necessary. In the following, we show that the dependence of the force on the distance follows from Kepler's third law.

Let us return now to the polygonal path that arises under the influence of brief impulses and attempt to derive the quantitative relationships for circular motion as a limiting case, where we shall assume as known the geometric and kinematic characteristics of this path. Following Figure 3.123, we inscribe a square in the circle and assume that the body moves along the path described by this square with constant speed. At the vertices of the square, our object will presumably collide with an elastic circular ring, from which it will be reflected according to the laws of collision. Both the magnitude and direction of the force exerted by the elastic ring are easily specified. It is directed toward the center of the circle, and its magnitude can be determined by geometric considerations. From the similarity of the triangles *OPP'* and *BP'A*, we obtain

$$\frac{\Delta(mv)}{mv} = \frac{a}{r},$$

and therefore, for a collision at the vertex of the square, we have

$$\Delta(mv) = \frac{a}{r} mv.$$

The net effect of the force exerted during a complete circumnavigation of the square path can now be determined from

$$4\Delta(mv) = \frac{4a}{r} mv. \qquad (1)$$

Rewriting this relationship in today's formalism, instead of the "effect of the force," we employ the product $F\Delta\tau$, where F is the force itself, and $\Delta\tau$ is the time over which the force acts, where we have made the simplifying assumption that the force is constant during the time of action. This assumption is valid for a collision only if we understand by "force" the average force. With these assumptions, equation (1) can be written as

$$F(4\Delta\tau) = \frac{4a}{r} mv.$$

Figure 3.124 shows a circle with an inscribed polygon of n sides. Because of the

similarity of the triangles, we derive here as well the relationship

$$\frac{\Delta(mv)}{mv} = \frac{a}{r},$$

from which follows

$$\Delta(mv) = \frac{a}{r} mv,$$

and in today's usual notation, we have

$$F(n\Delta\tau) = \frac{na}{r} mv.$$

This relationship and formula (1) can be summarized thus: In a complete circumnavigation, the net effect of the force—that is, the force multiplied by the sum of its times of action—is equal to the momentum multiplied by the quotient of the total length of the path and the radius of the circle.

If we increase the number of vertices without bound, in the limit we arrive at the circle. The validity of the conclusions that we have reached for polygonal paths is independent of the number of sides of the polygon, so that they remain true for arbitrary n and therefore for the circular path as well. The length of the path, of course, coincides with the circumference of the circle, and equation (1) becomes

$$F\tau = \frac{2\pi r}{r} mv = 2\pi mv.$$

From this equation, we obtain the constant magnitude of the force directed toward the center of the circle:

$$F = \frac{2\pi}{\tau} mv = \frac{2\pi}{2\pi r / v} mv = \frac{mv^2}{r}. \qquad (2)$$

Because of its direction, Huygens called this force a *centripetal force*. It is transmitted by the circular ring to the body, and it forces the body to move along a circular path. Of course, the body traversing the circular path exerts an equal and opposite force outwardly against the ring.

Figure 3.125 shows a page from Newton's *waste book* in which the square inscribed into the circle can be recognized as the starting point for the entire train of thought sketched above.

According to his *waste book*, Newton treated the problem of circular motion in perfect analogy with Huygens, but independently of him.

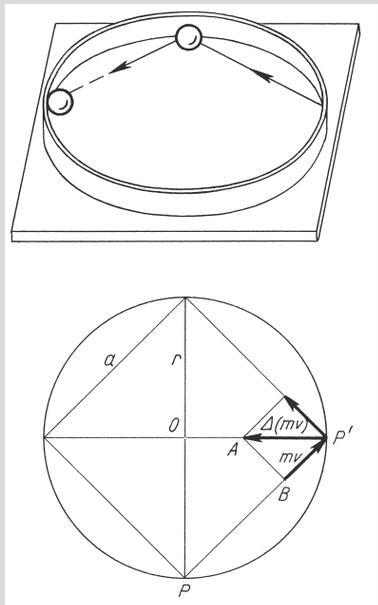
We have seen previously that circular motion can also be understood as motion with constant acceleration, given by the expression

$$a = \frac{v^2}{r},$$

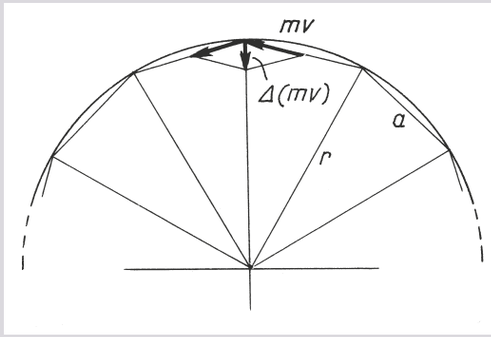
so that the force can be written as

$$F = ma = m\frac{v^2}{r}. \qquad (3)$$

This expression is identical to formula (2), but now it can be interpreted according to the relationship *force* = *mass* × *acceleration*.



◀ **Figure 3.123** For the quantitative treatment of uniform motion along a circular path, we begin with motion along a square inscribed in the circle. Because of the similarity of the triangles *ABP'* and *OPP'*, we obtain $\Delta(mv) : mv = a : r$.



▲ **Figure 3.124** For motion along an inscribed polygon with n sides, the proportion $\Delta(mv) : mv = a : r$ is still correct.

Quotation 3.44

The weight of a body is not understood, here, to be the tendency which makes it move towards the centre of the Earth, but rather to be its volume together with a certain solidity or condensation of the parts of its material which is probably the cause of its heaviness.

—Edme Mariotte, *Traité de la percussion ou choc des corps* [Dugas 1957, p. 199]

Quotation 3.45

There will be seen in it demonstrations of those kinds which do not produce as great a certitude as those of Geometry, and which even differ much therefrom, since whereas the Geometers prove their Propositions by fixed and incontestable Principles, here the Principles are verified by the conclusions to be drawn from them; the nature of these things not allowing of this being done otherwise. It is always

continued on next page

Quotation 3.45, ***continued***

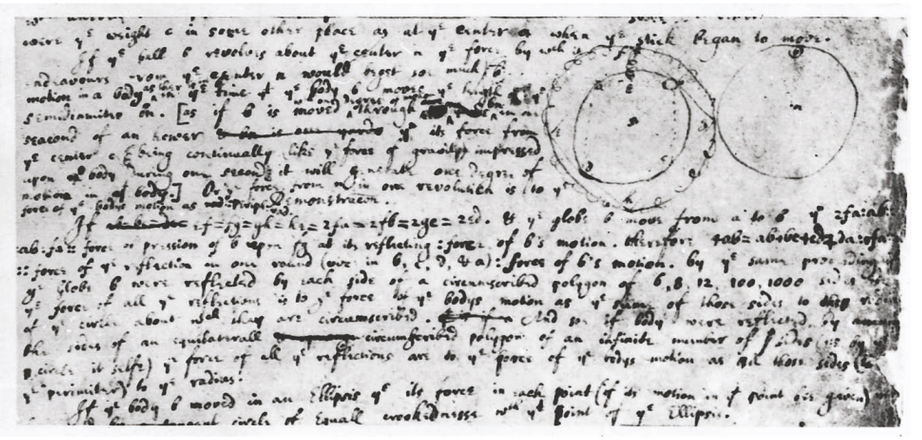
possible to attain thereby to a degree of probability which very often is scarcely less than complete proof. To wit, when things which have been demonstrated by the Principles that have been assumed correspond perfectly to the phenomena which experiment has brought under observation; especially when there are a great number of them, and further, principally, when one can imagine and foresee new phenomena which ought to follow from the hypotheses which one employs, and when one finds that therein the fact corresponds to our prevision.

—Christiaan Huygens, *Treatise on Light* [pp. vi–vii]

Quotation 3.46

After dinner, the weather being warm, we went into the garden and drank tea, under the shade of some appletrees, only he and myself. Amidst other discourse, he told me, he was just in the same situation, as when formerly, the notion of gravitation came into his mind. It was occasion'd by the fall of an apple, as he sat in a contemplative mood. Why should that apple always descend perpendicularly to the ground, thought he to himself. Why should it not go sideways or upwards, but constantly to the earth's centre? Assuredly, the reason is, that the earth draws it. There must be a drawing power in matter: and the sum of the drawing power in the matter of the earth must be in the earth's center, not in any side of the earth. Therefore does the apple fall perpendicularly, or towards the center. If matter thus draws matter, it must be in proportion to its quantity. Therefore, the apple draws the earth, as well as the earth draws the apple. That there is a power, like that we here call gravity, which extends its self thro' the universe.

—William Stukeley, *Memoirs of Sir Isaac Newton's Life*, 1936 [pp. 19–20]



▲ **Figure 3.125** Here in Newton's *waste book* these ideas appear for the first time.

Newton applied his results to circular motion—and that was the very first application of Newtonian dynamics—to investigate the relationship between the gravitational force on a body at the surface of Earth and the force that results from the rotation of Earth. Already in antiquity, the objection against Earth's rotation was raised—logically enough—that objects would be flung away from its surface. Newton showed that the force resulting from Earth's rotation is too small and that this objection is invalid. At the same time, this force is still strong enough to be measured experimentally; that is, Earth's rotation plays a role in the precise determination of the acceleration due to gravity (Figure 3.126).

3.7.3 *The Law of Universal Gravitation*

By the second half of the seventeenth century, a number of approaches toward a law of gravitational attraction had already appeared; it had even been generally formulated that all bodies mutually attract one another and that this attraction is responsible for the weight of objects on Earth's surface and therefore for motion in free fall, and also for the motion of the celestial bodies. Furthermore, the conjecture had been advanced that this force should be inversely proportional to the square of the distance separating two bodies. All this, however, was mere conjecture unsupported directly by experimental observation, and what is most important is that the supposed law of forces could not be harmonized with the elliptical planetary orbits; all such attempts had been unsuccessful.

Legend has it that the famed Newtonian apple provided the first impulse for Newton's formulation of the law of universal gravitation (Quotation 3.46).

According to Newton's own recollections and those of his friends, in deriving the law of forces, Newton proceeded as follows (Quotation 3.47): Let us assume that the planets move in circular orbits, which for most of the planets is, in fact, a good approximation. A planet moving along such a path of radius R must be acted on by the centripetal force

$$F^{\mathrm{cf}} = m\frac{v^2}{R} = m\left(\frac{2\pi R}{T}\right)^2 \frac{1}{R} = m\frac{(2\pi)^2 R}{T^2}.$$

To determine the dependence on distance of the force arising from the Sun, we must compare the orbital data of the planets in their different orbits. Such a comparison is given by Kepler's third law, according to which the squares of the orbital

periods of the planets are in the same relationship as the cubes of the orbital radii:

$$T_1^2 : T_2^2 : T_3^2 : \ldots = R_1^3 : R_2^3 : R_3^3 : \ldots,$$

or

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{R_1}{R_2}\right)^3.$$

From this, we obtain the following relationship for the forces acting on the planets:

$$\frac{F_1}{F_2} = \frac{m_1 (2\pi)^2 R_1 / T_1^2}{m_2 (2\pi)^2 R_2 / T_2^2} = \frac{m_1}{m_2} \frac{R_1}{R_2} \frac{T_2^2}{T_1^2},$$

and since

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{R_1}{R_2}\right)^3,$$

we may conclude that

$$\frac{F_1}{F_2} = \frac{m_1 R_2^2}{m_2 R_1^2},$$

or

$$F_1 : F_2 : F_3 : \ldots = \frac{m_1}{R_1^2} : \frac{m_2}{R_2^2} : \frac{m_3}{R_3^2} : \ldots.$$

From this relationship, it is clear that the force is inversely proportional to the square of the distance.

If the force exerted by Earth also satisfies this law, then we can compare the attractive forces of Earth on the Moon and on bodies near Earth's surface, or simply the acceleration measured at Earth's surface and the acceleration of the Moon in its orbit about Earth. The acceleration of the Moon in its orbit is given by the small value

$$a_{\text{Moon}} = \frac{g}{60^2} = \frac{9.8}{3600} = 2.73 \times 10^{-3} \text{ ms}^{-2},$$

since the force decreases as the square of the distance, and the average distance of the Moon from Earth is about 60 Earth radii. From a knowledge of the orbital period of the Moon, the acceleration can be calculated as follows:

$$a_{\text{Moon}} = \frac{v_{\text{M}}^2}{R_{\text{M}}} = \frac{(2\pi)^2 R_{\text{M}}}{T_{\text{M}}^2} = \frac{(2\pi)^2 3.84 \times 10^8}{(27.3 \times 24 \times 3600)^2} = 2.73 \times 10^{-3} \text{ ms}^{-2}.$$

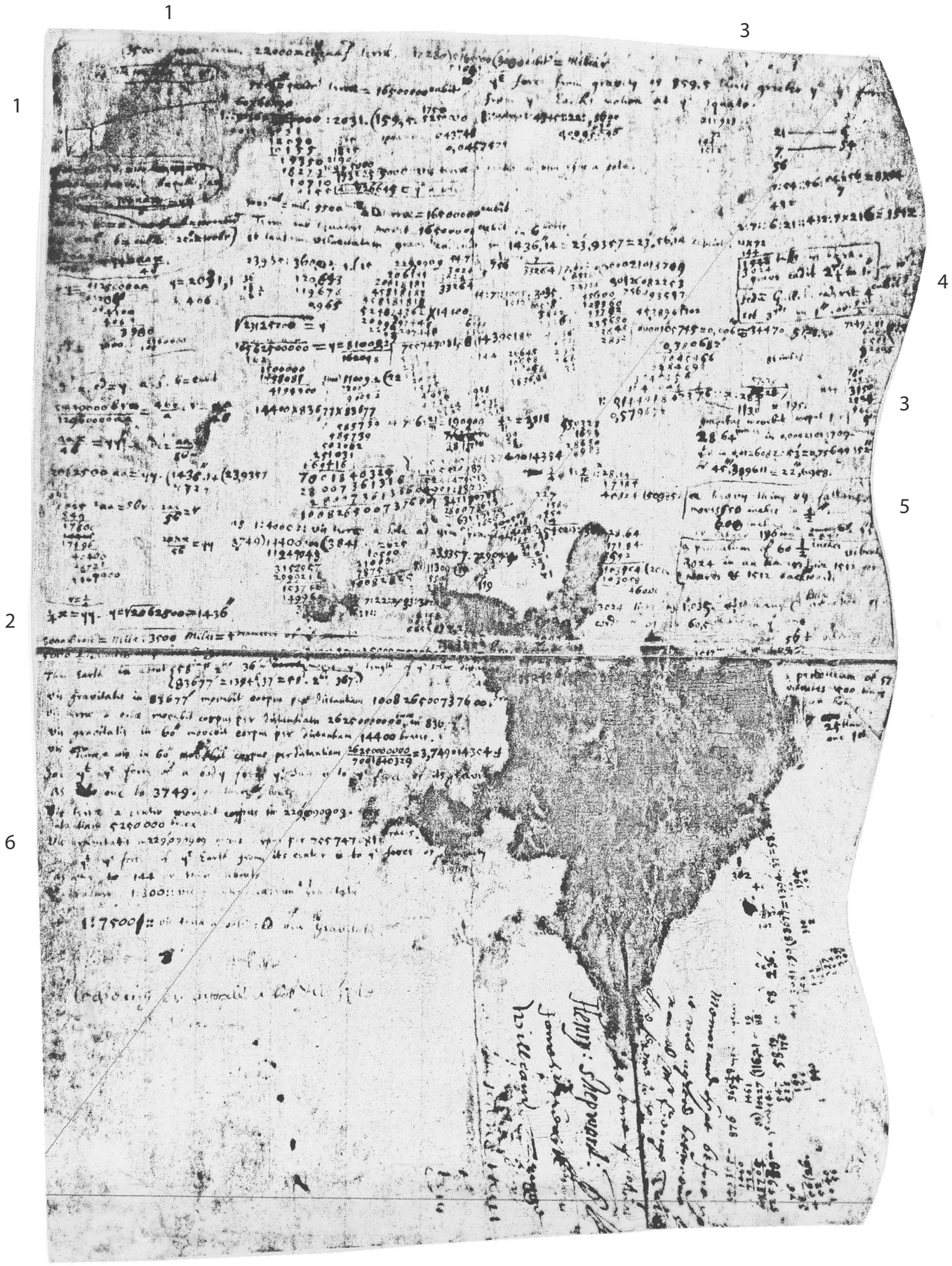
We can see that results of both computations agree, and according to his reminiscences, Newton saw this agreement as very good. However, it would be closer to the truth (Quotation 3.47) to say that Newton was unable to obtain a satisfactory agreement with the imprecise data available to him and therefore put off further investigation of the problem for a long time (almost 15 years).

It was only a single step from here to the formulation of the law of universal gravitation: If an attractive force from a body of mass m_A acts upon a body of mass m_B, then this force must be proportional to the mass m_A. However, in the process of interaction, the body of mass m_B attracts the body of mass m_A, and this force is proportional to the mass m_B. But because both forces are the same, it follows that

Quotation 3.47

[Newton's] first thoughts, which gave rise to his *Principia*, he had, when he retired from Cambridge in 1666 on account of the plague. As he sat alone in a garden, he fell into a speculation on the power of gravity: that as this power is not found sensibly diminished at the remotest distance from the center of the earth, to which we can rise, neither at the tops of the loftiest buildings, nor even on the summits of the highest mountains; it appeared to him reasonable to conclude, that this power must extend much farther than was usually thought; why not as high as the moon, said he to himself? and if so, her motion must be influenced by it; perhaps she is retained in her orbit thereby. However, though the power of gravity is not sensibly weakened in the little change of distance, at which we can place our selves from the center of the earth; yet it is very possible that so high as the moon this power may differ much in strength from what it is here. To make an estimate, what might be the degree of this diminution, he considered with himself, that if the moon be retained in her orbit by the force of gravity, no doubt the primary planets are carried round the sun by the like power. And by comparing the periods of the several planets with their distances from the sun, he found, that if any power like gravity held them in their courses, its strength must diminish in the duplicate proportion of the increase of distance. This he concluded by supposing them to move in perfect circles concentrical to the sun, from which the orbits of the greatest parts of them do not much differ. Supposing therefore the power of gravity, when extended to the moon, to decrease in the same manner, he computed whether that force would be sufficient to keep the moon in her orbit. In this computation, being absent from books, he took the common estimate in use among geographers and our seamen, before Norwood had measured the earth, that 60 English miles were contained in one degree of latitude on the surface of the earth. But as this is a very faulty supposition, each degree containing about 69½ of our miles, his computation did not answer expectation; whence he concluded, that some other cause must at least join with the action of the power of gravity on the moon. On this account he laid aside for that time any farther thoughts upon this matter.

—Henry Pemberton, *A View of Sir Isaac Newton's Philosophy*, 1728



◀ ▶ **Figure 3.126** The "vellum manuscript" (after Herivel). In the years 1665 and 1666, Newton used a sheet of parchment that had been used to draw up a rental contract as scrap paper. The numerical comparison of the gravitational and centrifugal forces and the path to this comparison can be read on this scrap.

1. Framed: *100 cubits in 5²* (100 ells in 5 seconds), data for free fall taken from the *Dialogo*.
2. $\frac{1}{4}x = yy$. The relationship for the distance traveled by a body in free fall; x is the distance and y is the time. If we write this formula as $x = \frac{1}{2}8y^2$ and compare it with today's notation $x = \frac{1}{2}gt^2$, we then see that g = 8 ells (= cubits = braces)/s² = 8 ´ 0.685 m/s² = 5.480 m/s². This is a very poor approximation, for the correct value (g = 9.8 m/s²) is almost twice as big. The formula $x = \frac{8}{2}y^2$ corresponds exactly to the data given by Galileo (y = 5, $y^2 = 25 \rightarrow x = 100$).
3. Newton determines g more precisely. A conical pendulum of length 81 inches and half conical angle of 45° executes…
4. … in one hour 1512 oscillations (1512 *ticks in hora*). With the help of this experiment, Newton determines the value of g in two different ways: He first assumes that in the case of a conical pendulum with half conical angle of 45°, the gravitational and centrifugal forces are equal ($g = v^2 / R$ with $R = l \sin\alpha$, l = 81 inches, and he uses, moreover, the fact that the period of the conical pendulum agrees with that of the mathematical pendulum with length $l \cos\alpha$; that is, it is equal to the projection of the pendulum length on the vertical. Thus Newton arrives…
5. … at the conclusion: *A heavy thing in falling moves 50 inches in 1* (crossed out and corrected) *1/2 second, 200 inches in one second, or rather 196 inches = 5 yds.* That is already quite a good approximation ($g \approx 10$ m/s²). The last conclusion is that …
6. … *vis terrae a centro movebit corpus in* 229.09 *minutes per distantiam* 5,250,000 *braces*. *Vis gravitatis in* 229.09 *minutes movet corpus per* 755,747,081 *braces*.

So that the force of the earth from its centre is to the force of gravity as one to 144 or thereabout.

Or rather as 1 : 300 : vis a centro terrae : vim gravitatis.

Newton uses here the following properties of circular motion: If the force acting on a body moving along a circular path with radius R is used to accelerate this body along a linear path, then the body will traverse in the time that it takes to traverse a distance R on the circular path, the distance $R/2$. (If one sets in the formula $s = \frac{a}{2}t^2$ the values $a = v^2 / R$ and $t = R / v$, then one obtains $s = R / 2$.)

The value 229.09 minutes = 3.818 hours (= 24/2p) is the time in which a point on the Earth's equator traverses the distance R_{Earth}. A free-falling body under the influence of centrifugal acceleration would traverse in this time the distance $R_{\text{Earth}} / 2 = 5250.000$ ells.

If for the case of actual free fall on Earth's surface one substitutes the value for g and the time t = 229.09 minutes into the formula $s = \frac{g}{2}t^2$, then one obtains 755,747,081 (ells). The ratio of the two path lengths is equal to the ratio of the centrifugal force to the gravitational force: 1:144. (The corrected value 1:300 came from the recognition of the error made in the determination of g.)

With this, one can refute the argument against the rotation of Earth that the centrifugal force resulting from a rotating Earth would fling objects off the surface of Earth (1 cubit = brace = ell = ¾ yard = 0.685 m; 1 inch = 2.54 cm).

the attractive force must be proportional to both masses; thus, *the gravitational force is proportional to the product of the masses of the two bodies and inversely proportional to the square of the distance between them*. In mathematical form,

$$F = G\frac{m_A m_B}{R^2},$$

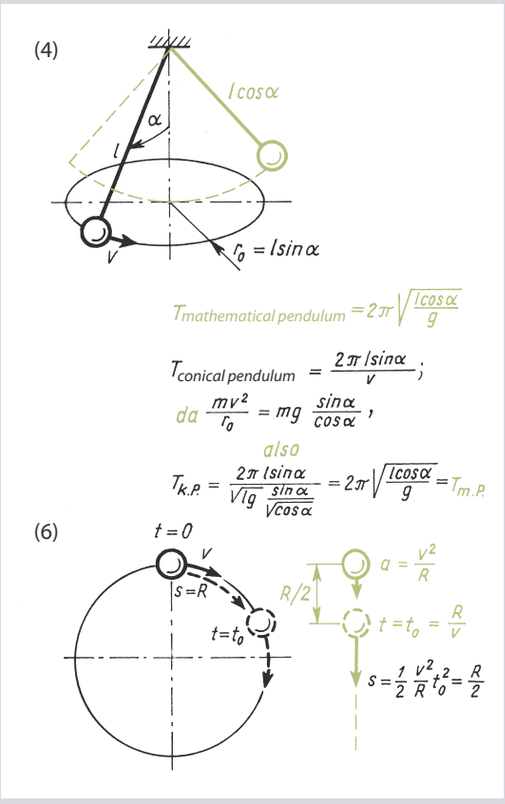
where G is a universal constant.

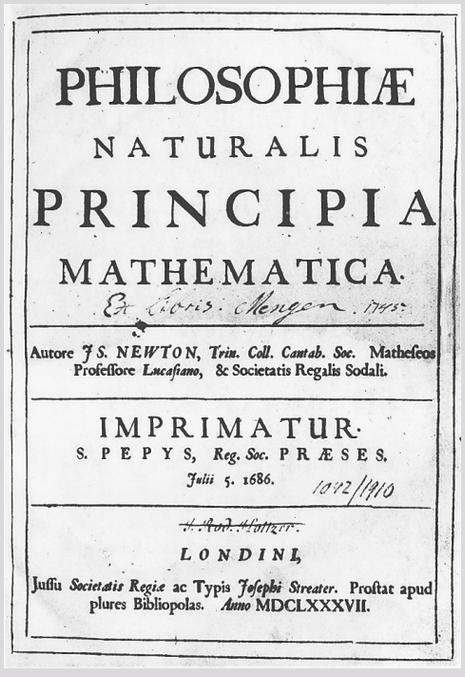
The proportionality factor G can be expressed in terms of the Keplerian constants and the Sun's mass (see Figure 3.31). To this end, we write Kepler's third law in a somewhat different form

$$\frac{R_1^3}{T_1^2} = \frac{R_2^3}{T_2^2} = \cdots = \frac{R^3}{T^2} = k_{\text{Sun}}.$$

However, the equation of motion for a planet of mass m_p is

$$G \times \frac{m_{\text{P}} m_{\text{S}}}{R^2} = m_{\text{P}} \frac{v^2}{R} = m_{\text{P}} \left(\frac{2\pi R}{T}\right)^2 \frac{1}{R},$$




PHILOSOPHIÆ
NATURALIS
PRINCIPIA
MATHEMATICA.
Autore J. S. NEWTON, Trin. Coll. Cantab. Soc. Matheseos Professore Lucasiano, & Societatis Regalis Sodali.
IMPRIMATUR.
S. PEPYS, Reg. Soc. PRÆSES,
Julii 5. 1686.
LONDINI,
Juffu Societatis Regiæ ac Typis Josephi Streater. Prostat apud plures Bibliopolas. Anno MDCLXXXVII.

▲ **Figure 3.127** The book *Philosophiae Naturalis Principia Mathematica* is often cited as the most significant work in the history of science, or even in the history of mankind.

Newton wrote it, as he later recounted, in a period of 17–18 months. In 1684, he began a series of lectures in Cambridge on this subject. The book appeared in 1687 with the financial support of Halley.

continued on next page

Figure 3.127 *continued*

The goal of the book is to provide answers finally to the questions of celestial mechanics.

> Dr. Vincent presented to the Society a manuscript treatise entitled *Philosophiae Naturalis Principia Mathematica*, and dedicated to the Society by Mr. Isaac Newton, wherein he gives a mathematical demonstration of the Copernican hypothesis as proposed by Kepler, and makes out all the phaenomena of the celestial motions by the only supposition of a gravitation towards the center of the sun decreasing as the squares of the distances therefrom reciprocally.
>
> —Minutes of the Royal Society, April 28, 1886 [Westfall 1980, pp. 444–445]

The work begins in the introduction with definitions and axioms; there follow three books. The first book deals with the motion of bodies, where above all, motions along a conic section and under the influence of a central force (not only in the form $1/r^2$) are considered. Essentially, the motion of a single body (mass point) is examined, although Newton also investigated the attraction between bodies of finite dimensions. The second book has as its subject motion in a viscous medium. Newton investigates here the resistance of a medium that can depend linearly, quadratically, or in a more complex manner on the velocity.

With its description of vortical motion of viscous fluids, the presentation reaches its goal of refuting Descartes's theory of vortices.

In the introduction to the third book, we meet the often quoted rules for philosophical thought. In the section *Phenomena*, we encounter tables of data for the moons of Jupiter and Saturn as well as for the five planets, with information contributed by a number of observers being compared. The section *Theses* of this book is the most important of the entire *Principia*. Here we find the law of universal gravitation and the inverse proportionality of the gravitational force to the square of the distance, an extensive description of the motion of the Moon, and an explanation of the precessional motion of Earth and an interpretation of the tides.

The *Principia* makes for quite difficult reading. It is tersely formulated, and theorems once stated are not repeated when they are applied, being merely cited by number. Thus, for example, Newton proves the assertion that the force acting on the moons of Jupiter acts in the direction of Jupiter and is inversely proportional to the square of the distance as follows:

> The first part of the theorem follows from phenomenon 1 and from Theorems 2 and 3 of the first book; the second part follows from phenomenon 1 as well as Corollary VI of the fourth theorem in the same book.

The second edition appeared in 1713 under the editorship of Roger Cotes with a large number of changes. The third edition (1726, published by Henry Pemberton) contains only a few further changes. The first English version appeared in 1729 with the title *Mathematical Principles of Natural Philosophy*.

from which follows the gravitational constant

$$G = \frac{(2\pi)^2}{m_S}\frac{R^3}{T^2} = \frac{(2\pi)^2}{m_S}k_{Sun} \Rightarrow m_S = \frac{(2\pi)^2}{G}k_{Sun},$$

where k_{Sun} is Kepler's constant for the Sun.

Available to Newton were not only data for the planets, but also for the orbits of four of Jupiter's moons and five of Saturn's. With these data, he was able to determine the masses of Jupiter and Saturn. We note that

$$G = \frac{(2\pi)^2}{m_{Jupiter}}k_{Jupiter} = \frac{(2\pi)^2}{m_{Saturn}}k_{Saturn} = \frac{(2\pi)^2}{m_{Sun}}k_{Sun} \Rightarrow \frac{m_{Jupiter}}{m_{Sun}} = \frac{k_{Jupiter}}{k_{Sun}}.$$

The gravitational constant G was determined in 1797–1798 by Cavendish with the help of a torsion balance. Its value is

$$G = 6.67 \times 10^{-11}\ \mathrm{Nm^2kg^{-2}}.$$

With this value, it is now possible to specify the masses of the Sun and the planets numerically.

3.7.4 *Selections from the Principia*

Anyone who picks up Newton's *Principia* (Figure 3.127) today will find two things greatly surprising. First, it will be seen that the law known as Newton's fundamental law of mechanics, *force = mass × acceleration*, does not appear anywhere in the book, neither in words nor in today's usual form

$$\mathbf{F} = m\frac{\mathrm{d}\,\mathbf{v}}{\mathrm{d}\,t}$$

or even as

$$\mathbf{F} = \frac{\mathrm{d}}{\mathrm{d}\,t}(m\mathbf{v})$$

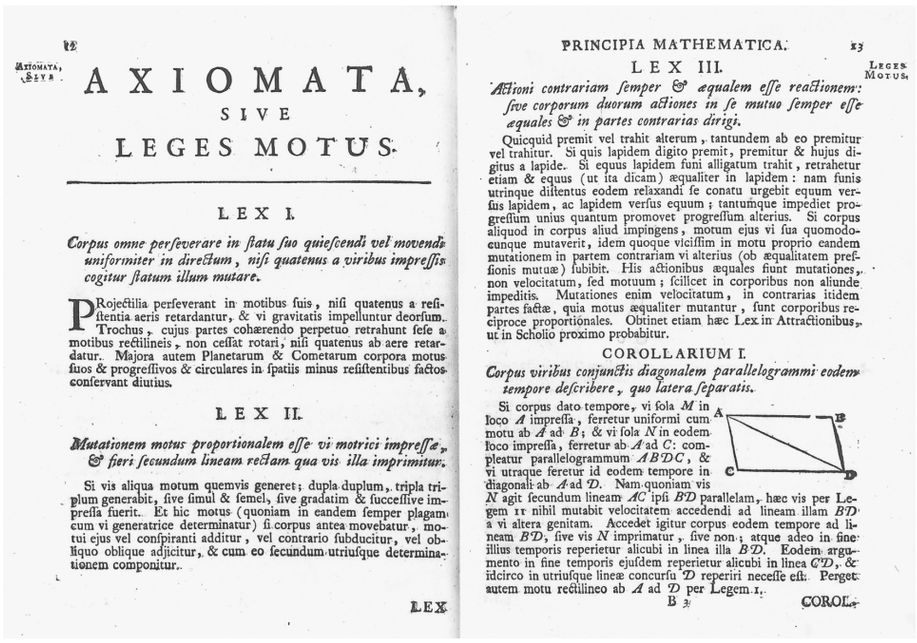
(Quotation 3.48).

The second surprising observation is that Newton, as one of the fathers of calculus, does not make use of this mathematical innovation anywhere in his book; all of his ideas are worked out on the basis of classical geometry.

We meet Newton's second axiom in the *Principia* in precisely the form already quoted: *The change in motion is proportional to the effective force and takes place along the line along which the force acts* (Figure 3.128). We have also mentioned already that by "motion," Newton means "momentum." There is no mention of changes per unit of time, nor of limits, and so the "effective force" here is not what we call force today. Accordingly, there is no mention of acceleration, mass, or force (in today's sense) in this law. In dealing with concrete cases, however, Newton uses the formulation of the law of motion familiar to us, and therefore it is apparent that he was aware of the relationship in that form as well (Quotation 3.49). The law of motion as it appears in the *Principia* can be written in today's notation as

$$\int_{t_1}^{t_2} \mathbf{F}\,\mathrm{d}\,t = \Delta(mv).$$

Regarding the lack of infinitesimal calculus, the general opinion is that while Newton made extensive use of the calculus of "fluxions" in his own calculations,



AXIOMATA, SIVE LEGES MOTUS.

LEX I.

Corpus omne perſeverare in ſtatu ſuo quieſcendi vel movendi uniformiter in directum, niſi quatenus a viribus impreſſis cogitur ſtatum illum mutare.

Projectilia perſeverant in motibus ſuis, niſi quatenus a reſiſtentia aeris retardantur, & vi gravitatis impelluntur deorſum. Trochus, cujus partes cohærendo perpetuo retrahunt ſeſe a motibus rectilineis, non ceſſat rotari, niſi quatenus ab aere retardatur. Majora autem Planetarum & Cometarum corpora motus ſuos & progreſſivos & circulares in ſpatiis minus reſiſtentibus factos conſervant diutius.

LEX II.

Mutationem motus proportionalem eſſe vi motrici impreſſæ, & fieri ſecundum lineam rectam qua vis illa imprimitur.

Si vis aliqua motum quemvis generet; dupla duplum, tripla triplum generabit, ſive ſimul & ſemel, ſive gradatim & ſucceſſive impreſſa fuerit. Et hic motus (quoniam in eandem ſemper plagam cum vi generatrice determinatur) ſi corpus antea movebatur, motui ejus vel conſpiranti additur, vel contrario ſubducitur, vel obliquo oblique adjicitur, & cum eo ſecundum utriuſque determinationem componitur.

LEX

PRINCIPIA MATHEMATICA. 13

LEGES MOTUS.

LEX III.

Actioni contrariam ſemper & æqualem eſſe reactionem: ſive corporum duorum actiones in ſe mutuo ſemper eſſe æquales & in partes contrarias dirigi.

Quicquid premit vel trahit alterum, tantundem ab eo premitur vel trahitur. Si quis lapidem digito premit, premitur & hujus digitus a lapide. Si equus lapidem funi alligatum trahit, retrahetur etiam & equus (ut ita dicam) æqualiter in lapidem: nam funis utrinque diſtentus eodem relaxandi ſe conatu urgebit equum verſus lapidem, ac lapidem verſus equum; tantumque impediet progreſſum unius quantum promovet progreſſum alterius. Si corpus aliquod in corpus aliud impingens, motum ejus vi ſua quomodocunque mutaverit, idem quoque viciſſim in motu proprio eandem mutationem in partem contrariam vi alterius (ob æqualitatem preſſionis mutuæ) ſubibit. His actionibus æquales fiunt mutationes, non velocitatum, ſed motuum; ſcilicet in corporibus non aliunde impeditis. Mutationes enim velocitatum, in contrarias itidem partes factæ, quia motus æqualiter mutantur, ſunt corporibus reciproce proportionales. Obtinet etiam hæc Lex in Attractionibus, ut in Scholio proximo probabitur.

COROLLARIUM I.

Corpus viribus conjunctis diagonalem parallelogrammi eodem tempore deſcribere, quo latera ſeparatis.

Si corpus dato tempore, vi ſola *M* in loco *A* impreſſa, ferretur uniformi cum motu ab *A* ad *B*; & vi ſola *N* in eodem loco impreſſa, ferretur ab *A* ad *C*: compleatur parallelogrammum *ABDC*, & vi utraque feretur id eodem tempore in diagonali ab *A* ad *D*. Nam quoniam vis *N* agit ſecundum lineam *AC* ipſi *BD* parallelam, hæc vis per Legem II nihil mutabit velocitatem accedendi ad lineam illam *BD* a vi altera genitam. Accedet igitur corpus eodem tempore ad lineam *BD*, ſive vis *N* imprimatur, ſive non; atque adeo in fine illius temporis reperietur alicubi in linea illa *BD*. Eodem argumento in fine temporis ejuſdem reperietur alicubi in linea *CD*, & idcirco in utriuſque lineæ concurſu *D* reperiri neceſſe eſt. Pergetautem motu rectilineo ab *A* ad *D* per Legem I.

B 3 COROL.

▲ **Figure 3.128** The two pages of the *Principia* on which the three Newtonian axioms are to be found.

he feared that the *Principia* contained enough difficult material as it was, so including the new method would simply serve to frighten readers away. Otherwise, Newton himself did not consider the fundamental law that bears his name to be of cardinal importance, so that in his reminiscences, in recounting his most important works, he lists the theory of color and the law of universal gravitation, but he does not even mention the fundamental law of mechanics (Plate XVII); moreover, he quite erroneously attributes the basic idea to Galileo (Quotation 3.50).

As we have already mentioned, the *Principia*, with its definitions, axioms, propositions, and lemmas, is written according to the patterns of antiquity. With respect to the clarity of definitions and concepts as well as their arrangement in logical sequence, we note a certain carelessness of the creative genius. Thus, as we have mentioned, Newton used the concept of force in a sense different from that used by succeeding generations, up to the present day. Taken alone, there would be no cause for objection; but the trouble is that Newton himself did not use the concept of force implicit in his own law of motion in a consistent manner.

Already the very first definition in the *Principia*, that of mass, is questionable, and indeed, it is seen by many as a *circulus vitiosus*: Newton defines quantity of matter, or mass, as the product of volume and the mass per unit volume, that is, as the product of density and volume. This definition is certainly a vicious circle if density is defined as mass per unit volume. However, if we regard the matter as being built up of identical atoms, then density could also be defined in such a way as to avoid this difficulty. Namely, density can be understood as a quantity proportional to the number of atoms per unit volume, or, more simply, as the relative occupation of space by the material under investigation.

Clarifying the concepts used by Newton, primarily his notion of force, is a popular topic for many contemporary historians of science.

It is a bit difficult, but nevertheless worth the effort, to follow Newton's method of determining the nature of the central force from the motion of a body along an elliptical path as set down in the *Principia*. Our goal here is of course once again to derive

Quotation 3.48

Previously the relation $F = m \times a$ was in fact generally inferred from the Definitions VII and VIII and Axiom II. However, this case resembles that of the Emperor's clothes in the fairy tale: all people saw them because they were convinced of their existence, until a child said that the Emperor had nothing on. Similarly Axiom II of the introductory chapter of Newton's *Principia* always used to be interpreted in the sense that a constant force produces a constant acceleration, and that their magnitudes are proportional, but if one looks at it impartially, nothing of the kind can be discovered. In order to interpret it this way, one has to assume that by change (*mutatio*) Newton means rate of change. Only then does it become possible to formulate the statement in a modern way as

$$F = d/dt\,(mv)\ \text{[where } mv \text{ multiplies } (d/dt)\text{]}$$

and if m is constant, this indeed amounts to

$$F = m \times a.$$

It is, however, extremely unlikely that Newton, who was quite capable of expressing his thoughts, should have committed the very serious mistake of confounding a magnitude and the rate of change of that magnitude in so fundamental a passage as that containing the axioms. Before making such a charge, we should consider whether it is not possible to make sense of the statement as it stands. ...

That he does not explicitly argue anywhere that a constant continuous force causes a uniformly accelerated motion (which after all is really the fundamental principle of the new dynamics which radically broke with the ancient and medieval conception) can probably be explained by the fact that he considered it perfectly self-evident (just as Huygens had done) that if a particle is acted upon by a constant force, the velocity necessarily increases by equal amounts in equal times. Both scientists were already imbued with the new dynamic conceptions to such a degree that they did not even think it necessary to mention this most cardinal point of difference with the old notions: a striking illustration of the rapidity with which new conceptions that have at first been paradoxical become commonplace. However, from the point of view of axiomatization—and that is the main point here—the omission of the proportionality of momentum and time, i.e. of the constancy of the acceleration, amounts to a flaw: in fact, while the momentum increases as the time, the kinetic energy increases as the distance, and how is one to know either one or the other without either postulating or proving it? Moreover, according to the Aristotelian conception of axiomatization the fact that a thing is evident is precisely a reason for stating it as an axiom, not for omitting it.

—E. J. Dijksterhuis, *The Mechanization of the World Picture* [pp. 471, 473]

▶ **Figure 3.129** The pages of the *Principia* that discuss the question of how the force law can be derived when the path is known.

Quotation 3.49

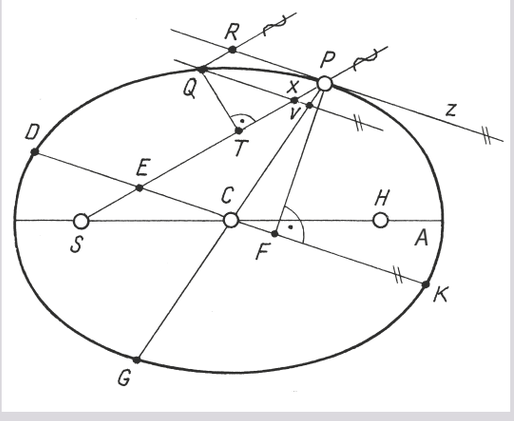
For the velocity which a given force can generate in a given matter in a given time is directly as the force and the time, and inversely as the matter. The greater the force or the time is, or the less the matter, the greater the velocity generated. This is manifest from the second Law of Motion.

—Isaac Newton, *Principia*, Book II, Section VI

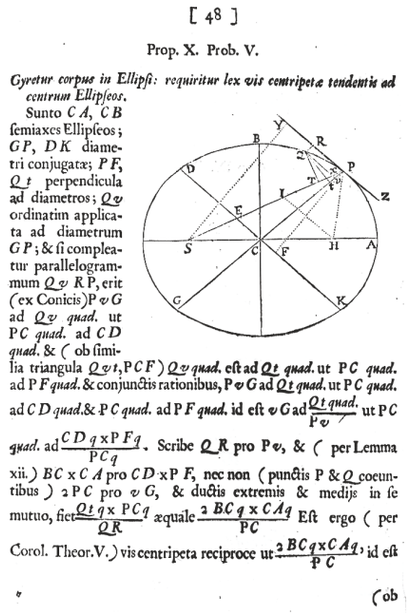
Quotation 3.50

Hitherto I have laid down such principles as have been received by mathematicians, and are confirmed by abundance of experiments. By the first two Laws and the first two Corollaries, Galileo discovered that the descent of bodies varied as the square of time (*in duplicate ratione temporis*) and that the motion of projectiles was in the curve of a parabola; experience agreeing with both, unless so far as these motions are a little retarded by the resistance of the air.

—Isaac Newton, *Principia*, Axioms, or Laws of Motion, Corollary VI



▲ **Figure 3.130** The diagram from Figure 3.129 enlarged for clarity.



[48]

Prop. X. Prob. V.

Gyretur corpus in Ellipſi: requiritur lex vis centripetæ tendentis ad centrum Ellipſeos.

Sunto *C A*, *C B* ſemiaxes Ellipſeos; *G P*, *D K* diametri conjugatæ; *P F*, *Q t* perpendicula ad diametros; *Q v* ordinatim applicata ad diametrum *G P*; & ſi compleatur parallelogrammum *Q v R P*, erit (ex Conicis) *P v G* ad *Q v quad.* ut *P C quad.* ad *C D quad.* & (ob ſimilia triangula *Q v t*, *P C F*) *Q v quad.* eſt ad *Q t quad.* ut *P C quad.* ad *P F quad.* & conjunctis rationibus, *P v G* ad *Q t quad.* ut *P C quad.* ad *C D quad.* & *P C quad.* ad *P F quad.* id eſt *v G* ad $\frac{Qt\ quad.}{Pv}$ ut *P C quad.* ad $\frac{CD\,q \times PFq}{PCq}$. Scribe *Q R* pro *P v*, & (per Lemma xii.) *B C* x *C A* pro *C D* x *P F*, nec non (punctis P & *Q* coeuntibus) 2 *P C* pro *v G*, & ductis extremis & mediis in ſe mutuo, fiet $\frac{Qt\,q \times PCq}{QR}$ æquale $\frac{2\,BC\,q \times CAq}{PC}$. Eſt ergo (per Corol. Theor. V.) vis centripeta reciproce ut $\frac{2\,BC\,q \times CAq}{PC}$, id eſt

(ob

[49]

(ob datum 2 *B C q.* x *C A q.*) ut $\frac{1}{PC}$, hoc eſt, directe ut diſtantia *P C*. *Q. E. I.*

Corol. 1. Unde viciſſim ſi vis ſit ut diſtantia, movebitur corpus in Ellipſi centrum habente in centro virium, aut forte in circulo, in quem Ellipſis migrare poteſt.

Corol. 2. Et æqualia erunt revolutionum in Figuris univerſis circa centrum idem factarum periodica tempora. Nam tempora illa in Ellipſibus ſimilibus æqualia ſunt per Corol. 3 & 7 Prop. IV: In Ellipſibus autem communem habentibus axem majorem, ſunt ad invicem ut Ellipſeon areæ totæ directe & arearum particulæ ſimul deſcriptæ inverſe; id eſt ut axes minores directe & corporum velocitates in verticibus principalibus inverſe, hoc eſt ut axes illi directe & ordinatim applicatæ ad axes alteros inverſe, & propterea (ob æqualitatem rationum directarum & inverſarum) in ratione æqualitatis.

Scholium.

Si Ellipſis, centro in infinitum abeunte, vertatur in Parabolam, corpus movebitur in hac Parabola, & vis ad centrum infinite diſtans jam tendens, evadet æquabilis. Hoc eſt Theorema *Galilei*. Et ſi Coniſectio Parabolica, inclinatione plani ad conum ſectum mutata, vertatur in Hyperbolam, movebitur corpus in hujus perimetro, vi centripeta in centrifugam verſa.

H SECT

the gravitational force field, which decreases as the square of the distance from the central body. However, this time we wish to arrive at it not starting from a simple circular orbit and using Kepler's third law, but on the basis of the general law of motion.

We begin with the following general theorem (*Principia, De motu corporum*, Liber I, Propositio VI, Corollarium V):

Suppose a body is moving, as depicted in Figures 3.129 and 3.130, along a curvilinear path *APQ* around the central point *S*, and let *ZPR* be the tangent to the path at the point *P*. Let us draw the segment *RQ* parallel to the segment *SP*, and the segment *QT* perpendicular to it. Then the centripetal force is inversely proportional to the quantity

$$\frac{SP^2 \times QT^2}{QR},$$

or, more precisely, to the limiting value of this quantity as the point *Q* approaches *P*.

The proof of this relationship is simple using today's methods. We specify the path by

$$\mathbf{r} = \mathbf{r}(t),$$

and then we expand this equation into a power series around time $t = 0$ and the place $\mathbf{r} = \mathbf{r}(0) = \mathbf{r}_p$:

$$\mathbf{r} = \mathbf{r}_p + \left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{t=0} t + \frac{1}{2}\left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{t=0} t^2 + \cdots.$$

If we now take into account the fundamental law

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \frac{\mathbf{F}}{m}$$

as well as the relationship $SP \times QT \propto t$, then we obtain the given expression for the centripetal force (Figure 3.131(a,b)).

We return next to the original problem, keeping to the formulation that appears in the *Principia* (Propositio IX).

Given that a body is moving along an elliptical path, let us find the law for the centripetal force directed toward the focus of the ellipse.

Suppose that at a certain moment the body is located at point *P*. We begin by drawing the diameter *DCK* conjugate to the diameter *GCP* (Figures 3.129 and 3.130). Newton proves that the distance of the point *P* from the point *E*, which is the intersection of the conjugate diameter with $SP \equiv \mathbf{r}_p$, is equal to the length *a* of the major semiaxis, that is,

$$EP = a.$$

We now draw the line Qxv parallel to the tangent RPZ. It can then be shown by making use of the similar triangles that

$$\frac{QR}{Pv} = \frac{PE}{PC} = \frac{a}{PC}.$$

Since we also have that QT is perpendicular to SP, and PF is perpendicular to the tangent, we obtain

$$\frac{Qx}{QT} = \frac{a}{PF}.$$

From a theorem of Apollonius on conic sections, we have

$$\frac{a}{PF} = \frac{CD}{b},$$

and from that, we obtain the relationship

$$QT = \frac{PF \times Qx}{a} = \frac{bQx}{CD},$$

as well as

$$QR = \frac{aPv}{PC}.$$

We now let the point Q approach the point P. Then the quotient Q_v/Q_x approaches 1, so that we obtain

$$\lim QT = \frac{bQv}{CD}.$$

We now calculate the expression appearing in the theorem given above as

$$\lim \frac{SP^2 \times QT^2}{QR} = \lim SP^2 \frac{Qv^2 \times b^2 \times PC}{CD^2 \times a \times Pv}. \qquad (4)$$

The equation of the ellipse in the coordinate system determined by the conjugate semiaxes CD and CP is

$$\frac{Qv^2}{CD^2} + \frac{Cv^2}{CP^2} - 1 = 0,$$

and with this, we can eliminate the quotient Q_v^2/CD^2 appearing in equation (4):

$$\frac{Qv^2}{CD^2} = \frac{CP^2 - Cv^2}{CP^2} = \frac{(CP + Cv)Pv}{CP^2},$$

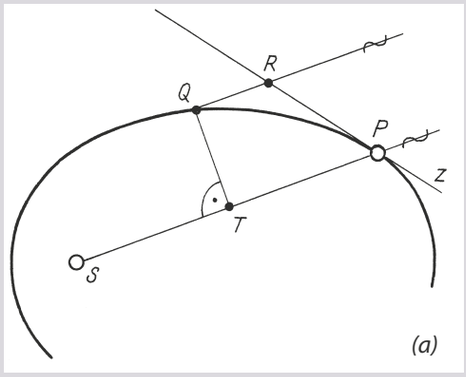
and thus we have

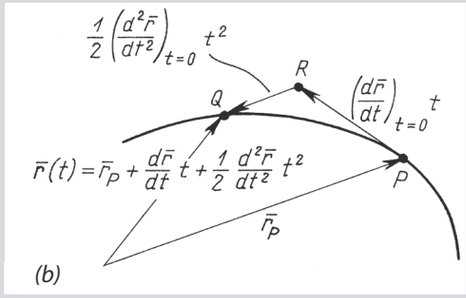
$$\lim \frac{SP^2 \times QT^2}{QR} = \lim SP^2 \frac{(CP + Cv)Pv \times b^2 \times PC}{CP^2 \times a \times Pv} = \frac{2PC^2 \times b^2}{a \times PC^2} \times SP^2 = \frac{2b^2}{a} SP^2$$
$$= L \times SP^2.$$

We can state the conclusion of the above derivation in the original words of the *Principia*: The centripetal force is reciprocal to the quantity $L \times SP^2$ and therefore inversely proportional to the square of the distance SP.

In this way, Newton, beginning with the general elliptical path, arrived at the law of the force being proportional to the inverse square of the distance.

Later, Newton raised the inverse question: given a central force that diminishes by the square of the distance, what is the path of a body in the general case? He





▲ **Figure 3.131** Since

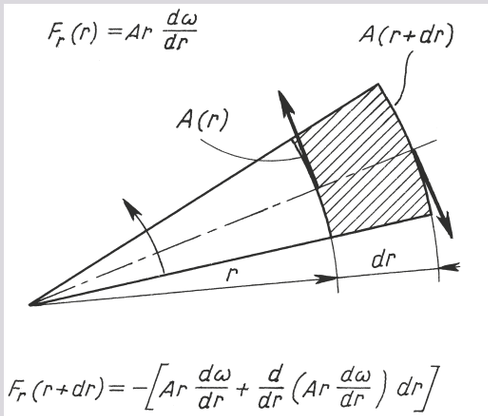
$$\mathbf{QR} = \frac{1}{2}\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\bigg|_{t=0} t^2, \quad \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \frac{\mathbf{F}}{m},$$

and the area $SP \cdot QT$ is proportional to the time t (in a neighborhood of $\mathbf{r}_p$, that is, for small QT), it follows that

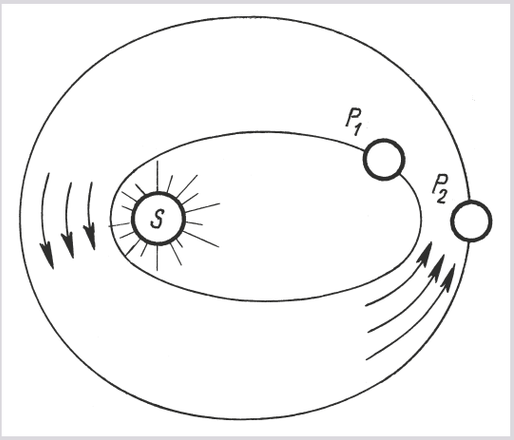
$$QR \propto F \cdot SP^2 \cdot QT^2,$$

and finally, the proportion

$$\frac{1}{F} = \frac{SP^2 \cdot QT}{QR}.$$



▲ **Figure 3.132** Cylindrically symmetric vortical flow of a viscous fluid.



▲ **Figure 3.133** For the planetary orbits shown, DESCARTES's vortex theory would give higher velocities when the planets are further from the Sun, in contradiction to KEPLER's law.

Quotation 3.51

I have endeavored in this Proposition to investigate the properties of vortices, that I might find whether the celestial phenomena can be explained by them; for the phenomenon is this, that the periodic times of the planets revolving about Jupiter are as the 3/2th power of their distances from Jupiter's centre; and the same rule obtains also among the planets that revolve about the sun. And these rules obtain also with the greatest accuracy, as far as has been yet discovered by astronomical observation. Therefore if those planets are carried round in vortices revolving about Jupiter and the sun, the vortices must revolve according to that law. But here we found the periodic times of the parts of the vortex to be as the square of the distances from the centre of motion; and this ratio cannot be diminished and reduced to the 3/2th power, unless either the matter of the vortex be more fluid the farther it is from the centre, or the resistance arising from the want of lubricity in the parts of the fluid should, as the velocity with which the parts of the fluid are separated goes on increasing, be augmented with it in a greater ratio than that in which the velocity increases. But neither of these suppositions seems reasonable. The more gross and less fluid parts will tend to be the circumference, unless they are heavy towards the centre. And though, for the sake of demonstration, I proposed, at the beginning of this Section, an Hypothesis that the resistance is proportional to the velocity, nevertheless, it is in truth probable that the resistance is in a less ratio than that of the velocity; which granted, the periodic times of the parts of the vortex will be in a greater ratio than the square of the distances from its centre. If, as some think, the vortices move more

continued on next page

showed that such a path will be a conic section and gave conditions under which an elliptical, parabolic, or hyperbolic path would be realized.

We select three topics from the second book of the *Principia*, of which the first two are of particular historical interest. NEWTON worked intensively on vortices in fluids, obviously with the goal of proving that the theory of vortices, which plays such a great role in Cartesian cosmology, was absurd.

In the case of a cylindrically symmetric vortex, NEWTON begins with the correct assumption that the frictional force deriving from the internal friction of the fluid "is proportional to the velocity with which the fluid elements move against one another."

We describe, following HUND [1972], a simplified description of NEWTON's train of thought.

Since the velocity at a particular point is given by the product of the radius and angular velocity, the relative velocity of the fluid elements to one another with respect to the relative velocities of the particles of a rotating rigid body are given by the first term of the total differential of the velocity,

$$\mathrm{d}(\omega r) = r\,\mathrm{d}\omega + \omega\,\mathrm{d}r.$$

The second term describes the portion that is the result of rigid rotation. It follows then that the frictional force is proportional to the surface area $A(r)$ and to the relative velocity gradient $r\frac{\mathrm{d}\omega}{\mathrm{d}r}$:

$$A(r)r\frac{\mathrm{d}\omega}{\mathrm{d}r}.$$

According to Figure 3.132, the equilibrium condition is

$$F_r(r) + F_r(r + \mathrm{d}r) = 0,$$

or

$$\frac{\mathrm{d}}{\mathrm{d}r}\left[A(r)r\frac{\mathrm{d}\omega}{\mathrm{d}r}\right] = 0,$$

from which follows

$$A(r)r\frac{\mathrm{d}\omega}{\mathrm{d}r} = \text{constant}.$$

If we further consider that the area A is proportional to the radius, the angular velocity must satisfy the relationship

$$\frac{\mathrm{d}\omega}{\mathrm{d}r} = \frac{\text{constant}}{r^2}, \quad \omega \propto \frac{1}{r}.$$

However, NEWTON's argumentation for the derivation of the relationship $\omega \propto 1/r^2$ between angular velocity and radius for the vortical flow around a sphere is not so convincing. But if the planets are carried along in their orbits by such vortices, then the relationship

$$T \propto r^2$$

for the orbital periods of the planets and their orbital radii would have to be satisfied, in contradiction to the relationship

$$T \propto r^{3/2}$$

required by Kepler's third law. From this contradiction, NEWTON concludes that the motion of the planets cannot be explained by vortices.

More convincing is the qualitative argument that the theory of vortices is incapable of leading to Kepler's second law. To see this, consider two planets as in Figure 3.133, one of which moves in a nearly circular orbit, while the orbit of the other planet is an ellipse of greater eccentricity; that is, the second planet moves along an elongated ellipsis. The streaming substance carrying the planets along must have a greater velocity farther from the Sun than closer in, due to the smaller cross section of the stream's channel; but then the same must hold for the planets,

in contradiction to Kepler's second law. Newton concluded from this that the vortex theory could not be brought into accord with astronomical observation. The theory explained nothing and just caused confusion (Quotation 3.51).

In the second problem that we have selected, Newton deals with the question of how deeply a moving solid cylindrical body would penetrate into a fluid or another solid body. He has this to say about the problem (Liber II, Propositio XXXVII): The resistance generated by the cross section of a cylinder that moves uniformly along its axis in a dense, unbounded, and inelastic liquid is in the same proportion to the force that could cause or stop its movement during the time during which it could traverse four times its length, as the density of the medium is to the density of the cylinder, approximately.

The meaning of this difficult proposition can be restated in a simplified form: As the cylinder moves in the fluid, its front face makes way by pushing the material ahead of it sideways. Today we would say that it continuously transfers its kinetic energy to the surrounding material. This continuous transfer of impulse is the breaking force, whose magnitude is specified by Newton's statement. Let us investigate, says Newton, what force would be required to accelerate the cylinder of the given mass to its present speed within a well-defined time, namely, the time in which the cylinder traverses a path that corresponds to four times its length, that is,

$$\frac{4l}{v_0}.$$

This force should be equal to the breaking force. This assertion is noteworthy in that from it, one can derive the surprising conclusion that the length of the path traversed by the cylinder in the fluid (until it comes to rest) depends not on the initial velocity but only on the length of the cylinder. Namely, to a good approximation, the ratio between the length of the path traversed by the cylinder to the cylinder's own length is the same as that of the density of the cylinder's material to that of the fluid. In connection with this, Gamow relates an interesting story: During the Second World War, it was surprising how little the depth to which bombs penetrated the soil depended on the height from which they were dropped. A theorem from Newton's *Principia* provided a solution to the mystery.

To investigate this problem in somewhat greater detail, we cast Newton's assertion in the form

$$R : F \propto \rho_{\text{fluid}} : \rho_{\text{metal}}.$$

The magnitude of the force is given by

$$F = m\frac{v_0}{4l / v_0} = \frac{mv_0^2}{4l},$$

from which we obtain the resistive force

$$R = F\frac{\rho_{\text{fluid}}}{\rho_{\text{metal}}} = \frac{mv_0^2}{4l}\frac{\rho_{\text{fluid}}}{\rho_{\text{metal}}}.$$

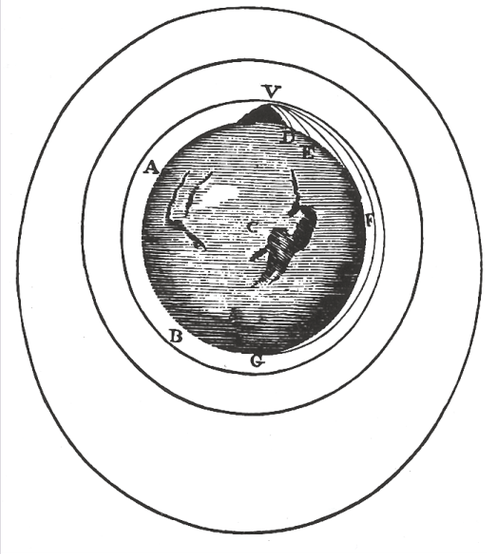
If now assume, by way of a crude approximation, that this force remains constant until the cylinder comes to rest, then the distance traveled by the cylinder is given by

$$s = \frac{v_0^2}{2a},$$

Quotation 3.51, ***continued***

swiftly near the centre, then slower to a certain limit, then again swifter near the circumference, certainly neither the 3/2th power, nor any other certain and determinate power, can obtain in them. Let philosophers then see how that phenomenon of the 3/2th power can be accounted for by vortices.

—Isaac Newton, *Principia*, Book II, Section IX



▲ **Figure 3.134** We meet here for the first time in the history of physics the idea of an artificial satellite and the associated theory that is still valid today. (*De mundi systemate liber Isaaci Newtoni*, 1728.)

A body moves in a circular path about the Earth if the attractive force GMm/R^2 yields precisely the necessary acceleration V^2/R. Therefore, for this (horizontal) initial

continued on next page

Figure 3.134 *continued*

velocity—called the first cosmic velocity—one has the equation

$$\frac{GMm}{R^2} = \frac{mv_1^2}{R} = mg \rightarrow v_1 = \sqrt{\frac{GM}{R}} = \sqrt{Rg} = 7900 \text{ m/s}$$

(R = 6380 km, g = 9.8066 m/s^2).

If the body's kinetic energy $\frac{1}{2}mv^2$ is greater at the surface of Earth than its potential energy GMm/R, then the body will leave Earth's gravitational field along a parabolic path. The equation $\frac{1}{2}mv^2 = GMm / R$ determines the second cosmic velocity, the escape velocity

$$v_2 = \sqrt{\frac{2GM}{R}} = \sqrt{2Rg} = \sqrt{2}v_1 = 11190 \text{ m/s}.$$

Newton's theory could, of course, also be applied to the "mysterious" comets. Newton himself gave a graphical method for an approximate determination of comets' paths and of the times of their reappearances, all from a small number of observations. Edmond Halley (1656–1742), publisher and supporter of the book *Principia*, applied this method to the comet appearing in 1681/1682 and was able to identify it with other historical comets. The correct prediction of the comet's return in 1758 was a triumph for Newton's theory of great scientific and indeed psychological consequence.

Although Halley's predictions were made more precise by the French mathematician Alexis Clairaut (1713–1765), the comet is known today as Halley's comet.

The first historical record of this comet comes from the year 240 BCE Since then, it has been regularly observed at periods of 74 to 79 years (with a single exception). The most recent years of return were 1531, 1607, 1682, 1758, 1835, 1910, 1986. Historical associations are linked with the years 451 (battle against Attila), 1066 (Battle of Hastings), and 1456 (Siege of Belgrade by the Turks).

Quotation 3.52

And hence appears a method both of comparing bodies one with another, as to the quantity of matter in each; and of comparing the weights of the same body in different places, to know the variation of its gravity. And by experiments made with the greatest accuracy, I have always found the quantity of matter in bodies to be proportional to their weight.

—Isaac Newton, *Principia*, Book II, Section VI

and using the relationship given above, we conclude that

$$s = \frac{v_0^2}{2(R/m)} = \frac{4l\rho_{\text{metal}}Rm}{m\rho_{\text{fluid}}2R} = 2\frac{\rho_{\text{metal}}}{\rho_{\text{fluid}}}l,$$

that is,

$$\frac{s}{l} = 2\frac{\rho_{\text{metal}}}{\rho_{\text{fluid}}}.$$

We see that the ratio of the two lengths indeed depends only on that of the two densities.

Our crude approximation can be justified in that when a bomb penetrates a solid medium, only the impulses transferred at high speeds play a significant role, since with reduced velocity, the bomb simply "gets stuck."

In Theorem XXIV of the second book, we encounter a statement of major theoretical significance: The concepts of weight (W) and mass (m) are separated here—in an experimentally measurable manner. If we do not use the relationship $W = mg$ in the equation of the pendulum, then for the period, we obtain

$$T = 2\pi\sqrt{\frac{ml}{W}},$$

which yields the relationship

$$m = \frac{T^2}{(2\pi)^2} \times \frac{W}{l}$$

for the mass. At this point (Quotation 3.52), Newton points out the proportionality between the weight W and mass m.

The third part of the *Principia* treats planetary motion on the basis of the general law of attraction between masses. In addition to the assertion that the same laws hold for both terrestrial and celestial phenomena, Newton also specifies the precise conditions under which a terrestrial body can become a celestial one. With reference to Figure 3.134, let us consider the paths of projectiles fired horizontally with varying velocities from the peak of a high mountain. Of course, with increasing velocity, the distance between the launching and landing sites of the projectile increases, and if the launching velocity is sufficiently great, then the projectile could—in principle—orbit Earth, landing at the spot from which it was launched. Let us suppose that the motion takes place in a vacuum. Then in this last case, the launching and landing velocities will be the same, so that the projectile can orbit Earth repeatedly like an artificial satellite. For today's reader, this is perhaps the most interesting part of the *Principia*, since we are so familiar with satellites and Moon landings as memorable events of our modern era and therefore are astonished that not only the requisite theoretical tools, but also the explanatory qualitative illustration, can already be found in the *Principia*.

For Newton's contemporaries it was of greater significance that besides the regular motion of the planets he also qualitatively explained a host of other phenomena whose interpretations had been attempted in vain by outstanding scholars before him. These include the tides, which, as we have mentioned, were investigated intensively by both Galileo and Descartes without either of them arriving at a satisfactory explanation. A similar once baffling problem was the precession of Earth's axis, which is manifested—as we have already mentioned in connection with ancient natural science (Section 1.4)—in the shifting of the equinoxes along the ecliptic.

3.7.5 *Newton as Philosopher*

Newton's significance as a philosopher is threefold: he formulated and therefore determined the method of research for the natural sciences, for a long time to come; he set the basic goals of scientific research; and finally, he established a coherent, unified cosmology.

These contributions remained unchallenged in the three centuries that have elapsed since then—the Newtonian conception was at most refined in certain points.

Newton's method of scientific investigation was already established in the foreword to the *Principia* (Quotation 3.53). The following sentence there deserves special emphasis: "[F]or the whole burden of philosophy seems to consist in this—from the phenomena of motions to investigate the forces of nature, and then from these forces to demonstrate the other phenomena."

Our first glance into the third book of the *Principia* is sufficient to convince us of how much Newton respected facts and shows that his considerations started from facts and led back to facts. The subject of the book is the structure of the universe, and it starts with a most precise and detailed collection of observations. The sheer volume of tables and data in such a theoretical work is astounding.

Today, we are so accustomed to the scientific method as laid down by Newton that we consider it to be self-evident. However, we should recall that even Galileo set out from considerations of a completely different sort. He argued, for example, that motion with uniform acceleration must be realized in nature because it is the simplest form of change in velocity, or even that circular motion is the natural form of motion for celestial bodies. Or recall that Descartes forced his entire cosmology into a rationally clear and easily understandable starting point. In this sense, Newton realized Bacon's program, but in contrast to Bacon, he gave mathematics its rightful place in scientific investigation.

Due to Newton's influence, the criterion for truth in the natural sciences is no longer the logical derivability from some set of simple axioms, but rather agreement with the conclusions of fundamental laws that were read from nature, with experiment having the final say. If theoretical conclusions and experimental observations cannot be brought into agreement, then the underlying theory needs to be revised; Newton expresses this categorically in the *Philosophical Rules* that preface the third book of the *Principia* (Figure 3.135). We quote here the following of the rules:

Rule I: We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances. …

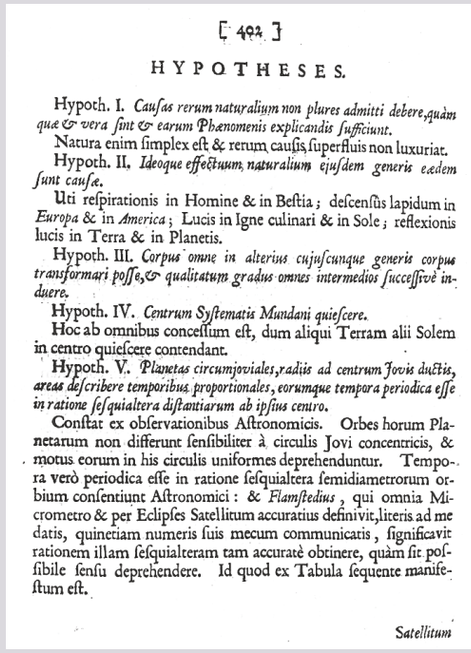
Rule II: Therefore to the same natural effects we must, as far as possible, assign the same causes.

> As to respiration in a man and in a beast; the descent of stones in Europe and in America; the light of our culinary fire and of the Sun; the reflection of light in Earth, and in the planets. …

Rule IV: In experimental philosophy, we are to look upon propositions inferred by general induction from phenomena as accurately or very nearly true, notwithstanding any contrary hypotheses that may be imagined, till such time as other phenomena occur, by which they may either be made more accurate, or liable to exceptions.

> This rule we must follow, so that the argument of induction may not be evaded by hypotheses.

Even in his book *Optics*, Newton dealt with questions of natural philosophy. "In the beginning, God created the universe and the atoms." These atoms were considered by Newton to be fixed and "indestructible," and he reformulated the



[402]

HYPOTHESES.

Hypoth. I. *Cauſas rerum naturalium non plures admitti debere, quàm quæ & veræ ſint & earum Phænomenis explicandis ſufficiunt.*
Natura enim ſimplex eſt & rerum cauſis ſuperfluis non luxuriat.
Hypoth. II. *Ideoque effectuum naturalium ejuſdem generis eædem ſunt cauſæ.*
Uti reſpirationis in Homine & in Beſtia; deſcenſûs lapidum in *Europa* & in *America*; Lucis in Igne culinari & in Sole; reflexionis lucis in Terra & in Planetis.
Hypoth. III. *Corpus omne in alterius cujuſcunque generis corpus transformari poſſe, & qualitatum gradus omnes intermedios ſucceſſivè induere.*
Hypoth. IV. *Centrum Syſtematis Mundani quieſcere.*
Hoc ab omnibus conceſſum eſt, dum aliqui Terram alii Solem in centro quieſcere contendant.
Hypoth. V. *Planetas circumjoviales, radiis ad centrum Jovis ductis, areas deſcribere temporibus proportionales, eorumque tempora periodica eſſe in ratione ſeſquialtera diſtantiarum ab ipſius centro.*
Conſtat ex obſervationibus Aſtronomicis. Orbes horum Planetarum non differunt ſenſibiliter à circulis Jovi concentricis, & motus eorum in his circulis uniformes deprehenduntur. Tempora verò periodica eſſe in ratione ſeſquialtera ſemidiametrorum orbium conſentiunt Aſtronomici: & *Flamſtedius*, qui omnia Micrometro & per Eclipſes Satellitum accuratius definivit, literis ad me datis, quinetiam numeris ſuis mecum communicatis, ſignificavit rationem illam ſeſquialteram tam accuratè obtinere, quàm ſit poſſibile ſenſu deprehendere. Id quod ex Tabula ſequente manifeſtum eſt.

Satellitum

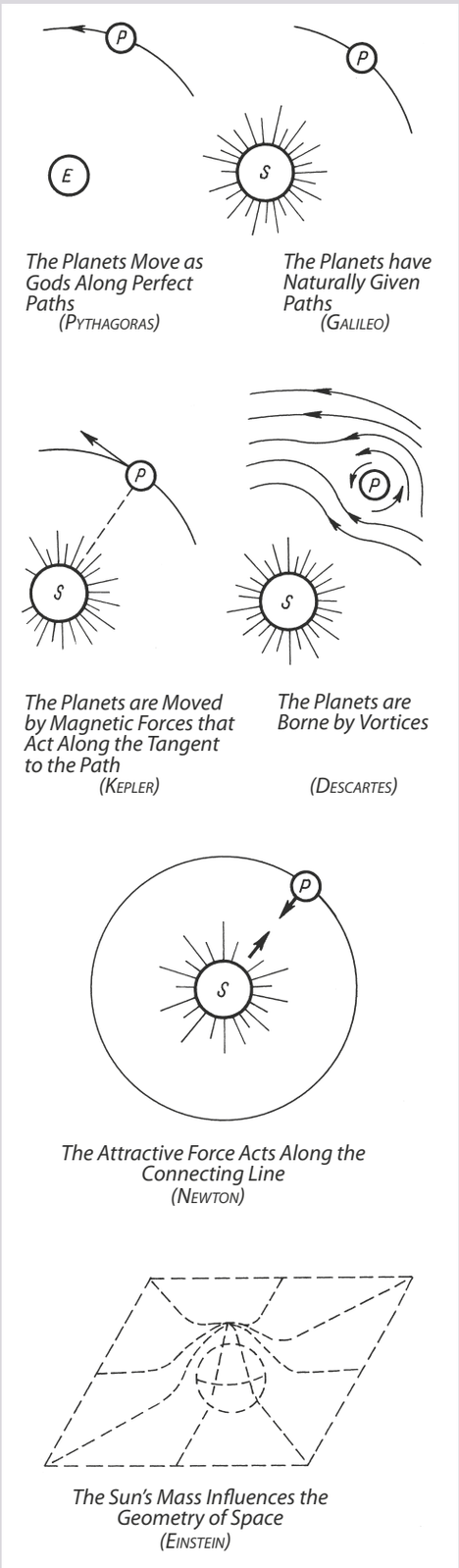
▲ **Figure 3.135** The initial hypotheses.

Quotation 3.53

… the whole burden of philosophy seems to consist in this—from the phenomena of motions to investigate the forces of nature, and then from these forces to demonstrate the other phenomena ….

I wish we could derive the rest of the phenomena of Nature by the same kind of reasoning from mechanical principles, for I am induced by many reasons to suspect that they may all depend upon certain forces by which the particles of bodies, by some causes hitherto unknown, are either mutually impelled towards one another, and cohere in regular figures, or are repelled and recede from one another. [These forces being unknown, philosophers have hitherto attempted the search of Nature in vain; but I hope the principles here laid down will afford some light either to this or some truer method of philosophy.]

—Isaac Newton, *Principia*, preface to the first edition, 1638



▲ **Figure 3.136** Explanations for the motion of planets at different times in history.

old principle of Democritus that stated that changes in macroscopic bodies are a result of the association, disassociation, and motion of atoms.

Although Newton used almost the same words as those used by the atomists of antiquity, their meaning was now deeper, in the sense that atomic theory had now become a program that could be quantitatively formulated. Specifically, if we can find the forces acting between particles, then the phenomena can also be described quantitatively. In the Newtonian world, we are dealing with centers of force and mechanistic motion that takes place under the influence of forces. Today, we say that Newton wanted to reduce all processes to mechanics, and the Newtonian worldview represents the completeness of the mechanistic worldview. However, we must note that Newton's contemporaries—and here we are thinking of the most astute and critical contemporaries, such as Huygens and Leibniz—understood something quite different by a mechanical explanation because, in their view, an interaction could be caused only by direct contact. They saw Newton's introduction of attractive forces acting at a distance as a step backward that would smuggle back into physics such ancient occult qualities as affinity, desire, and affection. Newton recognized the legitimacy of Huygens's objections, but he justified himself, correctly, with the argument that with the help of the force acting at a distance that he had postulated, "the phenomena of the heavens and the oceans" could be correctly described. Newton also felt that there was something there that needed explaining; he himself had long thought about this and, being unable to come up with the cause of gravitation, declared, "*Hypotheses non fingo*" (I frame no hypotheses). By "hypothesis," Newton meant an assumption that was unsupported by observation and could not be derived from it (Quotation 3.54). According to Newton, the Cartesian vortices are such a hypothetical concept, whereas the law of gravitation should be seen not as a hypothesis, but as fact.

Ultimately, the results of Newton's investigations and the creation of the Newtonian worldview are not only scientific but also philosophical achievements. We must therefore note with a measure of surprise and disappointment that despite his achievements, Newton is accorded no mention or at best a minor place in the history of philosophy. Many books on the history of philosophy can be found in which Newton's name does not appear in the index at all. This is all the more surprising when one considers that Aristotle and the Aristotelian cosmology is one of the favored topics in every such book. But it was exactly that cosmology that was replaced by the Newtonian worldview in the consciousness of mankind.

The old problem of a finite, closed universe was essentially restricted to the solar system: the regular motion of the sphere of fixed stars raised no particular questions, but the apparently irregular wanderings of the planets gave cause for much speculation. In Figure 3.136 we can trace the visions of our solar system up to Newton, and then beyond Newton to the picture given by Einstein's theory of general relativity. We should add that the corrections made to Newton's theory by the most modern theories are almost imperceptible even today, at least on the scale of the solar system. But the Newtonian cosmology goes beyond the solar system and attempts to describe a homogeneous and infinite universe through the connection of the laws of force and motion. By homogeneous we should understand a universe that is everywhere composed of the same matter and that is subject to the same laws, whether the matter is on Earth's surface, even is a planet, or is the Sun itself.

The carrier of the phenomena of the Newtonian world and the arbiter of their rhythm are Newtonian absolute space and absolute time. Newton defined these two notions at the beginning of the *Principia* (Figure 3.137):

Absolute space, in its own nature, without relation to anything external, remains always similar and immovable.

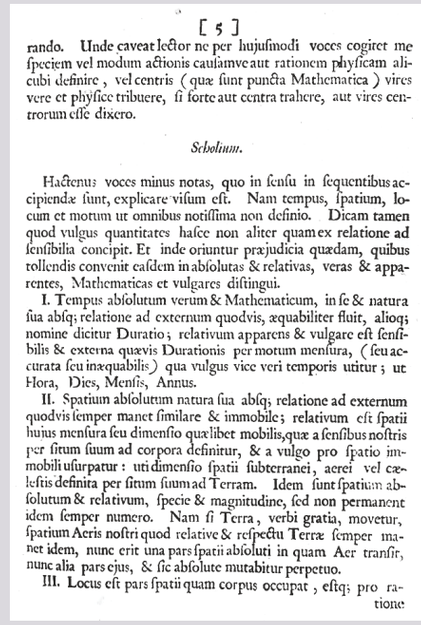
Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external.

Absolute motion is the translation of a body from one location in absolute space to another.

Newton also defined a relative space and a relative, perceptible, time, which we are able to measure in some manner with our instruments. Newton's notion of absolute space and time stands in sharp contrast to that of Descartes, who spoke only of relative position and relative motion. When Descartes discussed motion, he always meant a change of position of a body relative to the other bodies surrounding it. In this way, he could maintain that Earth does not move because its position does not change in relation to the surrounding vortical material (Section 3.4). However, Newton required the concepts of absolute space and time to relate the interaction of bodies separated by large reaches of space with the motion of those bodies. Newton clearly recognized the problems raised by his postulate of an absolute space. Leibniz sharply attacked this concept from the philosophical point of view and argued that we have no possibility of verifying any uniform motion of absolute space because it is not permitted to relate absolute space to any other object. Newton attempted to support his assumption of absolute motion—at least with respect to circular motion—with his well-known bucket analogy, which we can also demonstrate experimentally (Figure 3.138). In sharp contradiction to his own principle not to fabricate hypotheses, he even surmised that in distant reaches of the universe there are large masses that fix absolute space.

Although Newton strove to establish even his very general principles on a scientific and rational basis, he was deeply religious, and at many places in the *Principia*—especially in the later editions—he made reference to the necessity of divine influence, which, however, he limited to the setting of the universe into motion. Without this divine impulse, a host of observations made no sense to him, for example, the fact that all the planets orbit the Sun in the same direction and that the orbital planes are almost all coplanar. The Creator is even assigned the task in the Newtonian cosmology of intervening before a phenomenon could deviate from the laws. Leibniz commented sarcastically on this that God is apparently a poor mechanic, since he constructed a faulty machine that has to be repaired from time to time, and Dijksterhuis correctly remarked that Newton's god is an engineer who designed the world and set it in motion, and now God may still be an engineer, but He is in retirement.

Two great scientific revolutions of our era, the theory of relativity and quantum theory, call the Newtonian cosmology into question from two different points of view. With its concepts of quantities that are the same from any point of reference, relativity theory clears up the ideas of space and time, whereas quantum mechanics provides new equations of motion for particles of the micro world that replace those given by Newton. We have already mentioned that there is a great difference between the "wounding" of the Newtonian theory and attacks on the older, say the Aristotelian, theories in that Newtonian cosmology completely displaces the Aristotelian, taking on none of its results, whereas the Newtonian theory is a valid approximation of the theory of relativity and quantum theory in the limiting case of small velocity and large mass and therefore *remains a solid component of the natural sciences*.



[5]

rando. Unde caveat lector ne per hujusmodi voces cogiret me speciem vel modum actionis caulamve aut rationem physicam alicubi definire, vel centris (quæ sunt puncta Mathematica) vires vere et physice tribuere, si forte aut centra trahere, aut vires centrorum esse dixero.

Scholium.

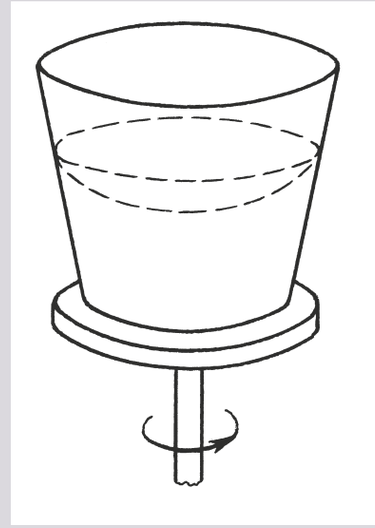
Hactenus voces minus notas, quo in sensu in sequentibus accipiendæ sunt, explicare visum est. Nam tempus, spatium, locum et motum ut omnibus notissima non definio. Dicam tamen quod vulgus quantitates hasce non aliter quam ex relatione ad sensibilia concipit. Et inde oriuntur præjudicia quædam, quibus tollendis convenit easdem in absolutas & relativas, veras & apparentes, Mathematicas et vulgares distingui.

I. Tempus absolutum verum & Mathematicum, in se & natura sua absq; relatione ad externum quodvis, æquabiliter fluit, alioq; nomine dicitur Duratio; relativum apparens & vulgare est sensibilis & externa quævis Durationis per motum mensura, (seu accurata seu inæquabilis) qua vulgus vice veri temporis utitur; ut Hora, Dies, Mensis, Annus.

II. Spatium absolutum natura sua absq; relatione ad externum quodvis semper manet similare & immobile; relativum est spatii hujus mensura seu dimensio quælibet mobilis, quæ a sensibus nostris per situm suum ad corpora definitur, & a vulgo pro spatio immobili usurpatur: uti dimensio spatii subterranei, aerei vel cælestis definita per situm suum ad Terram. Idem sunt spatium absolutum & relativum, specie & magnitudine, sed non permanent idem semper numero. Nam si Terra, verbi gratia, movetur, spatium Aeris nostri quod relative & respectu Terræ semper manet idem, nunc erit una pars spatii absoluti in quam Aer transit, nunc alia pars ejus, & sic absolute mutabitur perpetuo.

III. Locus est pars spatii quam corpus occupat, estq; pro ratione

▲ **Figure 3.137** The page of *Principia* with the Newtonian definition of absolute space and absolute time.



▲ **Figure 3.138** If a bucket of water is rotated, then the surface of the water takes the form of a paraboloid of revolution, which does not depend on the relative motion of the bucket and the water. Newton argued that the rotation is to be understood relative to absolute space.

Berkeley already pointed out that rotation can be imagined only with respect to something else, such as the system of fixed stars, for only the relative motion of the two makes sense. The correct comparison is therefore the following: rotating bucket, resting universe ↔ resting bucket, rotating universe. The evident physical and epistemological problem that arises here was investigated by Mach (1872) and later Einstein (1916, Quotation 5.2.7).

continued on next page

Figure 3.138 *continued*

The effects which distinguish absolute from relative motion are the forces of receding from the axis of circular motion. For there are no such forces in a circular motion purely relative, but in a true and absolute circular motion, they are greater or less, according to the quantity of the motion. If a vessel, hung by a long cord, is so often turned about that the cord is strongly twisted, then filled with water, and held at rest together with the water; thereupon, by the sudden action of another force, it is whirled about the contrary way, and while the cord is untwisting itself, the vessel continues for some time in this motion; the surface of the water will at first be plain, as before the vessel began to move; but after that, the vessel, by gradually communicating its motion to the water, will make it begin sensibly to revolve, and recede by little and little from the middle, and ascend to the sides of the vessel, forming itself into a concave figure (as I have experienced), and the swifter the motion becomes, the higher will the water rise, till at last, performing its revolutions in the same times with the vessel, it becomes relatively at rest in it. This ascent of the water shows its endeavor to recede from the axis of its motion; and the true and absolute circular motion of the water, which is here directly contrary to the relative, becomes known, and may be measured by this endeavor. At first, when the relative motion of the water in the vessel was greatest, it produced no endeavor to recede from the axis; the water showed no tendency to the circumference, nor any ascent towards the sides of the vessel, but remained of a plain surface, and therefore its true circular motion had not yet begun. But afterwards, when the relative motion of the water had decreased, the ascent thereof towards the sides of the vessel proved its endeavor to recede from the axis; and this endeavor showed the real circular motion of the water continually increasing, till it had acquired its greatest quantity, when the water rested relatively in the vessel. And therefore this endeavor does not depend upon any translation of the water in respect of the ambient bodies, nor can true circular motion be defined by such translation.

—Isaac Newton, *Principia,* Book I, Definitions, Scholium

Newton's experiment with the rotating vessel of water simply informs us, that the relative rotation of the water with respect to the sides of the vessel produces no noticeable centrifugal forces, but that such forces are produced by its relative rotation with respect to the mass of the earth and the other celestial bodies. No one is competent to say how the experiment would turn out if the sides of the vessel increased in thickness and mass till they were ultimately several leagues thick.

—Ernst Mach, *The Science of Mechanics: A Critical and Historical Account of Its Development* [p. 284]

Newton's personal achievements did more than fill his educated contemporaries with admiration, many of whom devoted their lives to working out only a small portion of his contributions. Even those such as Huygens and Leibniz, who while understanding Newton's work, criticized it, respected the greatness of his accomplishments. Newton's apotheosis began already in his lifetime. Maupertuis, whom we shall meet again later, asked people who had had direct contact with Newton, "Does he walk, eat, drink like other mortals?" Poets composed verses in his honor, of which the following couplet by Pope is the most famous:

All Nature and its laws lay hid in Night
God said, let Newton be, and all was light.

The following inscription appears beneath the bust of Newton at Trinity Chapel, Cambridge: *Newton qui genus humanum ingenio superavit* (Newton, who in genius towered above the human race).

In France, no less a personality than Voltaire propagated Newton's teachings. Ten years after Newton's death, in 1737, a book appeared in Italian with the title *Neutonianismo per le donne*, which was then translated into English with the title *Sir Isaac Newton's Philosophy explain'd for the use of the Ladies*. The author was one Algarotti, on whose grave, for him the highest honor, appears the epitaph, "A student of Newton" (Figure 3.139).

The seventeenth century has been called the century of geniuses, and indeed, no other century seems to have brought forth so many great thinkers. In Table 3.4 we provide a listing of the most important of these, and with this table we take leave of this remarkable century. All who are mentioned here knew of one another, frequently with mutual regard, seldom with love, often criticizing and even despising. To show this, we have provided in the table a characteristic quotation from each about the others and, when possible, how they regarded themselves.

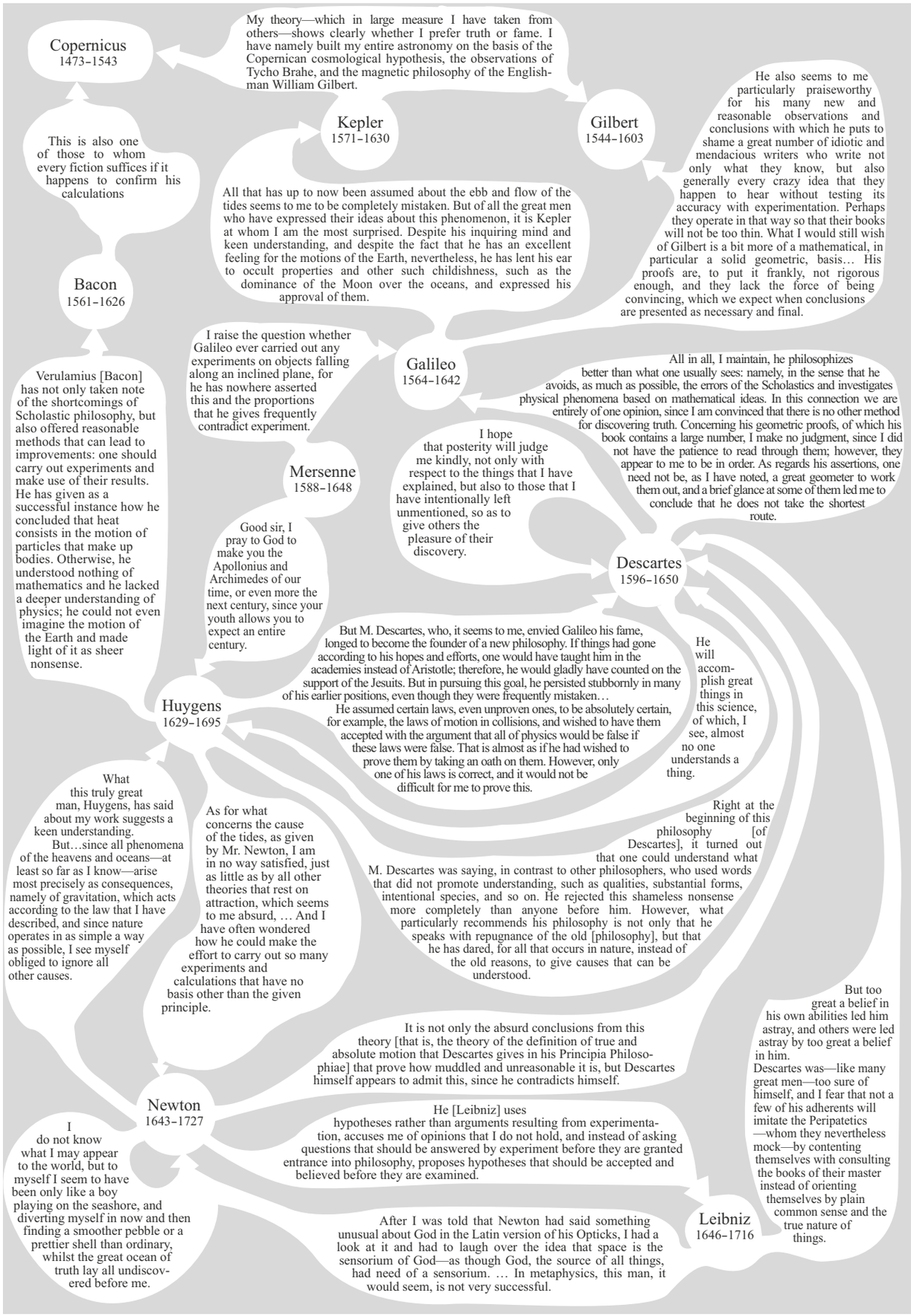
We conclude with the observations of a towering figure of the eighteenth century on two giants from the seventeenth:

The famous Newton, this destroyer of the Cartesian system, died in March, anno 1727. His countrymen honoured him in his lifetime, and interred him as though he had been a king who had made his people happy.

The English read with the highest satisfaction, and translated into their tongue, the Elogium of Newton, which Fontenelle spoke in the Academy of Sciences. Fontenelle presides as judge over philosophers; and the English expected this decision, as a solemn declaration of the superiority of the English philosophy over that of the French. But when it was found that this gentleman had compared Descartes to Newton, the whole Royal Society in London rose up in arms. So far from acquiescing with Fontenelle's judgment, they criticised his discourse. And even several (who, however, were not the ablest philosophers in that body) were offended at the comparison, and for no other reason but because Descartes was a Frenchman.

It must be confessed that these two great men differed very much in conduct, in fortune, and in philosophy.

Nature had indulged Descartes with a shining and strong imagination, whence he became a very singular person both in private life and in his manner of reasoning. This imagination could not conceal itself even in his philosophical works, which are everywhere adorned with very shining, ingenious metaphors and figures. Nature had almost made him a poet; and indeed he wrote



◀ **Table 3.4** How the great figures of the seventeenth century judged one another.

Quotation 3.54

Hitherto we have explained the phenomena of the heavens and of our sea by the power of gravity, but have not yet assigned the cause of this power. This is certain, that it must proceed from a cause that penetrates to the very centres of the sun and planets, without suffering the least diminution of its force; that operates not according to the quantity of the surfaces of the particles upon which it acts (as mechanical causes used to do), but according to the quantity of the solid matter which they contain, and propagates its virtue on all sides to immense distances, decreasing always as the inverse square of the distances. Gravitation towards the sun is made up out of the gravitations towards the several particles of which the body of the sun is composed; and in receding from the sun decreases accurately as the inverse square of the distances as far as the orbit of Saturn, as evidently appears from the quiescence of the aphelion of the planets; nay, and even to the remotest aphelion of the comets, if those aphelions are also quiescent.

But hitherto I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypotheses, for whatever is not deduced from the phenomena is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy. In this philosophy particular propositions are inferred from the phenomena, and afterwards rendered general by induction. Thus it was that the impenetrability, the mobility, and the impulsive force of bodies, and the laws of motion and of gravitation, were discovered. And to us it is enough that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea.

—Isaac Newton, *Principia*, preface to the first edition, 1638

a piece of poetry for the entertainment of Christina, Queen of Sweden, which however was suppressed in honour to his memory.

He embraced a military life for some time, and afterwards becoming a complete philosopher, he did not think the passion of love derogatory to his character. He had by his mistress a daughter called Froncine, who died young, and was very much regretted by him. Thus he experienced every passion incident to mankind.

He was for a long time of opinion that it would be necessary for him to fly from the society of his fellow creatures, and especially from his native country, in order to enjoy the happiness of cultivating his philosophical studies in full liberty.

Descartes was very right, for his contemporaries were not knowing enough to improve and enlighten his understanding, and were capable of little else than of giving him uneasiness.

He left France purely to go in search of truth, which was then persecuted by the wretched philosophy of the schools. However, he found that reason was as much disguised and depraved in the universities of Holland, into which he withdrew, as in his own country, for at the time that the French condemned the only propositions of his philosophy which were true, he was persecuted by the pretended philosophers of Holland, who understood him no better; and who, having a nearer view of his glory, hated his person the more, so that he was obliged to leave Utrecht. Descartes was injuriously accused of being an atheist, the last refuge of religious scandal: and he who had employed all the sagacity and penetration of his genius, in searching for new proofs of the existence of a God, was suspected to believe there was no such Being. ...

At last Descartes was snatched from the world in the flower of his age at Stockholm. His death was owing to a bad regimen, and he expired in the midst of some literati who were his enemies, and under the hands of a physician to whom he was odious.

The progress of Newton's life was quite different. He lived happy, and very much honoured in his native country, to the age of fourscore and five years.

It was his particular felicity, not only to be born in a country of liberty, but in an age when all scholastic impertinences were banished from the world. Reason alone was cultivated, and mankind could only be his pupil, not his enemy.

One very singular difference in the lives of these two great men is that Newton, during the long course of years he enjoyed, was never sensible to any passion, was not subject to the common frailties of mankind, nor ever had any commerce with women—a circumstance which was assured me by the physician and surgeon who attended him in his last moments.

We may admire Newton on this occasion, but then we must not censure Descartes.

The opinion that generally prevails in England with regard to these new philosophers is that the latter was a dreamer, the former a sage.

Very few people in England read Descartes, whose works indeed are now useless. On the other side, but a small number peruse those of Newton, because to do this the student must be deeply skilled in the mathematics, otherwise those works will be unintelligible to him. But notwithstanding this, these great men are the subject of everyone's discourse. Newton is allowed every advantage, whilst Descartes is not indulged a single one. According to some,

it is to the former that we owe the discovery of a vacuum, that the air is a heavy body, and the invention of telescopes. In a word, Newton is here as the Hercules of fabulous story, to whom the ignorant ascribed all the feats of ancient heroes.

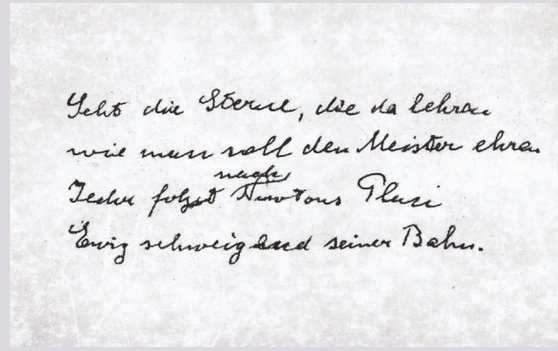
In a critique that was made in London on Fontenelle's discourse, the writer presumed to assert that Descartes was not a great geometrician. Those who make such a declaration may justly be reproached with flying in their master's face. Descartes extended the limits of geometry as far beyond the place where he found them, as Newton did after him. The former first taught the method of expressing curves by equations. This geometry which, thanks to him for it, is now grown common, was so abstruse in his time, that not so much as one professor could undertake to explain it; and Schotten in Holland, and Fermat in France, were the only men who understood it. ...

Geometry was a guide he himself had in some measure fashioned, which would have conducted him safely through the several paths of natural philosophy. Nevertheless, he at last abandoned this guide, and gave entirely into the humour of forming hypotheses; and then philosophy was no more than an ingenious romance, fit only to amuse the ignorant. He was mistaken in the nature of the soul, in the proofs of the existence of a God, in matter, in the laws of motion, and in the nature of light. He admitted innate ideas, he invented new elements, he created a world; he made man according to his own fancy; and it is justly said, that the man of Descartes is, in fact, that of Descartes only, very different from the real one.

He pushed his metaphysical errors so far, as to declare that two and two make four for no other reason but because God would have it so. However, it will not be making him too great a compliment if we affirm that he was valuable even in his mistakes. He deceived himself, but then it was at least in a methodical way. He destroyed all the absurd chimeras with which youth had been infatuated for two thousand years. He taught his contemporaries how to reason, and enabled them to employ his own weapons against himself. If Descartes did not pay in good money, he however did great service in crying down that of a base alloy.

I indeed believe that very few will presume to compare his philosophy in any respect with that of Newton. The former is an essay, the latter a masterpiece. But then the man who first brought us to the path of truth was perhaps as great a genius as he who afterwards conducted us through it.

—Voltaire, "Letters on England" ("Lettres Anglaises") XIV



▲ **Figure 3.139** It was not only his contemporaries who were in awe of Newton. This verse appears on the reverse side of a manuscript page by Einstein filled with complex formulas. Einstein wrote a number of ironic verses for his own and his friends' amusement, but this poem shows his great respect toward Newton:

Seht die Sterne, die da lehren
Wie man soll den Meister ehren
Jeder folgt nach Newtons Plan
Ewig schweigend seiner Bahn

(Look at the stars, which instruct us
How to honor the master
Each follows according to Newton's plan
Eternally silent its course)



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

CHAPTER 4

THE COMPLETION OF CLASSICAL PHYSICS

All events, even those which on account of their insignificance, do not seem to follow the great laws of nature, are a result of it just as necessarily as the revolutions of the sun. In ignorance of the ties which unite such events to the entire system of the universe, they have been made to depend upon final causes or upon hazard, according as they occur and are repeated with regularity, or appear without regard to order; but these imaginary causes have gradually receded with the widening bounds of knowledge and disappear entirely before sound philosophy, which sees in them only the expression of our ignorance of the true causes. ...

We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence. Its discoveries in mechanics and geometry, added to that of universal gravity, have enabled it to comprehend in the same analytical expressions the past and future states of the system of the world. Applying the same method to some other objects of its knowledge, it has succeeded in referring to general laws observed phenomena and in foreseeing those which given circumstances ought to produce. All these efforts in the search for truth tend to lead it back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed. This tendency, peculiar to the human race, is that which renders it superior to animals; and their progress in this respect distinguishes nations and ages and constitutes their true glory.

—Pierre Simon Laplace, *A Philosophical Essay on Probabilities* [pp. 4–5]

DOI: 10.1201/9781032697697-5
This chapter has been made available under a CC-BY-NC-ND license.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

The Completion of Classical Physics

4.1 Starting Capital for the Eighteenth Century

4.1.1 *Prior Results*

The confidence with which European intellectuals of this era looked to the future might best be characterized in the words of Fontenelle (Quotation 4.1). Important and less-important scholars expounded on the capacity of human reason to understand the laws of nature and of human society and thus to employ the forces of nature in the service of mankind; and they believed furthermore that by uncovering the causes of social ills, mankind could take control of its fate. They expected that the new century would see superstition and fanaticism supplanted by reason.

If we consider the objective accomplishments on which such a belief in progress might have been based, then we may point, first of all, to the concrete results of physics: Newtonian mechanics and the comprehensive mechanistic explanation of all celestial and terrestrial phenomena. Progress in optics had not only provided the laws necessary for the construction of optical instruments, but also taken the first serious steps toward the discovery of the nature of light. But it is not only these theoretical fundamentals on which the eighteenth century would build that were impressive; the seventeenth century had also bequeathed a number of other aids to progress, both for the theoretician—for example, coordinate geometry and the calculus—and for the experimentalist—the telescope, microscope, barometer, pendulum clock, and air pump.

Learned societies were formed, such as the Royal Society in 1660 (Quotation 4.2) and the Académie des Sciences in 1666; and the first scientific journals appeared (Figure 4.1).

Nor should we forget to mention the new attitude of the natural scientists: Above all authority stand reason and experience as the supreme arbiters in the search for truth. The study of epistemology not only by scientists but also by philosophers made it possible for researchers of the eighteenth century to decide on questions of truth in a way that is fully in the spirit of our approach today.

We have already spoken extensively of the development of Newtonian mechanics as well as of the experimental apparatus that arose as a result. We have not, however, discussed the development of optics in the last decades of the seventeenth century; even though this area was intertwined with the main lines of development of physics through the personalities involved, it was of only minor importance in the construction of a new worldview. Nor have we spoken of the development of the calculus or its philosophical background in the narrow sense of the word. To round out the inventory of intellectual capital at the beginning of the eighteenth century, we shall now say a few words about these issues.

4.1.2 *Waves or Particles?*

The investigation into the nature of light touched upon our discussion only to the extent that there was no Aristotelian impediment needing to be cast aside. In other words, the study of optics had no "worldview impact"; its findings, however, were just as important for the development of science in the following centuries as those of mechanics.

Quotation 4.1

And, in effect, the chief element in Philosophy, and that which expands over all others—I am referring to the method of reasoning—has in this century perfected itself to the utmost degree. I highly doubt that the majority of people will concur with the remark I am about to make; I will nonetheless make it for those who are adept at reasoning; and I can boast that it is a sign of courage when, for the sake of the truth, one exposes oneself to the criticism of all others—whose number is assuredly not negligible. No matter what the subject, the Ancients are generally liable not to reason to the ultimate degree of perfection. Often, weak preferences, slight similarities, insubstantial witticisms, and vague and confused disquisitions pass for proofs with them; furthermore, they have no stake in what they prove: but what an Ancient demonstrates in jest would, at the present time, give great pains to a poor Modern; for what rigor do we not demand in reasoning? We wish it to be intelligible, we wish it to be accurate, we wish it to be conclusive. We would have the cold-heartedness to disentangle the slightest ambiguity, whether of ideas or of words; we would have the toughness to condemn the most ingenious thing in the world, if it did not correspond to reality. Before M. Descartes, we reasoned more complacently; the earlier centuries are very lucky not to have known that particular man. It is he, as it seems to me, who brought in this new method of reasoning, which is much more to be esteemed even than his Philosophy, which has turned out to be in large measure false or greatly uncertain, according to the very rules he taught us. Finally, there reigns, not only in our good Treatises of Physics and Metaphysics, but also in those of Religion, of Morals, and of Criticism, a precision and accuracy which, up until now, had been entirely unknown.

—Bernard le Bovier de Fontenelle, *Digression sur les Anciens et les Modernes* [p. 149; translated by Paul Franz]

Quotation 4.2

The business and design of the Royal Society is:

To improve the knowledge of naturall things, and all useful Arts, Manufactures, Mechanick practices, Engynes and Inventions by Experiments—(not meddling with Divinity, Metaphysics, Moralls, Politicks, Grammar, Rhetorick, or Logick).

To attempt the recovering of such allowable arts and inventions as are lost.

To examine all systems, theories, principles, hypotheses, elements, histories, and experiments of things naturall, mathematicall, and mechanicall,

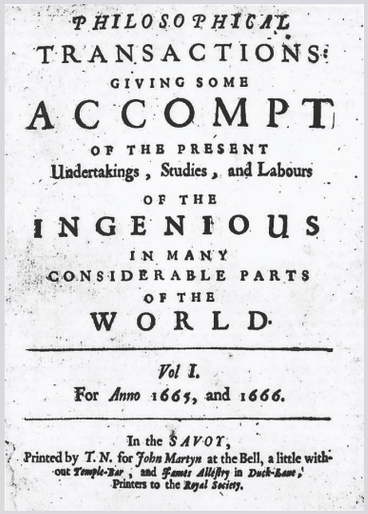
continued on next page

Quotation 4.2, *continued*

invented, recorded, or practiced, by any considerable author ancient or modern. In order to the compiling of solid philosophy for explicating all phenomena produced by nature or art, and recording a rationall account of the causes of things.

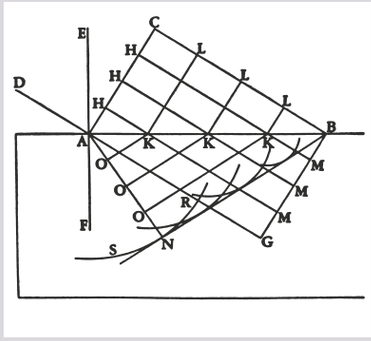
All to advance the glory of God, the honour of the King, the Royall founder of the Society, the benefit of his Kingdom, and the generall good of mankind.

—Robert Hooke, Memorandum of 1663 [Hall 1963]


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.
Vol I.
For Anno 1665, and 1666.
In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and Jamez Allestry in Duck-Lane, Printers to the Royal Society.

◀ **Figure 4.1** Title page of the first volume of Philosophical Transactions.

The first issue of *Journal des Savants* appeared in May 1665, while *Acta Eruditorum* was put out by Mencke beginning in 1682 (Leibniz was a regular contributor). The learned societies mentioned in the text have distinguished predecessors: *Academia Platonica* (Florence, around 1470, existed for half a century), *Academia antiquaria* (Rome, 1498), *Academia Secretorum Naturae* (Naples, 1560), *Accademia dei Lincei* (Rome, 1603), *Accademia del Cimento* (Florence, 1657), *Académie Française* (1635). They also had famous successors: *Kurfürstlich-Brandenburgische Sozietät der Wissenschaften* (Berliner Akademie, 1700), Swedish Academy (Uppsala, 1710), Petersburg Academy (1725).



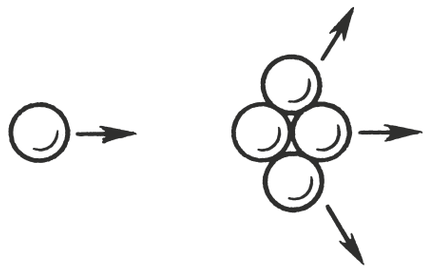
◀ **Figure 4.4** Explanation of the law of diffraction. This figure, like the previous one, is taken from Huygens's book *Traité de la lumière*, but we can find similar figures in today's physics books as well.



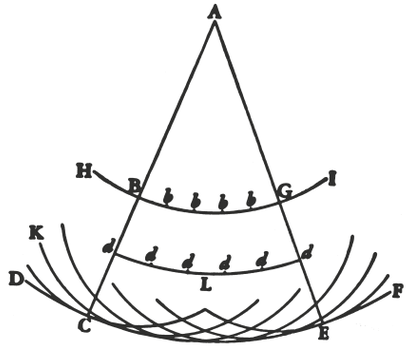
▲ **Figure 4.5** Explanation of the finite but very great speed of light.

While the two towering figures, Huygens and Newton, set the tone for work in optics, Hooke also played an important role, and perhaps his contribution is undervalued because he stood in the shadow of the two giants. Hooke himself came to sense this, for his life was one long battle, and in his bitter dispute with Newton about the theory of light, made both of their lives miserable.

Newton and Huygens each wrote a book on their work in optics, and both books belong among the classics of the history of science. The two books had a similar history even prior to their publication. Both authors began their work in the 1670s, but publication was delayed for a number of reasons. Huygens had originally written his book in French and then decided to translate it into Latin in order to make it available to a wider readership; "but the happy feeling of novelty passed, and I put off further and further the realization of my plans." His book finally appeared in 1691 under the title *Traité de la lumière*. Newton submitted his findings to the Royal Society in 1672 and 1675, but was attacked so vigorously, especially by Hooke, that he did not publish his *Opticks or a Treatise of the Reflections, Refractions and Colours of Light* until 1704, a year after Hooke's death. The investigations of Huygens and Newton are to a great extent interconnected, and indeed each refers to the other several times. Newton was of course universally hailed, and Huygens was one of the few scientists whom Newton recognized and respected.



▲ **Figure 4.2** Huygens' explanation of the propagation of light.



▲ **Figure 4.3** Explanation for why light travels in a straight line.

Today, it is customary to associate the name of Huygens with the wave theory of light, and that of Newton with the corpuscular theory. We may even read in physics textbooks that Newton's incorrect conception on the nature of light may well have retarded progress in optics for almost a century because of his great authority. And that is in fact the case; however, it is worth discussing briefly the extent to which Newton bears responsibility for such influence, as opposed those of his epigones who unduly simplified his ideas.

While Huygens criticized Descartes in a number of areas, he nevertheless remained true to the fundamental Cartesian view that every interaction is caused by a mechanical contact and that all natural phenomena must be reducible to mechanical processes. He sets forth this conception in the introduction to his book (Quotation 4.3). We note in passing that the books of both Newton and Huygens are of special interest because in each of them we can find a wealth of observations on natural philosophy.

Following Huygens, we may conceive of the propagation of light (Figure 4.2) in such a way that the light-emitting body strikes the particles in the surrounding ether, and that these then transmit their state of motion as though they were elastic spheres.

According to this model, light is transmitted much as sound is (this comparison was made by Huygens), and the ether as medium for the transmission of light corresponds to air as the medium for the transmission of sound. The great challenge in Huygens' theory is, of course, to explain why light travels in a straight line.

Huygens himself argued as follows: Previously, it had been taken as obvious that light travels in a straight line, and no explanation for the phenomenon was sought; but now his mechanical interpretation allows for an explanation. Following this, we can find in Huygens' book the illustrations for how light travels and how it refracts—illustrations that also appear today in every physics textbook (Figures 4.3, 4.4). According to Huygens' conception, every point on a developing wave front can be viewed as the midpoint of an elementary wave expanding in all directions, and the envelope of all the elementary waves forms a new wave front. To explain the law of refraction, we consider the boundary between the different materials as the series of midpoints from which the new elementary waves expand in succession, corresponding to the times of excitation.

To explain why a ray of light entering a denser medium is bent toward the normal point of incidence, Huygens was compelled—in contradiction to Descartes and Newton, but in agreement with Fermat—to assume that light is transmitted more slowly in the denser medium. Thus Huygens arrived at the form of the law of refraction given by Fermat:

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{v_1}{v_2}.$$

We may summarize the key points of Huygens' conception of light as follows:

1. Light travels in a vacuum or in air, as well as in other media, at a finite velocity. Huygens explains this assertion with a model (Figure 4.5) consisting of a row of elastic spheres. If the row of spheres is struck at one end, the impulse travels through the spheres with a finite velocity, and its effect becomes manifest when the last sphere at the other end bouncesa away. Huygens gives a detailed description of Ole Christensen Rømer's method for determining the velocity of light (Quotation 4.4; Figure 4.6; see also Figure 5.6).
2. Light consists in the propagation of a state of motion. With this thesis, Huygens broke with the Cartesian idea, according to which light was the result of pressure resulting from a vortical motion of a material medium. Huygens rejected decisively the Newtonian idea that light consists of rapidly moving corpuscles. He argued that such particles of light would mutually interact if their directions of motion were opposed or if their paths crossed, so that, for example, two people looking at each other would not be able to see each other at the same time. Using the device depicted in Figure 4.5, Huygens showed that in accord with his theory, light can propagate in the same medium in both directions at the same time; if the row of spheres is struck at both ends simultaneously, the impulse from the right end will reach the left end just as the effect of the leftmost sphere reaches the right end.

The most significant accomplishment of Huygens' theory of light—which we may call, with only slight exaggeration, the greatest result of seventeenth-century physics if we are considering only explanations of concrete complex phenomena—is his explanation of double refraction. In 1669, Bartholin was the first to report on the interesting property of Iceland spar (a form of calcite) that objects viewed through this transparent crystal appear double. To explain the phenomenon, Huygens assumed

Quotation 4.3

It is inconceivable to doubt that light consists in the motion of some sort of matter. For whether one considers its production, one sees that here upon the Earth it is chiefly engendered by fire and flame which contain without doubt bodies that are in rapid motion, since they dissolve and melt many other bodies, even the most solid; or whether one considers its effects, one sees that when light is collected, as by concave mirrors, it has the property of burning as a fire does, that is to say it disunites the particles of bodies. This is assuredly the mark of motion, at least in the true Philosophy, in which one conceives the causes of all natural effects in terms of mechanical motions. This, in my opinion, we must necessarily do, or else renounce all hopes of ever comprehending anything in physics.

And as, according to this Philosophy, one holds as certain that the sensation of light is excited only by the impression of some movement of a kind of matter which acts on the nerves at the back of our eyes, there is here yet one reason more for believing that light consists in a movement of the matter which exists between us and the luminous body.

Further, when one considers the extreme speed with which light spreads on every side, and how, when it comes from different regions, even from those directly opposite, the rays traverse one another without hindrance, one may well understand that when we see a luminous object, it cannot be by any transport of matter coming to us from this object, in the way in which a shot or an arrow traverses the air; for assuredly that would too greatly impugn these two properties of light, especially the second of them. It is then in some other way that light spreads; and that which can lead us to comprehend it is the knowledge which we have of the spreading of sound in the air.

We know that by means of the air, which is an invisible and impalpable body, sound spreads around the spot where it has been produced, by a movement which is passed on successively from one part of the air to another; and that the spreading of this movement, taking place equally rapidly on all sides, ought to form spherical surfaces ever enlarging and which strike our ears. Now there is no doubt at all that light also comes from the luminous body to our eyes by some movement, impressed on the matter which is between the two; since, as we have already seen, it cannot be by the transport of a body which passes from one to the other. If, in addition, light takes time for its passage—which we are now going to examine—it will follow that this movement, impressed on the intervening matter, is successive; and consequently it spreads, as sound

continued on next page

Quotation 4.3, ***continued***

does, by spherical surfaces and waves: for I call them waves from their resemblance to those which are seen to be spread in water when a stone is thrown into it, and which present a successive spreading as circles, though these arise from another cause, and are only in a flat surface.

To see then whether the spreading light takes time, let us consider first whether there are any facts of experience which can convince us to the contrary. As to those which can be made here on the Earth, by striking lights at great distances, although they prove that light takes no sensible time to pass over these distances, one may say with good reason that they are too small, and that the only conclusion to be drawn from them is that the passage of light is extremely rapid. Mr. Descartes, who was of opinion that it is instantaneous, founded his views, not without reason, upon a better basis of experience, drawn from the eclipses of the Moon; which, nevertheless, as I shall show, is not at all convincing. I will set it forth, in a way a little different from his, in order to make the conclusion more comprehensible.

...

It is true that we are here supposing a strange velocity that would be a hundred thousand times greater than that of sound. For sound, according to what I have observed, travels about 180 toises in the time of one second, or in about one beat of the pulse. But this supposition ought not to seem to be an impossibility; since it is not a question of the transport of a body with so great a speed, but of a successive movement which is passed on from some bodies to others.

—Christiaan Huygens, *Treatise on Light* [pp. 3–5, 7]

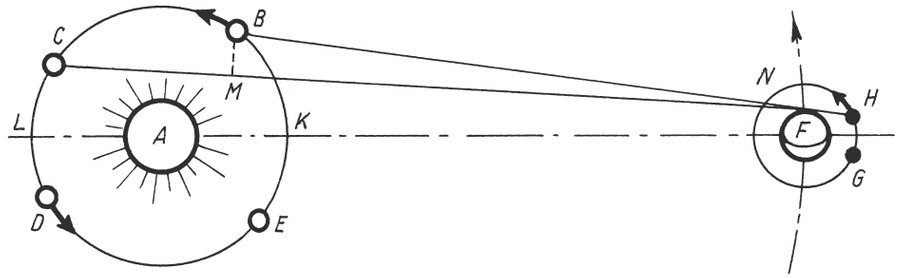
Quotation 4.4

But that which I employed only as a hypothesis, has recently received great seemingness as an established truth by the ingenious proof of Mr. Römer which I am going here to relate, expecting him himself to give all that is needed for its confirmation. It is founded as is the preceding argument upon celestial observations, and proves not only that light takes time for its passage, but also demonstrates how much time it takes, and that its velocity is even at least six times greater than that which I have just stated.

For this he makes use of the eclipses suffered by the little planets which revolve around Jupiter, and which often enter his shadow: and see what is his reasoning. Let A be the Sun, BCDE the annual orbit of the Earth, F Jupiter, GN the orbit of the nearest

continued on next page

that there are two forms of light transmission in this substance: In the ordinary ray, light is transmitted as in every other medium (that is, the elementary wave surfaces are spheres), while in the extraordinary ray, the transmission velocity of the state of motion proceeding from the centers of activation depends on the direction, so that the elementary wave surfaces end up as ellipsoids instead of spheres. If we draw, as in Figure 4.7, the envelope of all elementary wave surfaces of the extraordinary wave that are generated perpendicularly onto the surface of the crystal, we observe that this envelope remains parallel to the wave surface of the incident wave; but this shows that upon exiting the crystal, the extraordinary wave is shifted through a certain distance with respect to the incident wave. However, the directions of the incident wave and of the extraordinary wave exiting the crystal are again the same. Huygens pursued this simple interpretation of the effect further and finally suggested that the peculiarities of the transmission of light can be traced to the particular asymmetry of the crystal. Figure 4.8 is taken from Huygens' book, and we once again see that it looks just like a figure from a modern book on solid-state physics.



▲ **Figure 4.6** Rømer's ideas for measuring the speed of light as depicted by Huygens.

Huygens himself speaks in his book of waves, and today, we call Huygens' theory a wave theory of light. This is indeed justified if by "wave" we understand a transmission of some characteristic of a substance (here, its state of motion) in which the substance itself is not transmitted. We immediately see that Huygens' theory describes *longitudinal* waves, where the direction of motion of the particles responsible for the transmission of the wave agrees with the direction of motion of the wave itself. However, we know today that light is a transverse wave in which the oscillations (of the electromagnetic field) take place perpendicularly to the direction of transmission. In characterizing a wave, it is usual to give its spatial and temporal periods; the spatial period is given by the wavelength, and the temporal period is given by the oscillation period or by the associated frequency. A simple interpretation of the interference phenomenon characteristic of the wave nature of light can be given only with the knowledge of these quantities, but in Huygens' theory, we look for them in vain. With him, the impacts do not occur in a determined rhythm, but irregularly, so that we are not in fact dealing with waves comprising harmonic (that is, sinusoidal) wave trains whose interference could be used to explain a great variety of phenomena.

Newton gave considerable attention to the color spectrum in his optical research into the theory of color. The first part of his book is devoted to the characteristics of the spectrum as produced by a prism; the second part deals with the color phenomenon of thin films that is now called Newton's rings. In the third part, Newton briefly discusses diffraction phenomena; here we will also find, offered in the form of questions, all of Newton's thoughts on the nature of light, and in addition, some general thoughts about natural philosophy.

While the *Principia* contains a great number of mathematical derivations, it also contains the results of various experiments carried out with great care and described in great detail. It is no accident that Newton is viewed, particularly in his native land, as the greatest of the theoretical physicists, while he is also valued, particularly on the Continent, as a first-rate experimentalist.

His experiments on the colors of the spectrum have found their way into today's textbooks, where we may find the very same figures as in Newton's *Opticks*. Before we discuss these experiments, let us turn our attention briefly to Robert Hooke (1635–1703) and his ideas on the origin of color. We must first mention that Marco Antonio de Dominis (1611) and Jan Marek Marci (1648) performed experiments with light analogous to those carried out by Newton. They placed a prism in the path of a ray of light entering a darkened room through a small opening and investigated the resulting spectrum. Francesco Maria Grimaldi (1650) also observed the diffraction of light and the associated color phenomena and critically evaluated all the current theories on the origin of color. He considered most promising the model in which colors arise from the variety of motions of a very fine fluid.

Vibrations played an important role in the work of Hooke, and that he failed to achieve any lasting results in this area may be attributed to the fact that his exceptional experimental talent and physical sense (Figure 4.9) was not accompanied by a commensurate mathematical education and facility. Hooke saw both light and heat as oscillatory phenomena, and he also espoused the idea that the propagation of light entails a vibrational state. Huygens saw no way of accounting for color from this viewpoint; nevertheless, Hooke gave an interesting interpretation of the appearance of a color spectrum when light passes through a prism. As depicted in Figure 4.10, light strikes the interface between the two media (to this extent, the figure corresponds precisely to Huygens' idea), where it is diffracted. In his description of diffraction, Hooke, in contrast, took Descartes's point of view and assumed that the speed of propagation of light waves in the denser medium is greater than that in air, with the result that the normal to the wave surfaces is diffracted away from the normal to the interface between the media. Hooke knew, of course, that the light ray is diffracted toward the normal and therefore argued that in the case of a diffracted ray of light, the wave surface and direction of propagation form an acute angle, and it is this angle that can be made to account for the appearance of color: Where the wave surface moves ahead on the border of the ray, blue appears; where it holds back, red results.

Hooke quite correctly recognized that the color phenomena observable on thin films can be explained by the interaction between the light being reflected from the upper and lower layers. Beyond that, however, his ideas are completely off the mark. Today we know that what is going on in this case is simply the interference of the two light waves.

Newton originally became interested in colors because he noticed that the most annoying flaw of telescopes was a colored edge around the images. His goal was to construct a telescope without such problems.

In contrast to his predecessors, he chose a greater distance between a light-diffracting prism and the screen for observing the spectrum (22 feet) and thereby obtained a broad spectral image. Newton first noticed how the image of a light beam that is originally circular in cross section is not a circle after it is diffracted through the prism. Based on the experiments of his predecessors, he expected that on passing through a prism, the light beam would be diffracted according to the law of diffraction and would create a circular multicolored spot on the screen,

Quotation 4.4, ***continued***

of his satellites, for it is this one which is more apt for this investigation than any of the other three, because of the quickness of its revolution. Let G be this satellite entering into the shadow of Jupiter, H the same satellite emerging from the shadow.

Let it then be supposed, the Earth being at B some time before the last quadrature, that one has seen the said satellite emerge from the shadow; it must needs be, if the Earth remains at the same place, that, after 42½ hours, one would again see a similar emergence, because that is the time in which it makes the round of its orbit, and when it would come again into opposition to the Sun. And if the Earth, for instance, were to remain always at B during 30 revolutions of this satellite, one would see it again emerge from the shadow after 30 times 42½ hours. But the Earth having been carried along during this time to C, increasing thus its distance from Jupiter, it follows that if light requires time for its passage the illumination of the little planet will be perceived later at C than it would have been at B, and that there must be added to this time of 30 times 42½ hours that which the light has required to traverse the space MC, the difference of the spaces CH, BH. Similarly at the other quadrature when the earth has come to E from D while approaching toward Jupiter, the immersions of the Satellite ought to be observed at E earlier than they would have been seen if the Earth had remained at D.

Now in quantities of observations of these eclipses, made during ten consecutive years, these differences have been found to be very considerable, such as ten minutes and more; and from them it has been concluded that in order to traverse the whole diameter of the annual orbit KL, which is double the distance from here to the sun, Light requires about 22 minutes of time.

The movement of Jupiter in his orbit while the Earth passed from B to C, or from D to E, is included in this calculation; and this makes it evident that one cannot attribute the retardation of these illuminations or the anticipation of the eclipses, either to any irregularity occurring in the movement of the little planet or to its eccentricity.

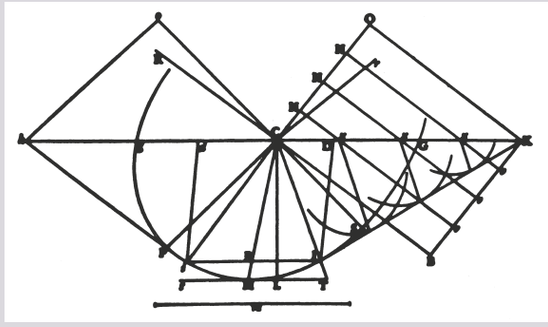
If one considers the vast size of the diameter KL, which according to me is some 24 thousand diameters of the Earth, one will acknowledge the extreme velocity of light. For, supposing that KL is no more than 22 thousand of these diameters, it appears that being traversed in 22 minutes this makes the speed a thousand diameters in one minute, that is 16⅔ diameters in one second or in one beat of the pulse, which makes more than 11

continued on next page

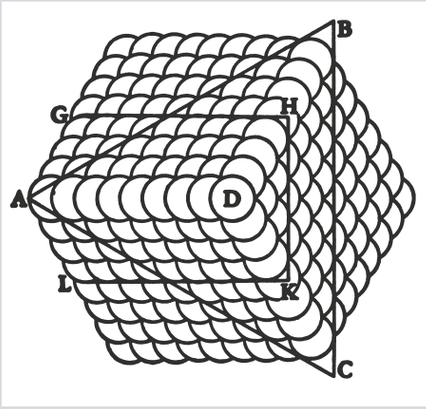
Quotation 4.4, ***continued***

hundred times a hundred thousand toises; since the diameter of the Earth contains 2,865 leagues, reckoned at 25 to the degree, and each league is 2,282 toises, according to the exact measurement which Mr. Picard made by order of the King in 1669. But sound, as I have said above, only travels 180 toises in the same time of one second: hence the velocity of light is more than six hundred thousand times greater than that of sound. This, however, is quite another thing from being instantaneous, since there is all the difference between a finite thing and an infinite. Now the successive movement of light being confirmed in this way, it follows, as I have said, that it spreads by spherical waves, like the movement of sound.

—Christiaan Huygens, *Treatise on Light* [pp. 7–10]



▲ **Figure 4.7** On the origin of the extraordinary ray, after Huygens.



▲ **Figure 4.8** The structure of Icelandic spar, after Huygens. This figure could as well be found in a modern text on crystallography.

given that the index of refraction was fixed for a given medium and for "light," by which sunlight was always meant. However, instead of a circular spot, Newton obtained a stripe, blurred at the top and bottom, which, as he put it, "appeared to take the form of a semicircle." It is worth mentioning that the length of the colored stripe was 13½ inches, and its width was 2⅝ inches. A series of experiments led Newton to the following conclusions:

1. White light is composed of light of various colors. It is not the case, as was previously supposed, that the material of which the prism is made affects the originally homogeneous and uniform white light in such a way that a spectrum results; rather, compound white light is broken by the prism into its components (Figure 4.11).
2. Spectral colors are homogeneous in the sense that they cannot be further decomposed; that is, if one isolates a particular color from the spectrum, then this color cannot be further broken down into a spectrum.
3. By mixing the elementary colors, various colors can be obtained that are no longer homogeneous. And of course, a mixture in the correct proportions will reproduce the original white light.
4. There is a unique relationship between the index of refraction and the various colors; that is, light rays have differing indices of refraction corresponding to their color. With these indices, the diffraction of a compound ray of light consisting of prescribed colors can be described for a given angle of incidence.

From among Newton's experiments we would like to highlight only a few characteristic ones. In one experiment, which Newton himself described as crucial (*experimentum crucis*), light that has been broken up by a prism into its spectral colors is allowed to pass through another prism. In this way, it was possible to show that a particular spectral color cannot be further divided into additional colors (Figure 4.11). With another experiment (Figure 4.12), it was possible to demonstrate that there is a fixed relationship between the color and the index of refraction for a given medium.

In a third experiment, Newton unites the spectral colors by employing a prism in reverse and thus proves that the mixing of colors can result in the re-formation of white light (Figure 4.13).

It is worth noting that Newton was unable to achieve his original goal of eliminating the chromatic aberration in his telescope; in fact, he believed that he had proved that the problem was insoluble. Namely, from his experimental data he drew the general conclusion that a medium with a large capacity for refraction, to wit, one with a large index of refraction, will also strongly break up white light into its spectral components, so that the spectrum is broadly spread out. In today's terminology, we would say that Newton assumed that the index of refraction and degree of dispersion (spectral decomposition) are proportional. If this assumption were correct, then indeed it would be impossible to construct an achromatic system of lenses (that is, one without chromatic aberration). After Euler pointed out the incorrectness of this assumption, the English lens maker J. Dollond in 1760 succeeded in constructing an achromatic pair of lenses. Newton, on his part, was inspired by his erroneous assumption to a successful workaround: To exclude chromatic aberration, he constructed a reflecting telescope (Figure 4.14).

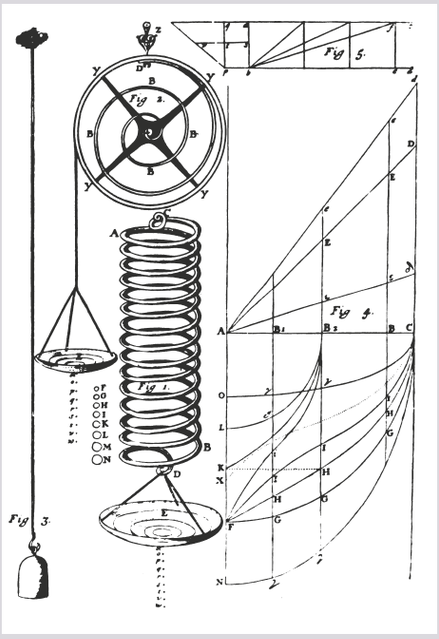
From the point of view of the history of physics, the observations of color phenomena on thin films are of even greater significance. It is well known how a drop of oil on water can form a thin film that shimmers in a rainbow of colors. Newton

created thin layers of air of various thicknesses by placing a slightly curved lens on a flat glass plate. The image that we perceive above, that is, in the reflected light, is different from the image that may be observed in the light that passes through (that is, from below), even though, of course, the two images are closely related. In any case, we perceive light and dark or colored rings that are concentric around the lens's point of contact. These are now called Newton's rings. Figure 4.15 is taken from Newton's *Opticks*. As understood today, the phenomenon is the result of the interference of the reflected light waves from the two surfaces. Newton investigated the relationship between the position of the individual rings and the thickness of the layer of air, selecting out a particular color in the reflected light. He found that the relationship between the locations of the rings for a given color and the thicknesses of the layers is given by the odd integers. Conversely, the rings for which color was absent belong to layer thicknesses corresponding to the even numbers. To explain this phenomenon, Newton assumed that a ray of light of a given color, as shown in Figure 4.15, is reflected at particular points, while at other places it passes through the layer of air. Newton then concluded that in its propagation, light periodically assumes various states in which its passage into a new medium is facilitated or hindered. Here Newton uses the word *fit*, meaning a state of adaptation.

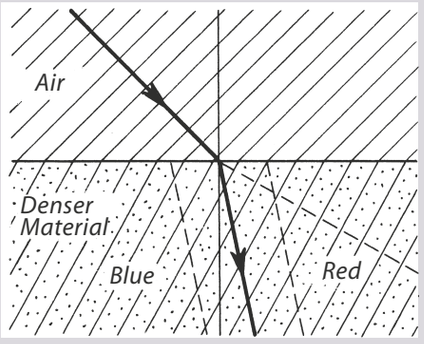
He was very cautious in conjecturing how such states arise (Quotation 4.5), and we are going beyond Newton when we consider Figure 4.16 as representing these states. Suppose a moving body in the form of a rugby ball rotates about an axis perpendicular to the direction of motion, so that the narrow and wide sides alternately point toward the direction in which the body is moving. In the first case, the body has a greater power of penetration; in the second case, a smaller. What is significant here is not the model, but that Newton, through experimentation, could conclude that light has a property that implies spatial periodicity, and, as he cautiously expressed in his thirteenth proposition, he thought that this periodic state is already present from the moment of emission. From his experimental results, Newton was able to give the measure of the periodicity for yellow light (Quotation 4.6), and as we see, this is in close agreement with the wavelength for yellow light.

A further surprising result: Newton also found how this periodicity changes with the index of refraction. Using today's terminology, he determined that the wavelength is inversely proportional to the index of refraction. Despite the recognition of the spatial periodicity, why was it that Newton couldn't arrive at the notion of wavelength and the wave theory of light in today's sense of the word, or more precisely, in that of the nineteenth century? The reason may be that Newton, together with Descartes, believed that in denser materials light travels *faster*, explaining the diffraction of light toward the normal to the interface by the stronger attraction of the denser material and the resulting accelerative effect. But the decrease of spatial periodicity in the denser material, which Newton correctly determined, is consistent only with a *decrease* of the velocity under the wave-theoretic equation $v = f\lambda$.

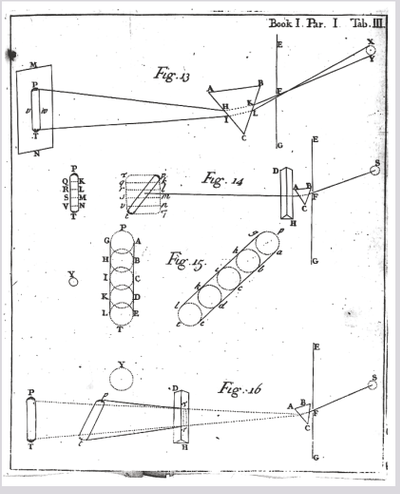
Newton, too, worked with Iceland spar crystals. Unlike Huygens, he was unable to offer a convincing explanation for the appearance of two distinct beams, but he further investigated the properties of the beams emitted from the crystal by allowing them to fall on a second such crystal. He discovered that for a particular orientation of the second crystal, the ordinary and extraordinary beams remain as such, while for a different orientation, the roles of the two beams are exchanged. The propagation of waves in crystals remains a complex problem even today, and



▲ **Figure 4.9** The law relating force and elastic extension is rightly called Hooke's law. A page from Hooke's 1687 *Lectures de Potentia Restitutiva*.



▲ **Figure 4.10** On the origin of color according to Hooke's theory. The basic idea here is original, but erroneous.

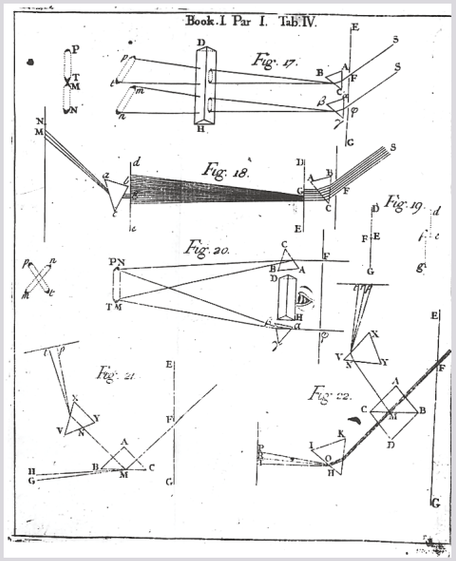


◀ **Figure 4.11** The origin of the spectrum (Fig. 13) and the *experimentum crucis*, the decisive experiment (Figs. 14 and 15): The prism ABC decomposes white light; bundles of a particular color are refracted by the prism DH but are not further split.

continued on next page

Figure 4.11 *continued*

> The Newtonian idea [on the origin of color] appears on calm examination so paradoxical, contradicting what is convincing from the immediate observation of nature, that one can hardly believe that it could spring from the best brain of his century, develop there, occupy him for an entire lifetime, and then take root in first-rate brains of posterity. One could almost dejectedly maintain that we are born to error; but it is in fact the great productive and constructive power of mankind that reveals itself in action here. For just as he reshapes mountains in order to erect palaces according to his own ideas, fells forests to furnish and roof his buildings, just so does the physicist make himself lord over nature's phenomena, collects experience, hammers and screws it together through artificial experiments and finally there stands a building in honor of its architect; only we meet the bold assertion that this too is but nature, at least with a silent smile, a soft shake of the head. After all, it never occurs to the architect to pretend that his palaces are really mountains and forests. …
>
> The entire first two books of Newton's *Opticks*—for it is in these that his teachings are most thoroughly presented—are a fundamentally untruthful, artificially confabulated hocus pocus, an affront to the eyes and the judgment.
>
> —Goethe, *Die Schriften zur Naturwissenschaft* [pp. 118, 140]



▲ **Figure 4.12** Newton's experiment to determine the relationship between the index of refraction and color (Fig. 18). By turning the prism *ABC*, one can vary the color of the light that falls on the prism *abc*, and the deflection can be measured for each color.

we need not go into detail here. However, we would like to emphasize the practical significance of this experiment—for what we have here is the first appearance of the arrangement of a polarizer and an analyzer—and Newton's conclusions (Quotation 4.7), namely, that light is not symmetric with respect to its direction of propagation, with particular planes being distinguished. Today, we say that light can have a plane of polarization.

From the outset, Newton rejected Descartes's ideas about the nature of light, believing that light could not heat bodies merely by the application of pressure, rather that there must be some kind of motion involved. He also discarded Huygens' theory because that would have required the assumption of the existence of a medium of transmission, which was incompatible with the experimental observation that the celestial bodies move without any hindrance. For these reasons, Newton found only the corpuscular theory of light to be appealing (Quotation 4.8). We can see how tentatively Newton presents his thoughts in the form of questions. But he was certain that the corpuscles had their fit and unfit states, a spatial periodicity, and even a direction of polarization. Perhaps we are too greatly influenced by the physics of the twentieth century when we say that Newton's conception comes closest to what we today call the dual nature of light. The generation following Newton simplified his ideas and pictured light simply as a streaming forth of smooth spheres. They abandoned Newton's skeptical stance and ignored the most valuable conclusions regarding special periodicity. When a century later, the English physicist Young became the first to revive the wave theory—now, however, under the assumption of harmonic waves—and also gave interference its rightful important place, he was accused of lack of respect toward Newton. To counter such charges, Young collected those passages in the *Opticks* that demonstrate that Newton himself was not an opponent of the wave conception of light.

We have quoted here extensively from Newton's *Opticks* in order to show that we are not reading anything into his work—even out of respect—when we maintain that Newton made a contribution to the exploration of the wave nature of light.

Thus, at the beginning of the eighteenth century, researchers in the field of optics could choose from among three theories on the nature of light:

1. Descartes's theory, in which the triggering of the perception of light is ascribed to a pressure originating in the vortical motion of fine particles of matter that fill all of space.
2. Huygens' theory, according to which the propagation of light is accounted for by the propagation of a state of motion resulting from collisions among the particles of the ether.
3. The corpuscular theory, whereby light is caused by the motion of particles of light, or corpuscles, at a particular velocity through empty space (Figure 4.17).

To see how these theories play out, let us consider the law of refraction, which has a common explanation in the theories of Descartes and Newton, according to which the velocity of light in a dense medium is greater than that in air, whereas Huygens' theory yields exactly the opposite result (Figure 4.18). Huygens' theory agrees with experiments measuring the velocity of light, but the world would have to wait another 150 years for such experiments to be carried out successfully. Finally, we must mention Fermat's method of determining the path of light with the help of an extremal principle. Like Huygens, Fermat proceeds from the correct assumption that the speed of propagation of light in a material medium is less than that in a vacuum or in air.

In the following century, Newton's conception of light was simplified, or rather, vulgarized, and in the century after that, a refined version of Huygens' principle became the starting point for further thinking about optics. Closest to the view of our present era is the original Newtonian idea, in which the corpuscular aspect of light and its spatial periodicity must both be taken into account.

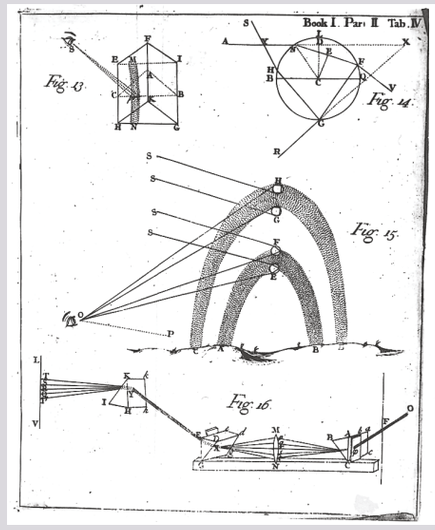
4.1.3 *Analytic Geometry*

While the two most significant areas of progress in mathematics in the seventeenth century—the introduction of coordinate geometry and the working out of the fundamentals of the calculus—were motivated by problems in the natural sciences, their full application had to wait until the eighteenth century. Almost as a summary, the fundamentals of the calculus were collected in a book by the Marquis de l'Hospital that appeared in 1696. Roughly one century earlier, the 1579 work of François Viète (1540–1603), *Canon mathematicus*, contains a collection of all the mathematical theorems and algorithms known at that time. Let us see, in broad strokes, how mathematics developed between these two dates, where we limit our attention to the branches of mathematics that had an organic relationship to the progress in the physics of the period in question. This focus is important because various other branches of mathematics—most notably probability theory, projective geometry, and mathematical logic—also were being developed in the seventeenth century but were gradually applied in the natural sciences only later.

We are reminded of the work of Descartes in mathematics primarily by the term "Cartesian coordinates." In every textbook of elementary mathematics, we are bound to find a diagram like Figure 4.19, which shows how a point in the plane can be characterized by its x- and y-coordinates; in other words, we project the segment running from the origin of the coordinate system to the given point perpendicularly to each of the coordinate axes and obtain the signed distances that are the coordinates. If we look through the book *La Géométrie*, written as an appendix to the *Discours*, which is the only published account by Descartes dealing with questions of geometry, we will be surprised to learn that in that book, despite its many figures, not a single description of the Cartesian coordinate system is to be found. We encounter similar surprises in the work of Kepler, in which his famous laws fail to make an appearance, and in Newton's *Principia*, in which the familiar form of Newton's equations is nowhere to be seen.

Consequently, as is not unusual in the history of science, the introduction of the coordinate system was accompanied by a heated battle over priority, this time between Fermat and Descartes. In 1629, Fermat wrote a work with the title *Ad locos planos et solidos isagoge*, and then, as was the custom of the time, circulated it privately among his circle of friends. It was not actually published until 1679. As we know, Descartes's book appeared in print in 1637, so that it is possible in principle at least that Descartes had prior knowledge of Fermat's results. However, today it is by and large undisputed that the two men took very different approaches to the problem, and that Descartes's book contained so much that was new in its finer points that one cannot speak of plagiarism even if one of the two had known of the work of the other before writing his own book. It is widely believed that Descartes was already at work on this problem before 1619, well before he could have known of Fermat's work.

The natural expectation that after the introduction of coordinate representation, there should follow—as in the mathematical curricula of schools today—the ana-



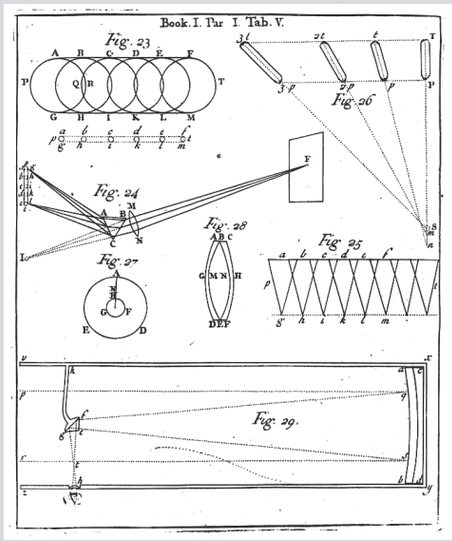
▲ Figure 4.13 The origin of the rainbow (Figs. 14 and 15) and the decomposition and restoration of white light (Fig. 16). (Newton. *Opticks*. Library of the Hungarian Academy of Sciences.)

Quotation 4.5

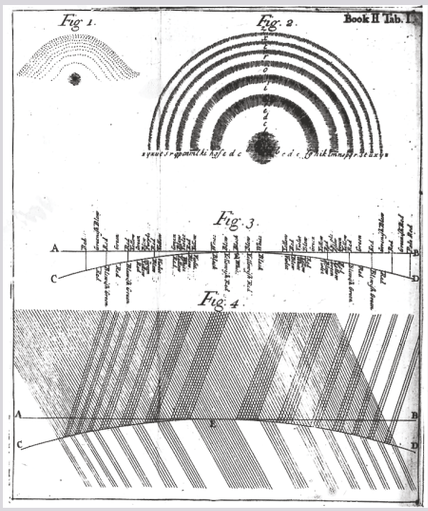
What kind of action or disposition this is; Whether it consists in a circulating or a vibrating motion of the Ray, or of the Medium, or something else, I do not here enquire. Those that are averse from assenting to any new Discoveries, but such as they can explain by a Hypothesis, may for the present suppose, that as Stones falling upon Water put the Water into an undulating Motion, and all Bodies by percussion excite vibrations in the Air; so the Rays of Light, by impinging on any refracting or reflecting Surface, excite vibrations in the refracting or reflecting Medium or Substance, and by exciting them agitate the solid parts of the refracting or reflecting Body, and by agitating them cause the Body to grow warm or hot; that the vibrations thus excited are propagated in the Air for causing Sound, and move faster than the Rays so as to overtake them; and that when any Ray is in that part of the vibration which conspires with its Motion, it easily breaks through a refracting Surface, but when it is in the contrary part of the vibration which impedes its Motion, it is easily reflected; and, by consequence, that every Ray is successively disposed to be easily reflected, or easily transmitted, by every vibration which overtakes it. But whether this Hypothesis be true or false I do not here consider. I content myself with the bare Discovery, that the Rays of Light are by some cause or other alternately disposed to be reflected or refracted for many vicissitudes. ...

And hence Light is in Fits of easy Reflexion and easy Transmission, before its Incidence on transparent Bodies. And probably it is put into such fits as its first emission from luminous Bodies, and continues in them during all its progress. For these Fits are of a lasting nature, as will appear by the next part of this Book.

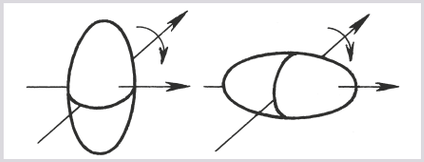
—Newton, *Opticks*, Book II, Propositions XII, XIII



▲ **Figure 4.14** Newton's reflector telescope (Fig. 29).



▲ **Figure 4.15** Newton's interpretation of the origin of color rings.



▲ **Figure 4.16** The *fit* and *unfit* states as we would interpret them today. Newton did not provide such a figure.

Quotation 4.6

If the Rays which paint the Colour in the Confine of yellow and orange pass perpendicularly out of any Medium into Air, the Intervals of their Hits of easy Reflexion are the 1/89000th part of an Inch. And of the same length are the Intervals of their Fits of easy Transmission.

—Newton, *Opticks*, Book II, Proposition XVIII

lytic representation of special curves (line, circle, the conic sections) was fulfilled by Fermat. The determination of a point in terms of a pair of coordinates was in itself nothing new; both Hipparchus and Ptolemy specified degrees of latitude and longitude in cataloguing the stars, thereby using coordinates like those used by Descartes and Fermat. Of course, what was new was not the coordinate representation per se, but the use of coordinates to translate problems from the language of geometry into that of algebra, so that well-developed algebraic methods could be used for solving concrete problems. The new representation was even more advantageous for dealing with problems that could not be solved at all by geometric means.

If, as Fermat argued, in solving a problem one arrives at an equation in two unknowns, then a value can be assigned to one of them. As shown in Figure 4.20, this quantity can be represented as the distance *OZ*. If one then substitutes this value into the equation in two unknowns, it becomes an equation in one unknown, and the unknown can then be determined. The solution obtained is marked off on the line *ZJ* that intersects the line *OZ* at some arbitrary angle. If one determines a large number of solution pairs in a similar manner, then the points thus obtained will all lie on some curve. Today, we would say that Fermat represented the curve in an oblique coordinate system. Fermat did not employ the notation introduced by Descartes, which remains today's usual notation; rather, he followed Viète and used uppercase letters, with consonants for constants and vowels for unknowns. Fermat gave the equations for a line, circle, ellipse, hyperbola, and parabola, although in an unusual form because of his notation. The equation of a line, which in today's usual notation would be written

$$y = ax, \text{ or with } a = \frac{D}{B}, \text{ as } Dx = By,$$

looks like this in Fermat's hands:

D in A aequatur B in E. Punctum J erit ad lineam rectam positione datam.

The equation of a circle, which today we recognize as

$$x^2 + y^2 = r^2, \text{ or with } B = r, \text{ as } B^2 - x^2 = y^2,$$

is given by Fermat as follows:

B quad – A quad aequatur E quad. Punctum J est ad circulum positione datum, quando angulus NZJ est rectus.

Descartes classified curves on the basis of the form of the associated equation in two unknowns. If the equation is, as we say today, algebraic, then Descartes speaks of *geometric curves*, calling other curves *mechanical*. Later, Leibniz called such nonalgebraic equations *transcendental*.

It took a relatively long time for analytic geometry to find a place in practical applications because Fermat's book was published only after a long delay. Descartes, on the other hand, deliberately left things somewhat vague in his book (Quotation 4.9). It is therefore of great significance that in 1649, van Schooten not only translated Descartes's *Géométrie* into Latin but also added commentaries and further developed Cartesian geometry. Thus, for example, he introduced the notion of coordinate transformation. He recognized the significance of this method for a simple, unified, and complete treatment of difficult geometric problems, and he went so far as to offer the opinion that the ancient Greeks must have used this method to achieve their impressive results, so that with the foreknowledge of the results they could easily apply their synthetic methods and astound the world. In his 1655 book *De sectionibus conicis*, Wallis defined the conic sections

as curves that could be described by quadratic equations, and algebraically derived their properties, which were already known from geometry. He was, moreover, the first to operate consciously with negative values of the coordinates.

A long time would pass before all the functions that appear in today's college textbooks were described. Thus, for example, the function $y = \sin x$, which had long since been specified in tables of values, appears first as the familiar curve in Wallis's 1670 work *Mechanica*, in which the two periods of the sine function are displayed.

We are familiar today with the functional notation $y = f(x)$; however, that notation was only introduced in 1734, by Euler. The notion of function itself was first encountered only in 1667, in James Gregory's book *Vera circuli et hyperboli quadratura*.

4.1.4 *Differential and Integral Calculus: The Battle of the Titans*

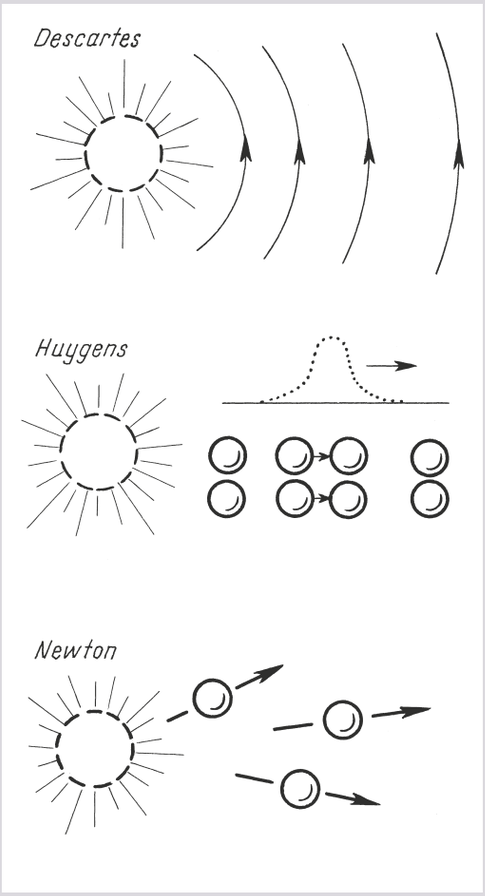
After the development of analytic geometry and the introduction of the concept of the function, it become possible with the development of differential and integral calculus to take one of the most decisive steps in the history of modern mathematics and to lay the foundations of higher analysis.

The calculus is intimately linked with a variety of questions in physics:

- *Investigation of nonuniform motion*. In motion that is not uniform, the average velocity is a well-defined measurable physical quantity. But in such motion, how is one to understand the meaning of instantaneous velocity? According to the definition of this concept, the question naturally arises of how to determine velocity and distance from acceleration, and conversely, given the distance function, how to determine velocity and acceleration. These questions had already been considered in the Middle Ages by Nicole Oresme, and at the beginning of the seventeenth century by Galileo.
- *Construction of tangent lines*. The study of tangents starts out as a problem of geometric interest, but it also has a special role in optics because the law of refraction involves the angle between the ray of light and the perpendicular, or *normal*, to the tangent to the interface surface. The construction of normals is, of course, a complementary task to the construction of tangents.
- *Determination of extreme values*. Galileo already gave a concrete answer to the question of what firing angle will result in the maximal range for a cannon shot at a given muzzle velocity if air resistance is ignored.
- *Determination of arc length, area, and volume*. Determining lengths, areas, and volumes is, of course, a problem of geometric interest. However, such questions also arise in kinematics and in the formulation of laws of motion, such as Kepler's laws.

For each of these problems, we present a characteristic attempt at a solution; then we see how, finally, Newton and Leibniz supplied a general answer to all problems of this kind.

A procedure for the construction of tangents was presented by Gilles Personne de Roberval (1602–1675) and, building on this method, by Torricelli. Both men invoked the work of Galileo on kinematics (Figure 4.21). Fermat came up with a solution that already contained the idea of the calculus in embryonic form. This method was then refined by Isaac Barrow (1630–1677), Newton's predecessor as Lucasian professor of mathematics at the University of Cambridge, and was published in his book *Lectiones geometricae* in 1670. Barrow explained



▲ **Figure 4.17** Theories of light at the beginning of the eighteenth century.

Quotation 4.7

Have not the Rays of Light several sides, endued with several original Properties? ... Every Ray of Light has therefore two opposite sides, originally endued with a Property on which the unusual Refraction depends, and the other two opposite Sides not endued with that Property. And it remains to be enquired, whether there are not more Properties of Light by which the Sides of the Rays differ, and are distinguished from one another.

—Newton, *Opticks*, Book III, Query 26

Quotation 4.8

Are not all Hypotheses erroneous, in which Light is supposed to consist in Pression or Motion, propagated through a fluid Medium? For in all these Hypotheses the Phaenomena of Light have been hitherto explain'd by supposing that they arise from new Modifications of the Rays; which is an erroneous Supposition. ...

continued on next page

Quotation 4.8, ***continued***

Are not the Rays of Light very small Bodies emitted from shining Substances? For such Bodies will pass through uniform Mediums in right Lines without bending into the Shadow, which is the Nature of the Rays of Light. They will also be capable of several Properties, and be able to conserve their Properties unchanged in passing through several Mediums, which is another Condition of the Rays of Light. Pellucid Substances act upon the Rays of Light at a distance in refracting, reflecting, and inflecting them, and the Rays mutually agitate the Parts of those Substances at a distance for heating them; and this Action and Re-action at a distance very much resembles an attractive Force between Bodies. If Refraction be perform'd by Attraction of the Rays, the Sines of Incidence must be to the Sines of Refraction in a given Proportion, as we shew'd in our Principles of Philosophy: And the Rule is true by Experience. The Rays of Light in going out of Glass into a *Vacuum*, are bent towards the Glass; and if they fall too obliquely on the *Vacuum*, they are bent backwards into the Glass, and totally reflected; and this Reflexion cannot be ascribed to the Resistance of an absolute *Vacuum*, but must be caused by the Power of the Glass attracting the Rays at their going out of it into the *Vacuum*, and bringing them back. For if the farther Surface of the Glass be moisten'd with Water or clear Oil, or liquid and clear Honey, the Rays which would otherwise be reflected will go into the Water, Oil, or Honey; and therefore are not reflected before they arrive at the farther Surface of the Glass, and begin to go out of it. If they go out of it into the Water, Oil, or Honey, they go on, because the Attraction of the Glass is almost balanced and rendered ineffectual by the contrary Attraction of the Liquor. But if they go out of it into a *Vacuum* which has no Attraction to balance that of the Glass, the Attraction of the Glass either bends and refracts them, or brings them back and reflects them.

—Newton, *Opticks*, Book III, Queries 28, 29

Quotation 4.9

As for the *analysis*, I have left out some details to make it more difficult for the malevolent. Namely, had I published my thoughts in all detail, they would have boasted that all of this had long ago been known. But this way, they dare not utter a single word so as not to reveal their ignorance.

—Descartes

that the tangent and curve segment PP' coincide (Figure 4.22) if the points P and P' are close and the curve segment is very small. If the quotient a/e is determined, then because of the similarity of the triangles $PP'R$ and NMP, the quotient PM/MN is known as well, and so, for a given PM, the location of the point N can be found. That is, the tangent can be drawn. Let us demonstrate the procedure whereby the quotient a/e is determined on the parabola

$$y^2 = px.$$

If we increase x by the quantity e, then y becomes $y + a$. Substituting these values into the equation of the parabola yields

$$y^2 + 2ay + a^2 = px + pe,$$

and since $y^2 = px$, we obtain

$$2ay + a^2 = pe.$$

The term a^2 is small to second order and can be ignored, so that we have

$$\frac{a}{e} = \frac{p}{2y}.$$

Because of the similarity of the triangles mentioned above, we then, however, have

$$NM = 2\frac{PM^2}{p},$$

and we can once again draw the tangent.

As we have seen, the quotient a/e is, up to a certain approximation, equal to the direction of the tangent, or as we would say today, to the derivative. The approach is quite similar to the presentations that one might meet—although less and less frequently—in technically oriented textbooks.

Extreme value problems appeared early in the history of mathematics. A property of extreme values first noticed by Kepler made possible an important step forward in solving such problems with the methods of higher mathematics. In 1615, Kepler published the book *Nova stereometria doliorum vinariorum*, which was a treatise on determining the volume of wine barrels. During the time he spent in Linz, Kepler decided to purchase a large quantity of wine following a good grape harvest. He thereupon became interested in the methods in use at the time for measuring the volume of barrels, and eventually, he tried to find a more precise system. His approach follows fully the Archimedean method of exhaustion, although it is somewhat sloppier, even reaching incorrect conclusions at times. In the process, the question arose as to how one could enclose the maximum volume with a given quantity of material of a specified surface area. As the simplest exercise of this kind, he looked for the square-based parallelepiped of maximal volume that can be inscribed in a sphere and showed that the answer is a cube. What is of importance for us here is Kepler's conclusion that as we approach the dimension corresponding to the maximal volume, the change in volume corresponding to a given change in the dimension becomes smaller and smaller. Today we find extreme values by looking for values of one or more variables where the derivative is zero. It is for this reason that Kepler's notes are interesting, because they imply something similar.

Fermat came even closer to the method in use today. He demonstrated his procedure with the following simple example: As in Figure 4.23, a segment of length

B is to be divided into two subsegments such that the area of the rectangle spanned by the two subsegments is maximal.

In FERMAT's demonstration of the solution, the given segment has length B, and one of the subsegments has length A, so that the area of the rectangle is

$$A(B-A) = AB - A^2.$$

We are to find the maximum such area. Let the length of the subsegment of length A be increased by E. Then the length of the other subsegment becomes smaller by E, having length $B - A - E$, and the length of the new rectangle is

$$(A+E)(B-A-E) = AB - A^2 + EB - 2EA - E^2.$$

FERMAT now argued that the two areas must be the same in a neighborhood of the maximum. This is stated a bit roughly, but it does correspond to the requirement that the derivative must vanish. The relevant equation is therefore

$$AB - A^2 + EB - 2EA - E^2 = AB - A^2,$$

from which it follows that

$$B = 2A + E.$$

If we now let E go to zero, then the condition becomes $B = 2A$. In other words, the solution to the problem requires that we divide the given segment in half in order to obtain the rectangle of maximal area, which is then, of course, a square.

At first, the method of exhaustion used by the Greeks was employed for determining arc length, surface area, and volume. As early as 1635, CAVALIERI, and in 1636, FERMAT, gave formulas for the areas under curves of the form $y = x^n$ (other than the case $n = -1$), thereby essentially discovering the relationship that we would write today as

$$\int_0^a x^n \, \mathrm{d}x = \frac{a^{n+1}}{n+1}.$$

The step forward here consists in using elementary rectangles of equal width to exhaust the area under the curve and in taking the limit, even if not rigorously. To see this, let us determine the area under the parabola $y = x^2$ using this method (Figure 4.24). The sum of the areas of the elementary rectangles shown in the figure is

$$dd^2 + d(2d)^2 + d(3d)^2 + \cdots + d(nd)^2 = d^3\left(1 + 2^2 + 3^2 + \cdots + n^2\right).$$

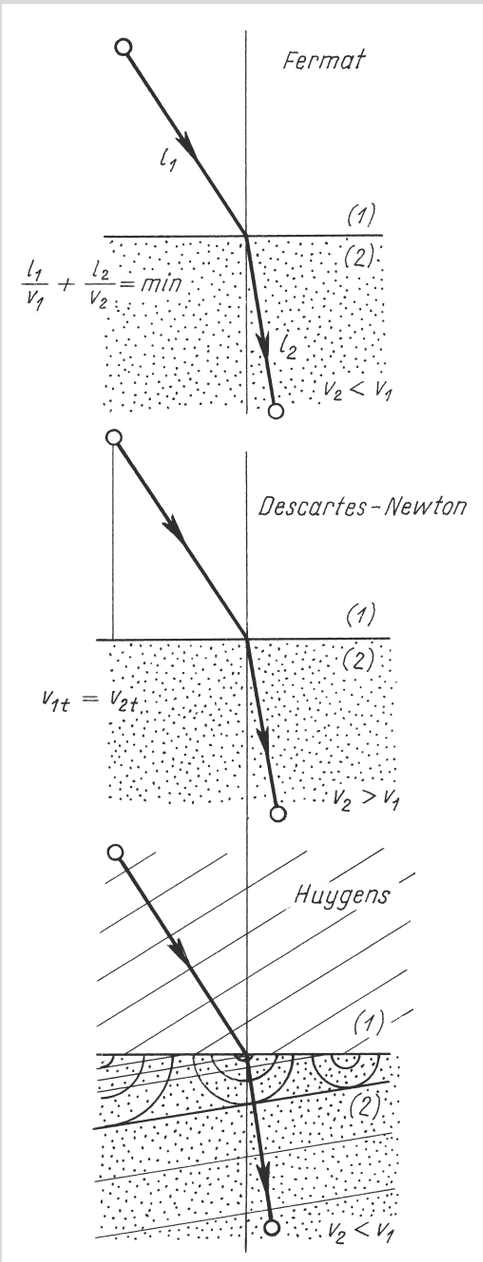
Sums of squares (or higher powers) of successive integers were either long known (see Sections 1.2, 2.3) or were calculated by FERMAT, PASCAL, and others in the process of determining the areas of regions such as the one under consideration, so that, in general, such sums presented no difficulty. In our example, we have

$$\sum_{k=1}^{n} k^2 = \frac{2n^3 + 3n^2 + n}{6},$$

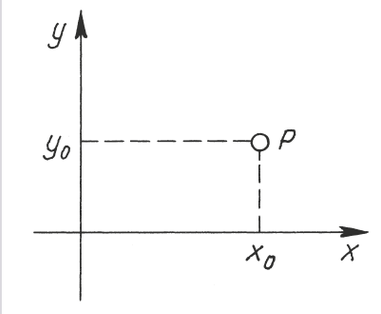
from which it follows that the sum of the elementary areas reduces to

$$a^3\left(\frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}\right), \quad d^3 = \left(\frac{a}{n}\right)^3.$$

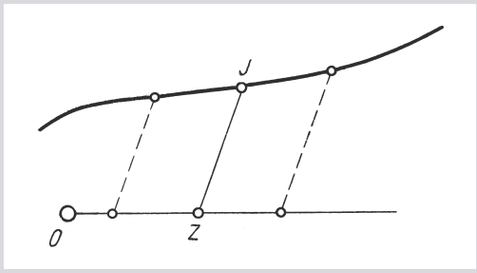
If n is taken to be a sufficiently large number, then the summands $1/(2n)$ and



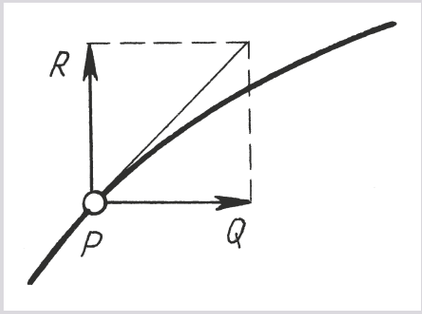
▲ **Figure 4.18** Interpretations of the law of refraction, according to different theories, each of which correctly gives the relationship $\sin \alpha_1 / \sin \alpha_2 = n$.



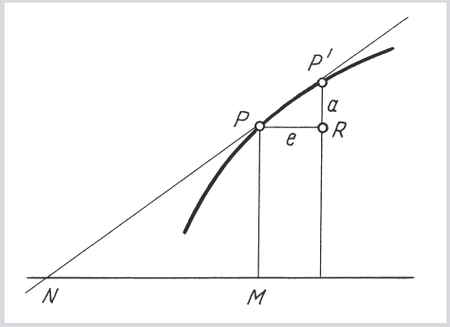
◀ **Figure 4.19** This is how the Cartesian coordinate system is introduced in today's schools.



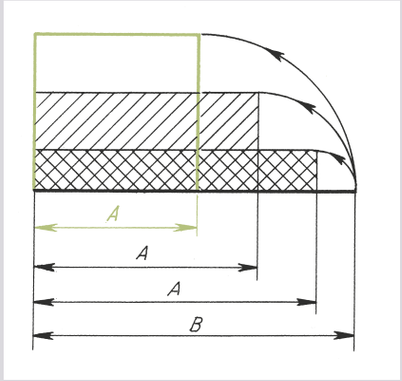
▲ **Figure 4.20** FERMAT employed—if we use today's terminology—an oblique-angled coordinate system to represent associated pairs of an equation in two unknowns.



▲ **Figure 4.21** ROBERVAL's method for constructing a tangent to a curve.



▲ **Figure 4.22** BARROW's characteristic triangle *PP'R*, in which the side *PP'* is simultaneously an arc of the curve and a segment of the tangent line.



▲ **Figure 4.23** FERMAT's extremal value problem: The length *B* is to be divided into two segments in such a way that the area of the rectangle determined by the two segments is a maximum.

$1/(6n^2)$ are small in comparison to 1/3 and may be ignored. We thus finally obtain the known result

$$\frac{a^3}{3}.$$

It was of decisive importance for the further development of this idea that the two problems we have just discussed—construction of tangents and determination of areas—are not independent of each other because, speaking in today's language, differentiation and integration may be seen as inverse operations. This insight may be found in the 1668 book *Geometriae pars universalis* by JAMES GREGORY; however, it was not taken note of at the time. Thus the connecting of the different threads, the great synthesis, remained for NEWTON and LEIBNIZ to achieve. NEWTON circulated his work *De analysi per aequationes numero terminorum infinitas* among his friends in 1669; but it was not printed until 1711. In this work, we find the following thought: Suppose we are given a curve (Figure 4.25), and let z denote the area of the region under the curve (the hatched area). Suppose further that the area depends on the x-coordinate according to

$$z = ax^m.$$

If we increase x by an infinitesimal amount (NEWTON calls this value the *moment* of x and denotes it by the letter o), then the area increases by the amount oy; that is, we have

$$z + oy = a(x + o)^m. \qquad (1)$$

If we now develop the right-hand side of (1) as a binomial series (Newton had by this time worked out his binomial theorem), then we obtain

$$z + oy = ax^m + oamx^{m-1} + o^2\frac{am(m-1)}{2}x^{m-2} + \cdots.$$

Subtracting $z = ax^m$ and ignoring higher powers of o leads to

$$y = amx^{m-1}.$$

This train of thought led NEWTON to two results: First, he determined the derivative of the function $y = ax^m$, for today we would represent the left-hand side of (1) as

$$z + \frac{dz}{dx}\Delta x.$$

Second, he showed that calculation of area and construction of tangents are indeed inverse operations. In fact, if we determine the area of the region under a curve given by some functional relationship and then study the change in area there, we return to the original function.

With the help of the binomial theorem that he had derived, NEWTON was able to both integrate and differentiate functions that up to that time had not been amenable to such processes. Once a function was expanded as a finite or infinite power series and the rules given above for power functions were applied, neither differentiation nor integration posed any difficulties. NEWTON did not concern himself much with issues of convergence; we find in his work only notes that this or that formula may be used only with small or large values of x.

Newton wrote his *Methodus fluxionum et serierum infinitarum* in 1671. However, it only became generally accessible in book form posthumously, in 1736. In this work, Newton explicitly investigates the continuous motion of a point, with time as the independent variable. He notes, however, that this is not a significant limitation. He studies the values of x and y as functions of time, calling these *fluents*. Changes in these values are denoted by $\dot{x}$ and $\dot{y}$, which are called *fluxions*. The inverse of this operation, that is, the fluents that have fluxions x and y, are denoted by x' and y'.

Newton treated the following problem in great generality: Given a relationship between two fluents, we seek the relationship that obtains between the associated fluxions. He also took up the inverse problem of deriving a relation between the fluents given the relationship between two fluxions. Newton illustrates his method with a simple example: Let o again denote an infinitely small interval of time. Then the quantities $\dot{x}o$ and $\dot{y}o$ are the moments of the fluents x and y. We now determine the relationship between the two fluxions when the relationship between the fluents is given by $y = x^n$. To do so, we write the known relationship

$$y + \dot{y}o = (x + \dot{x}o)^n,$$

and applying the binomial theorem while ignoring higher powers of o, we obtain the relationship

$$\dot{y} = nx^{n-1}\dot{x}.$$

In today's notation we would write

$$\frac{\mathrm{d}y}{\mathrm{d}t} = nx^{n-1}\frac{\mathrm{d}x}{\mathrm{d}t},$$

or

$$\frac{\mathrm{d}y}{\mathrm{d}x} = nx^{n-1}.$$

We have thereby once again derived one of the simplest and best-known relationships of the differential calculus.

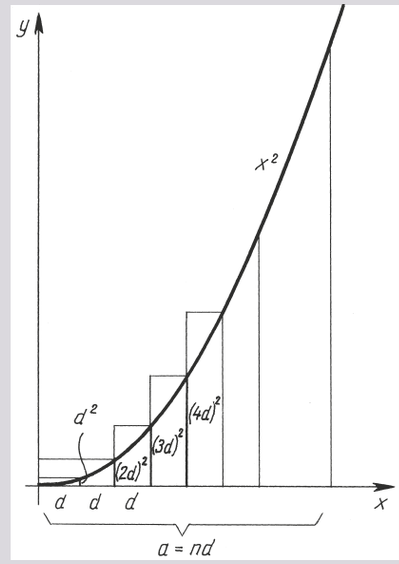
The task set by Newton of deriving a relationship between the fluents from a relationship between the fluxions raises the problem of differential equations in a general form. Quite naturally, Newton dealt with only the simplest of these equations, namely, those leading to a simple integration.

We first meet the derivative in the sense envisioned by Barrow—that is, derived from characteristic triangles in the limit of vanishing triangular area, in Newton's *Tractatus de quadratura curvarum*, written around 1676 but not published until 1704 as an appendix to his *Opticks*. Moreover, the chapters of the *Principia* devoted to this topic are written in the same spirit. In the *Tractatus*, we can already find all the types of questions considered today, such as determining maxima, minima, and inflection points, the determination of the radius of curvature

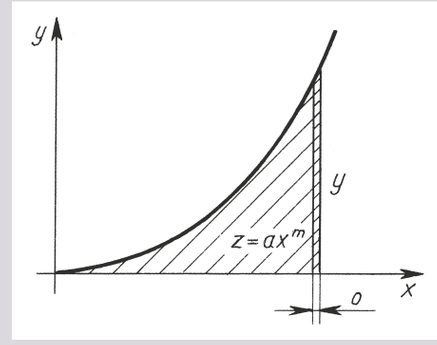
$$r = \frac{\left(1 + \dot{y}^2\right)^{3/2}}{\ddot{y}},$$

and also tables of integrals.

At the same time as Newton, Leibniz (Figure 4.26) arrived at essentially the same results. This is the cause of the priority dispute that was one of the bitterest and most consequential in the history of science—one that, needless to say, had damaging



▲ **Figure 4.24** Determination of the area under a parabola using the limiting value of the sums of the areas of the rectangles.



▲ **Figure 4.25** Newton's idea on the connection between the differential and the integral calculus.



◀ **Figure 4.26** Gottfried Wilhelm Leibniz (1646–1716): one of the most brilliant and multifaceted figures in modern philosophy and the natural sciences. 1666: doctor of jurisprudence. 1672: in the diplomatic service in Paris. The four years spent in Paris were extremely successful; he took part in the intellectual life in France and

continued on next page

Figure 4.26 *continued*
met Huygens, who was living in Paris at the time. In Paris, he built a model for a calculating machine, on the basis of which he was elected to the Royal Society in 1673. From 1677: librarian to the duke of Hanover. 1675: worked out the fundamentals of the differential calculus; he achieved these results ten years after Newton, but demonstrably independently of Newton's work. The notation that Leibniz introduced is better than Newton's, and is still used today. 1684: publication of his compendium *Nova methodus pro maximis et minimis itemque tangentibus* in the *Acta Eruditorum*. In 1700, encouraged by Leibniz, the Berlin *Societas Regia Scientiarum* was founded. In 1710 and 1714, his most important philosophical works, the *Essais de théodizée* and the *Monadologie*, appeared. (The philosophical ideas expressed in these works were mocked by Voltaire in his *Candide*.)

Leibniz's most important contribution to physics can be found in his 1686 work *Brevis demonstratio memorabilis erroris Cartesii*. Here Leibniz suggests that the conservation law postulated by Descartes rests not on the *quantitas motus*, but on the *vis viva*, that is, it is not the absolute value of the quantity of motion (mv) but the living force (mv^2) that remains constant over the course of the motion.

Leibniz also asserted that in determining the *vis viva*, one must also consider the *potentia motrix*, what we would today call the potential energy. The absolute *vis viva* thus obtained (*force vive absolue*) is the physical quantity that is conserved in the course of the motion. Furthermore, he generalized the conservation law of the living force to inelastic collisions and asserted that a loss in the living force of a body is manifested in the gain in the living force of the particles that make up the body.

Leibniz is also considered the founder of formal logic.

He died in 1716, completely alone and out of favor.

repercussions. The advantage of the method found by Leibniz was above all in its notation. He denoted the changes in x and y by dx and dy and chose as the integral sign the familiar elongated S, derived from the word *summa*, which is used to this day. It is doubtlessly correct that Leibniz was the first to publish his results in printed form (1684 and 1686), in *Acta Eruditorum* (Figure 4.27). It can also be firmly established that Newton's work in this area is of an earlier date, so that any possible charge of plagiarism would have to be directed at Leibniz if at anyone. The suspicion that plagiarism may be involved appears to be supported by the fact that Leibniz was in Paris and London in 1672 and 1676 and during that time met a circle of Newton's friends. According to entries in Leibniz's diary, he began working systematically on the calculus only from 1675 on, that is, after his stay in London.

The priority dispute became so rancorous that the Royal Society, at Leibniz's instigation, formed an investigative commission to adjudicate the case. The commission ruled in favor of Newton, essentially accusing Leibniz of plagiarism. However, in hindsight it is quite clear that the critical letter on which the Royal Society based its argument was not, in fact, in Leibniz's possession. Today, everyone acknowledges the contributions of both men because by now we have seen many times that when a science reaches a certain level of development, great ideas can arise simultaneously in different places. This was particularly likely to occur when two such geniuses as Newton and Leibniz were active at the same time.

The damage caused by this conflict affected English science most of all. Newton's victory was in fact a Pyrrhic one because the English, out of national pride, stuck to the path marked out by Newton, including his choice of notation. The result was that they ignored the extraordinarily successful further development of Leibniz's work—in particular, by the Basel school: the brothers Jakob and Johann Bernoulli, Johann's son Daniel, and his friend Leonhard Euler. Compared to this group and the great French mathematicians such as Lagrange, the mathematicians of the British Isles remained of only secondary importance, and it was more than a century before British mathematics was again able to assume a leading role.

Physicists and others employing mathematics in practical applications who consider the mathematicians' "epsilon–delta" nitpicking superfluous will be pleased to learn that the giants Newton and Leibniz, and even Euler, treated the methods of higher mathematics very much as engineers and physicists do today in daily practice—that is sloppily—inviting the disapprobation of mathematicians. The scholars of the seventeenth century, such as Pascal, who were trained in the rigor of Greek mathematics, believed in the importance of proof and believed that it should be possible to establish all results obtained with "indivisibles" with the methods and rigor of the ancients. In contrast, Cavalieri boldly declared that whereas philosophy worries about rigor, such was of no interest to geometry. In describing the work of Archimedes, we mentioned that the rigor of his methodology could be favorably compared with today's standards, so that he far surpasses in exactitude the mathematicians working at the turn of the eighteenth century. The need for a more solid foundation was recognized at the time, but the possibilities for progress in problem solving were so great that mathematicians neglected foundational questions. There were only a few who noted that in establishing such a foundation, it would not be possible to rely on the methods of antiquity, which would have imposed too great a restriction; instead, the foundations of the new analysis would have to rest on a clarification of the notions of limits and convergence. Still, at the given historical moment the new methods were applicable for the solutions of practical problems even with their lax foundations.

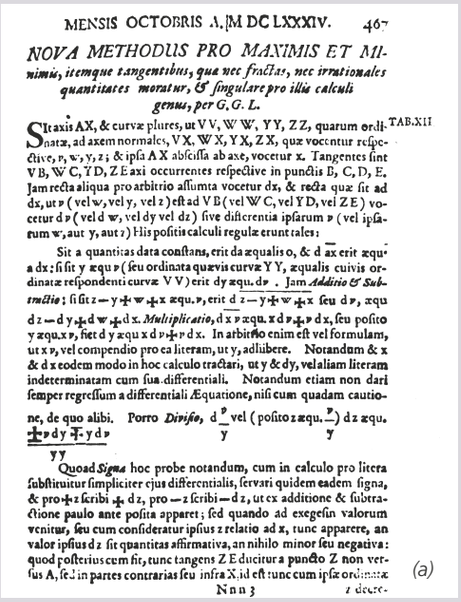
4.1.5 *For and against Descartes*

Perhaps the most valuable inheritance at the start of the eighteenth century was that scientific research did not have to overcome ideological and religious hurdles. At the end of the seventeenth century, the motto *ratio et experientia*, that is, "reason and experience," had generally triumphed, and neither the starting point nor end results of an investigation had to heed the historically sanctified authority of Aristotle, nor was it any longer necessary to take religious dogma into consideration to any great extent. The motto of the eighteenth century, as we shall see, could have been something more like *ratio vel experientia*, that is, "reason *or* experience," because now the essential question had become whether thinking or experience should have a primary role in the seeking of truth. During this period, the center of gravity shifted from an empirical approach, characteristic of science today, to one that favors theorizing. Two causes for such a shift might be mentioned: First, the belief that the world could be seen as a mathematical structure describable with mathematical methods led to a marvelous upsurge in scientific activity and to many concrete findings. Second, the Cartesian philosophy that had become generally accepted in the second half of the seventeenth century saw the guarantee of true knowledge precisely in the belief that the facts accumulated through research in the natural sciences could be arranged *more geometrico* (by the methods of geometry) into a unified and coherent structure derived from fundamental axioms.

The balance between experiment and experience is discussed most beautifully in a letter by Pascal, but it is most consistently realized in the work of Huygens. At the turn of the century, a new program for natural philosophy had also crystallized. Effects caused by direct contact were replaced by Newtonian centers of forces, but the fundamental worldview continued to be determined based on Cartesian principles: The world arose according to the laws of mechanics, and all processes—including biological processes—are governed by those laws.

However, the Cartesian worldview is dualistic. Extent and motion determine material bodies as a substance, while thinking determines another substance, the soul. And behind these substances there stands a third one—God. With the separation of body and soul as independent substances, it became possible to extend the Cartesian ideas in a variety of directions. In this sense, Descartes's ideas can indeed be seen as the starting point for all modern philosophical systems. Some further developed the Cartesian worldview, often believing that they remained true to the fundamental concepts, clarifying only the unclear points. Others consciously disagreed with Descartes at individual details while remaining loyal to the Cartesian edifice. Yet others rejected the entire conceptual basis and constructed new systems, in which, however, the contours of the Cartesian system could be recognized, although now in their denial, as on a photographic negative.

If we consider the three substances that play a role in the Cartesian system (God, soul, and material bodies) and look at later philosophical systems, we see that they can all be characterized by the interactions among these substances. The basic problem inherent in the Cartesian system, which Descartes himself had to answer in his lifetime, is how the spiritual and material substances interact, or more simply how the soul and the body can affect each other. In Figure 4.28 we have attempted to represent graphically, to the extent possible, the relationships among the three substances in various philosophical systems. In Descartes's immediate successors (Geulincx and Malebranche)—who held that continuing direct divine intervention was a necessity—we already encounter the basic idea,



MENSIS OCTOBRIS A. [M DC LXXXIV. 467

NOVA METHODUS PRO MAXIMIS ET MINIMIS, itemque tangentibus, quæ nec fractas, nec irrationales quantitates moratur, & singulare pro illis calculi genus, per G. G. L.

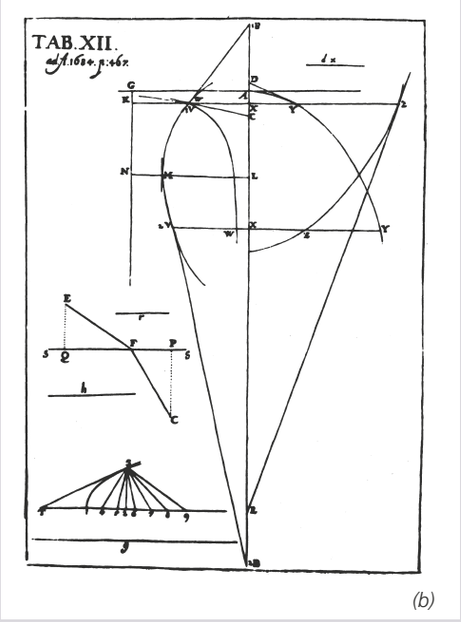
Sit axis AX, & curvæ plures, ut V V, W W, Y Y, Z Z, quarum ordinatæ, ad axem normales, V X, W X, Y X, Z X, quæ vocentur respective, v, w, y, z; & ipsa A X abscissa ab axe, vocetur x. Tangentes sint V B, W C, Y D, Z E axi occurrentes respective in punctis B, C, D, E. Jam recta aliqua pro arbitrio assumta vocetur dx, & recta quæ sit ad dx, ut v (vel w, vel y, vel z) est ad V B (vel W C, vel Y D, vel Z E) vocetur dv (vel dw, vel dy vel dz) sive differentia ipsarum v (vel ipsarum w, aut y, aut z). His positis calculi regulæ erunt tales: TAB. XII

Sit a quantitas data constans, erit da æqualis 0, & d ax erit æqua dx; si sit y æqu v (seu ordinata quævis curvæ Y Y, æqualis cuivis ordinatæ respondenti curvæ V V) erit dy æqu. dv. Jam *Additio & Subtractio*; si sit z = y + w + x æqu. v, erit dz = dy + dw + dx seu dv, æqu dz = dy + dw + dx. *Multiplicatio*, dx v æqu. x dv + v dx, seu posito y æqu. x v, fiet dy æqu x dv + v dx. In arbitrio enim est vel formulam, ut x v, vel compendio pro ea literam, ut y, adhibere. Notandum & x & dx eodem modo in hoc calculo tractari, ut y & dy, vel aliam literam indeterminatam cum sua differentiali. Notandum etiam non dari semper regressum a differentiali Æquatione, nisi cum quadam cautione, de quo alibi. Porro *Divisio*, $\frac{d\,v}{y}$ vel (posito z æqu. $\frac{v}{y}$) dz æqu. $\frac{+v\,dy\,-y\,dv}{yy}$

Quoad *Signa* hoc probe notandum, cum in calculo pro litera substituitur simpliciter ejus differentialis, servari quidem eadem signa, & pro + z scribi + dz, pro − z scribi − dz, ut ex additione & subtractione paulo ante posita apparet; sed quando ad exegesin valorum venitur, seu cum consideratur ipsius z relatio ad x, tunc apparere, an valor ipsius dz sit quantitas affirmativa, an nihilo minor seu negativa: quod posterius cum fit, tunc tangens Z E ducitur a puncto Z non versus A, sed in partes contrarias seu infra X. id est tunc cum ipsæ ordinatæ

Nnn 3 z decre-

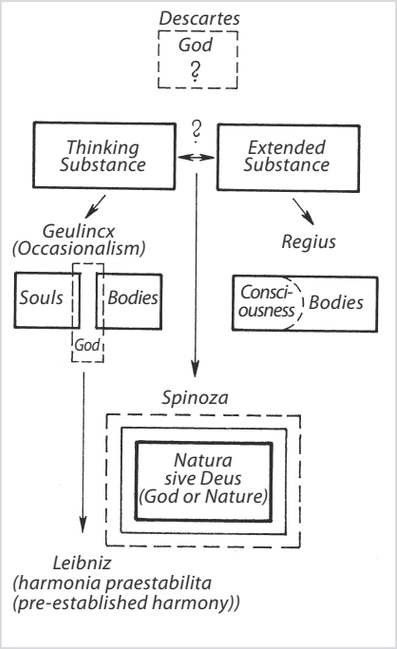
(a)



(b)

▲ **Figure 4.27** **(a)** Page from Leibniz's article on extreme values and tangents in the 1684 volume of *Acta Eruditorum*.

(b) Figure referenced in the article.



◀ **Figure 4.28** The philosophical currents for and against Descartes.

Quotation 4.10

You can say that you do not understand how I can prove what I have advanced about the communication or harmony of two substances as different as the soul and the body. It is true that I thought that I had given the way to do this. And here is how I hope to satisfy you.

Consider two clocks or watches in perfect agreement. Now this can happen in *three ways*: the *first* is that of a natural influence. This is what Huygens experienced, to his great surprise. He had suspended two pendula from the same piece of wood, and the constant swinging of the pendula transmitted similar vibrations to the particles of wood. But since these vibrations could not continue in an orderly way without interfering with each other, at least while the two pendula were not in accord with one another, it happened in a marvelous way that even when the swings of the pendula had been intentionally disturbed, they came to swing together again, almost as if they were two strings in unison. *The second way* to make two faulty clocks always agree would be to have them watched over by a competent workman, who would adjust them and get them to agree at every moment. *The third way* is to construct these two clocks from the start with so much skill and accuracy that one can be certain of their subsequent agreement.

Let us now put the soul and the body in place of these two watches; their agreement or sympathy will also come about in one of these three ways. *The way of influence* is that of the common philosophy; but since we can conceive neither material particles

continued on next page

later perfected by Leibniz, of the theory of monadism, according to which souls and bodies have no effect on each other, and so the observed interactions between the spiritual and material worlds are merely apparent. Corporeal and spiritual processes each behave according to their own laws without any interaction. The third substance, God, coordinates all processes, each proceeding in its own way, so that the appearance of an interaction results. Leibniz compares this interplay to two precisely synchronized clocks, one of which shows the time, while the other, which has no hands, merely strikes the hour. We observe that the second clock begins to strike when the big hand on the first clock reaches the twelve and might assume that there is a causal connection between the two clocks, although we know that the two clocks in fact run independently of each other, and even their inner workings can be entirely different (Quotation 4.10; for Euler's comments on this theory see Quotation 4.11).

Spinoza purged the Cartesian system of logical contradictions and with the most coherent logic took it to its ultimate level. In his work, the three substances merged into one: *Natura sive Deus*—God and nature are one. All of nature consists of a single substance, and this substance possesses properties of the divine. Spinoza's philosophy is therefore of interest because he applied a method of reasoning that begins with fundamental laws—according to Descartes, the only acceptable method—to domains, such as ethics, that go beyond mathematics and the natural sciences. Spinoza's ethics has exactly the logical structure that we would expect to encounter in a textbook on geometry. He begins with definitions, presents his axioms, and finally proves all of his assertions with the help of these axioms. In the field of ethics we are, of course, dealing with human qualities and passions, which, however, Spinoza treats as one would treat points, lines, and planes in geometry.

From the point of view of the natural sciences, what is of primary—almost exclusive—interest in a given philosophical system are the epistemological statements intended to deal with the question of how to arrive at truth. Ever since the seventeenth century, when work in the natural sciences was becoming increasingly intensive, attempts, both conscious and unconscious, were made to answer this question. Continental philosophers followed Descartes, while the British fell into line behind Bacon. In comparison to that of Descartes, Bacon's immediate influence was negligible; he made no concrete contribution to the natural sciences and also failed to appreciate the importance of mathematics or, more generally, the role of the deductive method. One hundred and fifty years later, the situation on the Continent had changed: the encyclopédistes considered Descartes passé and now swore by Bacon.

Among Bacon's followers, we should name first of all Hobbes, who, despite being a supporter of the inductive method, recognized the importance of mathematics and indeed attempted, though without success, to solve mathematical problems. John Locke (Figure 4.29), who was a friend of Boyle and Newton, had a greater effect on epistemology. Although he belongs to the English empirical school, he was also influenced by Descartes in the manner indicated above, in that his theses are frequently inspired by his disagreements with Descartes. A sympathetic aspect of his nature and his theory is that he does not set forth any prescriptions for scientific research but simply wishes to clear the surroundings of the monumental structure created by natural scientists of construction debris. Locke begins with a critique of the innate ideas (*idea innata*) at the root of Cartesian philosophy that are assumed to be inborn in mankind. He maintains that when all is said and done, such ideas can enter our understanding only via the sensory organs because reason is at birth a blank slate (*tabula rasa*) on which only experience acquired through our sense organs

can leave any traces. Locke thus sees experience as the sole source of all knowledge: There is nothing in our understanding that was not first in the sense organs (*Nihil est in intellectu, quod non fuerit in sensu*). However, for Locke, experience means more than sensory input, for he distinguishes external experience (*sensation*) and inner experience (*reflection*). To Locke's statement, Leibniz adds something that would later become the starting point for Immanuel Kant's philosophy: Indeed, there is nothing in our understanding that was not previously in our sensory organs, except for our ability to reason (*nisi intellectus ipse*).

Berkeley and Hume pick up directly where Locke leaves off; we shall speak of them later. The philosophical atmosphere at the beginning of the eighteenth century can be best illustrated for us by Voltaire (Section 4.1.6).

Giovanni Battista Vico was not a follower of any of these philosophical schools. His work focused on the philosophy of history—which would make him a precursor of Hegel and Marx—but his epistemology is also of interest to us. He raised the question of why knowledge obtained in the field of mathematics is precise and reliable. According to Vico, this is because mathematics, with its axioms and rules of inference, is a human creation. Nature, on the other hand, is hard to understand because it was created by God, and therefore only God can know the laws of nature. It follows that knowledge of the truth is based on creation itself (*verum ipsum factum*).

History—so goes Vico's sharp retort to Descartes—is accessible to human reason and can be considered an exact science, precisely because it is a creation of mankind.

Earlier, we mentioned that it was no longer necessary for those engaged in scientific investigation to concern themselves with questions of religion. This assertion holds, of course, only with some qualifications. In 1600, Giordano Bruno was burned at the stake. By 1633, it was enough for Galileo to recant his teachings, and in 1656, Spinoza in turn was punished "metaphysically" for his metaphysical deviation: He was expelled from the synagogue and struck with a "great curse."

Most seventeenth-century philosophers and scientists did not openly oppose religion. Although in some cases, this may have been for reasons of opportunism or exaggerated caution, it was generally the result of inner conviction. What else could have compelled Descartes to carry out an earlier vow and in 1623 make a pilgrimage to Loretto to thank the Virgin for freeing him of doubt? We also have no reason to doubt the sincerity of Newton's religious beliefs, and although he specifically rejected certain dogmas, such as that of the Holy Trinity, he was prudent enough to keep his "heretical" views to himself.

All the spiritual weapons sharpened during the previous century made it possible in the eighteenth century for open battle to be declared against the Church. The punishment for a particularly brash revolutionary spirit would now, in the worst case, be expulsion from one principality only to take up residence at the court of another competing prince.

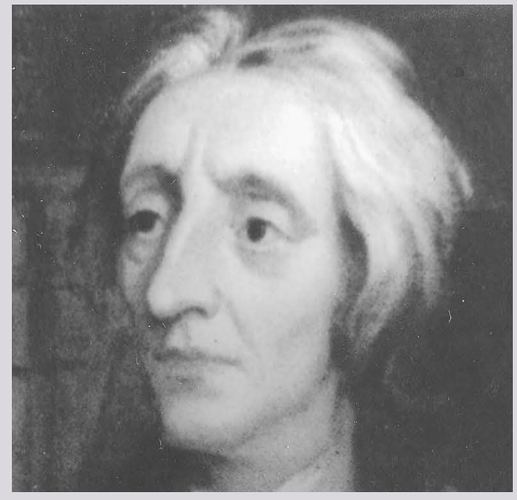
4.1.6 *Voltaire and the Philosophers*

> The traveler was stirred with pity for the little human race, in which he discovered such amazing contrasts. "Since you are of the small number of the wise," he said to these gentlemen, "and that apparently you do not kill people for money, tell me, I pray, how you employ yourselves."
>
> "We dissect flies," answered the philosopher, "we measure lines, we gather mathematical data. We agree on the two or three points we understand, and we argue about the two or three thousand we do not."

Quotation 4.10, ***continued***

nor immaterial qualities or species that can pass from one of these substances to the other, we must reject this opinion. *The way of assistance* is that of the system of occasional causes. But, I hold, that is to appeal to a *Deus ex machina* in a natural and ordinary matter, where, according to reason, God should intervene only in the sense that he concurs with all other natural things. Thus there remains only my hypothesis, that is, *the way of pre-established harmony*, through a prior divine artifice, which has formed each of these substances from the beginning in such a way that by following only its own laws, laws which it received with its being, it nevertheless agrees with the other, as if there were a mutual influence, or as if God always meddled with it, over and above his general concourse. Beyond this I do not think that I need to prove anything else, unless someone wishes me to prove that God is skillful enough to be able to make use of this prior artifice; we even see some examples of this among humans, in proportion to their ability. And assuming that God can do it, one sees clearly that it is the finest way, the way most worthy of him.

—Leibniz, Postscript of a Letter to Henri Basnage de Beauval, 1696



▲ **Figure 4.29** John Locke (1632–1704): Classical studies at Oxford; friendly contacts and collaboration with Robert Boyle. 1668: member of the newly founded Royal Society. From 1675 to 1679: residence in France, where Gassendi's philosophy exerted a great influence on him. Due to political turmoil, he spent five fruitful (and pleasant) years of banishment in the Netherlands.

Locke's best-known and also most influential work is his 1690 *An Essay Concerning Human Understanding*. However, he also published political (*Two Treatises of Government*, 1690) and pedagogical (*Some Thoughts Concerning Education*, 1693) works, as well as writings advocating religious tolerance (*Epistola de Tolerantia*, 1698), where he maintained:

continued on next page

Figure 4.29 *continued*

1. No man has so much wisdom and such wisdom that he can determine what another's religion should be.
2. Every being is morally responsible to God, which presupposes the assumption of his freedom.
3. Force against the will of the individual can ensure only external conformity.

Locke is often viewed as a key figure of the English, and even the French, Enlightenment.

The following quotation comes from Locke's *Essay Concerning Human Understanding*.

> The commonwealth of learning is not at this time without master-builders, whose mighty designs, in advancing the sciences, will leave lasting monuments to the admiration of posterity; but everyone must not hope to be a Boyle or a Sydenham; and in an age that produces such masters as the great Huygenius and the incomparable Mr. Newton, with some other of that strain, it is ambition enough to be employed as an under-labourer in clearing ground a little, and removing some of the rubbish that lies in the way to knowledge; which certainly had been very much more advanced in the world, if the endeavours of ingenious and industrious men had not been much cumbered with the learned but frivolous use of uncouth, affected, or unintelligible terms, introduced into the sciences, and there made an art of to that degree that philosophy, which is nothing but the true knowledge of things, was thought unfit or incapable to be brought into well-bred company and polite conversation. Vague and insignificant forms of speech, and abuse of language, have so long passed for mysteries of science; and hard and misapplied words, with little or no meaning, have, by prescription, such a right to be mistaken for deep learning and height of speculation, that it will not be easy to persuade either those who speak or those who hear them that they are but the covers of ignorance, and hindrance of true knowledge. …
>
> It is an established opinion amongst some men that there are in the understanding certain *innate principles*, some primary notions, … characters, as it were stamped upon the mind of man, which the soul receives in its very first being, and brings into the world with it. It would be sufficient to convince unprejudiced readers of the falseness of this supposition, if I should only show (as I hope I shall in the following parts of this Discourse) how men, barely by the use of their natural faculties, may attain to all the knowledge they have, without the help of any innate impressions, and may arrive at certainty, without any such original notions or principles. …
>
> Let us then suppose the mind to be, as we say, white paper, void of all characters, without any ideas; how comes it to be furnished? Whence comes it by that vast store which the busy and boundless fancy of

continued on next page

Immediately a fancy took the Sirian and the Saturnian to question these thinking atoms, in order to find out on what things they were agreed. "What do you reckon to be the distance," asked the latter, "between the Dog-star and Gemini?"

"Thirty-two and a half degrees," they all replied in concert.

"And from here to the moon?"

"In round numbers, sixty times the radius of the earth."

"What does your air weigh?" he continued, thinking to catch them. He could not succeed, however, for they all told him that the air weighs about nine hundred times less than an equal volume of the lightest water, and nineteen thousand times less than ducat gold. The little dwarf from Saturn, astounded at their replies, was tempted to take for sorcerers these same people to whom a quarter of an hour before he had refused a soul.

Finally, Micromegas put in a word. "Since you are so well acquainted with what is outside you," he said, "you doubtless know still better what is inside. Tell me what your soul is, and how you form your ideas."

As before, the philosophers replied in concert—but each held a different opinion. The oldest quoted Aristotle, another pronounced the name of Descartes, a third the name of Malebranche, a fourth that of Leibniz, a fifth that of Locke. An aged Peripatetic said loudly and confidently: "'The soul is an entelechy and a proof of its power to be what it is.' That is what Aristotle states expressly, page 653 of the Louvre edition: Ἐντελέχειαέστί etc."

"I do not understand Greek too well," remarked the giant.

"No more do I," said the philosophical maggot.

"Why then," asked the Sirian, "do you quote Aristotle in Greek?"

"Because," replied the scholar, "one should always quote what one does not comprehend at all in the language one understands least."

It was the Cartesian's turn. "The soul," he said, "is a pure spirit which has received in its mother's womb all metaphysical ideas and which, on leaving it, has to go to school to learn over again what it knew so well, and will never know again."

"It was not worth while," observed the animal eight leagues long," for your soul to be so wise in your mother's womb if it had to become so ignorant by the time your chin could grow a beard. But what do you mean by a spirit?"

"What a question!" said the man of abstractions. "I have not the remotest idea. A spirit, it is said, is not matter."

"But do you know at least what matter is?"

"Very well," replied the man. "For example, this stone is gray, has a certain shape and three dimensions, is heavy and divisible."

"Well," said the Sirian, "this thing which appears to you to be divisible, heavy and gray, will you kindly tell me what it is? You perceive some of its attributes, but do you know what it is fundamentally?"

"No," answered the other.

"Then you do not know at all what matter is."

Mr. Micromegas then asked another of the wise men he held on his thumb what his soul was, and what it did.

"Nothing at all," replied the disciple of Malebranche. "God does everything for me. I see everything in Him, I do everything in Him. It is He who does everything without my interfering."

"It would be well worth while not to exist," said the sage from Sirius. "And you, my friend," he continued to a Leibnitzian who was there, "what is your soul?"

"My soul," answered the Leibnitzian, "is a hand which points the hours while my body ticks: or, if you prefer, it is my soul which ticks while my body points the hour: or again, my soul is the mirror of the universe, and my body is the frame of the mirror. That is quite clear."

A humble partisan of Locke stood nearby, and when he was spoken to at last, replied: "I do not know how I think, but I do know that I have never thought save by virtue of my senses. That there are immaterial and intelligent beings I do not doubt: but that it is impossible for God to endow matter with mind I doubt very much. I hold the power of God in veneration: I am not free to set bounds to it: I predicate nothing: I am content to believe that more things are possible than we think."

The animal from Sirius smiled. He did not think the last speaker the least wise. The dwarf from Saturn would have embraced the follower of Locke had it not been for the difference in their proportions.

But, unluckily, there was present a minute animalcule in a clerical hat who interrupted the other animalcule philosophers. He said he understood the whole mystery; that the explanation was to be found in the Summa of St. Thomas. He looked the two celestial inhabitants up and down, and asserted that their persons, worlds, suns, and stars were created solely for man.

At this speech the two travelers fell on top of each other, suffocating with that inextinguishable laughter, which, according to Homer, is the lot of the gods. Their shoulders and stomachs heaved, and amid these convulsions the ship which the Sirian held in his thumb fell into one of the Saturnian's trousers pockets. These two good people tried to find it for a long time and, having recovered it at last, set everything very nicely in order. The Sirian picked the maggots up again and spoke to them once more with much kindness, although at the bottom of his heart he was rather angry that such infinitely small creatures should be possessed of an arrogance almost infinitely great. He promised to prepare for them a fine volume of philosophy, written very small so that they might be able to read it, and that in the volume they would find an explanation for everything. And to be sure, he did give them this book before he left them. They took it to Paris to the Academy of Science: but when the aged secretary opened it he found nothing but blank pages. "Ah!" said he. "I thought as much."

—Voltaire, *Micromegas* [pp. 432–435] (see also Figure 4.30)

4.2 Worthy Successors: d'Alembert, Euler, and Lagrange

4.2.1 *Possible Directions for Progress*

One cannot deny admiration and respect of a true genius, particularly one of the greatest, but it must have been depressing to be a contemporary or immediate successor. Newton had already discovered *the* law! What remains other than to understand the significance of his discoveries and begin working out the minor details to which the Great One had not had time to devote his attention? Such is, of course, the sad fate of all epigones. The reader therefore probably suspects that after the exciting battles of the titans of the seventeenth century during which they were giving shape to a new worldview, a boring chapter is about to follow: the listing of events of practical or theoretical importance in chronological order. In some sense, this is true, although we must not forget that in looking at a cathedral, we are first impressed by its monumentality, but on closer inspection, we may be captivated by the beauty of

Figure 4.29 *continued*

man has painted on it with an almost endless variety? Whence has it all the materials of reason and knowledge? To this I answer, in one word, from *experience*. In that all our knowledge is founded, and from that it ultimately derives itself. Our observation employed either about external sensible objects, or about the internal operations of our minds perceived and reflected on by ourselves, is that which supplies our understandings with all the materials of thinking. These two are the fountains of knowledge, from whence all the ideas we have, or can naturally have, do spring. …

First, our senses, conversant about particular sensible objects, do convey into the mind several distinct perceptions of things, according to those various ways wherein those objects do affect them. And thus we come by those *ideas* we have of *yellow, white, heat, cold, soft, hard, bitter, sweet,* and all those which we call sensible qualities; which when I say the senses convey into the mind, I mean, they from external objects convey into the mind what produces there those perceptions. This great source of most of the ideas we have, depending wholly upon our senses, and derived by them to the understanding, I call sensation. …

Secondly, the other fountain from which experience furnisheth the understanding with ideas is the perception of the operations of our own mind within us, as it is employed about the ideas it has got; which operations, when the soul comes to reflect on and consider, do furnish the understanding with another set of ideas, which could not be had from things without. And such are *perception, thinking, doubting, believing, reasoning, knowing, willing*, and all the different actings of our own minds; which we being conscious of, and observing in ourselves, do from these receive into our understandings as distinct ideas as we do from bodies affecting our senses. This source of ideas every man has wholly in himself; and though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called internal sense. But as I call the other sensation, so I call this reflection, the ideas it affords being such only as the mind gets by reflecting on its own operations within itself. By "reflection," then, in the following part of this discourse, I would be understood to mean that notice which the mind takes of its own operations, and the manner of them, by reason whereof there come to be ideas of these operations in the understanding. These two, I say, viz. external material things, as the objects of sensation, and the operations of our own minds within, as the objects of reflection, are to me the only originals from whence all our ideas take their beginnings.

—John Locke, *An Essay Concerning Human Understanding* [pp. 58–59, 67, 89–91]



◀ **Figure 4.30** In satirical writing one frequently encounters the misapprehensions of an epoch and its social mores as viewed by a foreigner (Montesquieu: *Lettres persanes*), by a savage (Voltaire: *l'ngénu*), or by giants, dwarves, and even horses (Swift). Voltaire met Swift in England; an echo of this meeting can be detected in Voltaire's book *Micromegas*. A gigantic denizen of Sirius visits Earth together with his dwarf friend from Saturn, who while small in comparison to the giant, is still huge compared to earthlings. The two see a ship plying the Mediterranean Sea and converse with the voyagers (illustration by Monet).

Voltaire worshiped Newton (*Éloge de Newton*), wrote a biting satire of Leibniz's philosophy (*Candide*), and launched an ad hominem attack against Maupertuis (*Dr. Akakia*). If one were to examine the treatise *Éléments de Philosophie de Newton*, one would be surprised at the extent of Voltaire's scientific knowledge, for here we find an extensive treatment of Kepler's laws and of the motions of the planets' moons as well as a thoroughly worked out theory of spectra.

Quotation 4.11

There was a time when the system of pre-established harmony had acquired such a high reputation over all Germany, that to dare to call it in question was to incur the imputation of ignorance, or bigotry. The supporters of this system boasted, that by means of it the omnipotence and omniscience of the Supreme Being were set in their clearest light, and that it was impossible for any one who believed in these exalted perfections of God to entertain a doubt of the truth of this sublime system.

In fact, say they, we see that poor pitiful mortals are capable of constructing machines so ingeniously as to fill the vulgar spectator with astonishment: how much stronger reason, then, have we to admit that God, having known from all eternity all that my soul would wish and desire, at every instant, should have been able to construct such a machine, which, at every instant, should produce motions conformable to the determinations of my soul! Now, this machine is precisely my body, which is united to my soul only by this harmony; so that if the organization of my body were deranged to such a degree as to be no longer

continued on next page

the details. If we look again at the building as a whole, we finally perceive the basic form as a vessel for a multiplicity of fascinating particulars.

It is only with distance in time or space that a great work can be understood as a challenge for the creation of something new; only then can one detect the "motes," that is, the defects and holes in the theory. Newton's gravitas and high standing were so great that for 150 years after his death, there was no room in England for anything but insignificant commentary on his work. However, on the Continent the situation was different. Here leading thinkers, even while no longer believing in Descartes, nevertheless attacked Newton with weapons drawn directly from Descartes's intellectual arsenal. On the problem of action at a distance, they admittedly took the completely wrong view, but on the positive side, criticism of Newton itself had become acceptable.

With the advantage of time, we have a clear picture of the weaknesses of Newton's theory, where we are thinking not so much of the gaps in the *Principia*, which were closed only by the theory of relativity, as of shortcomings that were apparent, or should have been apparent, already to critical observers at the beginning of the eighteenth century. "The *Principia* is … not a bible. It should be studied and weighed … but not sworn upon" (Quotation 4.12).

Let us now consider those aspects of classical Newtonian mechanics that are *not* contained in the *Principia*, even though today they are regarded as an integral part of it.

The first book of the *Principia* is devoted to the motion of a point mass, yet here we find no hint of general laws for systems of point masses. Nor is the motion of rigid bodies taken up, not even that of the physical pendulum, although a large number of scholars both before and after Newton dealt with the problem by using various fundamental principles of mechanics. In the second book, Newton investigates liquid flow, but most of his conclusions follow from ad hoc hypotheses that are frequently incorrect. Newton somehow never attempted to reconcile fluid flow with his fundamental law of motion. Consequently, it is no surprise that deformable bodies are not taken into consideration either.

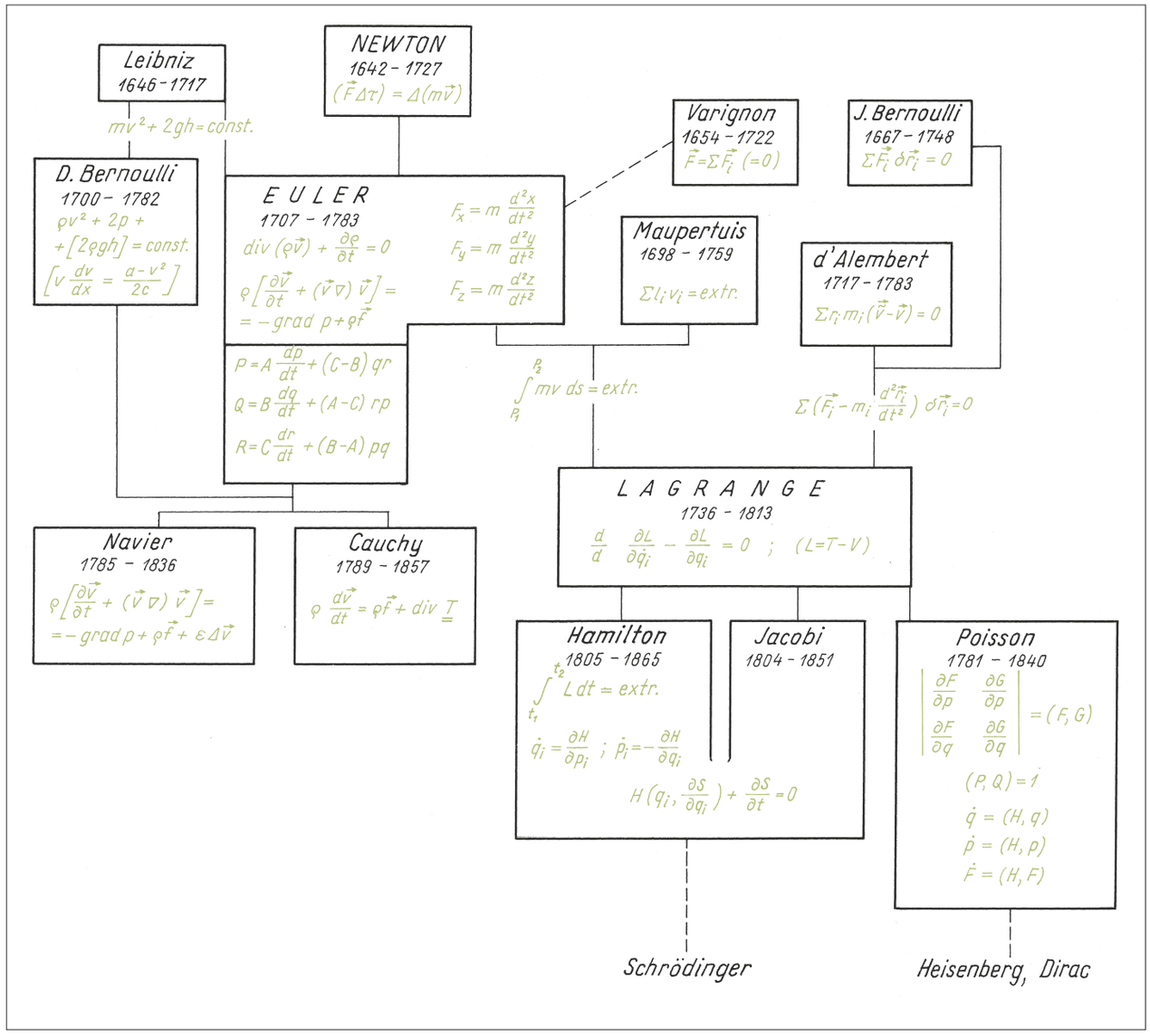
In the previous chapter, we mentioned the surprising fact that Newton did not use in the *Principia* the mathematical methods that he had developed. Instead, the derivations required his individual genius. Newton's successors were therefore faced with the additional task of bringing Newton's equations into a form—or more precisely, to represent Newton's ideas in such equations—that would allow anyone to work with them using the easily mastered methods of calculus.

Let us then examine what possibilities were available in the eighteenth century—at least in principle—for the further development of mechanics (Figure 4.31).

The simplest path began directly with Newton. As we have mentioned, Newton's explanations of phenomena needed to be translated into the language of the most up-to-date mathematical methods of the time and then applied to systems of point masses as well as to continua such as rigid and elastic solids or ideal and viscous fluids. This was the task taken up by Euler at the age of 19, which he would essentially fulfill in the course of a long and productive life.

But there were also attempts to do without Newton's axioms and formulate entirely new principles of mechanics. One of these principles, which turned out to be of lesser importance, was that of d'Alembert. It stands in conscious opposition to one of Newton's basic ideas, because it seeks to eliminate force from mechanics as a primary quantity (Quotation 4.13).

Of much greater importance is a development that begins with Maupertuis's principle. Using this principle, the laws of mechanics as well as the processes of



◀ **Figure 4.31** Paths to the perfection of mechanics after Newton.

We shall have more to say later in the text on the equations of motion for systems of mass points and of rigid bodies.

The mechanics of continua begins with the Bernoulli and Euler equations. These relate, respectively, to incompressible and frictionless ideal fluids. Navier proposed his equation in 1822, in which he considers internal friction using the coefficient of viscosity ε. This equation is usually referred to as the Navier–Stokes equation, although Stokes presented the equation in a more general form in 1845. Cauchy's equation describes the motion of deformable solid bodies. Here T is the stress tensor. Turbulent flow was investigated by Osborne Reynolds (1842–1912; introduction of the Reynolds number, which measures the presence of turbulence), Ludwig Prandtl (1875–1953; theory of interfaces), and Theodore von Kármán (1881–1963). Today, these topics are of great interest: Chaos, a new scientific discipline, provides a method for treating problems in turbulent flow.

motion are derived from requirements that certain quantities related to motion realized in nature must assume minimal or maximal values. We met the extremal principle earlier when we discussed how Fermat determined the path of a ray of light. As we shall see, Maupertuis consciously began with Fermat's principle. We take particular note of the fact, as it also can be also seen from Figure 4.31, that Euler played a decisive role in the formulation of all these methods. Many solutions to problems in mechanics and to mathematical questions were brought by him into the form that we know today.

Finally, some branches of mechanics for various reasons were at the center of interest before Newton or developed independently of him during his lifetime: Here we point to the contributions of Pierre Varignon (1654–1722) on statics and the methods devised by Johann Bernoulli for solving problems in statics using the concept of virtual work. Moreover, the concept of *vis viva* or *living force* had been the subject of heated discussion ever since the time of Galileo. It is Leibniz who formulated the problem precisely and partially solved it. The term is due to him as well.

Of course, these parallel developments eventually were either integrated into the general field of mechanics or continued to live an independent life within the field,

Quotation 4.11, ***continued***

in harmony with my soul, this body would no more belong to me than the body of a rhinoceros in the heart of Africa: and if, in the case of a derangement of my body, God should adjust that of a rhinoceros so that its motions were in such harmony with the determinations of my soul as to raise its paw at the moment I willed it, this body would then be mine, and would belong to my soul, as my present body now belongs to it, without having undergone itself, on that account, any change whatever.

Mr. Leibnitz himself has compared the soul and the body to two clocks, which continually indicate the same hour. A clown who should see this beautiful harmony of these two clocks would undoubtedly conclude that they acted upon each other; but he

continued on next page

Quotation 4.11, ***continued***

would be under a mistake, for the one performs its motions independently of the other. The soul and the body are likewise two machines totally independent, the one being spiritual, the other material; but their operations are always in a harmony so complete, that we are induced to believe them to belong to each other, and that the one has a real influence on the other, which is, however, a mere illusion.

In order to form a judgment of this system, I remark, first, that it cannot be denied to be possible for God to create a machine which should be always in the operations of my soul; but it appears to me that my body belongs to me by other rights than such a harmony, however beautiful it may be: and I believe you will not be disposed hastily to adopt a system which is founded on this principle alone, that no spirit can act upon a body; and that, reciprocally, a body cannot act upon, or supply ideas to, a spirit. This principle is, besides, destitute of all proof. And if God, who is a spirit, has the power of acting upon bodies, it is not absolutely impossible that a spirit, such as the human soul, should be able likewise to act upon a body. Accordingly, we do not pretend to say that our soul acts upon all bodies, but only upon a small particle of matter, with respect to which it has received the power of God himself, though to exercise it in a manner which we are utterly unable to comprehend.

Further, the system of pre-established harmony labors under other great difficulties. According to it the soul derives all its knowledge from its own proper fund, without any contribution on the part of the body and the senses. Thus, when I read in the gazette, that the pope is dead, and I came to the knowledge of the pope's death, the gazette and my reading have nothing to do with the communication of this knowledge, as these circumstances respect only my body and my senses, which have no manner of connection with my soul. ...

In like manner, when you do me the honor to read these letters, and derive the knowledge of some truth from them, it is your soul which extracts that truth from its own proper fund, without my contributing at all to it by my letters. The reading of them serves only to maintain the harmony which the Creator meant to establish between the soul and the body. It is only a formality altogether superfluous with respect to the knowledge itself. I shall nevertheless continue to tender you my instructions.

—Leonhard Euler, *Letters ... to a German Princess,* Volume I, Letter 83

as in the case of extremal principles. The integration of these methods came about as a result of the clarification of the relationship between the extremal principles and Newton's equations of motion.

Lagrange's work *Mécanique Analitique*, which appeared in 1788, just about one hundred years after the publication of the *Principia*, is generally regarded as the culmination of eighteenth-century mechanics. In it, the equations of Newtonian mechanics are presented in the form that is to be found today in textbooks on theoretical mechanics.

In the nineteenth century, classical mechanics advanced along two fronts. For practical applications, the equations for the motion of viscous fluids (by Navier and Stokes) as well as the fundamental equations of deformable solids (by Cauchy) were important developments. From the standpoint of theory, the formulations of the fundamental equations of mechanics by Poisson, Hamilton, and Jacobi were of great significance. By the beginning of the nineteenth century, they had already laid the groundwork in classical mechanics for the quantum mechanics of the twentieth century. Both Heisenberg and Schrödinger would later refer back to them directly; Heisenberg's commutation relation derives ultimately from the Poisson bracket, and Schrödinger worked out the fundamental equation of wave mechanics with the help of the Hamiltonian equations (Section 5.3).

4.2.2 *Results in Statics*

From among the work that was carried out in parallel to that of Newton, we would like here to take a closer look at statics. Varignon's book *Projet d'une nouvelle mécanique*, which appeared in 1687, at about the same time as the *Principia*, served for a long time as the standard work in this area. Varignon solved the problem of the equilibrium of rigid bodies using the method of the decomposition of forces as well as the equations for the rotational moment. Figure 4.32 shows the title page and another page from Varignon's book. We see that a solution is given for a variety of complicated cases, just as it would be done today. Johann Bernoulli (Figure 4.33) in a letter written in 1715 attempted to convince Varignon that it would be much more useful to investigate statics using the principle of virtual work. As we know, this principle goes all the way back to Aristotle and was even used in the Middle Ages to solve problems based on the principle of the lever (see Section 2.4). Bernoulli generalized the formulation used in the Middle Ages so that it could be applied to arbitrary forces: One may consider an arbitrary displacement of a rigid body, or, more generally, of a system of independently movable masses that satisfies the given constraints.

Let us consider—said Bernoulli—an arbitrary displacement of a rigid system, or, more generally, a displacement of a system of a connected set of masses that is consistent with the constraints imposed on the system by the connections. Let us then take the components of the displacements in the direction of the forces and multiply them by the forces, with the appropriate sign for each product depending on whether the displacement is in the direction of the force or is opposite to it. The sum of all such products must be zero at equilibrium. In today's notation we would write

$$\sum \mathbf{F}_i \delta \mathbf{r}_i = 0.$$

In the interest of historical completeness, it should be mentioned that Bernoulli was never able to convince Varignon to abandon his old method. It sufficed for

Varignon to demonstrate that the principle of virtual work leads to precisely the same results as those he had obtained with his own method.

4.2.3 *Newtonian Mechanics in Euler's Hands*

We now examine in somewhat greater detail how mechanics was perfected following Newton; in this we can essentially restrict our attention to Euler (Figure 4.34) and his program. Euler's first book appeared in 1736, its goal clearly announced in the title: *Mechanics, or the analytic treatment of kinematics* (*Mechanica sive motus scientia analytice exposita*). This book was the first in a series, so it deals with only a limited topic—the motion of a point mass. Every one of Euler's books is written with great care and pedagogical skill, so much so that one could profitably study the basic laws of mechanics from these books even today; the countless illustrative examples therein are still used in college-level textbooks. Euler improves on the *Principia* by systematically replacing the concept of body, which Newton uses in a number of different senses, with that of mass point. Euler also clearly understood the concept of acceleration for arbitrary motion along a curved path, and he decomposed it into two components: one parallel to the tangent of the curve and one parallel to the normal. Furthermore, he clarified the vectorial character of velocity and acceleration independently of the vectorial character of force. Immediately after the publication of this book, Euler began his study of forced oscillation and became the first to describe the phenomenon of *resonance*.


DANIELIS BERNOULLI Joh. Fil.
Med. Prof. Basil.
ACAD. SCIENT. IMPER. PETROPOLITANÆ, PRIUS MATHESEOS SUBLIMIORIS PROF. ORD. NUNC MEMBRI ET PROF. HONOR.
HYDRODYNAMICA
SIVE
DE VIRIBUS ET MOTIBUS FLUIDORUM COMMENTARII.
OPUS ACADEMICUM
AB AUCTORE, DUM PETROPOLI AGERET, CONGESTUM.
ARGENTORATI,
Sumptibus JOHANNIS REINHOLDI DULSECKERI,
Anno M D CC XXXVIII.
Typis Joh. Henr. Deckeri, Typographi Basilienſis.

SECTIO TERTIA. 37

De his quæ pertinent ad effluxum aquarum ex Cylindris verticaliter poſitis, per Lumen quodcunque, quod eſt in fundo horizontali.

§. 13.

Geometræ, quibus de aquis ex vaſe erumpentibus ſermo fuit, conſiderare potiſſimum ſolent cylindros verticaliter poſitos: Igitur haud abs re erit ex theoria noſtra generali confectaria illa, quæ huc pertinent, deducere. Sit amplitudo cylindri ad amplitudinem foraminis ut m ad n; altitudo aquæ ſupra foramen, cum fluxus incipit $= a$; altitudo aquæ reſiduæ $= x$, altitudo velocitati aquæ internæ debita $= v$; erit in æquatione canonica paragraphi octavi $y = m$, $N = m\,x$ (per §. 6.) quæ adeoque abit in hanc æquationem.

$$mxdv - \frac{m^1}{nn} vdx + mvdx = -mxdx, \text{ vel}$$

$$\left(1 - \frac{mm}{nn}\right) vdx + xdv = -xdx$$

multiplicetur hæc poſterior æquatio per $x^{-\frac{mm}{nn}}$, ut habeatur

$$\left(1 - \frac{mm}{nn}\right) x^{-\frac{mm}{nn}} vdx + x^{1-\frac{mm}{nn}} dv = -x^{1-\frac{mm}{nn}} dx.$$

Poteſt jam hæc æquatio integrari: obſervanda autem eſt in Integratione conſtantis additio, talis nempe, ut a fluxus initio, id eſt, cum $x = a$, ſit velocitas fluidi nulla, ipſaque proin v pariter $= 0$: ita vero oritur:

$$x^{1-\frac{mm}{nn}} v = \frac{nn}{2nn-mm}\left(a^{2-\frac{mm}{nn}} - x^{2-\frac{mm}{nn}}\right) \text{ vel}$$

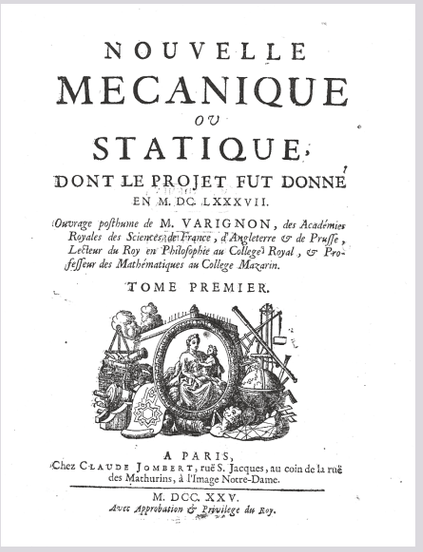
$$v = \frac{nna}{2nn-mm}\left(\left(\frac{a}{x}\right)^{1-\frac{mm}{nn}} - \frac{x}{a}\right)$$

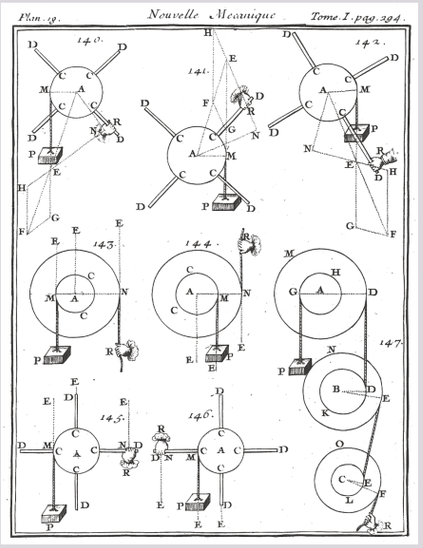
§. 14. Ex hâc igitur æquatione cognoſcitur altitudo generans velocitatem aquæ internæ; ubi notari meretur, ſi vas ſit ampliſſimum, mox poſſe cenſeri $v = \frac{nn}{mm}x$, poſtquam ſcilicet vel tantillum deſcendit aqua, id eſt,

E 3 cum

Euler gradually extended his investigations to systems of mass points, and then to systems of arbitrary numbers of mass points coupled by springs. He determined the frequencies of possible vibrational states. In 1752, he published an article with the surprising title of *The Discovery of a New Principle of Mechanics* (*Découverte d'un nouveau principe de la mécanique*). The new principle of which he is writing is precisely what today is called Newton's second law of motion, that is, the analytic representation of the relationship

$$\text{force} = \text{mass} \times \text{acceleration}.$$


NOUVELLE
MECANIQUE
OU
STATIQUE,
DONT LE PROJET FUT DONNÉ
EN M. DCC. LXXXVII.
Ouvrage poſthume de M. VARIGNON, des Académies Royales des Sciences de France, d'Angleterre & de Pruſſe, Lecteur du Roy en Philoſophie au College Royal, & Profeſſeur des Mathématiques au College Mazarin.
TOME PREMIER.
A PARIS,
Chez Claude Jombert, ruë S. Jacques, au coin de la ruë des Mathurins, à l'Image Notre-Dame.
M. DCC. XXV.
Avec Approbation & Privilege du Roy.



◀ **Figure 4.32** Two pages from Varignon's book on statics. (Library of the Hungarian Academy of Sciences.)

◀ **Figure 4.33** Three generations of Bernoullis produced eight outstanding mathematicians and physicists, the majority of whom worked at the University of Basel. Of greatest importance for the history of physics are the brothers Jakob (1654–1705) and Johann (1667–1748), and above all, the latter's son Daniel (1700–1782).

Daniel Bernoulli studied medicine, but was more drawn to mathematics. In 1725, he went to St. Petersburg to lecture on mathematics. In 1733 he returned to Basel, where he assumed a chair in anatomy and botany; after the death of his father, he succeeded him as a professor of physics. It is for him that the Bernoulli equation, which he derived in 1733 and published in his 1738 book *Hydrodynamica*, is named. The figure shows the title page and a typical page from this book (Library of the University for Heavy Industry, Miskolc). The Bernoullis' kinetic gas theory is written in what is already a quite modern formulation as evidenced, for example, by the following treatment of elastic and inelastic collisions:

> From this it follows that every *vis viva* possesses a well-defined quantity, and if it appears that this quantity in

continued on next page

Figure 4.33 *continued*

part disappears, then it reveals itself in reality in the effects that emanate from it. The living force thus always remains conserved in the sense that the same living force that was present in one or several bodies before they began to interact with one another exists in the same quantity in the one or other body or in the entire system after the interaction. This assertion is expressed in what I call conservation of the *vis viva*. … If the bodies are not fully elastic, it appears that a part of the living force is lost, since after the compression, the original state is not fully restored. In such cases we must assume that the process corresponds to the compression of an elastic spring whose return is hindered by a barrier; thus the body cannot return the living force that it receives from the colliding body; rather, it retains it, so that no loss of force occurs.

—Johann Bernoulli, *Acta Eruditorum*, 1735

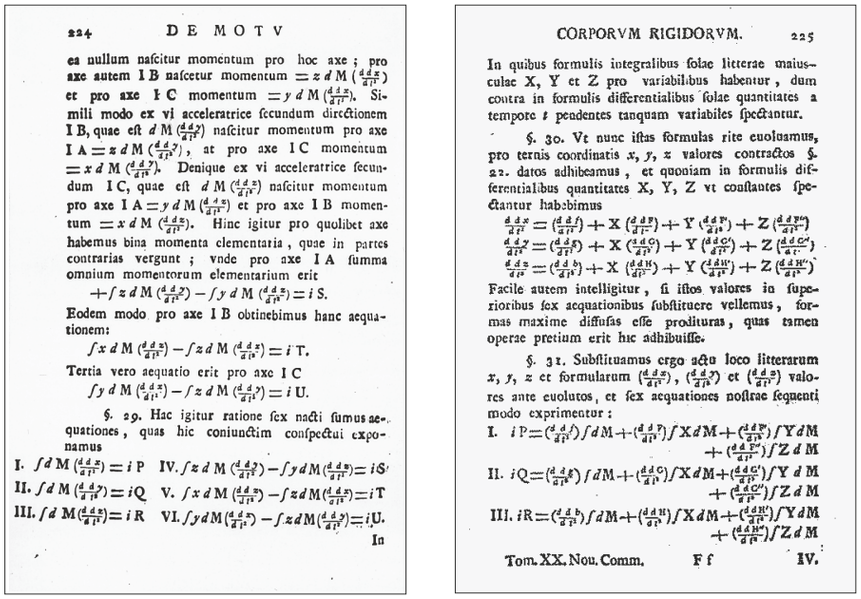


▲ **Figure 4.34** Leonhard Euler (1707–1783): revered by his contemporaries as *Mathematicus acutissimus*. This Swiss scholar was the leading mathematician of the eighteenth century. In physics, together with d'Alembert, he is one of the founders of analytic mechanics. The Euler equations of motion of rigid bodies and of hydrodynamics are named for him. As a founder of the calculus of variations, he contributed to the formulation of the mechanical minimality principle (the Maupertuis principle was formulated by Euler earlier and more precisely than by Maupertuis himself). He investigated in depth the vibrations of strings and membranes, and in opposition to the generally accepted Newtonian theory, he interpreted light as a wave state, although with a longitudinal direction of vibration. He refuted Newton's assertion that an achromatic lens could not be constructed. Euler spent the greater part of his life in St. Petersburg as a member of the Russian Academy of Sciences; it is there that he died.

His view of Leibniz's philosophy is made clear in Quotation 4.11.

Figure 4.35 shows the page of Euler's paper on which Newton's equation appears in the form well-known today. Why did Euler describe this equation as a new principle? As we mentioned in our discussion of the *Principia*, Newton's equation of motion did not appear in this general symbolic form, but was simply described in words for the case of a single body. Euler emphasizes that in defining force, one must employ the concept as it is used in statics. Moreover, he considered the equation to be valid for an arbitrary mass element, so that the equation is applicable to systems of mass points and even to continua, in contrast to Newton, who, as we have mentioned, never attempted to apply the laws of motion to fluid flow. Using the new principle, Euler was now in a position to derive the equations—which now bear his name—for fluid flow and the motion of rigid bodies, equations that remain valid to this day and are taught in this form (Quotation 4.14).

Let us focus now on a brief description of Euler's work on fluid flow. We meet the theorems that Euler derived in the most diverse branches of physics in a large number of guises. Euler investigated ideal fluids, but did not limit his attention to incompressible fluids. For this work, he formulated the continuity equation (Figure 4.36).



224 DE MOTV

ea nullum naſcitur momentum pro hoc axe ; pro axe autem I B naſcetur momentum $= z\,dM\left(\frac{d\,dx}{dt^2}\right)$ et pro axe I C momentum $= y\,dM\left(\frac{d\,dx}{dt^2}\right)$. Simili modo ex vi acceleratrice ſecundum directionem I B, quae eſt $dM\left(\frac{d\,dy}{dt^2}\right)$ naſcitur momentum pro axe I A $= z\,dM\left(\frac{d\,dy}{dt^2}\right)$, at pro axe I C momentum $= x\,dM\left(\frac{d\,dy}{dt^2}\right)$. Denique ex vi acceleratrice ſecundum I C, quae eſt $dM\left(\frac{d\,dz}{dt^2}\right)$ naſcitur momentum pro axe I A $= y\,dM\left(\frac{d\,dz}{dt^2}\right)$ et pro axe I B momentum $= x\,dM\left(\frac{d\,dz}{dt^2}\right)$. Hinc igitur pro quolibet axe habemus bina momenta elementaria, quae in partes contrarias vergunt ; vnde pro axe I A ſumma omnium momentorum elementarium erit

$$+\int z\,dM\left(\frac{d\,dy}{dt^2}\right) - \int y\,dM\left(\frac{d\,dz}{dt^2}\right) = i\,S.$$

Eodem modo pro axe I B obtinebimus hanc aequationem:

$$\int x\,dM\left(\frac{d\,dz}{dt^2}\right) - \int z\,dM\left(\frac{d\,dx}{dt^2}\right) = i\,T.$$

Tertia vero aequatio erit pro axe I C

$$\int y\,dM\left(\frac{d\,dx}{dt^2}\right) - \int z\,dM\left(\frac{d\,dy}{dt^2}\right) = i\,U.$$

§. 29. Hac igitur ratione ſex nacti ſumus aequationes, quas hic coniunctim conſpectui exponamus

I. $\int dM\left(\frac{d\,dx}{dt^2}\right) = iP$ IV. $\int z\,dM\left(\frac{d\,dy}{dt^2}\right) - \int y\,dM\left(\frac{d\,dz}{dt^2}\right) = iS$

II. $\int dM\left(\frac{d\,dy}{dt^2}\right) = iQ$ V. $\int x\,dM\left(\frac{d\,dz}{dt^2}\right) - \int z\,dM\left(\frac{d\,dx}{dt^2}\right) = iT$

III. $\int dM\left(\frac{d\,dz}{dt^2}\right) = iR$ VI. $\int y\,dM\left(\frac{d\,dx}{dt^2}\right) - \int x\,dM\left(\frac{d\,dy}{dt^2}\right) = iU.$

In

CORPORVM RIGIDORVM. 225

In quibus formulis integralibus ſolae litterae maiuſculae X, Y et Z pro variabilibus habentur, dum contra in formulis differentialibus ſolae quantitates a tempore t pendentes tanquam variabiles ſpectantur.

§. 30. Vt nunc iſtas formulas rite euoluamus, pro ternis coordinatis x, y, z valores contractos §. 22. datos adhibeamus, et quoniam in formulis differentialibus quantitates X, Y, Z vt conſtantes ſpectantur habebimus

$$\frac{d\,dx}{dt^2} = \left(\frac{d\,df}{dt^2}\right) + X\left(\frac{d\,dx}{dt^2}\right) + Y\left(\frac{d\,dy}{dt^2}\right) + Z\left(\frac{d\,dz}{dt^2}\right)$$
$$\frac{d\,dy}{dt^2} = \left(\frac{d\,dg}{dt^2}\right) + X\left(\frac{d\,dG}{dt^2}\right) + Y\left(\frac{d\,dG'}{dt^2}\right) + Z\left(\frac{d\,dG''}{dt^2}\right)$$
$$\frac{d\,dz}{dt^2} = \left(\frac{d\,dh}{dt^2}\right) + X\left(\frac{d\,dH}{dt^2}\right) + Y\left(\frac{d\,dH'}{dt^2}\right) + Z\left(\frac{d\,dH''}{dt^2}\right)$$

Facile autem intelligitur, ſi iſtos valores in ſuperioribus ſex aequationibus ſubſtituere vellemus, formas maxime diffuſas eſſe prodituras, quas tamen operae pretium erit hic adhibuiſſe.

§. 31. Subſtituamus ergo actu loco litterarum x, y, z et formularum $\left(\frac{d\,dx}{dt^2}\right)$, $\left(\frac{d\,dy}{dt^2}\right)$ et $\left(\frac{d\,dz}{dt^2}\right)$ valores ante euolutos, et ſex aequationes noſtrae ſequenti modo exprimentur:

I. $iP = \left(\frac{d\,df}{dt^2}\right)\int dM + \left(\frac{d\,dg}{dt^2}\right)\int X\,dM + \left(\frac{d\,dh}{dt^2}\right)\int Y\,dM + \left(\frac{d\,dH}{dt^2}\right)\int Z\,dM$

II. $iQ = \left(\frac{d\,dg}{dt^2}\right)\int dM + \left(\frac{d\,dG}{dt^2}\right)\int X\,dM + \left(\frac{d\,dG'}{dt^2}\right)\int Y\,dM + \left(\frac{d\,dG''}{dt^2}\right)\int Z\,dM$

III. $iR = \left(\frac{d\,dh}{dt^2}\right)\int dM + \left(\frac{d\,dH}{dt^2}\right)\int X\,dM + \left(\frac{d\,dH'}{dt^2}\right)\int Y\,dM + \left(\frac{d\,dH''}{dt^2}\right)\int Z\,dM$

▲ **Figure 4.35** In this work of Euler, we encounter for the first time Newton's fundamental law in the form that we know it today. (Library of the Hungarian Academy of Sciences.)

The continuity equation, or equation of conservation of matter, or mass, is of great significance; we find similar equations in countless other branches of physics. Thus, for example, in electrodynamics, the equation for conservation of charge has exactly the same form, except that mass density is replaced by charge density. To mention a more modern application, among the basic equations of the theory of semiconductors we can also find an equation of this type.

Euler also examined the total change in velocity of mass points in a flow; its first description can be seen in Figure 4.36. In this way he arrived at the equation of motion for fluids.

The general motion of rigid bodies is worth looking at in greater detail. Euler began his investigations by replacing the center of gravity by the center of mass (*Theoria motus corporum solidorum seu rigidorum*, 1760). In a homogeneous force field, these two points coincide, but whereas the center of gravity is defined as the

gravitational point of action, the center of mass can be defined without reference to a gravitational field. An important new quantity introduced is the moment of inertia, which we met earlier in HUYGENS' investigation of the physical pendulum. However, HUYGENS did not recognize the central importance of this quantity for the motion of rigid bodies and therefore saw no need to give a special name to the concept $\sum_i m_i r_i^2$ appearing in his calculations. EULER worked out the now familiar formulas for moments of inertia that can be found in today's handbooks of mechanics. He decomposed the most general motion of a rigid body into a translation of the center of mass and a rotation about an axis through the center of mass. Determining the path of the center of mass does not represent a new problem; the center of mass, under the influence of an external force, will move as if the mass of the entire body were concentrated there. For this problem, the forces acting on the various points of the rigid body can be transferred to the center of mass.

So the main problem EULER faced was to describe the motion of a rigid body when only the point at its center of mass is held fixed. EULER first determined that such a body will not rotate freely about an arbitrary axis, but that for a body of any form, there exist three axes through the center of mass about which free rotation is possible. These axes are called the principal axes of inertia. After considering these free rotations, EULER asked what sort of motion a body would execute if we were to start the spin about an arbitrary axis at a particular angle to the principal axes of inertia. In this case, the resulting motion can be described as a rotation about an axis through the center of mass, and the axis itself also moves. To determine the fundamental equations, EULER first looked at the problem from the inverse point of view: Given the angular velocity of the rotation of a rigid body around an axis through the center of mass, find the forces on the mass elements of the body that lead to a prescribed change in time interval dt of the angular velocity and the direction of the axis.

In other words, the inversion of the question means that instead of calculating the changes caused by a given force, we need to determine the forces necessary for a given change in the state of motion. Figure 4.37 shows how the location of the instantaneous axis of rotation of the rigid body can be specified in terms of the three principal axes of inertia *IA*, *IB*, *IC*.

The givens are therefore the quantities dα, dβ, dγ, dω; the unknowns are the components of force X, Y, Z. By the fundamental law of mechanics, the changes in the components of velocity du, dv, dw are proportional to the quantities

$$\frac{X\,\mathrm{d}t}{\mathrm{d}M}, \quad \frac{Y\,\mathrm{d}t}{\mathrm{d}M}, \quad \frac{Z\,\mathrm{d}t}{\mathrm{d}M}.$$

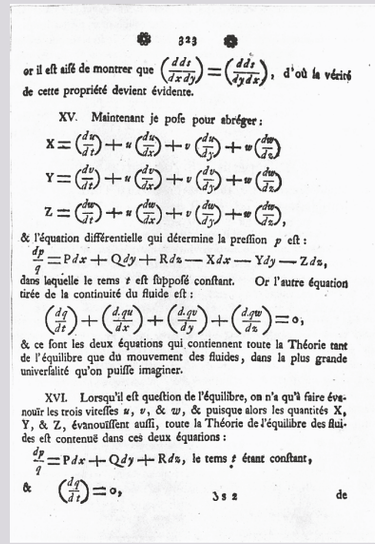
We must then express the changes in the components of velocity in terms of quantities characteristic of the instantaneous rotational axis, that is, in terms of ω, α, β, γ. From the figure one may easily deduce the relations

$$\begin{aligned} u &= \omega(z\cos\beta - y\cos\gamma), \\ v &= \omega(x\cos\gamma - z\cos\alpha), \\ w &= \omega(y\cos\alpha - x\cos\beta), \end{aligned} \qquad (1)$$

as well as

$$\begin{aligned} \mathrm{d}x &= u\,\mathrm{d}t = \omega\,\mathrm{d}t(z\cos\beta - y\cos\gamma), \\ \mathrm{d}y &= v\,\mathrm{d}t = \omega\,\mathrm{d}t(\omega(x\cos\gamma - z\cos\alpha), \\ \mathrm{d}z &= w\,\mathrm{d}t = \omega\,\mathrm{d}t(y\cos\alpha - x\cos\beta). \end{aligned} \qquad (2)$$

From these equations, the desired relation follows without much difficulty. To express the changes in velocity du, dv, dw in terms of the quantities dω, dα, dβ, dγ, we form the total differential of equation (1) and


323

or il eſt aiſé de montrer que $\left(\frac{d\,dt}{dx\,dy}\right)=\left(\frac{d\,dt}{dy\,dx}\right)$, d'où la vérité de cette propriété devient évidente.

XV. Maintenant je poſe pour abréger :

$$X=\left(\frac{du}{dt}\right)+u\left(\frac{du}{dx}\right)+v\left(\frac{du}{dy}\right)+w\left(\frac{du}{dz}\right)$$
$$Y=\left(\frac{dv}{dt}\right)+u\left(\frac{dv}{dx}\right)+v\left(\frac{dv}{dy}\right)+w\left(\frac{dv}{dz}\right)$$
$$Z=\left(\frac{dw}{dt}\right)+u\left(\frac{dw}{dx}\right)+v\left(\frac{dw}{dy}\right)+w\left(\frac{dw}{dz}\right),$$

& l'équation différentielle qui détermine la preſſion p eſt :

$$\frac{dp}{q}=P\,dx+Q\,dy+R\,dz-X\,dx-Y\,dy-Z\,dz,$$

dans laquelle le tems t eſt ſuppoſé conſtant. Or l'autre équation tirée de la continuité du fluide eſt :

$$\left(\frac{dq}{dt}\right)+\left(\frac{d.qu}{dx}\right)+\left(\frac{d.qv}{dy}\right)+\left(\frac{d.qw}{dz}\right)=0,$$

& ce ſont les deux équations qui contiennent toute la Théorie tant de l'équilibre que du mouvement des fluides, dans la plus grande univerſalité qu'on puiſſe imaginer.

XVI. Lorſqu'il eſt queſtion de l'équilibre, on n'a qu'à faire évanouïr les trois viteſſes u, v, & w, & puiſque alors les quantités X, Y, & Z, évanouiſſent auſſi, toute la Théorie de l'équilibre des fluides eſt contenuë dans ces deux équations :

$$\frac{dp}{q}=P\,dx+Q\,dy+R\,dz, \text{ le tems } t \text{ étant conſtant,}$$

& $\left(\frac{dq}{dt}\right)=0,$ 3 s 2 de

◀ Figure 4.36 The continuity equation plays a role in many areas of physics; here the equation appears for the first time [Truesdell 1968].

Quotation 4.12

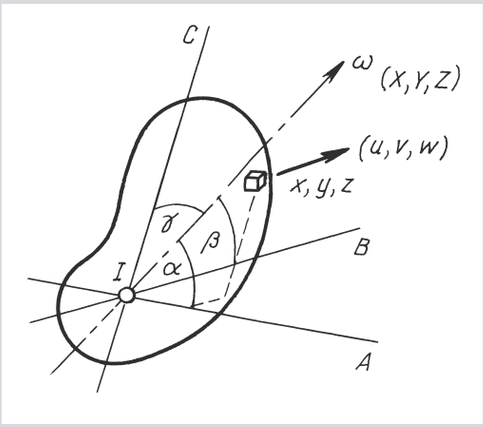
Since the *Principia* is one of those works everyone talks of but no one reads, anything said about it other than the usual honey-sauced eulogy must stand up against righteous indignation from all sides. But it is a work of science, not a bible. It should be studied and weighed—admired, indeed, but not sworn upon. It has its novelties and its repetitions, its elegant perfections and its errors, its lightning abbreviations and its needless detours, its extraordinary standards of rigor and its logical gaps, its elimination of stated hypotheses and its introduction of unstated ones.

—C. TRUESDELL, *Essays in the History of Mechanics* [p. 88]

Quotation 4.13

Why should we appeal to that principle used by everybody nowadays, that the accelerating or retarding force is proportional to the element [*i.e.* differential] of velocity, a principle resting only on that vague and obscure axiom that the effect is proportional to the cause? We shall not investigate at all whether this principle be a necessary truth; we shall assert only that the proofs of it given up to now do not seem very convincing to us; neither shall we adopt it, like some other geometers, as a purely contingent truth, for that would ruin the certainty of mechanics and reduce it to being nothing but an experimental science. Rather, we shall be content to remark that the principle [of accelerating force], be it true or be it dubious, be it clear or be it obscure, is useless to mechanics and ought therefore to be banished from it.

—D'ALEMBERT, *Traité de Dynamique* [Truesdell 1968, p. 113]



▲ **Figure 4.37** In his derivation of the equations of motion of a rigid body, EULER begins with the converse problem: Given a course of motion, find the forces that cause that motion.

Quotation 4.14

However sublime may be the investigations on fluids for which we are indebted to MM. BERNOULLI, CLAIRAUT, and D'ALEMBERT, they stem so naturally from our two general formulae that one cannot but admire this agreement of their profound meditations with the simplicity of the principles from which I have deduced my two equations, and to which I was directly led by the first axioms of mechanics.

—EULER, "Continuation des recherches sur la théorie du mouvement des fluids," 1755 [Dugas 1957, p. 304]

Quotation 4.15

In meditating deeply on this matter, I thought that, since light has already forsaken the shortest path when it goes from one medium to another—the path which is a straight line—it could just as well not follow that of the shortest line. Indeed, what preference can there be in this matter for time or distance? Light cannot at once travel along the shortest path and along that of the shortest time—why should it go by one of these paths rather than by the other? Further, why should it follow either of these two? It chooses a path which has a very real advantage—*the path which it takes is that by which the quantity of the action is the least*.

It must now be explained what I mean by the quantity of action. ...

It is the quantity of action which is Nature's true storehouse, and which it economizes as much as possible in the motion of light.

I know of the repugnance that several mathematicians have for *final causes* when applied to physics, and to a certain extent I am in accord with

continued on next page

substitute for dx, dy, dz the values given in equation (2). We thereby obtain the relationships

$$\begin{aligned}
\mathrm{d}u &= \mathrm{d}\omega(z\cos\beta - y\cos\gamma) - \omega z\,\mathrm{d}\beta\sin\beta + \omega y\,\mathrm{d}\gamma\sin\gamma \\
&\quad + \omega^2\,\mathrm{d}t(y\cos\alpha\cos\beta + z\cos\alpha\cos\gamma - x\sin^2\alpha), \\
\mathrm{d}v &= \mathrm{d}\omega(x\cos\gamma - z\cos\alpha) - \omega x\,\mathrm{d}\gamma\sin\gamma + \omega z\,\mathrm{d}\alpha\sin\alpha \\
&\quad + \omega^2\,\mathrm{d}t(z\cos\beta\cos\gamma + x\cos\beta\cos\alpha - y\sin^2\beta), \\
\mathrm{d}w &= \mathrm{d}\omega(y\cos\alpha - x\cos\beta) - \omega y\,\mathrm{d}\alpha\sin\alpha + \omega x\,\mathrm{d}\beta\sin\beta \\
&\quad + \omega^2\,\mathrm{d}t(x\cos\gamma\cos\alpha + y\cos\gamma\cos\beta - z\sin^2\gamma),
\end{aligned}$$

and we can now determine the quantities X, Y, Z from

$$\mathrm{d}u = \frac{X\,\mathrm{d}t}{\mathrm{d}M}, \quad \mathrm{d}v = \frac{Y\,\mathrm{d}t}{\mathrm{d}M}, \quad \mathrm{d}w = \frac{Z\,\mathrm{d}t}{\mathrm{d}M}.$$

EULER then calculated the torques P, Q, R generated by the forces around the three principal axes of inertia. From the definition of torque, the components of the differential torque created by the forces acting on the mass element dM are

$$\mathrm{d}P = \frac{1}{\mathrm{d}t}(y\,\mathrm{d}w - z\,\mathrm{d}v)\,\mathrm{d}M,$$

$$\mathrm{d}Q = \frac{1}{\mathrm{d}t}(z\,\mathrm{d}u - x\,\mathrm{d}w)\,\mathrm{d}M,$$

$$\mathrm{d}R = \frac{1}{\mathrm{d}t}(x\,\mathrm{d}v - y\,\mathrm{d}u)\,\mathrm{d}M.$$

Summing these differential torques for the entire body and considering the differential changes in velocity given above, we obtain

$$P = \frac{1}{\mathrm{d}t}\Big[A\,\mathrm{d}\omega\cos\alpha - \omega A\,\mathrm{d}\alpha\sin\alpha + \omega^2(C-B)\,\mathrm{d}t\cos\beta\cos\gamma\Big],$$

$$Q = \frac{1}{\mathrm{d}t}\Big[B\,\mathrm{d}\omega\cos\beta - \omega B\,\mathrm{d}\beta\sin\beta + \omega^2(A-C)\,\mathrm{d}t\cos\gamma\cos\alpha\Big],$$

$$R = \frac{1}{\mathrm{d}t}\Big[C\,\mathrm{d}\omega\cos\gamma - \omega C\,\mathrm{d}\gamma\sin\gamma + \omega^2(B-A)\,\mathrm{d}t\cos\alpha\cos\beta\Big].$$

The quantities A, B, C introduced here are the principal moments of inertia of the rigid body.

Now EULER can give an answer to the direct question of how the axis of rotation and angular velocity change over time if arbitrary forces acting on the rigid body are given.

EULER first observes that the equations given above can be simplified if the components of the angular velocity vector are given by

$$p = \omega\cos\alpha, \quad q = \omega\cos\beta, \quad r = \omega\cos\gamma.$$

Thus we obtain the equations for bodies that today bear EULER's name:

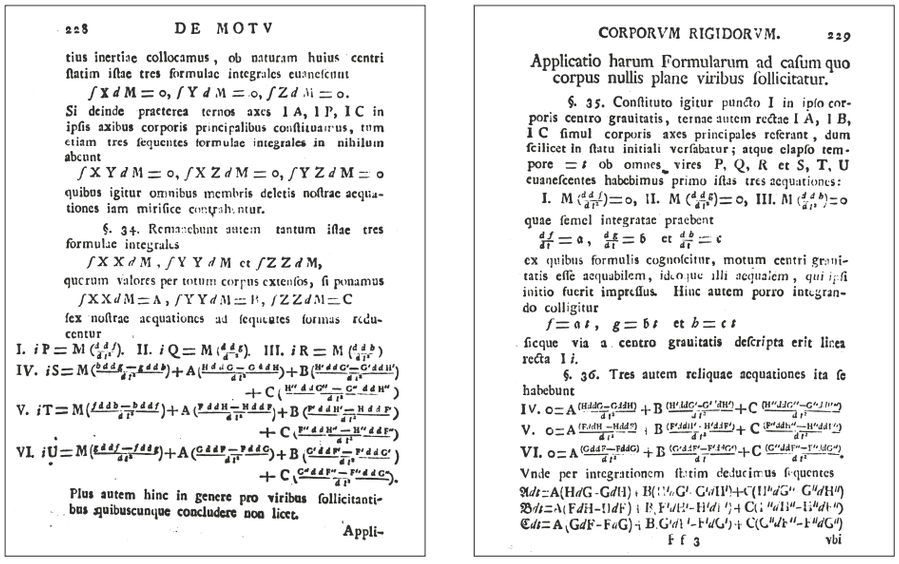
$$P = A\frac{\mathrm{d}p}{\mathrm{d}t} + (C-B)qr,$$

$$Q = B\frac{\mathrm{d}q}{\mathrm{d}t} + (A-C)rp,$$

$$R = C\frac{\mathrm{d}r}{\mathrm{d}t} + (B-A)pq.$$

EULER himself appreciated the significance of these equations; in his book mentioned above, he wrote that the entire theory of motion of rigid bodies is contained in these three simple formulas (Figure 4.38).

EULER also realized—although it was only in the final years of his life—that NEWTON's law of motion, even in the form given by EULER himself—that is, for an arbitrary mass element removed from the continuum, the relationship "force



228 DE MOTV

tius inertiae collocamus, ob naturam huius centri ſtatim iſtae tres formulae integrales euaneſcunt

$\int X\,dM = 0$, $\int Y\,dM = 0$, $\int Z\,dM = 0$.

Si deinde praeterea ternos axes 1 A, 1 P, 1 C in ipſis axibus corporis principalibus conſtituamus, tum etiam tres ſequentes formulae integrales in nihilum abeunt

$\int XY\,dM = 0$, $\int XZ\,dM = 0$, $\int YZ\,dM = 0$

quibus igitur omnibus membris deletis noſtrae aequationes iam mirifice contrahentur.

§. 34. Remanebunt autem tantum iſtae tres formulae integrales

$\int XX\,dM$, $\int YY\,dM$ et $\int ZZ\,dM$,

quarum valores per totum corpus extenſos, ſi ponamus

$\int XX\,dM = A$, $\int YY\,dM = B$, $\int ZZ\,dM = C$

ſex noſtrae aequationes ad ſequentes formas reducentur

I. $iP = M\left(\frac{d\,df}{dt^2}\right)$. II. $iQ = M\left(\frac{d\,dg}{dt^2}\right)$. III. $iR = M\left(\frac{d\,db}{dt^2}\right)$

IV. $iS = M\left(\frac{b\,d\,dg - g\,d\,db}{dt^2}\right) + A\left(\frac{H\,d\,G - G\,d\,H}{dt^2}\right) + B\left(\frac{H'\,d\,G' - G'\,d\,H'}{dt^2}\right) + C\left(\frac{H''\,d\,G'' - G''\,d\,H''}{dt^2}\right)$

V. $iT = M\left(\frac{f\,d\,db - b\,d\,df}{dt^2}\right) + A\left(\frac{F\,d\,H - H\,d\,F}{dt^2}\right) + B\left(\frac{F'\,d\,H' - H'\,d\,F'}{dt^2}\right) + C\left(\frac{F''\,d\,H'' - H''\,d\,F''}{dt^2}\right)$

VI. $iU = M\left(\frac{g\,d\,df - f\,d\,dg}{dt^2}\right) + A\left(\frac{G\,d\,dF - F\,d\,dG}{dt^2}\right) + B\left(\frac{G'\,d\,dF' - F'\,d\,dG'}{dt^2}\right) + C\left(\frac{G''\,d\,dF'' - F''\,d\,dG''}{dt^2}\right)$.

Plus autem hinc in genere pro viribus ſollicitantibus quibuscunque concludere non licet.

Appli-

CORPORVM RIGIDORVM. 229

Applicatio harum Formularum ad caſum quo corpus nullis plane viribus ſollicitatur.

§. 35. Conſtituto igitur puncto I in ipſo corporis centro grauitatis, ternae autem rectae I A, I B, I C ſimul corporis axes principales referant, dum ſcilicet in ſtatu initiali verſabatur; atque elapſo tempore $= t$ ob omnes vires P, Q, R et S, T, U euaneſcentes habebimus primo iſtas tres aequationes:

I. $M\left(\frac{d\,df}{dt^2}\right) = 0$, II. $M\left(\frac{d\,dg}{dt^2}\right) = 0$, III. $M\left(\frac{d\,db}{dt^2}\right) = 0$

quae ſemel integratae praebent

$\frac{df}{dt} = a$, $\frac{dg}{dt} = b$ et $\frac{db}{dt} = c$

ex quibus formulis cognoſcitur, motum centri grauitatis eſſe aequabilem, idoneque illi aequalem, qui ipſi initio fuerit impreſſus. Hinc autem porro integrando colligitur

$f = at$, $g = bt$ et $b = ct$

ſicque via a centro grauitatis deſcripta erit linea recta I i.

§. 36. Tres autem reliquae aequationes ita ſe habebunt

IV. $0 = A\left(\frac{HdG - GdH}{dt^2}\right) + B\left(\frac{H'dG' - G'dH'}{dt^2}\right) + C\left(\frac{H''dG'' - G''dH''}{dt^2}\right)$

V. $0 = A\left(\frac{FdH - HdF}{dt^2}\right) + B\left(\frac{F'dH' - H'dF'}{dt^2}\right) + C\left(\frac{F''dH'' - H''dF''}{dt^2}\right)$

VI. $0 = A\left(\frac{GdF - FdG}{dt^2}\right) + B\left(\frac{G'dF' - F'dG'}{dt^2}\right) + C\left(\frac{G''dF'' - F''dG''}{dt^2}\right)$.

Vnde per integrationem ſtatim deducimus ſequentes

$\mathfrak{A}dt = A(HdG - GdH) + B(H'dG' - G'dH') + C(H''dG'' - G''dH'')$

$\mathfrak{B}dt = A(FdH - HdF) + B(F'dH' - H'dF') + C(F''dH'' - H''dF'')$

$\mathfrak{C}dt = A(GdF - FdG) + B(G'dF' - F'dG') + C(G''dF'' - F''dG'')$

▲ **Figure 4.38** EULER's equations for the motion of rigid bodies as they appear in his book. (Library of the Hungarian Academy of Sciences.)

equals mass times acceleration"—is insufficient for a complete description of the state of motion of continua. To start with, one also needs the equations of torque. The simple view would be that the torque equations follow directly from the equation of motion and from NEWTON's third law of motion, in other words, that it should be sufficient to consider only the external forces because the internal forces make no net contribution. However, it is evident that if we take a small volume element from the continuum and consider the forces acting on it from its environment, then it would be very questionable, if not downright impossible, to reduce this force into pairwise canceling interactions. Moreover, in other instances, the torque itself is given as the primary effect, for example in the motion of magnetic dipoles in a homogeneous magnetic field, and its reduction to forces would falsify the physical picture. It therefore follows that we have to accept the torque equations as being independent of the equations of motion.

A further extension is necessary: In the discussion of continua in the context of mechanics, we do not, of course, go into the atomic microstructure; such an approach would be extremely complicated, if not in fact impossible. For a complete description of the motion of such a body, in addition to the equations of motion and the equations of torque, an additional relationship is necessary that characterizes the given continuum and specifies how it reacts to external effects. In simpler terms, are we talking about a rigid body, an elastic body, an ideal fluid, or a gas? Such a relationship characterizing a substance is usually called a constitutive equation.

4.2.4 *The First Variational Principle in Mechanics: Maupertuis*

The representation of the fundamental law of mechanics as an extremal principle is substantially different from NEWTON's, whose second law of motion—both in its original form and in the modern form given by EULER—brings the kinematic characteristics, that is, the acceleration and the force from the environment, into relationship with each other at a given point in space and time. In contrast, the extremal principle looks at the entire time interval during which the motion takes place; more precisely, it investigates the whole of the path between some arbitrary point in time and space and some later point in time, and states a requirement

Quotation 4.15, ***continued***

them. I believe that they are not introduced without risk. The error, which men like Fermat and those that followed him, have committed, only shows that, too often, their use is dangerous. It can be said, however, that it is not the principle which has betrayed them, but rather, the haste with which they have taken for the principle what is merely one of the consequences of it. It cannot be doubted that all things are regulated by a Supreme Being who, when he impressed on matter the forces which denote his power, destined it to effect the doings which indicate his wisdom. ...

After so many great men have worked on this matter, I hardly dare say that I have discovered the principle on which all the laws of motion are founded. ... What satisfaction to the human mind, in contemplating these laws—so beautiful and so simple—that they may be the only ones that the Creator and the Director of things has established in matter in order to accomplish all the phenomena of the visible world.

—MAUPERTUIS, "The agreement between the different laws of Nature that had, until now, seemed incompatible," read to the Académie des Sciences, April 15, 1744 [Dugas 1957, pp. 262–264, 268–269]

Quotation 4.16

The laws of equilibrium and of motion are necessary truths. A metaphysician would perhaps be satisfied to prove this by saying that it is the wisdom of the Creator and the simplicity of his intentions never to establish other laws of equilibrium and of motion than those which follow from the very existence of bodies and their mutual impenetrability. But we have considered it our duty to abstain from this kind of argument, because it has seemed to us that it is based on too vague a principle. The nature of the Supreme Being is too well concealed for us to be able to know directly what is, or is not, in conformity with his wisdom. We can only discover the effect of his wisdom by observation of the laws of nature, since mathematical reasoning has made the simplicity of these laws evident to us, and experiment has shown us their application and their scope.

It seems to me that this consideration can be used to judge the value of the demonstrations of the laws of motion which have been given by several philosophers, in accordance with the principle of final causes; that is, according to the intentions that the Author of nature might have formulated in establishing these laws. Such demonstrations cannot have as much force as those which are

continued on next page

Quotation 4.16, ***continued***

preceded and supported by direct demonstrations, and which are deduced from principles that are more within our grasp. Otherwise, it often happens that they lead us into error. It was because he followed this method, and because he believed that it was the Creator's wisdom to conserve the same quantity of motion in the Universe always, that Descartes has been misled about the laws of impact. Those who imitate him run the risk of being similarly deceived; or of giving as a principle, something that is only true in certain circumstances; or finally, of regarding something which is only a mathematical consequence of certain formulae as a fundamental law of nature.

—d'Alembert, *Traité de Dynamique, Discours préliminaire*, 1758 [Dugas 1957, pp. 269–270]

for this path. Behind a principle of this type lurks the assumption, or better, the notion, that in the beginning segment of its path, the body already knows—so to speak—that it will have to move in such a way to satisfy certain extremal conditions over the entire path. This means that the principle somehow incorporates the long-outmoded ideas of a final cause (*causa finalis*) and the influence of a higher power consciously prescribing the goal.

Maupertuis's principle is of interest in part because Maupertuis began from metaphysical considerations of just this sort (Quotation 4.15). The history of this principle can serve as an eloquent example of the circuitous paths that human reason follows on its way to knowledge, for not only was Maupertuis's metaphysical interpretation—which even in his own lifetime was seen as obsolete and was attacked sharply (Quotation 4.16)—without any basis, but in addition, the physical hypothesis that he took as his starting point was also false. Maupertuis wanted to bring the minimality principle of Fermat into accord with the generally accepted conception, going back to Descartes and sanctioned by Newton, that the velocity of light is greater in a denser medium than in a less-dense one. This goal is indicated in the title of his work *Accord de différentes lois de la Nature qui avaient jusqu'ici paru incompatibles* (agreement between the different laws of Nature that had, until now, seemed incompatible). We saw in Section 3.5 that Fermat derived the law of refraction from the consideration that light strives to travel from one point to another in minimal time. To arrive at the correct law of refraction, he had to assume—again correctly—that the velocity of light in an optically denser medium is less than in a less-dense medium. Nevertheless, Maupertuis wanted to bring the contrary assumption about the relationship of the velocities of light into accord with Fermat's principle, and he argued as follows: If we ask what goal nature is pursuing in a course of motion (or, expressed in everyday language, in what way nature is trying to be parsimonious), then we can immediately assume that it will prefer the shortest path. This is, in fact, the situation when we consider the propagation of light in a homogeneous medium or look at the path of light reflected in a mirror. But Fermat maintains that in refraction, nature selects the path that requires the least time; the refraction law holds for this path. Obviously, nature has deviated here from the principle of the shortest path. But what motive could nature have—continues Maupertuis—for such asymmetry and to prefer the shortest time over the shortest distance? It should be clear that neither of the two conceptions is correct. When something moves, nature will strive to minimize *effect*, which, according to Maupertuis, is the product of the path length and velocity. If we investigate the law of refraction in reference to Figure 4.39, then the total effect from start to finish is given here by

$$mAR + nRB$$

(here m and n are quantities proportional to the velocities). This expression should be minimal on the path actually taken. Maupertuis generalized the principle and used it to solve various problems. For example, he used it for dealing with collisions, where, we must add, he succeeds only by using a number of—in fact correct—additional assumptions.

Euler, who at that time already had extensive knowledge of the calculus of variations—which studies just such curves with extremal properties—had observed that paths obtained by integrating Newton's equations of motion have certain minimal properties. He thereby arrived at a precise formulation of Maupertuis's





▲ The revolving stage of the history of science: 2000 years BCE the sciences were flourishing in the great river cultures. After Greek, Hellenistic, and Arabic contributions, they could shine with full brilliance in the Europe of the seventeenth century. (Photo: Apollo 8, NASA.)

… So small a shade and Heaven's high human scheme
Be hemmed within the coasts yon arc implies?

Is such the stellar gauge of earthly show,
Nation at war with nation, brains that teem,
Heroes, and women fairer than the skies?

—Thomas Hardy, "At a Lunar Eclipse"

Socrates: The first thing of which I am convinced is that if the earth is a sphere in the middle of the heavens, it has no need of air or any other force to prevent it from falling. The homogeneous nature of the heavens on all sides and the earth's own equipoise are sufficient to hold it, for an object balanced in the middle of something homogeneous will have no tendency to incline more in any direction than any other but will remain unmoved. This … is the first point of which I am persuaded.

Simmias: And rightly so.

Socrates: Further, the earth is very large, and we live around the sea in a small part or portion of it between Phasis and the Pillars of Heracles, like ants or frogs around a swamp; many other peoples live in many such parts of it.

—Plato, *Phaedo* [p. 93]

We may show summarily that the earth is spheroidal … for here the evidence of the senses, and common observation, is alone requisite. The convexity of the sea is a further proof of this to those who have sailed; for they cannot perceive lights at a distance when placed at the same level as their eyes, but if raised on high, they at once become perceptible to vision … Sailors, as they approach their destination, behold the shore continually raising itself to their view; and objects which had at first seemed low, begin to elevate themselves.

—Strabo, *Geographica* [p. 19]



▲ The Acropolis of Athens with the temple to Athena Parthenos (the virgin), protectress of Athens.

The most beautiful and artistically coordinated parts of the Acropolis were built under the supervision of Pericles (ca. 495–ca. 429 BCE). Today, the Acropolis is seen as a symbol of European culture. Significant questions that have occupied mankind ever since were raised, and what was "conceivable was conceived," at the foot of the Acropolis.

Herald: What man is master in this land? To whom
Must I give the word I bring from Creon … ?

Theseus: One moment, stranger.
Your start was wrong, seeking a master here.
This city is free, and ruled by no one man.
The people reign, in annual succession.
They do not yield the power to the rich;
The poor man has an equal share in it.

Herald: That one point gives the better of the game
To me. The town I come from is controlled
By one man, not a mob. And there is no one
To puff it up with words, for private gain,
Swaying it this way, that way. Such a man
First flatters it with wealth of favors; then
He does it harm, but covers up his blunders
By blaming other men, and goes scot-free.
The people is no right judge of arguments;
Then how can it give right guidance to a city?
A poor man, working hard, could not attend
To public matters, even if ignorance
Were not his birthright. When a wretch, a nothing,
Obtains respect and power from the people
By talk, his betters sicken at the sight.

Theseus: What bombast from a herald! Waster of
words,
If it is argument you want—and you yourself
Have set the battle going—listen. Nothing
Is worse for a city than an absolute ruler.
In earliest times, before there are common laws,
One man has power and makes the law his own:
Equality is not yet. With written laws,
People of few resources and the rich
Both have the same recourse to justice. Now
A man of means, if badly spoken of,
Will have no better standing than the weak;
And if the lesser is in the right, he wins
Against the great. This is the call of freedom:
"What man has good advice to give the city,
And wishes to make it known?" He who
responds
Gains glory; the unwilling may hold their
peace.
For the city, what can be more fair than that?
Again, when the people is master in the land,
It welcomes youthful townsmen as its
subjects;
But when one man is king, he finds this
hateful,
And if he thinks that any of the nobles
Are wise, he fears for his despotic power
And kills them. How can a city become strong
If someone takes away, cuts off new ventures
Like ears of corn in a spring field? What use
To build a fortune, if your work promotes
The despot's welfare, not your family's?
Why bring up girls as gentlewomen, fit
For marriage, if tyrants may take them for
their joy—
A grief to parents? I would rather die
Than see my children forced to such a union.
These are the darts I shoot at what you say.

—Euripides, *The Suppliant Women* [pp. 153–154]

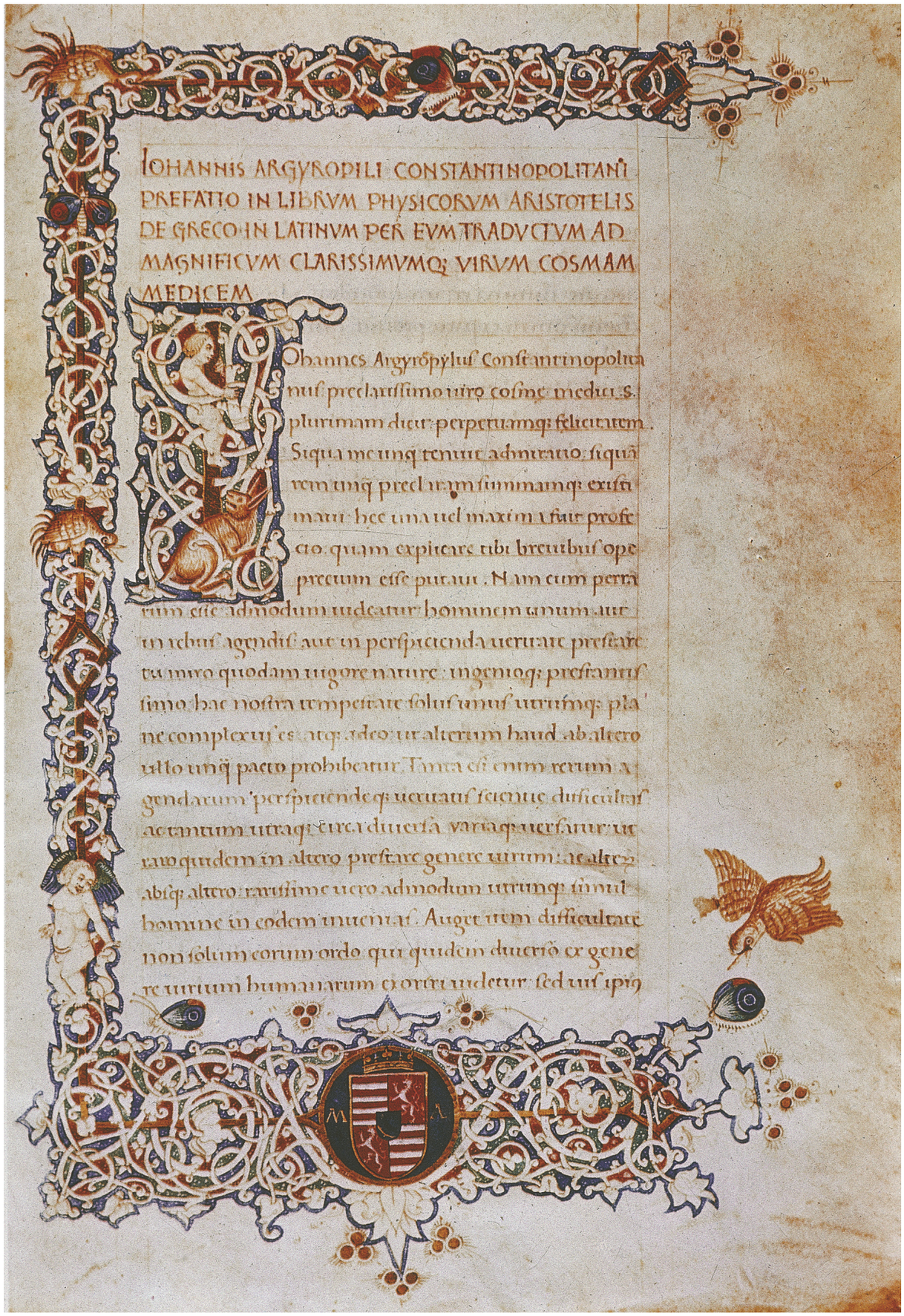
Chorus:
Many the wonders but nothing walks stranger than man,
This thing crosses the sea in the winter's storm,
making his path through the roaring waves.
And she, the greatest of gods, the earth –
ageless she is, and unwearied—he wears her away
as the plough go up and down from year to year
and his mules turn up the soil.

Gay nations of birds he snares and leads,
wild beast tribes and the salty brood of the sea,
with the twisted mesh of his nets, this clever man.
He controls with craft the beasts of the open air,
walkers on hills. The horse with his shaggy mane
he holds and harnesses, yoked about the neck,
and the strong bull of the mountain.

Language, and thought like the wind
and the feelings that make the town,
he has taught himself, and shelter against the cold,
refuge from rain. He can always help himself.
He faces no future helpless. There's only death
that he cannot find an escape from. He has contrived
refuge from illnesses once beyond all cure.

Clever beyond all dreams
The inventive craft that he has
Which may drive him one time or another to well or ill.
When he honors the law of the land and the gods'
sworn right
high indeed in his city; but stateless the man
who dares to dwell with dishonor. Not by my fire,
never to share my thoughts, who does these things.

—Sophocles, *Antigone*


IOHANNIS ARGYROPILI CONSTANTINOPOLITANI
PREFATIO IN LIBRVM PHYSICORVM ARISTOTELIS
DE GRECO IN LATINVM PER EVM TRADVCTVM AD
MAGNIFICVM CLARISSIMVMQ; VIRVM COSMAM
MEDICEM

Iohannes Argyropylus Constantinopolita
nus preclarissimo viro cosme medici s.
plurimam dicit perpetuamq; felicitatem.
Siqua me unq tenuit admiratio siqua
rem unq predicandam summamq; existi
mauit hec una uel maxime fuit profec
to quam expicare tibi breuibus ope
precium esse putauit. Nam cum perua
rum esse admodum uideatur hominem unum aut
in rebus agendis aut in perspicienda ueritate prestare
tu uiro quodam uigore nature ingenioq; prestantis
simo hac nostra tempestate solus unus utrumq; pla
ne complexus es atq; ideo ut alterum haud ab altero
ullo unq pacto probibeatur. Tanta est enim rerum a
gendarum perspicendaq; ueritatis scientie difficultas
ac tantum utraq; circa diuersa variaq; uersatur ut
raro quidem in altero prestare genere uirum: ac alte
ro ab altero rarissime uero admodum utrumq; simul
homine in eodem inueniat. Augent item difficultate
non solum eorum ordo qui quidem diuerso ex gene
re uirtutum humanarum exoriri uidetur sed uis ipsa

◀ A page from Aristotle's *Physics*, translated into Latin from the Greek by Johannes Argyropilus Constantinopolitanus. (Bibliotheca Corviniana, Niedersächsische Staats- und Universitätsbibliothek. Göttingen.)



Magne compositionis Claudii Ptolomei libri a Georgio trapezuntio traducti Incipiunt Capitula primi libri Feliciter

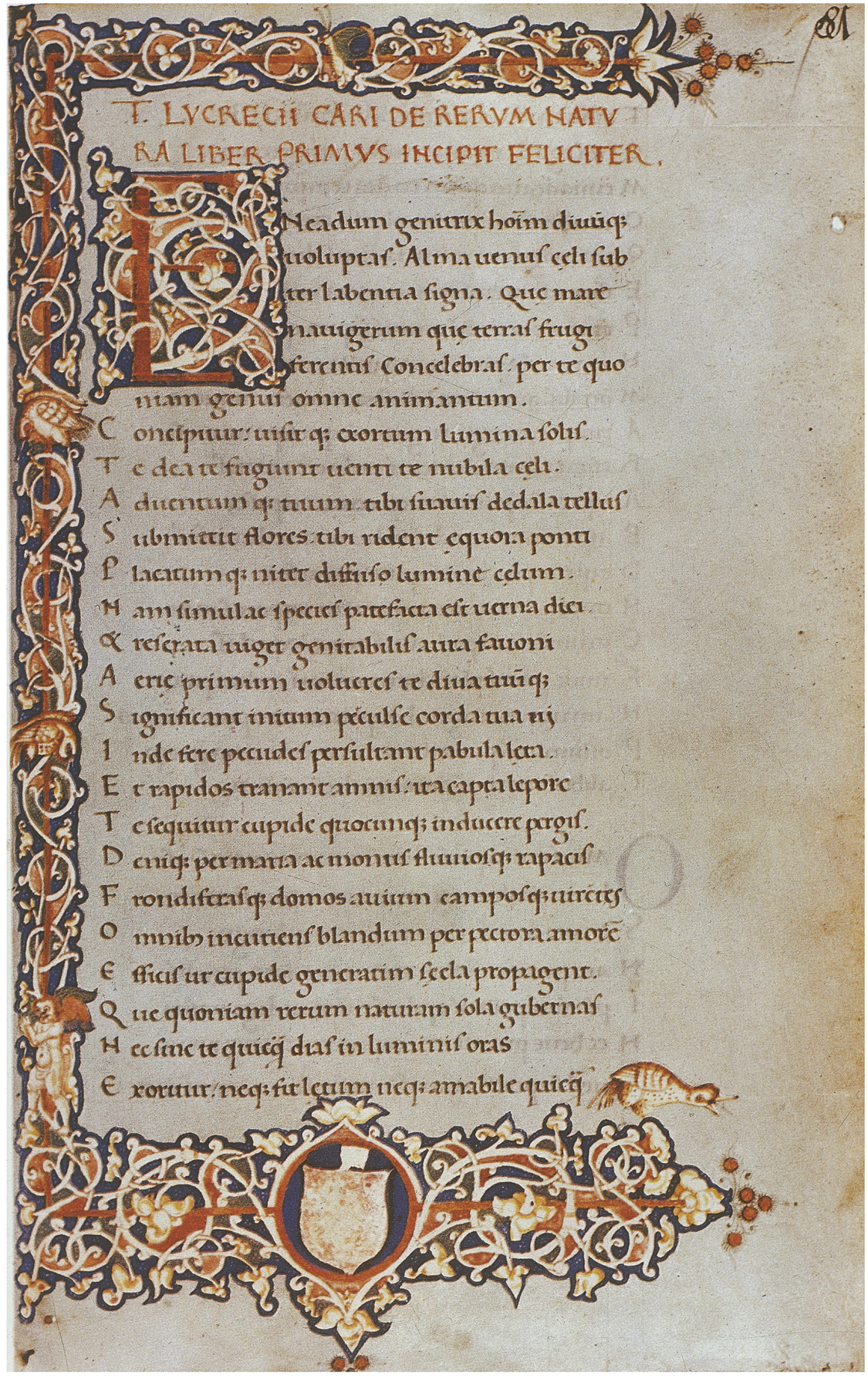
Prefatio auctoris
De ordine huius doctrine
Q: sphere ac globi modo celu conuoluitur
Q: terra quoq; spherica sit ad sensum quantum ad uniuersas partes
Q: terra in medio celi sita est
Q: terra si puncti ad celes[tia]
Q: tia nullo motu progressiuo mouetur
Q: dupplex in celo primorum motuum differentia sit
De quantitate rectarum linearum que in circulo producuntur
Tabula cordarum & arcuum
De arcu qui est inter tropicos
Theoremata que ad spericas demonstrationes premittunt
De arcubus qui sunt int[er] equinoctialem atq; obliquum circulum
Tabula solaris obliquationis
De ascensionibus in recta sphera

Sequitur Capitulum primum Prefatio Auctoris

ER OPTIME MIHI VI dentur o vere qui bene philosophari sunt speculatiuam philosophie partem ab actiua seperasse Nam & si actiue accidat parti ut prius speculatiua sit magnam tamen differentia in ipsis inesse non solum qua nonnulle virtutes morales absq; disciplina et multis inesse possunt cum speculatiuam scientia sine doctrina consequi impossibile sit Verum et qua maxima utilitas in altera ex frequenti actione que in ipsis rebus habetur in altera ex progressu speculationum fieri solet Hinc opus esse nobis putauimus ut actionem quidem cogitandi motibus sic temperemus q; ne in minimis quidem considerationis eius obtunsa minus que ad pulcram ordinatamq; mentis constitutionem perducat Otiu autem maxime ad doctrinam theorematum que plurima pulcherrimaq; sunt & precipue illorum que propie mathematica dicuntur conuertamus Commode namq; ad modum Aristotelis speculatiuam ptem in tria rursus genera partitur physicum Mathematicum Theologicum Nam cum res omnes ex materia & forma et motu constent quorum singula quis minime seorsum a subiecto inspici possint intelligi tamen sine reliquis possunt Primam quidem primi omnium motus causam si quis in suma simplicitate accipiat dm

◀ A page from the *Almagest* of Ptolemy; translated from the Greek into Latin by Georgius Trapezuntius. (Bibliotheca Corviniana, Nationalbibliothek, Vienna.)

▶ Lucretius's poem *De Rerum Natura*, in Latin. (Bibliotheca Corviniana, Nationalbibliothek, Vienna.)



▶ Justinian (482–565), Eastern Roman (Byzantine) emperor 527–565, was an important legal reformer, the creator of the *Corpus Juris Civilis* who sought to revive the glory of the ancient Roman Empire; but he also oversaw the closing of the philosophical school in Athens (529). (Ravenna, San Vitale; photo: Leonhard von Matt.)



Aristotle is an artist of perfect wisdom, adviser of nature and her remarkable partner in her deeds and masterworks; he was sent to us as deacon in matters of nature to proclaim each of her utterances successfully, truthfully, and precisely, exactly as she is to be found and according to what principles she was created.

Whosoever fails to hold Plato's or Aristotle's wisdom for a great, indeed the greatest, treasure and worthy of wonder, indeed more than wonderful, such a person—in my opinion—has no common sense for the entirety of human life. Indeed, every utterance of these masters is such as the utterances of Pythia, as they have noted, namely that by the fact alone that they have been uttered by them and determined by them, that alone suffices to hold us back from desiring anything greater, to attempt to add further thoughts or make additional discoveries. It should satisfy everyone to recognize them with resignation as teachers and explicators.

—Theodorus Metochites, *Hypomnema*

Roman law was the basis of Byzantine law. Previously, state law and common law existed together in the Greek Near East, but gradually they interpenetrated; the religious views of Christianity gained power in legal reasoning. Then Justinian brought about the grand summation of Roman law and made it the single basis of law in the Byzantine state. The effects were immeasurable. Though the Eastern Empire continued unbroken the ancient culture in almost every other direction, through Justinian's deed, there existed a state governed by the rule of law. This superiority over all barbarian peoples of the Middle Ages was never forgotten. Passed on by Byzantium, the effects of Roman law also extended to the Slavic peoples … .

The Church in the Byzantine Empire never forgot that an emperor, the founder of the empire, gave it security to exist. It almost willingly shut it eyes to the fact that it was closely linked to the state, indeed was to become part of the state organization and therefore must lose its internal freedom. The Church in the West later, after a difficult struggle, regained its independence, but such a goal could be achieved only with a complete break with the Eastern Empire. Constantine considered himself the *pontifex maximus* even in the Christianized state; Justinian, as a theologian, exercised decisive influence over the establishment of dogma.

Even in the Middle Ages, the Greeks always viewed education and attention to spiritual matters as a duty. Until well into the Justinian era, antiquity continued its influence, and while Christianity had the goal of education, at first it changed neither the methods nor the means. From the sixth century the picture began to change. Justinian consciously turned from the Hellenistic educational ideal. It is a mistake to view the closing of the philosophical school in Athens for an arbitrary act; it represents, however, a milestone and turning point in the history of Greek intellectual life. From this time forward, the education of the Byzantines must be exclusively Christian. The immediate effect was a period of decline following the era of Justinian. But intellectual life in the East did not become completely submerged as in Western Europe, and more quickly than here, there came a time when the knowledge of not quite forgotten antiquity gave new impetus to education …

Thus in Byzantium, science never came into serious conflict with the Church, but also never broke through to complete freedom. A scholastic streak ruled the thought of the Greek middle ages long before the Scholastic era began in the West. Aristotle was never unknown to the Greek world; his effect on the ideas of Greek theology can be seen already in the sixth century in the writings of Leontius of Byzantium. Plato was considered an enemy of the Church. Memories were long regarding the battles that Christianity had had to wage even in the time after Constantine against Neo-Platonism, and it took half a millennium before Platonic ideas again gained currency. Psellus considered himself the prophet of Plato and declared war against Aristotle; in the time of the Comnenus dynasty, Platonism again became a force. But this development destroyed the Latin conquest, and the thought of the previous century stands more under the sign of Aristotle. In this formulation it exercised a strong influence on the intellectual life of the West; the logic and physics of Nicephorus Blemmydes had a wide readership in the universities of France and Italy in the fifteenth and sixteenth centuries. Conversely, in the fourteenth century, Thomism found its way to Byzantium, and Gemistus Pletho was entrusted with the ideas of Neo-Platonism not in Constantinople or Mistra, but in the academy at Florence.

All higher education was an inseparable union of the religious–theological and humanistic. There was no prominent Byzantine who did not know the Holy Writ and the teachings of the Church fathers precisely and was just as familiar with the world of Homeric epic and a greater or lesser circle of classical literature. Instruction in the Middle Ages remained close to the fundamentals at the time of the Roman Empire both in method and object of study, with language instruction, rhetoric, and philosophy leading the field. In the natural sciences as well, no one went beyond the borders of Aristotelian knowledge, while in medicine and technology, the Byzantines were soon overtaken by the Arabs, and later by the West. The Greek tradition also remained predominant in this area and established the boundaries. It was disastrous for world history that the Slavic peoples who were offered a higher intellectual culture by the Byzantines accepted only Christian–theological instruction. The world of the Greek intellect remained closed to them and found an entry only centuries later in the West.

—A. Heisenberg, *Staat und Gesellschaft des byzantinischen Reiches*, 1923



▲ *Cistercian Monks at the Harvest*, by Jörg Breu the Elder, 1500. (Zwettl, collegiate church.)

XLVII. *On manual labor.* Idleness is the enemy of the soul. Therefore the brethren should be occupied at certain time in manual labor, and again at fixed hours in sacred reading. To that end we think that the times for each may be prescribed as follows. From Easter until the Calends of October, when they come out from Prime in the morning let them labor at whatever is necessary until about the fourth hour [from 6am to 10am], and from the fourth hour until about the sixth [noon] let them apply themselves to reading. After the sixth hour, having left the table, let them rest on their beds in perfect silence; or if anyone may perhaps want to read, let him read to himself in such a way as not to disturb anyone else. …

And if the circumstances of the place or their poverty should require that they themselves do the work of gathering the harvest, let them not be discontented; for then are they truly monks when they live by the labor of their hands as did our Fathers and the Apostles. Let all things be done with moderation, however, for the sake of the faint-hearted. …

On the days of Lent, from morning until the end of the third hour let them apply themselves to their reading, and from then until the end of the tenth hour let them do the work assigned them. And in these days of Lent they shall each receive a book from the library, which they shall read straight through from the beginning. These books are to be given out at the beginning of Lent.

But certainly one or two of the seniors should be deputed to go about the monastery at the hours when the brethren are occupied in reading and see that there be no lazy brother who spends his time in idleness or gossip and does not apply himself to the reading, so that he is not only unprofitable to himself but also distracts others. If such a one be found (which God forbid), let him be corrected once and a second time; if he does not amend, let him undergo the punishment of the Rule in such a way that the rest may take warning …

On Sundays, let all occupy themselves in reading, except those who have been appointed to various duties. But if anyone should be so negligent and shiftless that he will not or cannot study or read, let him be given some work to do so that he will not be idle.

—*St. Benedict's Rule for Monasteries* [pp. 67–70]

The river enters the abbey as much as the well acting as a check allows. It gushes first into the corn-mill, where it is very actively employed in grinding the grain under the weight of the wheels and in shaking the fine sieve which separates flour from bran. Thence it flows into the next building, and fills the boiler in which it is heated to prepare beer for the monks' drinking, should the vine's fruitfulness not reward the vintner's labor. But the river has not yet finished its work, for it is now drawn into the fulling-machines following the corn-mill. In the mill it has prepared the brothers' food and its duty is now to serve in making their clothing.... Thus it raises and alternately lowers the heavy hammers and mallets ... of the fulling-machines.... Now the river enters the tannery where it devotes much care and labor to preparing the necessary materials for the monks' footwear; then it divides into many small branches and, in its busy course, passes through various departments, seeking everywhere for those who require its services for any purpose whatsoever, whether for cooking, rotating, crushing, watering, washing, or grinding.... At last, to earn full thanks and to leave nothing undone, it carries away the refuse and leaves all clean."

—Description of Clairvaux Abbey [quoted in Mumford 1961, pp. 258–259]



▲ The "technological revolution" in the Middle Ages led to a spectacular increase in productivity in agriculture. The various pieces of equipment shown in the picture are still in use today, in some cases even in Europe. (*Grimani Breviarium*, Venice, Biblioteca Nazionale Marciana.)



◀ University life in the Middle Ages (just like today). (Miniature, Laurentius de Voltolina, *Liber ethicorum*, second half of the fourteenth century: Staatliche Museen Berlin.)

Tam pro papa, quam pro rege
Bibunt omnes sine lege. …
Bibit constans, bibit vagus,
Bibit rudis, bibit magus; …
Bibunt omnes sine meta,
Quamvis bibant mente laeta.
Sic nos rodunt omnes gentes
Et sic erimus egentes.

Rulers papal, rulers regal
Drink forgetting what is legal, …
Travelers drink and so do locals,
Learned people drink like yokels, …
All keep drinking without finish
While their joy does not diminish.
Thus we in the drinking nations
Drink ourselves up and our rations.

Bacchus tollat
Venus molliat
Vi bursarum pectora –
Et inmutet
Et computet
Vestes in pignora.

Let Wine awake
And let Love slake
Our hearts with gifts of legal tender,
And cash-in
Our clothes by pawning to a lender.

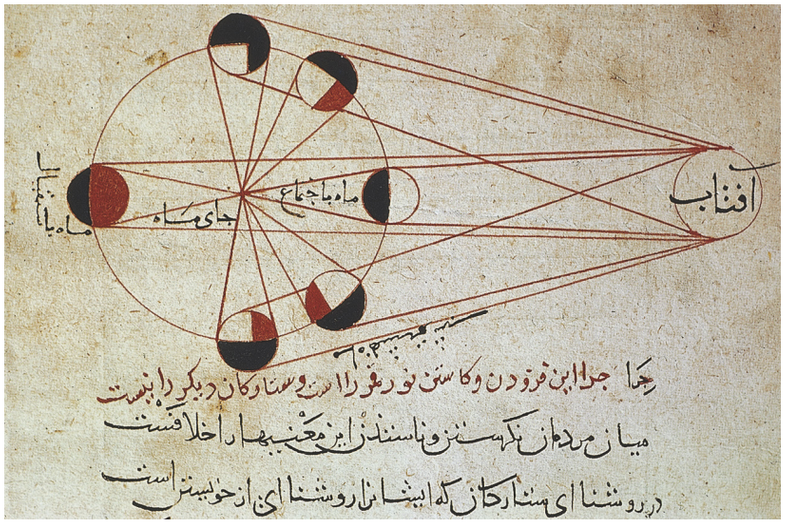
—*Carmina Burana* [translated by Paul Franz]

The master mounts the lectern, picks up a book by Aristotle, reads a short section, and then explains it—this is the process of "commentary" (as recorded in, for example, Thomas Aquinas' *Commentarius super libros De caelo et mundo Aristotelis* (Commentary on Aristotle's *On the Heavens*)). Many commentaries were even further from the original text, referring to another commentary instead: Averroes, "the commentator," was usually the model. But other teachers asked questions (Does there exist a vacuum in nature?) and sought an answer in Aristotle's works (Albert of Saxony, *Questiones in Aristotelis libros De caelo et mundo*). We may adjudge the intellectual activity here as quibbling and hair-splitting. Indeed, it does not seem the best framework for engendering creativity in students. We must also remember the mistrust with which the Church watched for any conclusions contrary to the dogmas of the faith …

We can also come to the opposite conclusion by observing how the works of Aristotle were evaluated: on one side, the joy in possessing the treasure of ancient knowledge and the enthusiasm for its general application and, on the other, the counterattack of the Church with intellectual tools of the same origin—namely Aristotelian logic—had within them the possibility of a dynamic development. At the beginning of the thirteenth century, in the year 1215, the teaching of the natural-philosophical views of Aristotle was forbidden at the University of Paris, but by 1255 all of his works were in the list of textbooks at the same university, indeed frequently in the materialistic–deterministic coloring of Averroes. In 1277, Pope John XXI instructed the bishop of Paris to open an investigation into the relationship between—as we would say today—religious ideology and science. In his overzealousness, the bishop, with the help of theologians, hastily put together 219 theses—among them even some from Thomas Aquinas—whose teaching was forbidden under the threat of excommunication. The list included a few theses maintaining the strict determinism of nature, others testified to the doctrine of the "dual truth." There were even some that questioned whether theology was a science, or whether it was at all useful. We must emphasize that the theologian scholars were not content with the administrative silencing of their opposition. Determinism called into doubt the omnipotence of God. The strategy of their spiritual counterattack was this: God created the world well ordered, but in his omnipotence, he could have created it differently had he wished. The world is therefore not of necessity the way that it, in fact, is. Its laws can therefore not be derived from some kind of first principles; no necessary relationships can exist between contingent things (Ockham, Nicholas of Autrecourt).

There are modern historians of science who maintain (P. Duhem), and others who doubt (A. Koyré), that it was really the theologians who arrived, through such considerations, at radical empiricism, which led, with the contributions of scientists (Buridan, Albert of Saxony), to one of the most important basic principles in Western science: The world is well ordered, but its laws cannot be understood through pure thought alone; empirical observations must also be made.

PLATE X





▲ Examples of contributions made by Islamic science: explanation of the phases of the Moon in the book by AL-BIRUNI (left); color, form, and geometry in Islamic art (right).
(From [Nasr 1976]; photo: Roland Michaud.)

Scientists from the Near East often speak of Islamic instead of Arab science, thereby emphasizing the basic spiritual and intellectual nature of this culture, which is the common heritage of a number of nations and peoples, rather than of a single people. The accompanying quotation gives some insight into the attitude of the contemporary Persian scholar SEYYED HOSSEIN NASR.

Islam became heir to the intellectual heritage of all the major civilizations before it save that of the Far East, and it became a haven within which various intellectual traditions found a new lease upon life, albeit transformed within a new spiritual universe. This point must be repeated, particularly since so many people in the West wrongly believe that Islam acted simply as a bridge over which ideas of Antiquity passed to mediaeval Europe. As a matter of fact nothing could be further from the truth, for no idea, theory or doctrine entered the citadel of Islamic thought unless it became first Muslimized and integrated into the total world view of Islam. Whatever could not make its peace (salām) with Islam was sooner or later dispelled from the arena of Islamic intellectual life or relegated completely to the margin of the tapestry of the Islamic sciences.

Before the rise of the Islamic sciences many civilizations had come and gone and produced sciences of various domains of reality and with different orders of perfection. The vast proto-history of science which leads to the deepest recesses of human history and which reveals with each new discovery astounding achievements need not concern us here. It is to the immediate predecessors of Islamic science that we must turn. The two outstanding river civilizations of Egypt and Mesopotamia had already produced medicine and mathematics of extraordinary quality before the Greek philosophers and scientists came upon the stage to theorize about them and to develop them further. Basing themselves upon this long tradition of the study of the heavens as well as the world of nature, the Greeks in turn produced THALES, PYTHAGORAS, PLATO and ARISTOTLE within a relatively short period of less than three centuries before the centre of their scientific activity shifted to Alexandria. There in the soil of Egypt, at the moment of the twilight of Greek power and the dying gasps of ancient Egyptian civilization, a new synthesis of Greek, Egyptian and Oriental learning was achieved leading to one of the most prolific periods of the history of science, which produced such men as EUCLID, PTOLEMY and indirectly GALEN, figures who entered Islamic civilization almost as if they had been Muslim teachers and masters. It is important for an understanding of Islamic science to realize that the Graeco-Hellenistic heritage reached Islam not directly from Athens but through Alexandria, that PLATO was seen mostly through the ideas of the Neoplatonists and ARISTOTLE through ALEXANDER APHRODISIAS and THEMISTIUS. Alexandrian science, in its combining of mystical elements with rigorous logic, in synthesizing various scientific traditions, in basing all the sciences upon a hierarchy related to the "mode of knowledge" and in many other ways, is a historical anticipation of Islamic science and in fact became transformed into Islamic science in the same manner that the Alexandria of PTOLEMY and ORIGEN became transformed into the jewel of Islamic Egypt, the home of such masters as IBN 'ATA ALLAH AL-ISKANDARI.

But the transmission of the Graeco-Hellenistic tradition to Islam was not a direct one. Several centuries of Christian history lie between the golden age of Alexandria and the rise of Islam. …

As far as the Persian world is concerned, it too transmitted to Islamic civilization many sciences, some of its own and some ultimately of Greek and Indian origin. During the Sassanid period the Persians developed Jundishapur, near the present day Persian city of Ahvaz, as a university centre which in fact grew steadily until it became heir to both Antioch and Edessa, a haven for men of learning everywhere, and by the time of the fall of the Sassanids without doubt the most important centre of learning in Western Asia, particularly in the field of medicine. Jundishapur was a cosmopolitan gathering place where Persian, Greek and Indian men of learning met and worked together. In many fields, especially in medicine, this school more than any other was the living link between Islamic science and the ancient world. Meanwhile the Persians themselves showed special interest and made important discoveries in astronomy on the one hand and pharmacology on the other at the same time that they cultivated avidly both the Indian and the Greek sciences. Persia made major contributions to nearly every facet of Islamic civilization, became one of its major centers and played a central role in its creation. But in the field of science Persia played a triple role as far as transmission was concerned: it transmitted its own scientific tradition to Islam, as seen in such works as the *Royal Astronomical Tables* (*Zīj-i shahriyār*); it made available certain aspects of Greek learning to Muslims which had been translated into Pahlavi or even Syriac but taught in centers of learning within Persia such as Jundishapur; and finally it passed on to Islam many of the Indian sciences, especially medicine, astronomy and natural history, which had been cultivated by the Sassanids. A particularly outstanding example of this latter role is the *Kalīlah wa Dimnah*, … becoming rapidly one of the masterpieces of Arabic literature and at the same time a source for Muslim natural history.

—*Islamic Science* [pp. 9–11]

Books whose appearance and content are typical of the sixteenth and seventeenth centuries: Augustine, *De Civitate Dei* (1527); a book of sermons; a book of popular science by Casparus Schott: *Physica curiosa* (1675); a book dedicated to Alfonso of Aragon on the rights and duties of kings (1608) (photo: Tamás Szigeti).

The typography and layout of the pages fully corresponds to the books of today. However, the bindings are more massive and more ornamental.

In the earliest times of book publishing, the printing imitated the decoration of the codices. We can see this in the so called 42-line Gutenberg Bible, illustrated in Plate XIII. The illuminations in these books were still painted in by hand. However, esteem for the codices did not prevent bookbinders from cannibalizing them for stiffening the spines of books or using them directly as a protective book cover. Such is the case with the book shown in the foreground of the photograph. Thus, some rare historical material can be found inside such books.

The purpose of the understanding, and perhaps also mocking, smile of the Chinese statuette is to remind the reader that book printing originated in China. Paper was used there as early as 200 bce. The production of this writing material was then taken up by the Arabs in Samarkand in the eighth century; in Spain, the first paper mill was put into service in the twelfth century. As to the art of book printing, the Chinese began with wood blocks, but after the eleventh century, they were using movable-type ideograms made of baked clay, and since the fourteenth century, movable metal type for the reproduction of texts. Printing with wood blocks was also known in Europe in the late Middle Ages (*Biblia pauperum*).

With the invention of the modern printing techniques by Johannes Gutenberg (ca. 1398–1468), the quality of printing not only equaled that in East Asia, but surpassed it. That was not only because of the advantage of our phonetic alphabet in comparison to the 40,000 characters of Chinese, but also due to the fact that the use of a printing press, which was Gutenberg's key innovation, was completely unknown in the Far East. Indicative of the demand for printed books is the rapid increase in the number of printing houses. Around the year 1500, there were approximately 250 in Europe, with several dozen, perhaps even one hundred or more, in the larger cities such as Cologne, Venice, Bologna, and Paris.

The spread of printing presses is primarily the contribution of the German printers, writes Fisher in his *History of Europe*, and he quotes a writer of the era:

> "As the apostles of Christ … formerly went through the world announcing the good news, so in our days the disciples of the new art spread themselves through all countries, and their books are as the heralds of the Gospel and the preachers of truth and of science." [pp. 465–466]

Concerning the number of incunabula—as books printed before 1500 are called; the word *incunabulum* means "cradle"—only a rough estimate can be given. The order of magnitude is about 30,000 distinct works for a total of a few hundred thousand incunabula.

In the border areas of the European cultural world, the situation was more or less as follows: In Hungary, the first printing house was founded in 1473 in Buda. In Poland, a press opened in Cracow in 1474, while interestingly, the first English establishment opened only two years later. In Russia, Ivan Fedorov's printing house opened in 1564, while the first press in North America was at Harvard in 1639.





From a work of Thomas Aquinas on a medieval manuscript from the Bibliotheca Corviniana (Nationalbibliothek, Vienna).

The thoughts of "doctor angelicus" still speak to us today, in contrast to his "philosophical" writings, which are closely linked to the Christian ideology and which we can view today as mental exercises of dubious utility, save perhaps for sharpening skills for argumentation and debate.

However, if we try to look at the issues in the context of the times, we might arrive at a different conclusion and better understand the ferocity—literally, at the point of a sword—of the fight over the problem of universals that continued though the entire medieval period. This is a seemingly abstract philosophical question about how a general concept (*mankind*, for example) is related to a concrete instance (a *concrete person* in this case). The two extremes are the realists and the nominalists. The *realists* answer this question in the spirit of Plato and Augustine: The general is what really exists, it is the ideal of the concrete instance, the thought of God. The substance of like things represents a unity, while the concrete is differentiated through accidental features, that is, through secondary inessential properties (*universalia ante rem*, universality before things). The *nominalists*, in contrast, maintain that it is only to the concrete that a real existence can be assigned; the general concept is only a summarizing name (*nomen*, whence the name; *universalia post rem*, universality after things). A middle of the road was represented by *conceptualism*: The universal also exists, but only in the individual things (*universalia in re*). The problem of universals flared up as early as the eleventh century. Gerbert, later Pope Sylvester II, was an enthusiastic nominalist; Roscellinus (teacher of Abelard) was led into heresy by his nominalism. So, for a time, the realists were winning; only in the fourteenth century could nominalism reassert itself (Ockham).

Among the many religious theses that affected, intellectually or emotionally, the daily lives of the people of the Middle Ages, we can highlight just one. If the general does not really exists, then mankind could not have fallen into sin through Adam, nor could Christ have redeemed mankind. It is no coincidence that Marxism considered the problem of universals as the dialectic fight during the period of feudalism between idealism (namely realism; be careful with the confusing nomenclature!) and materialism (namely nominalism).

Predestination and freedom. It would be an error to maintain that human affairs and events are not subject to divine foreknowledge and predestination. But it would be no less an error to teach that divine providence and predestination of human affairs imposes necessity. If so, with one stroke, freedom of choice, the power of counsel, the utility of laws, the desire to do good, and the righteousness of reward and punishment would cease.

Above all, we must consider that God's knowledge is different from that of men. Namely, mankind is subject to time, and so he recognizes things in relation to time; some things he sees as present, while of others he speaks as past, while yet others are seen as in the future. God, in contrast, stands above the flow of time; his existence is eternal, and thus his knowledge is not limited by time, but is eternal. In eternity there is no earlier and later, since eternity has no part in change. Thus eternity exists at always the same time. And so what to us seems present, past, or future is seen by God as uniformly present, known with certainty and without error, without compelling anything of chance in his existence. We may obtain an accurate picture by comparing the course of time with the traffic in the street. When someone goes out into the street where many people are walking, he can see those approaching him well, but he does not know who is approaching him from behind. But someone observing from above, who can see the entire road, sees all who pass. Thus, too, is man within time and cannot see the course of time; he sees only what occurs before his eyes, that is, present things and a small amount of the past; the future he cannot foresee with certainty. God, however, from the height of eternity, sees with certainty all things as present without forcing necessity on what might come to pass. …

Contemplation of truth. If now the greatest happiness of mankind is to be sought not in external things, which are called material goods, not in bodily superiority … and also not in intellectual superiority in daily affairs, that is, cleverness and astuteness, then nothing else remains but that the ultimate happiness of mankind lies in contemplation of truth.

This activity is namely that which belongs to mankind alone; no other living being takes part in it. And it offers no other goal, for the contemplation of truth contains within it its own goal. In this activity, mankind is like the higher beings, for it is only this among human activities in which God and the angels partake.

In this activity man is self-sufficient, for here he is least in need of external help.

It is clear that the entirety of all human activities is directed to this goal. The contemplation of truth demands an unspoiled state of the body: To this end all things serve that are necessary for the maintenance of life. This can be attained through moral virtue and through wisdom. As well, this activity requires freedom from all outside disturbance. All regulations of civil life aim in this direction. If we go into the matter fundamentally, we may confirm that all human activity and callings have to serve the search for truth.

—Thomas Aquinas

You call him a dumb ox; someday the whole world will pay attention to his bellowing.

—Albertus Magnus, on his pupil Thomas Aquinas



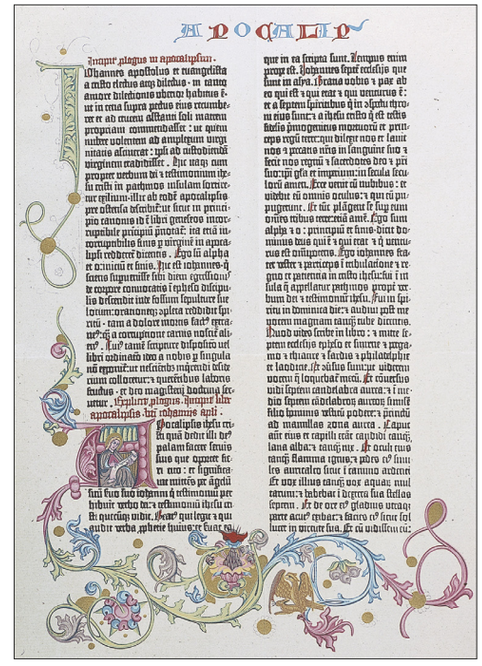
Luca Pacioli (ca. 1446–1517), painting by Jacopo de Barbari. The young man is assumed to be one of Pacioli's patrons, Guidobaldo da Montefeltro, duke of Urbino (the painting was completed in 1495; Capodimone, Naples).

The picture even shows two pieces from his collection of 60 models of regular and semiregular solids. Pacioli was no original thinker; yet he is very significant in that in his book *Summa de arithmetica geometria proportioni et proportionalita*—which appeared in 1494, in Italian, and was one of the first printed books in mathematics—he collected the works of his predecessors, carefully listing his sources as well. Pacioli drew from the work of ancient scholars—Plato, Aristotle, Euclid, Archimedes, Boethius—and from the medieval works of Thābit, Ahmed ibn Yusuf, Leonardo of Pisa (Fibonacci), Bradwardine, Albert of Saxony, Jordanus Nemorarius, and Sacrobosco. His 1509 book *De divina proportione* was illustrated by Leonardo da Vinci. Progress with respect to ancient science is evident in his notation, which is closer to what we use today, especially in the use of Indo-Arabic numerals, which greatly simplified practical computation. The introduction of double-entry bookkeeping is also due to Pacioli. In the final chapter of his book, he (incorrectly) states that the third-degree equation in the form $x^3 + px = q$ is just as unsolvable a problem as the squaring of the circle. But, in fact, Niccolò Tartaglia found the solution to equations of this form. His contemporary and rival Girolamo Cardano (1501–1576) published a method for reducing the general third-degree equation to this simplified type and even for the solution of equations of the fourth degree.

François Viète (1540–1603) made significant contributions in systematizing the solution of algebraic equations, in giving approximation procedures for the numerical solution of equations, in spherical trigonometry, and in the introduction of practical notations (*Canon mathematicus*, 1579; *In artem analyticam isagoge*, 1591). He found an exact expression for π as an infinite product of factors f_n for $n = 0, 1, 2, \ldots$,

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}\sqrt{\cdots} \qquad \text{where } f_0 = 0 \text{ and } f_n = \sqrt{\frac{1}{2}+\frac{1}{2}f_{n-1}},$$

and used it to calculate the value of π to ten decimal places. Christopher Clavius (1538–1612), the father of the Gregorian calendar reform, should also be mentioned here as the encyclopedic summarizer and great teacher of this era at the Jesuit college in Rome. Descartes, for instance, studied from his books. Jesuits working in China also used these books to try to convince Chinese scholars of the higher state of development of Western science.



▶ This is how the Gutenberg Galaxy started. In Mainz, in 1456 (or even earlier), the famous "42 line" Bible was printed: of the 180 copies, about 48 have survived to this day.

PLATE XIV

In a pamphlet by John Donne, Ignatius Loyola (1491–1556), founder of the Jesuit order, mocks the accomplishments of Copernicus, who—in the story—has applied to a high office with Lucifer, the prince of the damned.

> But for you, what new thing have you invented, by which our Lucifer gets anything? What cares hee whether the earth travell, or stand still? Hath your raising up of the earth into heaven, brought men to that confidence, that they build new towers or threaten God againe? Or do they out of this motion of the earth conclude, that there is no hell, or deny the punishment of sin? Do not men beleeve? do they not live just as they did before? Besides, this detracts from the dignity of your learning, and derogates from your right and title of comming to this place, that those opinions of yours may very well be true. … But your inventions can scarce bee called yours, since long before you, Heraclides, Ecphantus, & Aristarchus thrust them into the world: who notwithstanding content themselves with lower roomes amongst the other Philosophers, & aspire not to this place, reserved onely for Antichristian Heroes …. Let therefore this little Mathematitian (dread Emperour) withdraw himselfe to his owne company.
>
> —*Ignatius His Conclave*, 1611

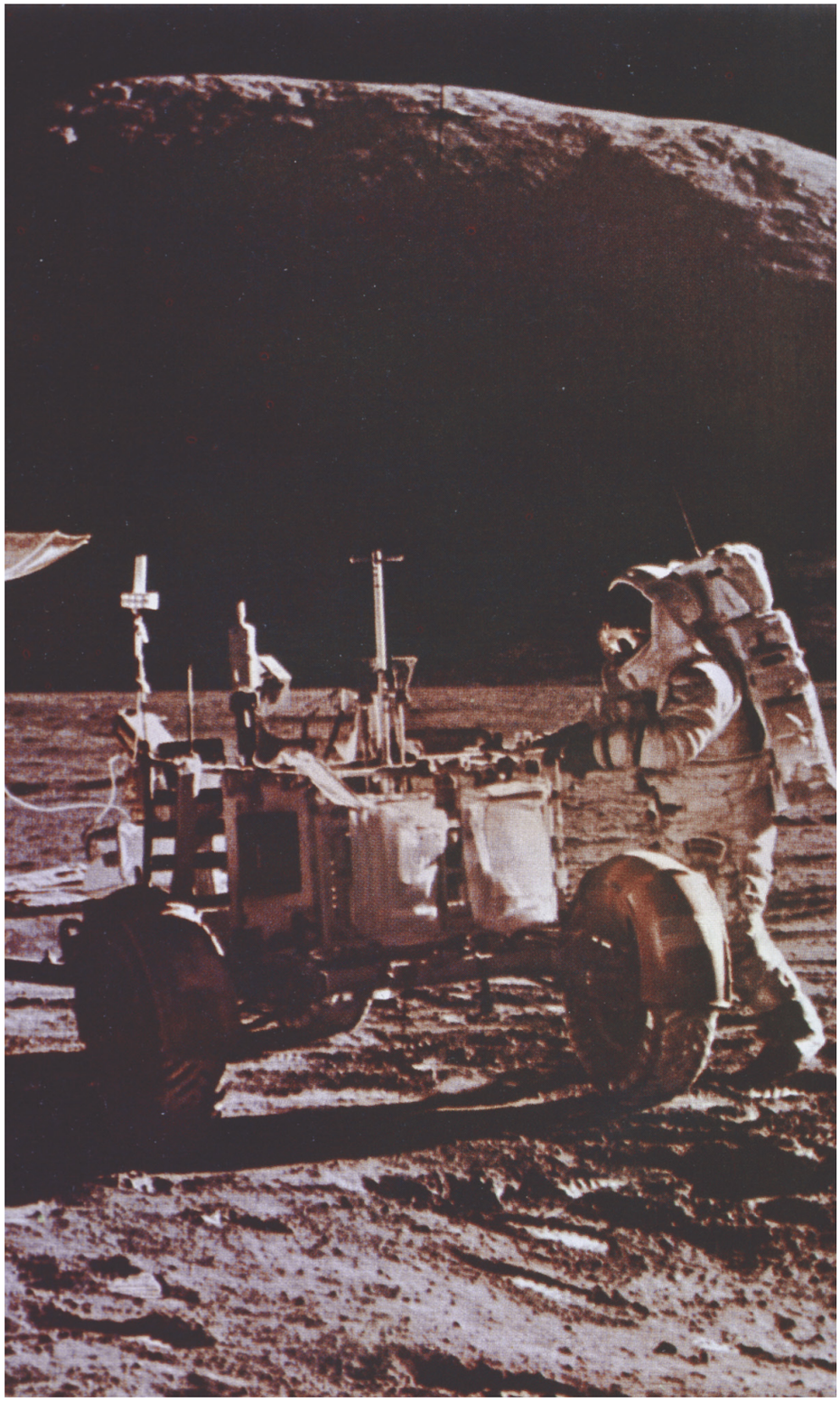


◀ The painting of Copernicus in the Cathedral of Torun. The poem—Copernicus's motto—is by Aeneas Sylvius Piccolomini, the later Pope Pius II:

> I crave not the Grace
> bestowed on Paul
> Nor the remission
> granted to Peter
> Only forgive me,
> I fervently pray
> As thou forgavest the
> crucified thieves.
>
> [Koestler 1990, p. 190]

Painting by Aleksander Lesser; photo: K. Jablonski.)

> But perhaps your Holiness will wonder less that I should have dared to bring these labors of mine into the light of day, after having taken on myself so much toil in working them out, as I have not hesitated even to commit to writing my thoughts on the motion of the Earth, than because you expect rather to hear from me how it entered my mind to dare against the received opinion of mathematicians, and almost against common sense, to imagine some motion of the Earth. I therefore wish your Holiness to know that I was impelled to think out another way of calculating the motions of the spheres of the universe by nothing else than the realization than the mathematicians themselves are inconsistent in investigating them. For first, the mathematicians are so uncertain of the motion of the Sun and Moon that they cannot represent it or even be consistent with the constant length of the seasonal year. Secondly, in establishing the motions both of the Sun and Moon and of the other five wandering stars they do not use the same principles or assumptions, or explanations of their apparent revolutions and motions. For some use only homocentric circles, others eccentric circles and epicycles, from which however the required consequences do not completely follow. For those who have relied on homocentric circles, although they have shown that diverse motions can be constructed from them, have not from that been able to establish anything certain, which would without doubt correspond with the phenomena. But those who have devised eccentric circles, although they seem to a great extent to have extracted from them the apparent motions, with numerical agreement, nevertheless have in the process admitted much which seems to contravene the first principle of regularity of motion. Also they have not been able to discover or deduce from them the chief thing, that is the form of the universe, and the clear symmetry of its parts. They are just like someone including in a picture hands, feet, head, and other limbs from different places, well painted indeed, but not modeled from the same body, and not in the least matching each other, so that a monster would be produced from them rather than a man. Thus in the process of their demonstrations, which they call their system, they are found either to have missed out something essential, or to have brought in something inappropriate and wholly irrelevant, which would not have happened to them if they had followed proper principles. For if the hypotheses which they assumed had not been fallacious, everything which follows from them could be indisputably verified. If what I am now saying is obscure, it will nevertheless become clearer in its proper place.
>
> Therefore on long pondering this uncertainty of mathematical traditions on the deduction of the motions of the mechanism of the universe, I began to feel disgusted that no more certain theory of the motions of the mechanisms of the universe, which has been established for us by the best and most systematic craftsman of all, was agreed by the philosophers who otherwise theorized so minutely with most careful attention to the details of this system. I therefore set myself the task of reading again all the books of all philosophers which were available to me, to search out whether anyone had ever believed that the motions of the spheres of the universe were other than was supposed by those who professed mathematics in the schools. And first I found in Cicero that Nicetus had supposed that the Earth moved. After that I discovered in Plutarch also that certain others held the same opinion. …
>
> I therefore took this opportunity and also began to consider the possibility that the Earth moved. Although it seemed an absurd opinion, nevertheless because I knew that others before me had been granted the liberty of imagining whatever circles they wished to represent the phenomena of the stars, I thought that I likewise would readily be allowed to test whether, by assuming some motion of the Earth's more dependable representations than theirs could be found for the revolutions of the heavenly spheres.
>
> —Copernicus, *On the Revolutions of the Heavenly Spheres* [pp. 24–26]



A long time had to pass before the visage of the goddess Selene, who was in love with the handsome Endymion, was recognized as a landscape of bare mountains. And another three centuries would pass before a human footstep would appear on its surface. In the picture, we see an astronaut of Apollo 14 (NASA, 1971). The Moon was already populated in the human imagination as expressed by the quotation that follows below, which is also noteworthy in that in 1672, when Molière wrote *The Learned Ladies*, Cartesian philosophy was a popular topic of conversation, meeting with both admiration and ridicule.

Trissotin: For Method, Aristotle suits me well.

Philaminte: But in abstractions, Plato does excel.

Armande: The thought of Epicurus is very keen.

Bélise: I rather like his atoms, but as between
A vacuum and a field of subtle matter
I find it easier to accept the latter.

Trissotin: On magnetism, Descartes supports my notions.

Armande: I love his falling worlds …

Philaminte: And whirling motions!

Armande: I can't wait for our conclaves. We shall proclaim
Discoveries, and they shall bring us fame.

Trissotin: Yes, to your keen minds Nature can but yield,
And let her rarest secrets be revealed.

Philaminte: I can already offer one such rarity:
I have seen men in the moon, with perfect clarity.

—Molière, *The Learned Ladies*, Act III, Scene ii

I hold [the Moon's] material to be very dense and solid, no less than the Earth's, of which a sufficiently clear proof to me is the unevenness of the major parts of its surface, evidenced by the many prominences and cavities revealed by the aid of the telescope. The prominences there are mainly very similar to our most rugged and steepest mountains, and some of them are seen to be drawn out in long tracts of hundreds of miles. Others are in more compact groups, and there are also many detached and solitary rocks, precipitous and craggy. But what occur most frequently there are certain ridges (I shall use this word because no more descriptive one occurs to me), somewhat raised, which surround and enclose plains of different sizes and various shapes, but for the most part circular. In the middle of many of these there is a mountain in sharp relief, and some few are filled with a rather dark substance similar to that of the large spots that are seen with the naked eye; these are the largest ones, and there are a very great number of smaller ones, almost all of them circular. …

What is clearly seen in the moon is that the darker parts are all plains, with few rocks and ridges in them, though there are some. The brighter remainder is all full of rocks, mountains, round ridges, and other shapes, and in particular there are great ranges of mountains around the spots. Quite apart from this, I consider the moon very different from the earth. Though I fancy to myself that its regions are not idle and dead, still I do not assert that life and motion exist there, and much less that plants, animals, or other things similar to ours are generated there. Even if they were, they would be extremely diverse, and far beyond all our imaginings. I am inclined to believe this because in the first place I think that the material of the lunar globe is not land and water, and this alone is enough to prevent generations and alterations similar to ours. But even supposing land and water on the moon … plants and animals similar to ours would not be produced there.

—Galileo, *Dialogue Concerning the Two Chief World Systems* [pp. 63, 99–100]



▲ The picture shows Queen Christina of Sweden being instructed by Descartes—a ritual that began at 5:00 am at the command of the queen.

Descartes had a very weak constitution, to the extent that the Jesuits at La Flèche had even allowed him to stay in bed until noon. It is possible that this change in his daily routine contributed to his early death.

Christina of Sweden (1626–1689), daughter of Gustavus Adolphus II, one of the main figures of the Thirty Years' War, wished to develop Sweden into a great cultural power as well. She contacted Descartes through the French ambassador in Stockholm, Chanut. Before his journey to Sweden, in letters to Chanut, Descartes answered a number of questions put to him by the queen. Perhaps the most interesting for us is where Descartes attempts to assuage the queen's fears that the idea of an infinite cosmos may be in conflict with religious faith. Descartes refers to the arguments of Nicholas of Cusa, and most remarkable of all is perhaps that he makes a distinction between the universe being infinite and it being unbounded. We will meet a distinction of a similar sort later in the General Theory of Relativity. (Painting by Dumesnil; Musée National du Château de Versailles.)

I found in [your letters] great proofs of your friendship and your tact. I was frightened, in reading the first pages, to learn that M. du Rier had spoken to the queen about me in one of my letters, and that she had asked to see it. Afterward, I was reassured when I came to the place where you write that she heard it read with some satisfaction. I do not know whether I was more filled with admiration that she had so easily understood things that the most learned consider very obscure, or with joy that my letters had not displeased her. But my admiration was redoubled when I saw the force and the weight of the [objections] that Her Majesty had noted concerning the size I attributed to the universe. …

In the first place, I remember that Cardinal Cusa and several other scholars have assumed the world to be infinite, without ever having been censured by the Church on this subject. On the contrary, it is thought to be honoring God to conceive of his works as very great. And my opinion is less difficult to accept than theirs, since I do not say that the world is infinite, but only indefinite. There is a remarkable enough difference between these two. For to say that something is infinite, we have to have some argument that makes us know it as such, and we can have that only in the case of God. But to say that it is indefinite, it is sufficient to have no argument by which the world can be proved to have limits. Thus it seems to me that we cannot prove, or even conceive, that there are limits in the matter of which the world is composed. … As for the prerogatives that religion accords to man, and which seem difficult to believe in if the extension of the universe is supposed to be indefinite, they deserve some explanation. For even though we can say that all created things are made for us insofar as we can derive some advantage from them, still I do not know that we are obliged to believe that man is the goal of the creation. Yet it is said that *all things are made for his (that is, for God's) sake*, that it is God alone who is the final as well as the efficient cause of the universe. And as far as creatures are concerned, insofar as they are reciprocally of use to one another, each one can claim this advantage, that all those of use to it are made for its sake. It is true that the six days of creation are described in Genesis in such a way that man seems to be its principal subject. But we can say that since this story of Genesis was written for man, it is chiefly things that concern him that the Holy Ghost wanted to specify and that he did not speak there of anything except as they related to man. …

And although I do not at all on that account infer that there are intelligent creatures in the stars or elsewhere, I do not see, either, that there is any argument by which it could be proved that there are none. But I always leave undecided questions of that sort, rather than denying or affirming anything about them.

—Descartes, letter to Chanut, June 6, 1647

▶ This portrait reveals Newton's incomparable intellect. In the picture no hint of human weakness can be seen, although Aldous Huxley could write, with justification:

> If we evolved a race of Isaac Newtons, that would not be progress. For the price Newton had to pay for being a supreme intellect was that he was incapable of friendship, love, fatherhood, and many other desirable things. As a man he was a failure; as a monster he was superb.

(Painting by G. Kneller, 1702; National Portrait Gallery, London.)



Here Lies Buried
Isaac Newton, Knight Bachelor,
Who with his nearly divine intellect
First demonstrated
The movements and Shapes of the Planets,
The paths of the Comets and the Tides of the Ocean,
By extending his torch of Learning;
He investigated what no one had even suspected before:
The differences among the Rays of Light,
And the properties of the colors arising therefrom.
A dedicated, wise, and faithful Interpreter
Of Nature, Antiquity, and Sacred Scripture,
He exalted the Majesty of God Almighty in his Philosophy,
And expressed the Simplicity of the Gospel in his habits.
Mortals congratulate themselves
To have produced such and so great
A GLORY OF THE HUMAN RACE.

BORN 25 DECEMBER A.D. 1642. DIED 20 MARCH. 1726

—Newton's epitaph in Westminster Abbey [translated from the Latin by Paul Franz]

Newton's credo …

> [God] is Eternal and Infinite, Omnipotent and Omniscient; that is, his duration reaches from Eternity to Eternity; his presence from Infinity to Infinity; he governs all things, and knows all things that are or can be done. He is not Eternity and Infinity, but Eternal and Infinite; he is not Duration and Space, but he endures and is present. He endures forever, and is every where present; and, by existing always and every where, he constitutes Duration and Space … In him are all things contained and moved; yet neither affects the other: God suffers nothing from the motion of bodies; bodies find no resistance from the omnipresence of God. 'Tis allowed by all that the supreme God exists necessarily; and by the same necessity he exists always and every where. Whence also he is all similar, all eye, all ear, all brain, all arm, all power to perceive, to understand, and to act; but in a manner not at all human, in a manner not at all corporeal, in a manner utterly unknown to us. As a blind man has no idea of colours, so have we no idea of the manner by which the all-wise God perceives and understands all things. He is utterly void of all body and bodily figure, and can therefore neither be seen, nor heard, not touched; nor ought he to be worshipped under the representation of any corporeal thing. We have ideas of his attributes, but what the real substance of anything is we know not. In bodies, we see only their figures and colours, we hear only the sounds, we touch only their outward surfaces, we smell only the smells, and taste the savours; but their inward substances are not to be known, either by our senses, or by any reflex act of our minds; much less then have we any idea of the substance of God.
>
> —Newton, "Scholium Generale," 1713

> In the beginning of the year 1665 I found the method of approximating Series and the Rule for reducing any dignity of any Binomial into such a series. The same year in May I found the method of tangents of Gregory and Slusius, and in November had the direct method of Fluxions, and the next year in January had the Theory of Colours, and in May following I had entrance into the inverse method of Fluxions. And the same year I began to think of gravity extending to the orb of the Moon, and having found out how to estimate the force with which [a] globe revolving within a sphere presses the surface of the sphere, from Kepler's Rule of the periodical times of the Planets being in a sesquialterate [3/2 power] proportion of their distances from the centers of their orbs I deduced that the forces which keep the Planets in their Orbs must [be] reciprocally as the squares of their distances from the centers about which they revolve: and thereby compared the force requisite to keep the Moon in her orb with the force of gravity at the surface of the earth, and found them to answer pretty nearly. All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded Mathematicks and Philosophy more than at any time since.
>
> —Isaac Newton, *A Catalogue of the Portsmouth Collection*

… and a critique by Leibniz.

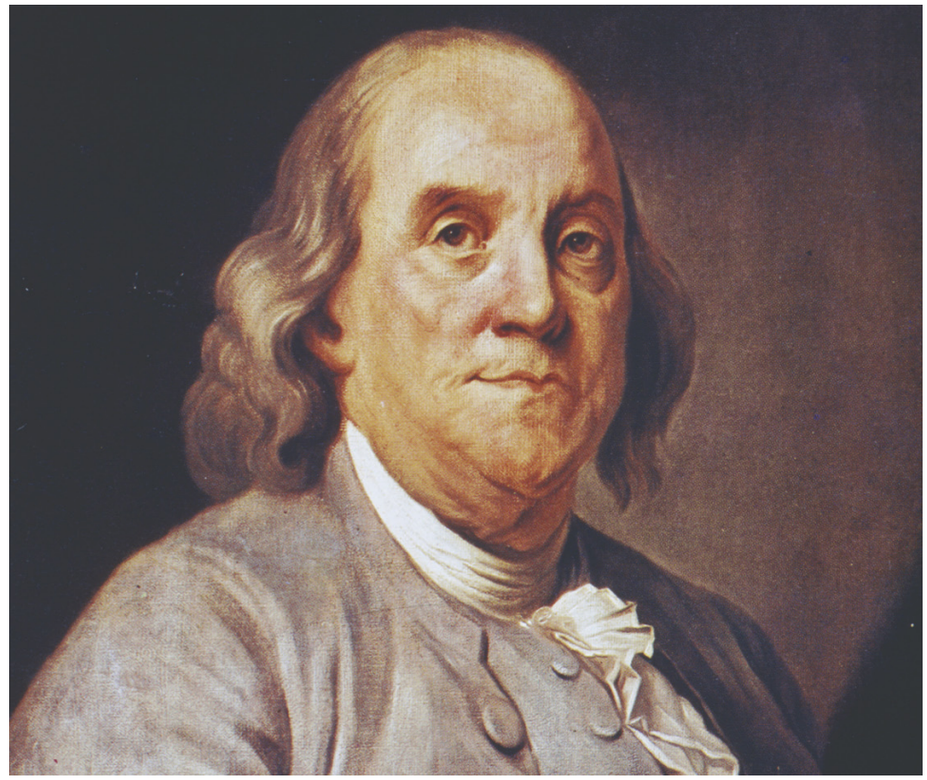
> Sir Isaac Newton and his followers also have a very odd opinion concerning the work of God. According to them, God Almighty needs to wind up his watch from time to time, otherwise it would cease to move. He had not, it seems, sufficient foresight to make it a perpetual motion. No, the machine of God's making is so imperfect, according to these gentlemen, that he is obliged to clean it now and then … and even to mend it, as a clockmaker mends his work, who must consequently be so much more the unskillful a workman as he is more often obliged to mend his work and to set it right. According to my opinion, the same force and vigor remains always in the world and only passes from one part of matter to another agreeably to the laws of nature and the beautiful pre-established order.
>
> —Leibniz, "From the Letters to Clarke"

Nearly every foreign traveler who visited the United States during these early years, carried away an impression sober if not sad. A thousand miles of desolate and dreary forest, broken here and there by settlements; along the sea-coast a few flourishing towns devoted to commerce; no arts, a provincial literature, a cancerous disease of negro slavery, and differences of political theory fortified within geographical lines,—what could be hoped for such a country except to repeat the story of violence and brutality which the world already knew by heart, until repetition for thousands of years had wearied and sickened mankind? Ages must probably pass before the interior could be thoroughly settled; even Jefferson, usually a sanguine man, talked of a thousand years with acquiescence, and in his first Inaugural Address, at a time when the Mississippi River formed the Western boundary, spoke of the country as having "room enough for our descendents to the hundredth and thousandth generation." No prudent person dared to act on the certainty that when settled, one government could comprehend the whole; and when the day of separation should arrive, and America should have her Prussia, Austria, and Italy, as she already had her England, France, and Spain, what else could follow but a return to the old conditions of local jealousies, wars, and corruption which had made a slaughter-house of Europe?

The mass of Americans were sanguine and self-confident, partly by temperament, but partly also by reason of ignorance; for they knew little of the difficulties which surrounded a complex society. The Duc de Liancourt, like many critics, was struck by this trait. Among other instances, he met with one in the person of a Pennsylvania miller, Thomas Lea, "a sound American patriot, persuading himself that nothing good is done, and that no one has any brains, except in America; that the wit, the imagination, the genius of Europe are already in decrepitude"; and the duke added: "This error is to be found in almost all Americans—legislators, administrators, as well as millers, and is less innocent there." In the year 1796 the House of Representatives debated whether to insert in the Reply to the President's Speech a passing remark that the nation was the "freest and most enlightened in the world"—a nation as yet in swaddling-clothes, which had neither literature, arts, sciences, nor history; nor even enough nationality to be sure that it was a nation. The moment was particularly ill-chosen for such a claim, because Europe was on the verge of an outburst of genius. Goethe and Schiller, Mozart and Haydn, Kant and Fichte, Cavendish and Herschel were making way for Walter Scott, Wordsworth, and Shelley, Heine and Balzac, Beethoven and Hegel, Oersted and Cuvier, great physicists, biologists, geologists, chemists, mathematicians, metaphysicians, and historians by the score. Turner was painting his earliest landscapes, and Watt completing his latest steam-engine; Napoleon was taking command of the French armies, and Nelson of the English fleets; investigators, reformers, scholars, and philosophers swarmed, and the influence of enlightenment, even amid universal war, was working with an energy such as the world had never before conceived. The idea that Europe was in her decrepitude proved only ignorance and want of enlightenment, if not of freedom, on the part of Americans, who could only excuse their error by pleading that notwithstanding these objections, in matters which for the moment most concerned themselves Europe was a full century behind America. If they were right in thinking that the next necessity of human progress was to lift the average man upon an intellectual and social level with the most favored, they stood at least three generations nearer than Europe to their common goal.

—Henry Adams, "American Ideals" [pp. 107–108].

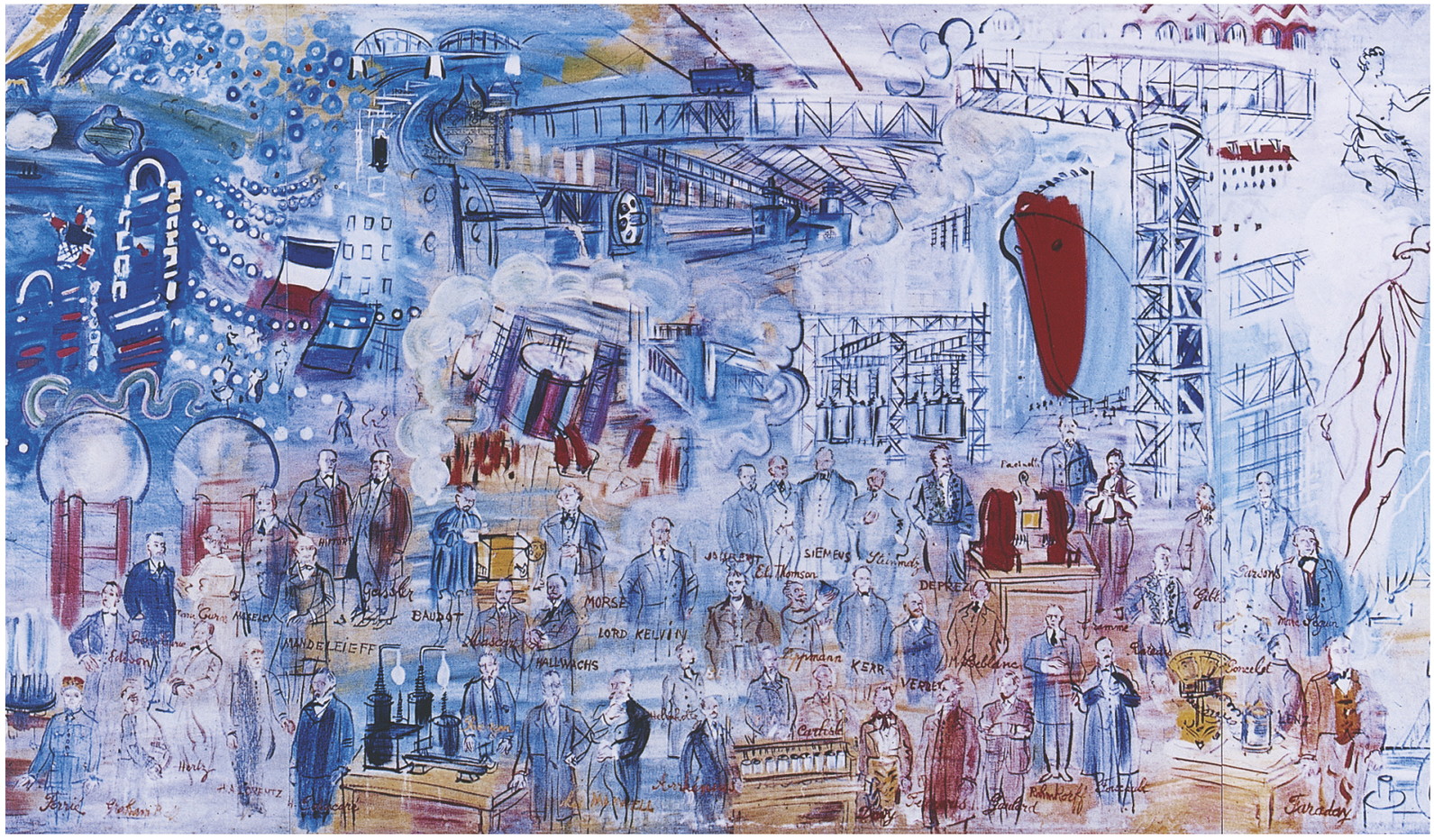




▲ Madame de Pompadour and Benjamin Franklin, contemporaries with similar intellectual backgrounds, but vastly different in their social and ethical attitudes.

Madame de Pompadour (1721–1764): Mistress of Louis XV, proud of her education, she considered it important that volumes of the *Encyclopédie* appear next to her in the painting. In her we find embodied a type of intellectual at the time of the French Enlightenment: broadly educated, ostentatious, clever in conversation, decadent in social relations, and doomed to the destruction they brought upon themselves. (Painting by Quentin de la Tour in the Louvre.)

Benjamin Franklin (1706–1790): "He snatched the lightning from the skies and the scepter from the tyrants" (*Eripuit coelo fulmen sceptrumque tyrannis*). Puritan with the highest moral qualities and social conscience. The first American scientist to achieve world renown. (National Gallery, London.)



La Fée Electricité (Raoul Dufy, Musée d'Art Moderne de la Ville de Paris).

This monumental painting (60 m by 10 m, 180 ft by 30 ft) adorned the wall of the Pavilion of Electricity at the 1937 World's Fair in Paris. The artist was striving to express the relationship between basic science and its applications from Thales to Lorentz and Edison. The segment shown here comprises the second half of the nineteenth century and first decades of the twentieth.

It is always instructive to compare how creative spirits are judged by people of different nationalities. The greatest names, maybe with a few exceptions, would be expected to be included. Franklin and Gibbs, the greatest American scientists of the eighteenth and nineteenth centuries, indeed appear in this mural. The great Germans Helmholtz and Gauss, and famous Englishmen such as Faraday, Joule, and Maxwell, as well as the Frenchmen Carnot and Laplace are all there. It is worth noticing—and it is probably no accident—that Einstein is absent, even though Lorentz and Poincaré are shown.

Some noteworthy names in the picture that are not mentioned elsewhere in this book: John Kerr (1824–1907) discovered the phenomenon that some isotropic gases, liquids, and solids become doubly refracting in the presence of an electric field, making it possible to convert changes in voltage to changes in light (Kerr effect). The observation by Wilhelm Hallwachs (1859–1922) that negatively charged metal electrodes lose their charge when illuminated with ultraviolet light served as the starting point for further investigations into the theoretically and practically significant photoelectric effect. Important names in the field of electrochemistry: A. Carlisle (1768–1842), gave the theoretical basis for breaking chemical bonds using electric current. The Nobel Laureate Svante Arrhenius (1859–1927) worked on the dissociation theory of solutions; Gaston Planté (1834–1889) constructed the first lead acid battery. In the field of electrical engineering, we find the following names: Heinrich Daniel Ruhmkorff (1803–1877), discoverer of the inductance coil that bears his name; Marcel Deprez (1843–1918), known primarily for the galvanometer constructed with d'Arsonval, but also worked successfully in the area of electrical energy transmission; Elihu Thomson (1853–1937) is one of the founders of the electrical products industry in the United States and holder of many patents (alternating current motor, watt-hour counter, x-ray tube, etc.); Galileo Ferraris (1847–1897), pioneer in the theory of alternating current, particularly multiphase alternating current; Jules-François Joubert, for a time Pasteur's colleague, wrote much-used books on electrical engineering; Lucien Gaulard constructed a new type of transformer that foreshadowed more economically viable versions. Samuel Morse (1791–1872), an American painter and physicist, became known for his telegraph and the Morse code. Jean-Maurice-Emile Baudot is the inventor of the telex machine.

One also sees the name Becquerel in the picture. The artist was likely referring to the renowned physicist and electrical engineer Edmond Becquerel (1820–1891), the father of one of the protagonists in this book, Henri Becquerel, but he could also have been thinking of the grandfather, Antoine Becquerel (1788–1878), who worked in the fields of electrochemistry and telegraphy.

Too many Frenchmen, one might say, but that is understandable at a world's fair in Paris. In Budapest, one certainly would have found in such a painting the names Ányos Jedlik (inventor of the dynamo with autoexcitation) and the triumvirate Déri–Bláthy–Zipernowsky (builders of economically viable transformers); in Moscow, one would be sure to find, among others, the names Pavel Nikolaevich Yablochkov (1847–1894), famous for his alternating-current arc lamp; Mikhail Osipovich Dolivo-Dobrovolsky (1862–1919), who further developed the three-phase motor, generator, and transformer. Boris Semenovich (Moritz Hermann) Jacobi (1801–1874) with his electric motor, and Aleksandr Stepanovich Popov (1859–1905), one of the pioneers of radio transmission.

▶ "Technology is just applied science" is what one frequently hears as the definition of technology. Here one sees a schematic drawing of a 1.5 GW generator for an atomic power plant built by the company Brown, Boweri and Cie. In a physics textbook such a generator would be pictured as a simple wire loop rotating in a homogeneous magnetic field. Theory and practical realization look very different.

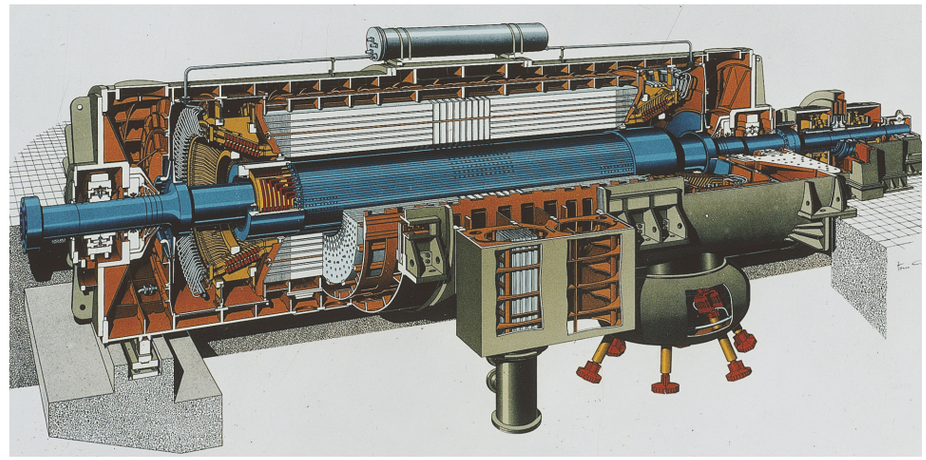
Thus the counterclaim that technology is no more applied science than music is applied acoustics is also true to some extent, especially if we consider that the engineers, while planning a technological object, must go beyond physics and think about aspects such as effectiveness, reliability, safety, environmental friendliness, competitiveness, or even aesthetics or fashion.

If we look into history, we can observe that technology existed long before there was any trace of physics. Technology was ahead of physics until the Industrial Revolution: until then people built machines whose functional principles and the physical laws underlying the principles had been far from clear.

With a bit of exaggeration, one could even claim that physics was only able to start along its own path of developmental when it consciously stopped speculating about immediate technological questions. We need only to think of the complex technical problems that Leonardo struggled with (aircraft construction, turbulent flow, motion with friction, etc.). Galileo, in contrast, succeeded in arriving at a precise foundation for mechanics because he dealt with phenomena that had no immediate practical significance (frictionless motion on an inclined plane, pendulum motion). The first technological applications of the newly discovered laws of physics were, in fact, the construction of new devices for further measurements and experiments. Newton and his successors, such as Euler, knew almost the entire system of theorems of classical mechanics and also the mathematical apparatus necessary for their treatment. However, the solution of concrete technological problems presented overwhelming difficulties even for them.

Frederick the Great was justifiably able to write to Voltaire in the year 1778 [Klemm 1959]:

> The English have built ships with the most advantageous section in Newton's opinion, but their admirals have assured me that these ships did not sail nearly so well as those built according to the rules of experience. I wanted to make a fountain in my garden. Euler calculated the output of the wheels which should have raised the water into a reservoir, from which it was to flow again through canals and again mount on high in the fountains at Sans Souci. My lifting-gear was carried out according to mathematical calculations but could not raise a drop of water to fifty paces from the reservoir. Vanity of vanities! Vanity of mathematics!



The inventors of the steam engine were certainly scientifically oriented men, even though they often did not have a formal education. Nevertheless, William Rankine, the well-known thermodynamicist, had good reason to say:

> The railroad, complete and entire, as left to us by Stephenson, is a product of necessity and the spirit of its age. The untutored talent, the healthy, practical thought of the people, the callused hand of the worker alone created it; book learning played no part in it. No formula was developed in the greatest technological creation of our time; no equation was solved.
>
> [Brinkmann 1946, p. 79]

The necessity for scientific underpinnings in technological processes appears with urgency in the practical application of electromagnetic phenomena. The impetus in this direction came mostly from the physicists. Work with electricity requires a great capacity for abstraction, to wit, a more exact, mathematically oriented way of thinking. And yet Edison, perhaps the greatest inventive mind of all time, had great difficulties with mathematics: At the age of fourteen he tried to read Newton's *Principia*, and for the rest of his life was terrified of mathematics. He was flattered by the legends that dub him the "wizard of Menlo Park," but, in fact, he organized an industrial research institute operating on a scientific basis with highly qualified employees. His favorite book that remained his bible all his life was Faraday's *Experimental Researches in Electricity*.

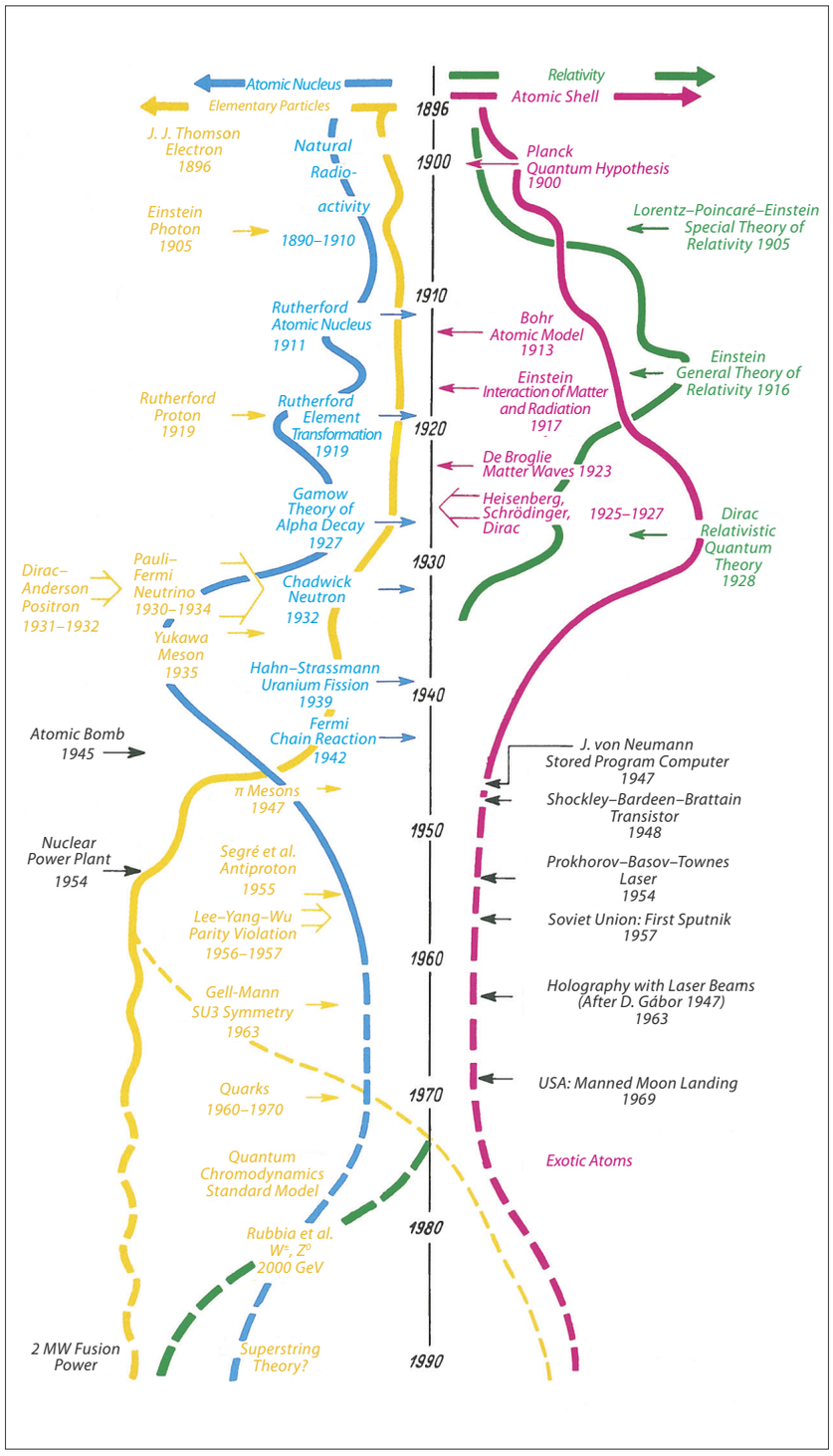
The methodology of engineering was fully developed through the turn of the century. We quote Whitehead:

> The great invention of the nineteenth century was the method of invention.

Engineering was fully recognized as an independent scientific discipline only as the first technical institutes were founded, on the same level as the universities. The importance of technical knowledge for science, and scientific knowledge for technology, was emphasized by Francis Bacon, but the spreading of these ideas was the contribution of the encyclopedists, primarily Diderot. We may consider the École des Ponts et Chaussées (1747) to be the first technical institute; Gaspard Monge, the famous mathematician, organized the École Polytechnique, one of the most important institutions of this type, somewhat later. It is not as well known or appreciated that the mining schools founded in 1763, in Banská Štiavnica (at the time Selmecbánya in Hungary) and in 1765, in Freiberg, served as models for the École Polytechnique. In England, the teaching of technological subjects at universities only started around 1840. Here we see confirmation of the opinion that at the beginning of the Industrial Revolution there was no direct connection between theoretical education and technological application. France had the best technical university with the best students, and yet England—where no such institution existed—was leading in technological invention. The leading role there was played by the newly formed engineering societies, such as the Institute of Civil Engineers (1820), the Institute of Mechanical Engineers (1846, Stephenson was the first president), and later (1889) the Institute of Electrical Engineers.

Germany took over the leading role in the optical, chemical, and electrical industries in the second half of the nineteenth century because these fields required a high level of theoretical knowledge. Education was available at numerous technical institutes (later technical universities), and the development and application proceeded in state research institutes or in research centers of private firms.

Today there is little significant difference between basic and applied research, at least with respect to the method (the interplay of theory and experiment,) the equipment and the expense. The difference is much more in the goals and therefore also the motivation of researchers: economically viable results (perhaps in the form of patents) on the one hand, scientifically significant results in the form of publications on the other. These two aspects are by no means antagonistic, in the sense that they are not mutually exclusive.



The shifts in the focus of physics research in the twentieth century (the further a line is from the center, the greater the emphasis on the corresponding area).

In the first two decades of this century it was the physics of the atomic shell, in the 1930s and 1940s nuclear physics, and from then on the physics of elementary particles that stood at the center of interest and was pursued as the path into the unknown.

This division corresponds to the following energy ranges: 0 to 10^4 eV: physics of the atomic shell; 10^4 to 10^7 eV: nuclear physics; 10^7 eV and above: particle physics. Yet the more recent results suggest that the processes taking place in the different energy ranges are intimately connected, and to explain or exploit them, the laws relevant to all the ranges have to be applied in unison. Thus the theory of nuclear forces had to be completely reformulated to fit the quark model. Another example: using exotic atoms—in which an electron is replaced by a negatively charged, relatively long-lived elementary particle (such as a muon or kaon)—the experimental investigation of some theoretically important questions about the atomic shell, such as the Lamb shift, can be carried out with great precision. To explain the big bang and the Planck era that immediately followed, the general theory of relativity and particle physics have to be combined in some way. The curves toward the bottom of the diagram indicate this overlap in research themes.

In this diagram, there is no trace of classical physics to be seen. We are reminded that classical mechanics lives on in the theory of relativity and quantum mechanics as a special case with well-determined limits of validity and remains useful in practical applications.

Since the 1960s, however, the equations of classical mechanics have moved to the foreground again and revealed themselves in another light. It turned out that nonlinear systems are extremely sensitive to initial conditions—small changes result is drastic changes in their path; even very simple systems with as few as two or three degrees of freedom can exhibit behavior so chaotic that even approximate prediction is impossible. The new discipline of chaos theory deals with such deterministic but unpredictable phenomena. It is interesting that the first scientist investigating this area was a meteorologist, Edward Lorenz (1917–2008), who was working on the problem of heat conduction in the atmosphere in 1963. The chaotic—unpredictable—behavior of a complex system is by no means surprising. What was new and surprising was the realization that even very simple systems can exhibit "chaotic" behavior. The significance of investigating such phenomena is that they continually appear in current research, such as in particle motion in very large accelerators, containment of plasma for the generation of fusion energy, or the motion of asteroids or moons in nonhomogeneous gravitational fields.

New geometric structures of surprising complexity and beauty have appeared that have captured the imagination of many physicists and mathematicians and encouraged them to seek out new areas of application. The state of a nonlinear dissipative (that is, energy-consuming) system may converge to certain geometric structures. It can be a point (attractor of dimension 0) or a line (attractor of dimension 1), but it can also possess a very involved structure that can be characterized only by a fractional dimension; such shapes are called fractals (strange attractor).

Although we cannot agree fully with the opinion of Joseph Ford about the fundamental importance of such investigations, we quote him for his concise formulation:

> The central point here is that the term "deterministic randomness" is not self-contradictory … Determinate because of existence and uniqueness for classical orbits, and random because chaotic orbits cannot be computed by any algorithmic rule appreciably simpler or shorter than the copy program. … Finally, it is enlightening, though somewhat audacious, to view chaos as the beginning of the scientific revolution in this century … Gregory Chaitin has proven that even simple questions can have answers so complicated that they contain more information than man's entire logical system, and nonlinear dynamics has proven the existence and uniqueness of chaotic orbits which no human can ever exhibit. … Relativity eliminated the Newtonian illusion of absolute space and time; quantum theory eliminated the Newtonian dream of a controllable measurement process; and chaos eliminates the Laplacian fantasy of deterministic predictability. But of the three, chaos deals perhaps the most devastating blow because chaos exposes the clay feet of Newtonian dynamics as it stands in that final nonrelativistic, nonquantal domain which serves as its last refuge.
>
> —"What Is Chaos, that We Should Be Mindful of It?" [pp. 350, 353–354]

Every energy has an associated mass and every mass, an associated energy, expressed by the "conversion formula" $E = mc^2$. This is the content of the mass–energy equivalence theorem. Therefore, the law of conservation of energy is identical to the law of conservation of mass. This equivalence principle is one of the most basic theoretical and practical laws of the twentieth century and at the same time understandable—also in a qualitative form—by the general public. The discovery of the relationship $E = mc^2$ is today generally attributed exclusively to EINSTEIN, with reference to the journal article reproduced here in its entirety (first three pages in the illustration).

PLANCK, in a footnote in his 1907 paper where he gave a more general derivation for the equivalence law, made the following remark:

> A. EINSTEIN drew essentially the same conclusion (*Ann. d. Phys.* 18 [1905], 639) from applying the relativity principle to a special radiation process, however under the assumption, permissible to a first approximation, that the total energy of a moving body is composed additively of its kinetic energy and the energy for it in a frame of reference stationary with respect to it. There we also find an idea for a possible verification of the theory through observations of radium salts.

That it was not only the relationship between mass and energy that was up in the air, but also the possibility of an experimental verification, is confirmed by the page, also shown here (fourth page in the illustration), from a book by SODDY published in 1904(!) in German(!). The following lines are underlined:

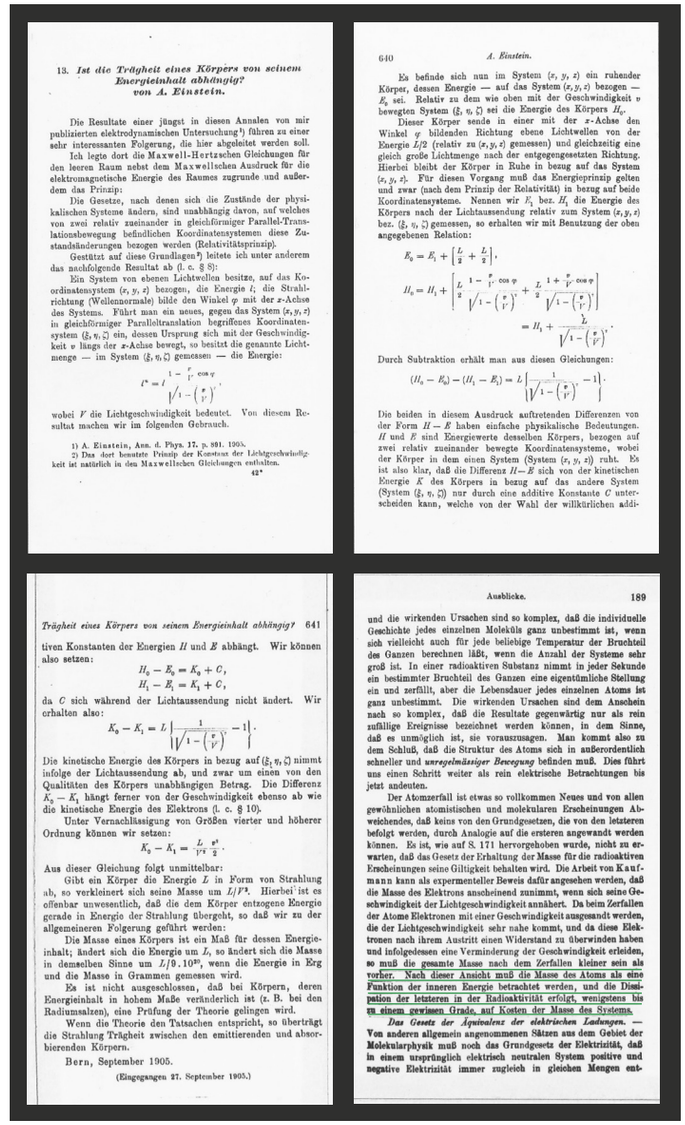
> According to this view, the mass of the atom must be seen as a function of the internal energy … the dissipation of energy … occurs … at the expense of the mass of the system.

We quote as well a present-day writer. JAMMER writes the following [1961, p. 177]:

> It is a curious incident in the history of scientific thought that EINSTEIN's own derivation of the formula $E = mc^2$, as published in his article in the *Annalen der Physik*, was basically fallacious. In fact, what for the layman is known as "the most famous mathematical formula ever projected" in science was but the result of a *petitio principi*, the conclusion of begging the question. This statement, of course, does not diminish in the least the importance of EINSTEIN's contribution to the problem, since the mass-energy relation is a necessary consequence of the theory of relativity and can be deduced from the fundamental assumptions of the theory by various methods, and not only by the method employed by EINSTEIN in his original publication. The logical illegitimacy of EINSTEIN's derivation has been shown by IVES.



▲ The formula $E = mc^2$ for the equivalence of mass and energy displayed as an array of tulips in the park of the Budapest University of Technology and Economics. (Photo: T. Szigeti.)



13. *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?* *von A. Einstein.*

Die Resultate einer jüngst in diesen Annalen von mir publizierten elektrodynamischen Untersuchung[1]) führen zu einer sehr interessanten Folgerung, die hier abgeleitet werden soll.

Ich legte dort die Maxwell-Hertzschen Gleichungen für den leeren Raum nebst dem Maxwellschen Ausdruck für die elektromagnetische Energie des Raumes zugrunde und außerdem das Prinzip:

Die Gesetze, nach denen sich die Zustände der physikalischen Systeme ändern, sind unabhängig davon, auf welches von zwei relativ zueinander in gleichförmiger Parallel-Translationsbewegung befindlichen Koordinatensystemen diese Zustandsänderungen bezogen werden (Relativitätsprinzip).

Gestützt auf diese Grundlagen[2]) leitete ich unter anderem das nachfolgende Resultat ab (l. c. § 8):

Ein System von ebenen Lichtwellen besitze, auf das Koordinatensystem (x, y, z) bezogen, die Energie l; die Strahlrichtung (Wellennormale) bilde den Winkel φ mit der x-Achse des Systems. Führt man ein neues, gegen das System (x, y, z) in gleichförmiger Paralleltranslation begriffenes Koordinatensystem (ξ, η, ζ) ein, dessen Ursprung sich mit der Geschwindigkeit v längs der x-Achse bewegt, so besitzt die genannte Lichtmenge — im System (ξ, η, ζ) gemessen — die Energie:

$$l' = l \frac{1 - \frac{v}{V}\cos\varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}},$$

wobei V die Lichtgeschwindigkeit bedeutet. Von diesem Resultat machen wir im folgenden Gebrauch.

1) A. Einstein, Ann. d. Phys. 17. p. 891. 1905.
2) Das dort benutzte Prinzip der Konstanz der Lichtgeschwindigkeit ist natürlich in den Maxwellschen Gleichungen enthalten.

42*

640 *A. Einstein.*

Es befinde sich nun im System (x, y, z) ein ruhender Körper, dessen Energie — auf das System (x, y, z) bezogen — E_0 sei. Relativ zu dem wie oben mit der Geschwindigkeit v bewegten System (ξ, η, ζ) sei die Energie des Körpers H_0.

Dieser Körper sende in einer mit der x-Achse den Winkel φ bildenden Richtung ebene Lichtwellen von der Energie $L/2$ (relativ zu (x, y, z) gemessen) und gleichzeitig eine gleich große Lichtmenge nach der entgegengesetzten Richtung. Hierbei bleibt der Körper in Ruhe in bezug auf das System (x, y, z). Für diesen Vorgang muß das Energieprinzip gelten und zwar (nach dem Prinzip der Relativität) in bezug auf beide Koordinatensysteme. Nennen wir E_1 bez. H_1 die Energie des Körpers nach der Lichtaussendung relativ zum System (x, y, z) bez. (ξ, η, ζ) gemessen, so erhalten wir mit Benutzung der oben angegebenen Relation:

$$E_0 = E_1 + \left[\frac{L}{2} + \frac{L}{2}\right],$$

$$H_0 = H_1 + \left[\frac{L}{2}\frac{1 - \frac{v}{V}\cos\varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} + \frac{L}{2}\frac{1 + \frac{v}{V}\cos\varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}}\right] = H_1 + \frac{L}{\sqrt{1 - \left(\frac{v}{V}\right)^2}}.$$

Durch Subtraktion erhält man aus diesen Gleichungen:

$$(H_0 - E_0) - (H_1 - E_1) = L\left[\frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} - 1\right].$$

Die beiden in diesem Ausdruck auftretenden Differenzen von der Form $H - E$ haben einfache physikalische Bedeutungen. H und E sind Energiewerte desselben Körpers, bezogen auf zwei relativ zueinander bewegte Koordinatensysteme, wobei der Körper in dem einen System (System (x, y, z)) ruht. Es ist also klar, daß die Differenz $H - E$ sich von der kinetischen Energie K des Körpers in bezug auf das andere System (System (ξ, η, ζ)) nur durch eine additive Konstante C unterscheiden kann, welche von der Wahl der willkürlichen addi-

Trägheit eines Körpers von seinem Energieinhalt abhängig? 641

tiven Konstanten der Energien H und E abhängt. Wir können also setzen:

$$H_0 - E_0 = K_0 + C,$$
$$H_1 - E_1 = K_1 + C,$$

da C sich während der Lichtaussendung nicht ändert. Wir erhalten also:

$$K_0 - K_1 = L\left[\frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} - 1\right].$$

Die kinetische Energie des Körpers in bezug auf (ξ, η, ζ) nimmt infolge der Lichtaussendung ab, und zwar um einen von den Qualitäten des Körpers unabhängigen Betrag. Die Differenz $K_0 - K_1$ hängt ferner von der Geschwindigkeit ebenso ab wie die kinetische Energie des Elektrons (l. c. § 10).

Unter Vernachlässigung von Größen vierter und höherer Ordnung können wir setzen:

$$K_0 - K_1 = \frac{L}{V^2}\frac{v^2}{2}.$$

Aus dieser Gleichung folgt unmittelbar:

Gibt ein Körper die Energie L in Form von Strahlung ab, so verkleinert sich seine Masse um L/V^2. Hierbei ist es offenbar unwesentlich, daß die dem Körper entzogene Energie gerade in Energie der Strahlung übergeht, so daß wir zu der allgemeineren Folgerung geführt werden:

Die Masse eines Körpers ist ein Maß für dessen Energieinhalt; ändert sich die Energie um L, so ändert sich die Masse in demselben Sinne um $L/9 \cdot 10^{20}$, wenn die Energie in Erg und die Masse in Grammen gemessen wird.

Es ist nicht ausgeschlossen, daß bei Körpern, deren Energieinhalt in hohem Maße veränderlich ist (z. B. bei den Radiumsalzen), eine Prüfung der Theorie gelingen wird.

Wenn die Theorie den Tatsachen entspricht, so überträgt die Strahlung Trägheit zwischen den emittierenden und absorbierenden Körpern.

Bern, September 1905.

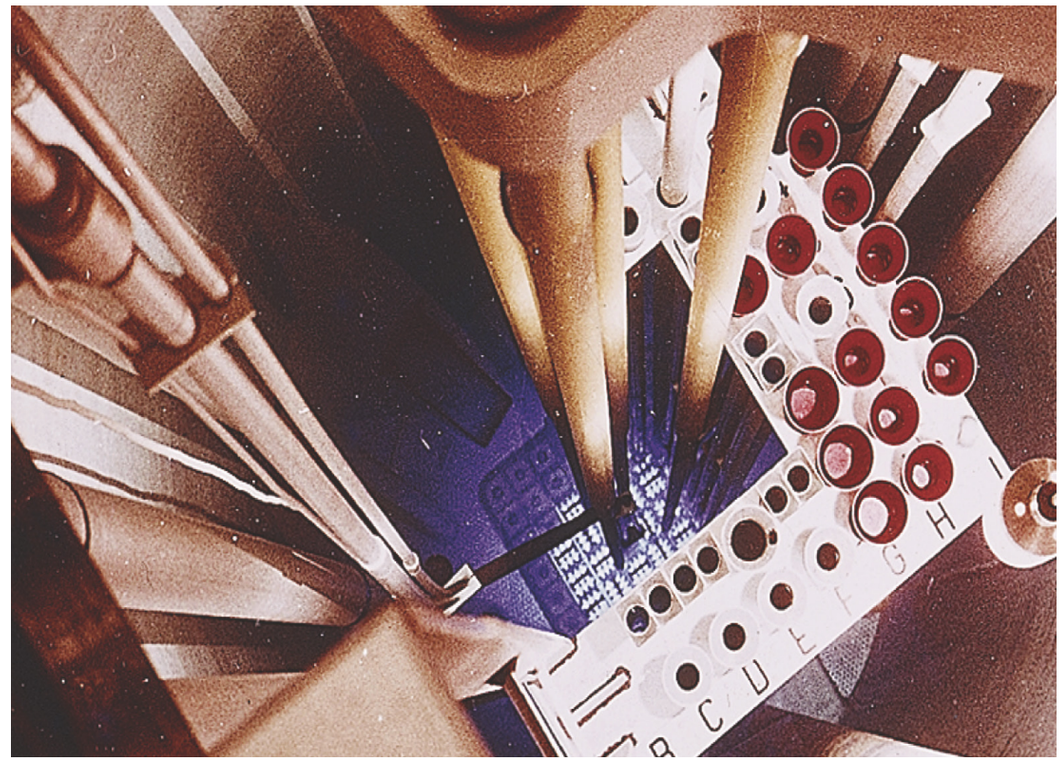
(Eingegangen 27. September 1905.)

Ausblicke. 189

und die wirkenden Ursachen sind so komplex, daß die individuelle Geschichte jedes einzelnen Moleküls ganz unbestimmt ist, wenn sich vielleicht auch für jede beliebige Temperatur der Bruchteil des Ganzen berechnen läßt, wenn die Anzahl der Systeme sehr groß ist. In einer radioaktiven Substanz nimmt in jeder Sekunde ein bestimmter Bruchteil des Ganzen eine eigentümliche Stellung ein und zerfällt, aber die Lebensdauer jedes einzelnen Atoms ist ganz unbestimmt. Die wirkenden Ursachen sind dem Anschein nach so komplex, daß die Resultate gegenwärtig nur als rein zufällige Ereignisse bezeichnet werden können, in dem Sinne, daß es unmöglich ist, sie vorauszusagen. Man kommt also zu dem Schluß, daß die Struktur des Atoms sich in außerordentlich schneller und *unregelmäßiger Bewegung* befinden muß. Dies führt uns einen Schritt weiter als rein elektrische Betrachtungen bis jetzt andeuten.

Der Atomzerfall ist etwas so vollkommen Neues und von allen gewöhnlichen atomistischen und molekularen Erscheinungen Abweichendes, daß keins von den Grundgesetzen, die von den letzteren befolgt werden, durch Analogie auf die ersteren angewandt werden können. Es ist, wie auf S. 171 hervorgehoben wurde, nicht zu erwarten, daß das Gesetz der Erhaltung der Masse für die radioaktiven Erscheinungen seine Giltigkeit behalten wird. Die Arbeit von Kaufmann kann als experimenteller Beweis dafür angesehen werden, daß die Masse des Elektrons anscheinend zunimmt, wenn sich seine Geschwindigkeit der Lichtgeschwindigkeit annähert. Da beim Zerfallen der Atome Elektronen mit einer Geschwindigkeit ausgesandt werden, die der Lichtgeschwindigkeit sehr nahe kommt, und da diese Elektronen nach ihrem Austritt einen Widerstand zu überwinden haben und infolgedessen eine Verminderung der Geschwindigkeit erleiden, so muß die gesamte Masse nach dem Zerfallen kleiner sein als vorher. Nach dieser Ansicht muß die Masse des Atoms als eine Funktion der inneren Energie betrachtet werden, und die Dissipation der letzteren in der Radioaktivität erfolgt, wenigstens bis zu einem gewissen Grade, auf Kosten der Masse des Systems.

Das Gesetz der Äquivalenz der elektrischen Ladungen. — Von anderen allgemein angenommenen Sätzen aus dem Gebiet der Molekularphysik muß noch das Grundgesetz der Elektrizität, daß in einem ursprünglich elektrisch neutralen System positive und negative Elektrizität immer zugleich in gleichen Mengen ent-



A modern university is distinguished from its medieval predecessor in that laboratory work is now much more prevalent. Our picture shows the inside of the teaching reactor at the Budapest University of Technology and Economics. The active zone of the reactor—where the fission and therefore the production of energy takes place in the fuel rods containing uranium enriched with the isotope U-235—is marked by the bluish Cherenkov light of the electrons released by the activity of the fission products. Ordinary water is used to moderate the neutrons. Also visible in the picture are the control rods and the equipment for exposing various materials to the radiation in the reactor. (Photo: Tamás Fényes, MTI.)

Among the manifold aspects of university education, the following quotation highlights one that is the closest to the theme of this book:

> The history of the sciences shows that scientists who have students of their own and work together with them are the most prolific of all. Many major scientists can be mentioned to prove this point. Thus, Mendeleev discovered the periodic system of elements while looking for a way to explain the properties of chemical elements in a manner best remembered by his students. The young Lobachevski, while teaching geometry in a school for adult secondary school students, found no satisfactory way of explaining the obviousness of the postulate stating that parallel lines never intersect; thus he discovered the non-Euclidean geometry. Stokes, while formulation problems for mathematics students, asked them in one of these problems to prove that the surface integral is directly related to the magnitude of flux or current passing through the contour. This is now known as the Stokes theorem, though in fact he never published its proof, leaving this to the students themselves. … The list of such examples can be continued to our own time. Thus, Schrödinger found his well-known equations while explaining de Broglie's work to a group of postgraduate students at the University of Zurich.
>
> —P. L. Kapitsa, "Some Principles of the Creative Upbringing and Education of Today's Youth" [p. 213]

This quotation from Kapitsa requires some context, however: Many great scientists (perhaps the majority) have considered teaching duties to be a burden. This was true for Newton, Gauss, and Einstein, to name only some of the most famous. It is also doubtful whether there is a correlation between pedagogical ability and creative genius; only a few of the greatest, such as Rutherford or Bohr, ever founded a scientific school. At the same time, we should not forget that the genius working alone is also an educator—at the very highest level. Bose sent his work to Einstein, with the following introduction:

> Though a complete stranger to you, I do not hesitate in making such a request. Because we are all your pupils though profiting only from your teachings through your writings.

The most pressing task for technology today is to satisfy mankind's energy needs. Disregarding the power of tides, which comes from gravitational energy, all energy available to us is, in the final analysis, the result of nuclear transformations. We are presently living off the energy of solar nuclear fusion accumulated in the form of fossil fuels. These supplies are estimated at 4×10^{23} J, and if the increase of the rate of consumption holds, it should last another 100 years or so (in 1970, energy use was at 2×10^{20} J, doubling every 20 years). Fossil chemical energy is transformed in power plants via heat and mechanical energy into the most practical form: electrical energy.

Hopes for the future:

1. Use of the energy of the sun continuously arriving to Earth (3×10^{24} J/year), either through its thermal energy or directly converting it to electricity using semiconductor solar cells, or though wind or hydro power.
2. Use of nuclear energy on Earth using controlled nuclear fission or fusion reactions: The fuels for nuclear-fission power plants are uranium and thorium, whose supplies (2×10^{24} J) are also insufficient in the very long term. The fuel for fusion reactors, in contrast, would be deuterium, which is available in the oceans in sufficient quantities to last for thousands of years.

For a few decades of the twentieth century the focus of electrical energy generation moved in the direction of atomic power plants. In the mid 1980s, about 15% of the world's electricity was obtained from nuclear energy. In some countries, such as France, the proportion has reached 80%. There have also been frequent objections against the use of atomic energy and the demand for the decommissioning of older reactors—especially after serious accidents such as at Chernobyl (1986). Even in normal operation, the burned-out fuel elements are very radioactive. Their safe storage and long-term disposal is as yet an unsolved problem.

The experts try to regain public support for nuclear energy by continually improving on the security of the systems and strengthening national and international oversight. It is also the case that all energy sources have drawbacks: fossil plants emit environmentally harmful gases, accidents are frequent in fossil fuel extraction, and even the cleanest of power sources, hydroelectric power, can be the cause of major environmental and societal disruption.

Radioactive isotopes, created in reactors, can be used for a variety of practical purposes. Some examples:

1. Strong gamma radiators, primarily Co-60 ($T_{1/2}$ = 5.3 years, E_γ = 1.17 to 1.33 MeV) can be used directly in therapy and radiography.
2. Many applications in industry and research use radioactive isotopes as *radioactive tracers*. Complex chemical or biological processes can be studied by replacing some atoms with a chemically identical radioactive isotope whose path can be followed by the emitted radiation. For organic materials the indicator isotopes are H-3 ($T_{1/2}$ = 11 years; E_β = 0.018 MeV) and C-14 ($T_{1/2}$ = 4700 years; E_β = 0.155 MeV.)
3. Medical diagnostics: Some cancerous tumors are rich in phosphorous and can therefore be localized with P-32 ($T_{1/2}$ = 14 days, E_β = 1.71 MeV). With I-131 ($T_{1/2}$ = 8 days, E_β = 0.6 to 0.32 MeV, E_γ = 0.64 to 0.36 MeV), the thyroid gland can be studied and thyroid illnesses treated.
4. Radioactive isotopes can supply light, long-lived, low-power energy sources for spacecraft and formerly for pacemakers: Pu-238 as energy source with a thermoelectric generator).



▲ The island Helgoland. (Photo: Uwe Muuss.)

Every year, many thousands of pilgrims journey to the places that mark important moments in the lives of the founders of their religions. One such pilgrimage site is the tree in Bodh Gaya, the "tree of enlightenment" (Bodhi), under which the Buddha sat in mediation until he achieved Enlightenment. A similar pilgrimage place for physicists could be the island Helgoland, where Heisenberg for the first time became certain that he was on the correct path. Let us hear from Heisenberg himself:

Toward the end of May 1925, I fell so ill with hay fever that I had to ask Born for fourteen days' leave of absence. I made straight for Helgoland, where I hoped to recover quickly in the bracing sea air, far from blossoms and meadows …

Apart from daily walks and long swims, there was nothing in Helgoland to distract me from my problem, and so I made much swifter progress than I would have done in Göttingen. A few days were enough to jettison all the mathematical ballast that invariably encumbers the beginning of such attempts, and to arrive at a simple formulation of my problem. Within a few days more, it had become clear to me what precisely had to take the place of the Bohr–Sommerfeld quantum conditions in an atomic physics working with none but observable magnitudes. It also became obvious that with the additional assumption I had introduced a crucial restriction into the theory. Then I noticed that there was no guarantee that the new mathematical scheme could be put into operation without contradictions. In particular, it was completely uncertain whether the principle of the conservation of energy would still apply, and I knew only too well that my scheme stood or fell by that principle … Hence I concentrated on demonstrating that the conservation law held, and one evening I reached the point where I was ready to determine the individual terms in the energy table, or, as we put it today, in the energy matrix, by what would now be considered an extremely clumsy series of calculations. When the first terms seemed to accord with the energy principle, I became rather excited, and I began to make countless arithmetical errors. As a result, it was almost three o'clock in the morning before the final result of my computations lay before me. The energy principle had held for all the terms, and I could no longer doubt the mathematical consistency and coherence of the kind of quantum mechanics to which my calculations pointed. At first, I was deeply alarmed. I had the feeling that, through the surface of atomic phenomena, I was looking at a strangely beautiful interior, and felt almost giddy at the thought that I now had to probe this wealth of mathematical structures nature had so generously spread out before me. I was far too excited to sleep, and so, as a new day dawned, I made for the southern tip of the island, where I had been longing to climb a rock jutting out into the sea. I now did so without too much trouble, and waited for the sun to rise.

—*Physics and Beyond* [pp. 60–61]



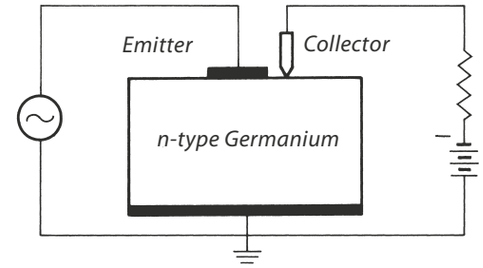
▲ A relatively common type of crystalline solid is the semiprecious stone amethyst pictured here. It consists of SiO_2, usually with additional Fe_2O_3 that gives its color. According to the ancients, it protects its possessor from alcoholism. From the picture, we can visualize how the process of crystallization took place.

After the laws of atomic microsystems had been worked out, the properties of interacting microsystems began to be investigated. The breakthrough in this area is due to W. Heitler and F. London. They were the first to apply quantum mechanics to interpret chemical bonding (H_2 molecule, 1927). The classical statistical description of macroscopic bodies consisting of weakly interacting particles—gases—has already been replaced by quantum statistics: the Bose–Einstein statistics (1924) and the Fermi–Dirac statistics (1926). Using the former, the laws of blackbody radiation are understood as the behavior of a photon gas in equilibrium, while the latter, when applied to the electrons of metals (Pauli, 1926; Sommerfeld, 1927) resolve the contradictions in the theory of specific heats and also bring the theory of thermoelectric and magnetic phenomena into better alignment with the experiments.

The spectacular practical results came about from the theory of solids. With the x-ray investigations of M. von Laue (1912) and W. H. and W. L. Bragg (1913), crystalline structure was made visible. In 1928, F. Bloch investigated the behavior of electrons in a field with a periodic potential on a quantum-mechanical basis and in this way worked out the energy band model of crystalline solids. The generalization and deepening of these ideas are the contribution of R. Peierls and L. Brillouin (1932).

Starting from these theoretical results, W. Shockley, J. Bardeen, and W. Brattain discovered the principle of the transistor in 1948. In 1955–1956, A. M. Prokhorov, N. G. Basov, C. Townes, and N. Bloembergen constructed the first solid-state quantum electronic devices: the maser and the laser. The same theory also provided the basis for the Mössbauer effect (1958).

The idea that the vibrational state of a crystalline lattice can be viewed as a "phonon" gas with a certain energy distribution turned out to be quite fruitful. It goes back to M. Born and Theodor von Kármán (1912); two decades later, Bloch, Houston, and Brillouin again took up the idea to calculate the temperature dependence of the resistivities of metals.



Normally, the number of holes in the n-type semiconductor is negligible so far as a contribution to the conductivity is concerned. However, when electrons are removed from the valence bonds and holes are injected, relatively large numbers of holes will be introduced. The conductivity of the semiconductor will be increased in the neighborhood of the emitter point … This disturbance in the electronic structure can be used to produce amplifying action in the transistor.

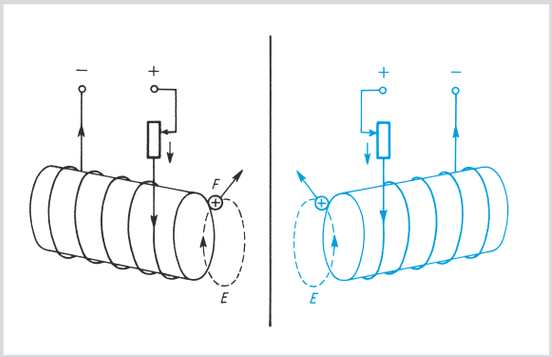
—W. F. Shockley, Nobel lecture, 1956

The investigation of the electron–phonon interaction led to the explanation of one of the most remarkable phenomena: superconductivity. It was discovered in 1911 by Kamerlingh-Onnes, and a phenomenological theory was given in 1934 by F. and H. London, and in 1950 by W. L. Ginzburg and L. Landau. In the more rigorous theory of J. Bardeen, L. N. Cooper, and J. R. Schrieffer, as well as N. N. Bogoliubov (1957), the interaction of phonons and pairs of electrons with opposite spin plays a significant role. A strong support for this theory was given in 1961 by an experiment showing that a magnetic flux surrounded by a superconducting ring will be quantized according to the formula $2e\Phi = nh$, where e is the electron charge, Φ the magnetic flux, h Planck's constant, and n an integer. Today, superconductors are essential for creating high magnetic fields in large volumes, and they are beginning to play a role in high-current technology (generators of high power, energy transfer). Through the superconducting Josephson effect (1962), very sensitive measurements of magnetic fields are possible and also oscillators can be constructed that are tunable over very wide frequency ranges. About the Josephson effect: As a consequence of quantum tunneling, a current can flow through a thin isolating layer (10^{-7} cm) separating two superconductors. When the current exceeds a certain critical value, a potential U will appear and electromagnetic radiation of frequency ν will be emitted. Between the frequency ν and the voltage U, one has the relationship $h\nu = 2eU$. Among its many uses, the effect allows for the determination of the ratio h/e of the two most important quantities of microphysics with what is currently the greatest possible accuracy.

Among the remaining properties of solids, magnetism should be mentioned. After the pioneering work of P. Curie (1895), in 1905, P. Langevin worked out an electron theory of para- and diamagnetism. In 1907, P. Weiss predicted the existence of magnetic domains that result from the interactions of elementary magnets. Heisenberg and Y. Frenkel (1928) were the first to use quantum mechanics to explain ferromagnetism, and finally, in 1935, L. Landau and E. Lifschitz developed a comprehensive phenomenological theory.

With the work of K. Müller and G. Bednorz in 1986, research on superconductivity achieved renewed momentum. Until that time, research was focused on increasing the critical temperature (T_c) of superconductors, from below the boiling point of liquid helium at 4.2 K (0 K = –273.15 °C), to above the boiling point of liquid hydrogen (21 K) using an alloy of niobium, silicon, and germanium. In the 1970s, this goal was in fact achieved but further progress along this path was not possible until Müller and Bednorz started their investigation of a ceramic made of barium, lanthanum, and cupric oxide and showed the possibility of superconductivity at much higher temperatures. Since then, many materials of this sort have been found with T_c far above the boiling point (77 K) of liquid nitrogen (T_c = 94 K for $YBa_2Cu_3O_7$) However, an explanation of superconductivity of ceramic materials remains a challenge for the theoreticians.

▶ In macrophysics, the mirror image of an observable process also represents a real, that is, realizable, process. It is difficult to determine whether what is seen in this picture is the direct or a reflected mirror image of the girl. From the letter B appearing backward on her sleeve, one could conclude that the picture is a mirror image. But a mirrored letter could have been prepared beforehand, which is indeed just what had been done here. It is the form of the snail shell that the girl is tracing with the marker that tells us the truth: the shell is right-handed, as it is in nature; in a mirror image it would look left-handed, therefore the picture shows the true direct image. We see therefore that in macrophysics also we can distinguish left and right; however we ascribe the distinction to some initial—and then inherited—asymmetry that could just as well have been the opposite.



▲ An everyday macro process and its reflection. On the left we see an experimental setup showing the effect of the electrical field induced by the changing magnetic field on an electric charge. The field created by the mirrored setup will act on the (mirrored) charge so that its path will correspond to the mirror image of the path of the real charge.

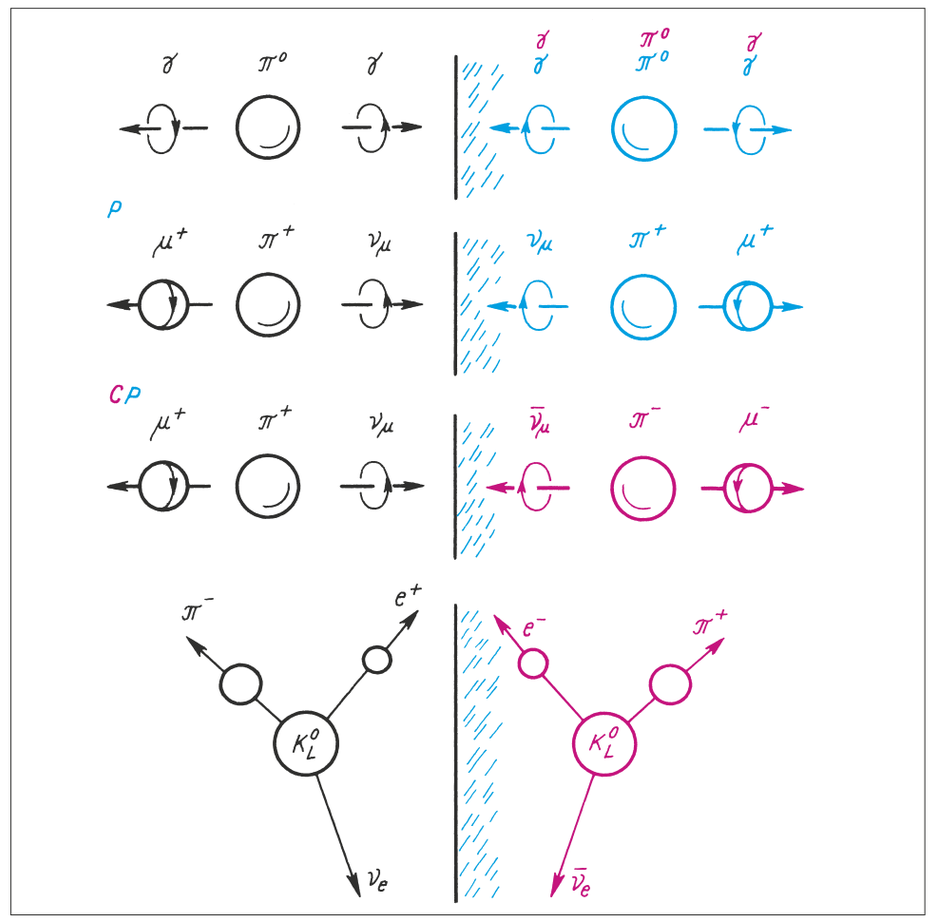
This microphysical process is also mirror-symmetric and obeys the law of conservation of parity; in other words it is P-invariant.

Through reflection of this process we would get a right-handed neutrino which does not exist in nature. The process therefore violates conservation of parity.

▶ If in the mirror image we replace particles with antiparticles, we again have a realizable process. This process is invariant with respect to reflection and simultaneous charge conjugation; it is thus PC-invariant.

The fourth process violates even PC-invariance.

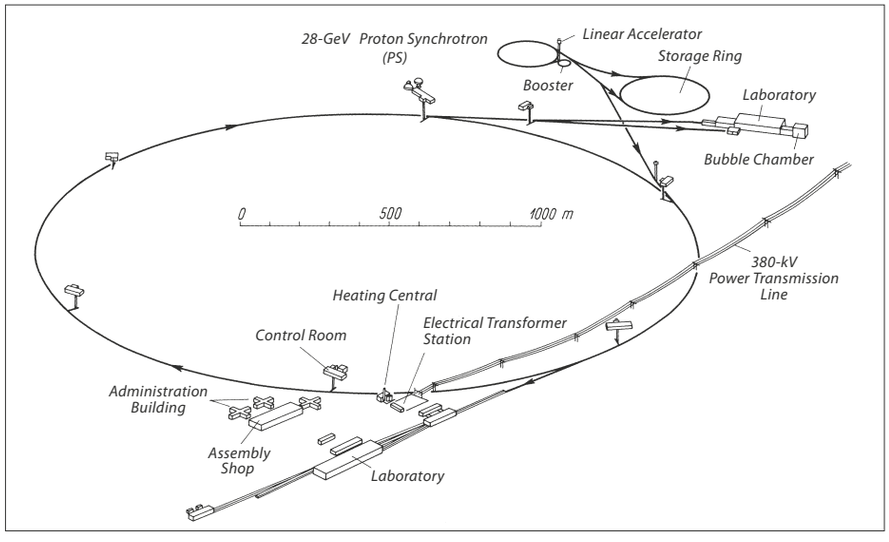






▲▼ A photograph of a portion of the 400-GeV Super Proton Synchroton (SPS) at CERN (Conseil Européen pour la Recherche Nucléaire), Geneva, and a map of the entire installation. Together with the Tevatron in Batavia, IL, and the Linear Accelerator at Stanford, these giant instruments made the discoveries leading to the Standard Model and paved the way for the LHC (Large Hadron Collider), which is expected to reveal the new physics of the twenty-first century. Today, the SPS serves as an auxiliary installation (injector) to the much larger LHC ring.

Protons orbited the nearly 7 kilometer (4 mile) long ring 200,000 times in a vacuum (residual pressure less than 10^{-10} pascal) in a strong magnetic field until they received their full energy from the high frequency alternating electric field. The cost for construction and equipment of the accelerator was about a billion dollars (1700 million Swiss francs), and the installation had its own transformer substation delivering a power of 170 megawatts. Construction began in 1971, and in 1976, the accelerator was placed in service.



FERMIONS
Spin ½

Quarks $\begin{pmatrix} u & u & u \\ d & d & d \end{pmatrix}\begin{pmatrix} c & c & c \\ s & s & s \end{pmatrix}\begin{pmatrix} t & t & t \\ b & b & b \end{pmatrix}$

Leptons $\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}\quad\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}\quad\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}$

BOSONS

γ Photon	Spin 1
$W^{\pm}$, Z_0	Spin 1
g Gluons	Spin 1
G Graviton	Spin 2
Higgs Particle	Spin 0

▲ According to the Standard Model, there are six quarks, each with its own special "flavor": *u* (up), *d* (down), *s* (strange), *c* (charm), *b* (bottom, beauty), *t* (top, truth). Each flavor is associated with three colors. The quarks and the lepton pairs (e–ν_e, μ–ν_μ, τ–ν_τ) are organized into three corresponding "generations." These are all the fermions, and they may be viewed as elementary particles in the true sense of the word. The members of the boson families play the role of mediators in interactions. The much awaited (but not yet seen) Higgs particles are the quanta of the Higgs field that gives mass to the $W^{\pm}$ and Z^0 particles.



▲ Nuclear test explosion. (Comité de l'énergie atomique, France.)

The infernal beauty of the light from these explosions of almost cosmic dimensions brings forth visions of the end of humanity. But for our everyday lives it is of practical importance—especially in the context of reactor accidents—to understand the effects of atomic radiation and the various ways the strength of the radiation is expressed.

The unit for the *intensity* of a radioactive source is connected with the radiation from radium, the historically first known radiation source: The intensity is 1 curie (Ci) when the number of decays per second is the same as in 1 gram of radium (3.7×10^{10} decays/sec.) Smaller—and today more usual—units are:

1 becquerel: one decay per second,

1 rutherford: 10^6 decays per second.

Therefore,

3.7×10^{10} becquerel = 3.7×10^4 rutherford = 1 curie.

Two examples:

2000 Ci represents the activity of a very strong cobalt-60 irradiator.

After the nuclear accident at Chernobyl (1986), the activity of one liter of milk contaminated with I-131 in Great Britain was measured as 500 Bq.

The *effect* of radiation on living organisms or inanimate objects (such as electronic components) depends on the absorbed energy, in other words, on the radiation dose. Again, the basic unit came from the first historical experience with x-ray dosimetry where ionization chambers were used: the unit energy dose 1 roentgen (R) will release through ionization in 1 kg of dry air mass the charge 2.58×10^{-4} coulombs (ampere-seconds). A dose of 1 R can be expressed in terms of energy (J/kg) using the ionization energy required to create a pair of ions. The result for water or for the human body is 0.93×10^{-2} J/kg. This value rounded up was taken to be the dosage unit: 1 rad = 10^{-2} J/kg, which is just the roentgen expressed as energy in human tissue. The more recent dosage unit is the gray (Gy), defined as 1 Gy = 100 rad = 1 J/kg.

The unit of the *effective* dosage is the sievert (Sv) = 100 rem, where 1 rem (roentgen equivalent man) has by definition the same biological effect as an x-ray dose of size 1 rad. To take into account the different effects of different forms of radiation, the factor RBE (relative biological effectiveness) has been employed; its values are 20 for α particles, 10 for neutrons, and 1 for β, γ, and x-rays. The biological effect depends as well on additional factors. For example, living tissue is more likely to regenerate if the same dose is received over a longer time. Conversely, some radioactive isotopes may concentrate in certain parts of the body and thus cause greater damage. In this regard, some long-lived isotopes among fission products are particularly dangerous: Sr-90 ($T_{1/2}$ = 28 years) is stored in bones and can cause cancer; Cs-137 ($T_{1/2}$ = 30 years) collects in soft tissues.

We are continually exposed to normal background radiation (cosmic rays, natural radioactivity, etc.) of about 2×10^{-3} sieverts per year. Mankind has borne this dose rate for millennia. Ten sieverts, if received within a few minutes, can be considered lethal.

A standardization of radiation dosage safety limits for the general public as well as for certain professions has been under way since 1986.

principle. His derivation did not rest on metaphysical considerations but on the mathematical properties of the paths. Euler brought Maupertuis's principle into the form that is still in use today:

$$\int_{P_1}^{P_2} mv\,\mathrm{d}s = \text{ extremum.}$$

We wish to emphasize once again that the equation for light with which Maupertuis began is incorrect. As Leibniz had conjectured a half century earlier, the correct relationship for light is

$$l_1 n_1 + l_2 n_2 = \text{ extremum.}$$

If we consider further the correct proportionality $n \propto 1/v$, then this extreme value requirement yields Fermat's principle; but, if we substitute the incorrect relationship $n \propto v$, then we get the initial equation used by Maupertuis. Nevertheless, the formulation of Maupertuis's principle given by Euler is valid in the entire domain of mechanics and is applicable to a great variety of problems—with only certain restrictions, which we shall discuss in the context of the Hamiltonian principle.

It is worthwhile at this point to recall the intellectually stimulating, but truly tedious, train of thought Fermat used to determine the minimum and thus the law of refraction (Section 3.5). We present here the derivation set forth by Maupertuis to show how simple and immediate the solution of extreme value problems had become two generations later with the application of differential calculus. We seek the minimum of the expression (Figure 4.39)

$$mAR + nRB,$$

and to this end, we employ the relationship

$$\mathrm{d}\left[m\sqrt{AC^2 + CR^2} + n\sqrt{BD^2 + DR^2}\right] = 0.$$

Since AC and BD are constants, it follows that

$$m\frac{CR\,\mathrm{d}(CR)}{\sqrt{AC^2 + CR^2}} + n\frac{DR\,\mathrm{d}(DR)}{\sqrt{BD^2 + DR^2}} = 0.$$

Since CD is also a given constant, we obtain

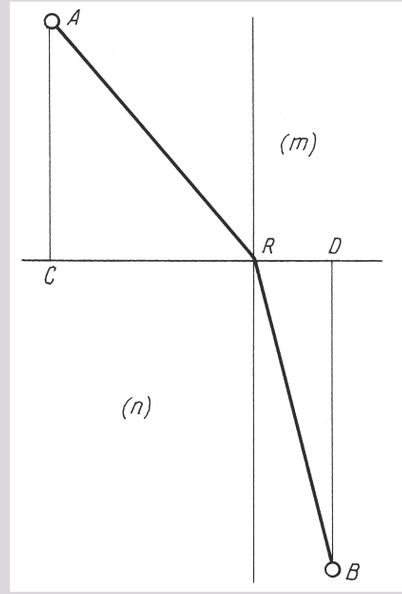
$$\mathrm{d}(CR) = -\mathrm{d}(DR),$$

and finally,

$$m\frac{CR}{AR} = n\frac{DR}{BR} \rightarrow \frac{CR}{AR}\Big/\frac{DR}{BR} = \frac{n}{m}.$$

We can thus see that the quotient of the sine of the angle of incidence and the sine of the angle of refraction is inversely proportional to the quotient of the velocities of light in the corresponding media.

When we come to speak of the mechanics of Lagrange and Hamilton, we shall discuss the close relationship, mentioned at the beginning of this section, between Newton's law of motion and the principle of least action.



▲ **Figure 4.39** Derivation of the law of refraction after Maupertuis. Here m and n are quantities proportional to the velocities.

4.2.5 *The First "Positivist": d'Alembert*

D'Alembert's principle is not much used today, especially not in the form in which it was given by d'Alembert himself (Figure 4.40) in the *Traité de Dynamique* (1743). D'Alembert's philosophical attitude toward the problem he investigated is interesting and characteristic. It is no accident that we are in the century of Reason, and it is no accident that d'Alembert, an outstanding and highly typi-



▲ **Figure 4.40** Jean Le Rond d'Alembert (1717–1783): illegitimate child, left by his mother, a well-known courtesan, on the steps of the church of Saint Jean le Rond. His foster mother was so well supported by the father that she was able to give the young Jean Le Rond a good education. In 1743, his most important work in physics, *Traité de Dynamique*, appeared. D'alembert was also concerned with problems in optics, and he won a prize in 1746 for his treatise on the origin of winds. In 1754, he became a member of the *Académie Française*, and from 1772 its permanent secretary. Of perhaps greater significance is his work in producing the great French *Encyclopédie*; his introduction (*Discours préliminaire*, 1750) is a declaration in support of the French Enlightenment. As a philosopher, he was an adherent of Bacon, Locke, and the English empirical school. He is often viewed as the first representative of positivism; in fact, he rejected minimality principles and even the energy conservation law as statements of metaphysics. Tired of controversy, he withdrew from his editorship of the *Encyclopédie* in 1757.

cal representative of this century, is one of the editors of the *Encyclopédie*. Indeed, throughout the entire eighteenth century, the question was debated as to whether the fundamental laws of mechanics were necessary, that is, derivable from generally acceptable logical principles, or whether they were accidental, which, simply expressed, would mean that while they are what they are in nature, logically they could well be otherwise. In the eighteenth century, one was inclined to emphasize the *ratio* over the *experientia* in the motto regarding the search for truth, going back to Huygens' *ratio et experientia*. It is again not by chance that in the middle of the century, the Prussian Academy of Sciences offered a prize for the best investigation into and—if possible—the best solution to the question of whether the fundamental principles of mechanics are necessary truths or accidental truths, thereby posing Huygens' motto as a question: *ratio vel experientia?* D'Alembert was firmly convinced that the fundamental principles could be derived with logical necessity, and that is why he wanted to eliminate the effects of force from mechanics (Quotation 4.13), the idea being that these effects make the principles less secure and tend to turn mechanics into a mere experimental science.

We note here that in 1703, Jakob Bernoulli formulated a similar principle, which could, however, only be applied to a much narrower range of phenomena; indeed, Bernoulli used his principle only to determine the motion of the physical pendulum. It is noteworthy that d'Alembert also turned to the physical pendulum to prove the applicability of his principle. This is not the only case in the history of science where in order to prove the power of a new theory, it is applied to the fashionable problem of the era that is nontrivial and that has already been solved exactly several times before. Similarly, in the twentieth century, we shall see that at each stage in the development of quantum theory—in Bohr's theory, Heisenberg's matrix mechanics, and Schrödinger's wave mechanics—the hydrogen atom was used as a touchstone. In the first half of the eighteenth century, up until the time when Euler presented a fully developed general theory of motion of rigid bodies, the physical pendulum was used to test whether a new physical principle could be placed among recognized theories, or at least among those that were disputed but still worthy of notice.

We now look at how d'Alembert formulated his general principle and how he applied it to the physical pendulum. We begin with a definition:

> By "motion" of a body I mean the velocity of the body, also considering its direction. By "magnitude of motion" I mean, as is customary, the product of mass and velocity.

D'Alembert formulated the general problem of mechanics as follows:

> Let there be given a system of bodies that are linked together in some way however they are arranged; we assume that a certain motion has been imparted to each body, which it is unable to fulfill on account of its linkage with the other bodies. One seeks the motion that each body must exhibit.

The solution given by d'Alembert is as follows:

> Let *A*, *B*, *C*, etc. be the system of associated bodies, and let us suppose that one has imparted to these bodies the motions *a*, *b*, *c*, etc., which because of the interactions are transformed into a, b, c, etc. It is clear that one may think of the motion *a* imparted to *A* as composed of the motion a, which it has assumed, and another motion α; and one may think of the motions *b*, *c*, etc. as composed of b and β, c and γ, etc., from which it follows that the motions of the bodies *A*, *B*, *C* would be the same if instead of the propulsions *a*, *b*, *c*, etc., one were to apply in pairs the propulsions a, α; b, β; c, γ; etc. According to our

assumption, the bodies *A*, *B*, *C*, etc. have assumed on their own the motions a, b, c, etc. The motions α, β, γ must therefore be such that they in no way alter the motions a, b, c, etc. That is, if the bodies had been given the motions α, β, γ, these would have to have canceled, resulting in zero net motion.

Thus one has the following principle for finding the motion of several bodies with mutual interaction: one decomposes the motions *a*, *b*, *c*, etc. applied to each body into two others, a, α; b, β; c, γ; etc., such that if one had applied to the bodies only the motions a, b, c, etc., then these bodies would have been able to maintain these motions without any mutual hindrance, and that if one had applied the motions α, β, γ, etc. to the bodies, the system would have remained at rest. It is then clear that a, b, c, etc. are the motions that these bodies would assume due to their interaction. Such is the solution to the problem.

—D'ALEMBERT, *Traité de Dynamique*

We now apply d'Alembert's principle to a physical pendulum (Figure 4.41) consisting of three bodies *A*, *B*, and *R* connected by a rigid straight rod. We assume that the three bodies, if not prevented from doing so by the rod, would have moved during the same period of time through three infinitesimally small distances *AO*, *BQ*, and *RT*. In fact, the actual path is *RS*, not *RT*, and from the size of *RS*, we can determine the values of *AM* and *BG*.

The motion (velocity) thus consists of the following parts:

$$RT = RS + ST, \quad BQ = BG - GQ, \quad AO = AM - MO,$$

and the principle asserts that the pendulum remains at rest if the bodies *A*, *B*, and *R* are given only the motions *ST*, *–GQ*, and *–MO*. We obtain the following equilibrium condition:

$$A \cdot MO \cdot AC + B \cdot QG \cdot BC = R \cdot ST \cdot CR.$$

Let us now follow D'ALEMBERT and introduce the notation

$$AO = a, \quad BQ = b, \quad RT = c, \quad RS = z, \quad CA = r, \quad CB = r', \quad CR = \rho.$$

Then the equilibrium equation assumes the form

$$R(c - z)\rho = Ar\left(\frac{zr}{\rho} - a\right) + Br'\left(\frac{zr'}{\rho} - b\right),$$

from which we finally obtain

$$z = \frac{Aar\rho + Bbr'\rho + Rc\rho^2}{Ar^2 + Br'^2 + R\rho^2}.$$

If now the forces *F*, *f*, and *φ* act on the bodies *A*, *B*, and *R*, then we obtain the accelerational force (acceleration) of the body *R* if in place of the quantities *a*, *b*, *c* we substitute the quantities *F*/*A*, *f*/*B*, *φ*/*C*. We then obtain the expression

$$\frac{Fr + fr' + \varphi\rho}{Ar^2 + Br'^2 + R\rho^2}\rho,$$

and the increase in velocity from passing through the arc element d*s* can be calculated from

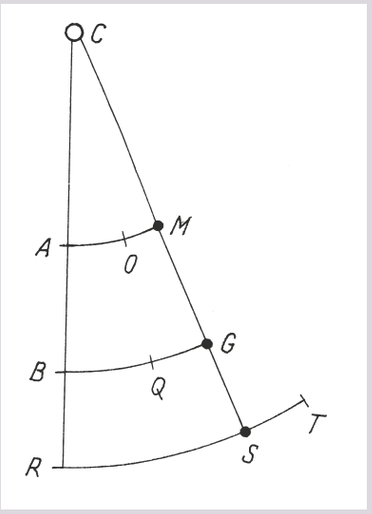
$$\frac{Fr + fr' + \varphi\rho}{Ar^2 + Br'^2 + R\rho^2}\rho\,\mathrm{d}s = u\,\mathrm{d}u.$$

Today, in such investigations one would start with the relationship "torque = moment of inertia × angular acceleration" (Figure 4.42); that is,

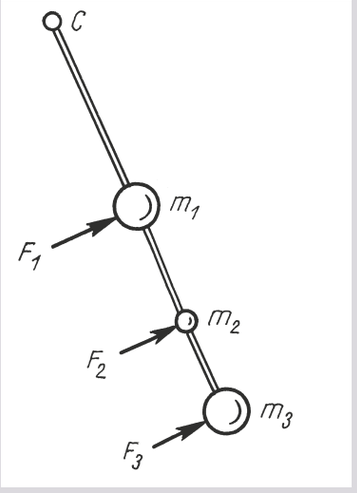
$$F_1r_1 + F_2r_2 + F_3r_3 = \left(m_1r_1^2 + m_2r_2^2 + m_3r_3^2\right)\frac{\mathrm{d}\omega}{\mathrm{d}t}.$$

This equation can then be brought into the form

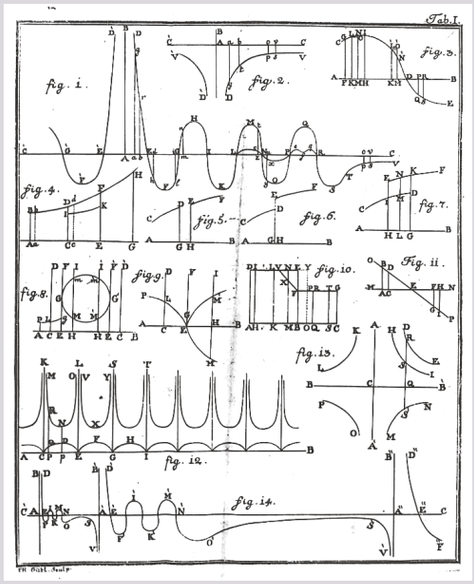
$$\frac{\sum F_ir_i}{\sum m_ir_i^2} = \frac{\mathrm{d}\omega}{\mathrm{d}t}.$$



▲ **Figure 4.41** Application of d'Alembert's principle to determining the period of a physical pendulum, how D'ALEMBERT himself imagined it. Our conception today is fundamentally different.



▲ **Figure 4.42** Restatement of d'Alembert's formula using modern concepts.



▲ **Figure 4.43** According to Boscovich, force depends on distance in the neighborhood of a "mass point" (fig. 1). (Library of the University for Heavy Industry, Miskolc.)



▲ **Figure 4.44** Mikhail Vassilyevich Lomonosov (1711–1765): born into the family of a fisherman not far from Arkhangelsk. In 1730, he left his place of birth to study in Moscow and then St. Petersburg. In order to do so, he had to hide his origins. From 1736, he studied with a scholarship in Marburg under the famous physicist and philosopher Christian Wolff. Later, he studied chemistry and metallurgy in Freiburg. In 1745, he was appointed professor of chemistry in Petersburg, the first Russian to be named to the academy, where previously only foreigners had taught. In 1748, he set up a chemical research laboratory. On his initiative, a university on the European model was founded in 1755 in Moscow

continued on next page

If this equation is now multiplied by $r_3\,\mathrm{d}s$ and the right-hand side transformed via

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} r_3 \mathrm{d}s = \frac{\mathrm{d}s}{\mathrm{d}t} \mathrm{d}\left(r_3\omega\right) = u_3\,\mathrm{d}u_3,$$

we obtain

$$\frac{\sum F_i r_i}{\sum m_i r_i^2} r_3\,\mathrm{d}s = u_3\,\mathrm{d}u_3,$$

in agreement with the relationship derived by d'Alembert.

4.2.6 *Modern Ideas*

We have seen that on the Continent, the introduction of a gravitational force acting at a distance was viewed as the most questionable proposition of the *Principia*. Even Huygens was of the view that this amounted to a smuggling back into science of the medieval occult qualities that had fortunately been banished. In this respect, everyone followed Descartes; they simply could not imagine any interaction other than through immediate contact. We should mention here the surprising fact that this Cartesian proposition actually goes back to Aristotle. Even Newton was inclined to believe that the law of action at a distance should somehow be explainable within the Cartesian framework. However, he avoided pursuing the matter. In any case, Newton held to the idea of two sorts of force: forces that are transmitted through direct contact and those, such as gravitation, acting at a distance, where a transmitting medium is excluded.

The most forceful advocate of the Newtonian point of view was the Jesuit Roger Joseph Boscovich (1711–1787), from Ragusa (Dubrovnik), who held a teaching post in Rome. He turned the question around: How is one to understand in detail the interaction that we call immediate contact? On the basis of clever considerations, he finally arrived at the notion that even in this case, the interaction can occur only as a result of force acting at a distance. With these ideas, the defensive arguments turned into an offensive against the Cartesians, as Boscovich went on to show that the arguments put forth by Descartes's followers were untenable. His arguments are very easy to understand and extremely convincing. Following Boscovich, we imagine two bodies, one of which is traveling at a speed of, say, 6 units, the other at a speed of 12. We imagine what transpires when the two bodies collide, where we naturally assume that the faster body followed the slower one of the same mass along the same straight path. By conservation of "quantity of motion," both bodies should continue after collision along the same path, each with a speed of 9 units (in the case of elastic collision in the first brief phase of the collision, and in the case of inelastic collision thereafter as well).

But how did the velocity of the faster body come to be reduced from 12 to 9, and that of the slower body increased from 6 to 9? Clearly, the time interval for the change in velocities cannot be zero, for then, argues Boscovich, the instantaneous change in speed would violate the law of continuity. Furthermore, we would have to say that at the moment of impact, the speed of one body is simultaneously 12 and 9, which is patently absurd. It is therefore necessary for the change in speed to take place in a small, yet finite, amount of time. But with this assumption, we arrive at yet another contradiction. Suppose, for example, that after a small interval of time, the speed of the faster body is 11, and that of the slower body is 7. But this would mean that they are not moving at the same velocity, and the front surface of the faster body would advance through the rear surface of the slower body, which

is impossible because we have assumed that the bodies are impenetrable. It therefore becomes apparent that the interaction must take place immediately before the impact of the two bodies and that this interaction can only be a repulsive one because it is expressed in the slowing down of one body and the speeding up of the other. Moreover, this argument is valid for arbitrary speeds, so one can no longer speak of definite dimensions for the particles that were until now thought of as impenetrable, namely, for the atoms. An atom should rather be viewed as a point source of force, with the force emanating from it acting in some complicated fashion that depends on distance.

Boscovich paints a picture very similar to those that we use today when we try to visualize the interactions of atoms. According to Boscovich, when bodies are far apart, they act on each other through a force corresponding to the gravitational force, which is inversely proportional to the square of the distance. But with decreasing distance, this law must be modified because, in accordance with the above considerations, the force changes sign and must become a repulsive force. Figure 4.43 shows a typical page from Boscovich's book with a drawing that represents the course of the action of the forces. As we can see, the force changes sign several times. To understand the significance of these sign changes, we need to go beyond Boscovich and consider the potential associated with this force. Mathematically, this is the quantity whose derivative gives the force; physically, the potential is equal to the potential energy associated with the force. We then find that the potential energy has a minimum at a number of places. Such a minimum corresponds to a state of equilibrium, meaning that no force acts on the body there; furthermore, this state is stable because in order to move a body out of its equilibrium in whichever direction, we have to use some additional force. The existence of energy minima or stable states of equilibrium means that a fixed, stable bond can come into existence between the particles. With this idea, Boscovich not only offered a new picture for interactions in place of the Aristotelian–Cartesian theory based on immediate contact, but also showed a way toward the understanding of the structure of matter, especially that of solid bodies.

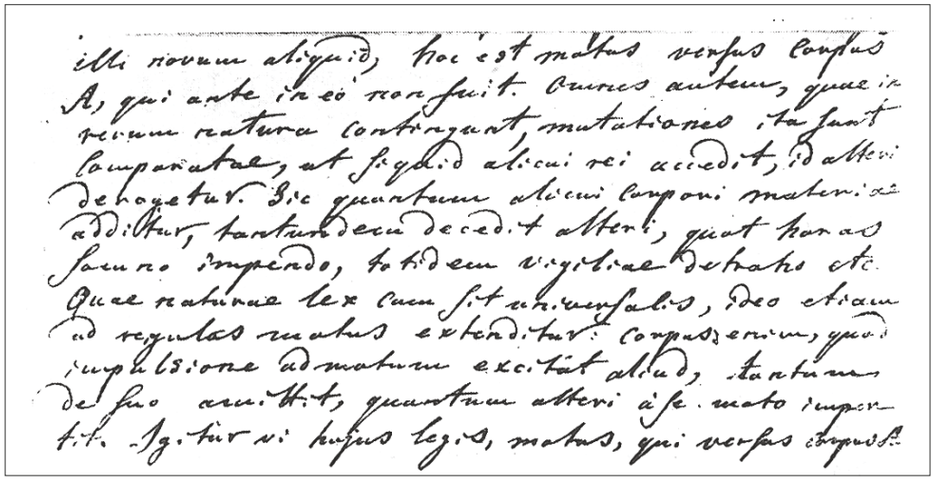
Boscovich had hardly any influence in his own time and had to be rediscovered in the second half of the twentieth century, in spite of the fact that he wrote his works in Latin and lived his life in the maelstrom of scientific activity. The reason may have been that his ideas were qualitative, and it was still much too early for them to be cast into a form that would have allowed for quantitative calculations.

Figure 4.44 *continued*

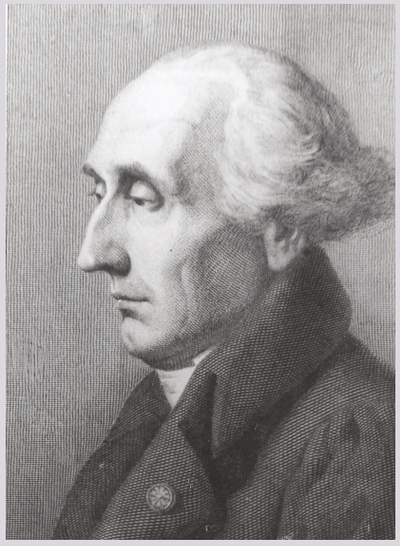
(Lomonosov Moscow State University). His scientific accomplishments earned him international recognition. He was elected to the Swedish Academy in 1760 and to the academy in Bologna in 1764.

Indicative of the extraordinary range of his interests and talents is the fact that he immediately made use of the technology that he developed for stained glass to compose his own mosaics, the most famous of which is perhaps the monumental (21 by 16 feet) representation of the Battle of Poltava (the decisive victory of Peter the Great over Sweden). In recognition of his artistic achievement, he was elected to the Academy of Art.

> Since we cannot find a speed of motion so great that we cannot imagine a greater, the highest possible temperature represents an unimaginable motion. On the other hand, motion can come to rest: The greatest possible degree of coldness is therefore an absolute cessation of thermal motion and can thus exist.
>
> From all the above, we conclude that it is completely unnecessary to attribute to the heat of bodies a subtle material invented just for the purpose; on the contrary, heat consists in an internal circular motion of the bound material of the warm body. We maintain that the finest material of the ether that fills all of space is sensitive to this motion and heat; this it is that receives warmth from the Sun, transmits it to our Earth and the other celestial bodies and thereby makes them warm; the ether is the medium by means of which bodies far distant from each other communicate heat without anything palpable.
>
> Having taken care of heat substance, we could close this article were it not that some scholars would ascribe a substance to cold as well, seeing its location in salts, since these create cold on being dissolved in water. However, since salts also frequently create heat, for example in mixing table salt with vitriol, we could just as well ascribe the cause of heat to salts if we did not hold such a senseless argument to be beneath our dignity.
>
> —Lomonosov, *Reflections on the Causes of Heat and Cold*, 1744–1747

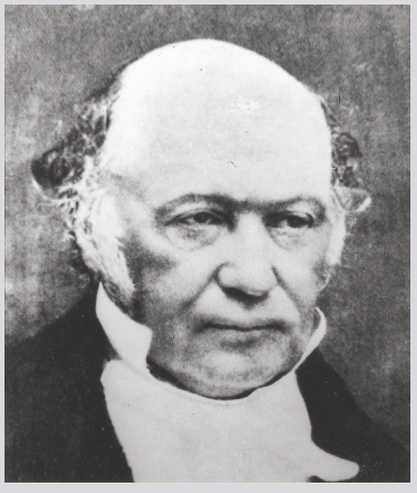


◀ **Figure 4.45** Lomonosov's letter to Euler (1748, written in Latin); he describes here his theory on the conservation of matter and motion.



◀ Figure 4.46 Joseph Louis Lagrange (1736–1813): Born and raised in Turin; at nineteen, professor of mathematics at the Turin artillery school. His first significant contributions to mathematics and physics (solution of the isoperimetric problem, theory of vibration of strings, propagation of sound) were published in the communications of the Royal Academy of Turin, which he founded. As successor to Euler, after the latter was called to St. Petersburg, he worked for 20 years in Berlin. 1787: move to Paris; there from 1795 lecturer at the École Normale and École Polytechnique. He was an outstanding teacher, and through his own personality and his books, he remained a revered figure for the following generations of scientists. His most important works are the *Mécanique analytique* (1788, Paris) and the *Théorie des fonctions analytiques* (1797).

The Lagrange function and Lagrange equations of the second kind remain of fundamental importance in the physics of our times.



◀ Figure 4.47 William Rowan Hamilton (1805–1865): Born in Dublin; after studies there, worked as teacher and scholar as professor of astronomy and was the Astronomer Royal of Ireland. He was a child prodigy, at age five translating texts from Latin, Greek, and Hebrew. At 16, he worked through Newton's *Principia* and Laplace's *Mécanique céleste*. At age 22, before finishing university, he was appointed to a professorship. In his later years he succumbed to alcoholism.

His most important works for the physics of the twentieth century are *Theory of Systems of Rays* (1827) and *On a General Method in Dynamics* (1835).

The Hamiltonian principle of least action, Hamiltonian canonical equations, and Hamilton's principal function of classical mechanics are named in his honor.

At this time of flourishing advances in mathematics, which on the one hand often concealed the physical essence but on the other achieved astounding successes, even the best qualitative theory—just think of the principle of least action—could achieve recognition only if it could be cast in the form of a differential equation whereupon, with its help, problems could be solved. It is quite possible that scholars such as Lagrange took passing note of the intellectual flights of the Jesuit priest with at best a measure of goodwill.

At this point we must bring up the name another important personality of the eighteenth century, Mikhail Vasilyevich Lomonosov (1711–1765). Lomonosov was one of the leading figures of Russian intellectual life in the eighteenth century. He saw it as his goal in life to invigorate the scientific activity of his homeland (Figures 4.44, 4.45). He wrote poetry and plays; his Russian grammar was a groundbreaking work. In 1755, he founded Moscow State University. Lomonosov was involved in every branch of the natural sciences, and he made pioneering investigations in physics, chemistry, astronomy, geology, and geophysics. For the history of physics, his two works, *Razmyshleniya o prichinakh tepla i kholoda* (*Reflections on the Causes of Heat and Cold)* (1744) and *Opyt teorii uprugosti vozdukha* (*An Experiment in the Theory of the Elasticity of Air*) (1748), are of particular importance. In the first of these, he describes how heat is due to the turbulent motion of imperceptible particles, and he explains the transmission of heat as a transfer of motion from faster-moving particles to slower-moving ones. He speaks of a "final low degree of cold" (which today we would call absolute zero), at which all motion ceases. In his second article, he attempts to use these ideas to derive the gas law. What is important is that he correctly understood the role of temperature. In addition, he formulated the law of the conservation of mass.

He wrote in Russian. Due to the vast scope of his interests, his ideas were not worked out in detail, and, although they were seen as important by Euler, they had no influence on the development of Western science.

4.2.7 *Mechanics as Poetry*

Newton's ideas were transformed by Euler into a tool suitable for daily use, enabling anyone with a mathematical background to solve practical technical problems. In the hands of Lagrange (Figure 4.46) and Hamilton (Figure 4.47), this tool was further refined, so that not only could it be used to solve complex problems, but it was indeed transformed into a work of art. The title of this section is a play on a remark of Hamilton's regarding Lagrange's work: "… a scientific poem …."

We begin with a brief description of the Lagrangian equations of motion. Given is a mechanical system of arbitrary complexity with f degrees of freedom whose spatial disposition can be described by f independent generalized spatial coordinates $q_1, q_2, \ldots, q_f$. We now determine the difference between potential and kinetic energy $T - U \equiv L$ as the function

$$L(q_1, q_2, \ldots, q_f;\ \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_f)$$

of the generalized spatial coordinates q_i and their derivatives $\dot{q}_i$ with respect to time, called the *generalized velocities*.

This function is called the *Lagrangian* of the system. The equations of motion are obtained from the Lagrangian in a form called the *Lagrangian equations of motion*, given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \quad i = 1, 2, 3, \ldots, f.$$

For a mass point moving under the influence of a force $-\partial U/\partial x$ derivable from the potential $U(x)$, we have

$$L = T - U = \frac{1}{2} m\dot{x}^2 - U$$

and therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} = \frac{\mathrm{d}}{\mathrm{d}t} m\dot{x} = m\ddot{x}, \quad \frac{\partial L}{\partial q_i} = -\frac{\partial U}{\partial x}.$$

Thus the Lagrangian equations of motion lead to the second law of motion in the usual form

$$m\ddot{x} = -\frac{\partial U}{\partial x}.$$

Of course, with such a simple example, the advantages of using the Lagrangian equations are not apparent.

The germ of the Hamiltonian principle can already be found here: if we take the Lagrangian $L = T - U$ introduced above, it can be proved in general that the system under consideration moves along a path between two fixed points for which the time integral over the Lagrangian satisfies the condition

$$\int_{t_1}^{t_2} (T - U)\,\mathrm{d}t = \text{extremum}.$$

This condition may be written more simply as

$$\delta \int_{t_1}^{t_2} L\,\mathrm{d}t = 0.$$

If we seek the value x_0 of the independent variable x that leads to an extreme value of a given function $y = y(x)$, that is, that makes $y_0 = y(x_0)$ either a maximum or a minimum, then we must study the first derivative $y'(x) = \mathrm{d}y/\mathrm{d}x$. Where the first derivative is zero, an extreme value *may* be present. The examination of the second (or further) derivatives can give information about whether the extreme is there and, if so, of what type (maximum or minimum).

However, if we are looking for a function $y = y(x)$ that makes the (definite) integral of a given function $F(x, y, y')$ an extremum, then we have at our disposal the methods of the calculus of variations. This simplest problem of this type is to find the shortest curve between two points. The solution is easy: a straight line. Earlier, we met a nontrivial problem of this type, namely, the brachistochrone problem: What is the curve in the vertical plane with given initial and terminal points that a body traverses solely under the influence of gravity in the least amount of time. This problem, posed first in this form by Johann Bernoulli (1696, in *Acta Eruditorum*), was correctly solved by several scientists of that era—Jakob Bernoulli, Leibniz, Newton, and Johann Bernoulli himself. A general solution method was given only a half century later (1744), by Euler, who reduced the problem to the solution of a particular differential equation.

Let $y = y(x)$ be the desired function that makes the value of the integral

$$\int_{x_1}^{x_2} F(x, y, y')\,\mathrm{d}x$$

an extremum. Then $y = y(x)$ must satisfy the following differential equation, now named after Euler:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0.$$

If several functions $y_i = y_i(x)$, $i = 1, 2, \ldots, f$, make up the given function F, that is,

$$F = F(x; y_1, y_2, \ldots, y_f; y_1', y_2', \ldots, y_f'),$$

then the Euler equations become

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y_i'}\right) - \frac{\partial F}{\partial y_i} = 0, \quad i = 1, 2, \ldots f.$$

We see immediately that the problem

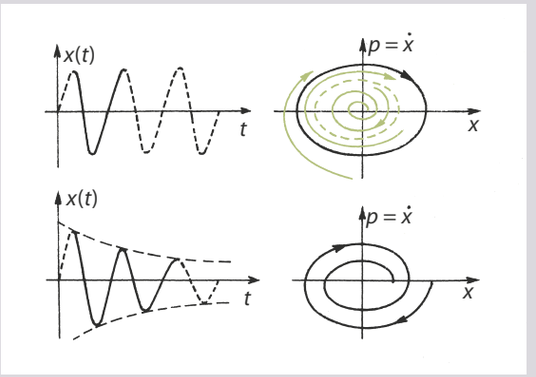
$$\int_{t_1}^{t_2} L\,\mathrm{d}t = \int_{t_1}^{t_2} (T - U)\,\mathrm{d}t = \text{ extremum}$$

leads to the following Euler equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0.$$

Here the Euler equations are identical to the Lagrangian equations of motion.

The Hamiltonian canonical equations can now be written as the sum of the kinetic and potential energies, but instead of using the quantities q_i and $\dot{q}_i$, now using q_i and its corresponding (conjugate)



▲ **Figure 4.48** The 2*f*-dimensional space specified by the coordinates

$$q_1, q_2, \ldots, q_f, \qquad p_1, p_2, \ldots, p_f$$

is called a phase space (see also Section 4.5.11). In this space, the state of a system can be fully characterized by a single point. By specifying such a point, the entire "path" of the system can be uniquely determined in the phase space because by giving the initial conditions, the motion is determined at later points in time. In the case of a time-independent potential, the point representing the system moves along a $(2f - 1)$-dimensional "surface," determined by the equation E = constant.

The calculations that result from these statements can be easily followed and visualized in the case of a simple but very important system, that of the linear oscillator. Suppose a body of mass m moves along the x-axis without friction under the action of a force $F = -kx$. The kinetic energy is given by

$$\frac{1}{2}mv^2 = \frac{1}{2}\frac{p^2}{m},$$

and the potential energy is given by $\frac{1}{2}kx^2$. Therefore, the Hamilton function becomes

$$H(p, x) = \frac{1}{2}\frac{p^2}{m} + \frac{1}{2}kx^2.$$

The equations of motion are therefore

$$p = m\dot{x}, \quad \frac{\mathrm{d}p}{\mathrm{d}t} = m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = -kx.$$

The frequency of this vibrational system is

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{m}}.$$

The phase point moves along an elliptical path in the phase plane (p, x) depicted in the figure, determined by the total energy E:

$$\frac{1}{2}\frac{p^2}{m} + \frac{1}{2}kx^2 = E, \quad \frac{p^2}{2mE} + \frac{x^2}{2E/k} = 1, \quad \frac{p^2}{a^2} + \frac{x^2}{b^2} = 1.$$

A fact that will be important in later sections is that

$$A = ab\pi = \sqrt{2mE}\sqrt{2E/k}\,\pi = 2E\sqrt{m/k}\,\pi = E/\nu.$$

continued on next page

Figure 4.48 *continued*

In quantum theory, the energy of a (linear) oscillator can assume only values that are integer multiples of the energy quantum *hv*: *E* = *nhv*. Thus, *A* = *E*/*v* = *nhv*/*v* = *nh*; that is, the area of the ellipse is quantized. This chain of ideas leads to the generalization of the simple BOHR quantization rule (see the Bohr–Sommerfeld model, Section 5.3.5).

If an oscillator is dissipative, that is, energy-consuming, we obtain—in the case of damping that is linearly dependent on the velocity—an exponentially decaying process. An oscillation with constant amplitude can be maintained in such cases only with the help of (external) excitation.

In the case of nonlinear dissipative systems with external excitation, new phenomena occur. The final state depends strongly on the initial conditions. It can be the zero point, a cyclic path, or even the infinite point in the phase space. This system exhibits a peculiarity in addition to this extraordinary sensitivity: The motion can be extremely complex. The paths cannot be pictured graphically; instead of curved paths, we obtain enormous stochastic sets of points. The investigation of such phenomena has moved into the foreground in recent years and has emerged as a new branch of science: chaos theory.

Quotation 4.17

The idea of necessity arises from some impression. There is no impression convey'd by our senses, which can give rise to that idea. It must, therefore, be deriv'd from some internal impression, or impression of reflection. There is no internal impression, which has any relation to the present business, but that propensity, which custom produces, to pass from an object to the idea of its usual attendant. This therefore is the essence of necessity. Upon the whole, necessity is something that exists in the mind, not in objects; nor is it possible for us ever to form the most distant idea of it, consider'd as a quality in bodies. ... The efficacy or energy of causes is neither plac'd in the causes themselves, nor in the deity, nor in the concurrence of these two principles; but belongs entirely to the soul, which considers the union of two or more objects in all past instances. 'Tis here that the real power of causes is plac'd, along with their connexion and necessity.

I am sensible, that of all the paradoxes, which I have had, or shall hereafter have occasion to advance in the course of this treatise, the present one is the most violent, and that 'tis merely by dint of solid proof and reasoning I can ever hope it will have admission, and overcome the inveterate prejudices of mankind. Before we are reconcil'd to this doctrine, how often must we repeat to ourselves, *that* the simple view of any two objects or actions, however related, can never give us any idea of power, or of a connexion betwixt them: *that* this idea arises from the repetition of their union: *that* the repetition neither discovers nor causes any thing in the objects, but has an influence only on the mind, by that customary transition it produces:

continued on next page

momentum p_i, defined as

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \quad i = 1,2,\ldots,f.$$

We thereby obtain the Hamiltonian function

$$H = H(q_1, q_2, \ldots, q_f; p_1, p_2, \ldots, p_f).$$

This is nothing other than the total energy, expressed in positional and momentum coordinates (Figure 4.48).

After these preparations, the Hamiltonian equations can be given as follows:

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H}{\partial p_i}.$$

As an illustration, let us again take the example given by LAGRANGE of a mass point. As a function of $q \equiv x$, $\dot{q} \equiv \dot{x}$, the energy can be written as

$$\text{energy} = \frac{1}{2} m\dot{x}^2 + U(x),$$

while the Lagrangian can be written as

$$L = \frac{1}{2} m\dot{x}^2 - U(x).$$

Therefore, the momentum conjugate to the positional coordinate x is

$$p = \frac{\partial L}{\partial \dot{x}} = m\dot{x},$$

while the Hamiltonian function is

$$H(p,x) = \frac{1}{2}\frac{p^2}{m} + U(x).$$

Now the Hamiltonian canonical equations for our concrete example can be written down:

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} \rightarrow m\ddot{x} = -\frac{\partial U}{\partial x},$$
$$\dot{q}_i = \frac{\partial H}{\partial p_i} \rightarrow \dot{x} = \frac{p}{m}.$$

It is clear that the first Hamiltonian equation in this simplest of cases becomes the usual Newtonian equation of motion. The second equation, on the other hand, simply gives the definition of momentum.

The following relationships play an important role in both classical physics and quantum mechanics.

The Hamiltonian action function is defined by

$$S(q_i, t) = \int_{t_0}^{t} L \, \mathrm{d}t.$$

The values of S are calculated along the path of integration as a function of time and the endpoints of the path. The Hamilton–Jacobi equation

$$\frac{\partial S}{\partial t} + H\left(q_i, \frac{\partial S}{\partial q_i}\right) = 0$$

is satisfied by the action function $S(q_i,t)$. Conversely, the solution of this equation, that is, the determination of $S = S(q_i,t)$, leads to the equation of motion of the system.

To solve the equation, we begin with the provisional solution

$$S = -Et + S(q_i).$$

Then the Hamilton–Jacobi equation assumes the simpler form

$$H\left(q_i, \frac{\partial S}{\partial q_i}\right) = E,$$

and we then have

$$\frac{\partial S}{\partial t} = -E.$$

We now introduce the hypersurfaces $S(q_i,t)$ = constant for various times in the f-dimensional configuration space of the positional coordinates q_i and observe that these hypersurfaces propagate like the equiphase surfaces of a wave process; indeed, one can specify the local speed of propagation (along the normals to the surfaces):

$$v = \frac{|\partial S / \partial t|}{|\text{grad}\, S|}.$$

The significance of all this can be seen in the fact that an analogy can be drawn between optical and mechanical processes. A light wave, for example, propagating in a non-homogeneous medium is described by the function

$$\varphi(x,y,z,t) = Ae^{j[-2\pi\nu t + S(x,y,z)]}.$$

Therefore, the Hamiltonian action function does, in fact, correspond to the phase function of a light wave. The family of surfaces $S(x,y,z)$ = constant was given the name "eikonal" by Bruns in 1895 (εἰκών = image).

In this way, Hamilton cleared the path for Schrödinger's wave mechanics (see Section 5.3.9).

Jacobi and Poisson, in contrast, posed and answered another question, that of how one might introduce, by suitable transformations $P = P(p, q)$, $Q = Q(p, q)$, the new canonical conjugate coordinates P, Q that again satisfy the canonical Hamiltonian equations

$$\dot{P}_i = -\frac{\partial H(P_i,\dot{Q}_i)}{\partial Q_i}, \quad \dot{Q}_i = -\frac{\partial H(P_i,Q_i)}{\partial P_i}.$$

For this to be achieved, the transformations $P = P(p, q)$, $Q = Q(p, q)$, written in the form of the *Poisson–Jacobi bracket*, must satisfy the following condition:

$$[P,Q] = \begin{vmatrix} \frac{\partial P}{\partial p} & \frac{\partial P}{\partial q} \\ \frac{\partial Q}{\partial p} & \frac{\partial Q}{\partial q} \end{vmatrix} = \frac{\partial P}{\partial p}\frac{\partial Q}{\partial q} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial q} = 1.$$

Poisson and Jacobi also proved that, in general, the time rate of change of an arbitrary function $F(p, q, t)$ is given by the equation

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial F}{\partial t} + [H,F].$$

If F depends on $p = p(t)$ and $q = q(t)$ but not explicitly on the time, then this equation simplifies to $\dot{F} = [H,F]$.

Heisenberg, Born, Jordan, and Dirac used these relationships to forge the link between quantum mechanics and classical mechanics.

4.3 The Century of Light

4.3.1 *The Enlightenment*

The eighteenth century is generally known as the Age of Enlightenment or the Age of Reason, and sometimes also as the Age of the Philosophers, and was dominated by French culture and the French language: *siècle des lumières, siècle de la raison, siècle des philosophes*. Earlier, we saw that a great breakthrough occurred in the seventeenth century, which we called therefore the century of genius. In mechanics, it remained for the eighteenth century to refine the great and rough ideas, with Euler as the central personality, even though he was not himself French. It is true, nevertheless, that the Frenchman Lagrange also had a lion's share in casting theo-

Quotation 4.17, ***continued***

that this customary transition is, therefore, the same with the power and necessity; which are consequently qualities of perceptions, not of objects, and are internally felt by the soul, and not perceiv'd externally in bodies? ...

But tho' this be the only reasonable account we can give of necessity, the contrary notion is so riveted in the mind from the principles above-mention'd, that I doubt not but my sentiments will be treated by many as extravagant and ridiculous. What! the efficacy of causes lies in the determination of the mind! As if causes did not operate entirely independent of the mind, and wou'd not continue their operation, even tho' there was no mind existent to contemplate them, or reason concerning them. Thought may well depend on causes for its operation, but not causes on thought. This is to reverse the order of nature, and make that secondary, which is really primary. To every operation there is a power proportion'd; and this power must be plac'd on the body, that operates. If we remove the power from one cause, we must ascribe it to another: But to remove it from all causes, and bestow it on a being, that is no ways related to the cause or effect, but by perceiving them, is a gross absurdity, and contrary to the most certain principles of human reason. ...

As to what may be said, that the operations of nature are independent of our thought and reasoning, I allow it; and accordingly have observ'd, that objects bear to each other the relations of contiguity and succession; that like objects may be observ'd in several instances to have like relations; and that all this is independent of, and antecedent to the operations of the understanding. But if we go any further, and ascribe a power or necessary connexion to these objects; this is what we can never observe in them, but must draw the idea of it from what we feel internally in contemplating them. ...

'Tis now time to collect all the different parts of this reasoning, and by joining them together form an exact definition of the relation of cause and effect, which makes the subject of the present enquiry. ...

There may be two definitions given of this relation, which are only different, by their presenting a different view of the same object, and making us consider it either as a *philosophical* or as a *natural* relation; either as a comparison of two ideas, or as an association betwixt them. We may define a *cause* to be "An object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedency and contiguity to those objects, that resemble the latter." If this definition be esteem'd defective, because drawn from objects foreign to the cause,

continued on next page

Quotation 4.17, ***continued***

we may substitute this other definition in its place, *viz*. "A *cause* is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other." Shou'd this definition also be rejected for the same reason, I know no other remedy, than that the persons, who express this delicacy, shou'd substitute a juster definition in its place. But for my part I must own my incapacity for such an undertaking.

—David Hume, *A Treatise of Human Nature* [pp. 111–114]

retical mechanics into its present form. In philosophy, two Englishmen—Berkeley and Hume—led the empirical school of Hobbes and Locke into a dead end. George Berkeley (1685–1753) shifted the emphasis in cognitive processes from a material world independent of the observer to a world of sensory impressions immanent in consciousness and granted a real existence only to the latter of these two worlds. He thus represented the subjectively idealistic viewpoint that objects exist only to the extent that they are perceived by us (*esse est percipi*). He escaped from solipsism only by arguing as follows: The objects of the world exist, even if I do not perceive them, since God is always thinking about them. In this way, his philosophy assumes an objectively idealistic air.

One of the most astute and critical minds in the history of philosophy, David Hume (1711–1776), accepted the objective existence of the external world as the source of all experience and perception. Yet he emphasized that all sensory impressions can at best serve as a guide to our daily lives because the senses do not provide a reliable picture of the actual relationships in the external world. Thus, for example, even the cause-and-effect relationship is just a habit formed by repetition and the result of the logical fallacy *post hoc, ergo propter hoc* (following this, therefore because of this). Hume's agnosticism awakened Kant from his "dogmatic slumber" and spurred him to complete the "Copernican Revolution." This became the starting point for nearly all philosophical movements of the nineteenth century (Quotation 4.17).

Would it not be proper for us, following these considerations, to view the *lumière* of the Enlightenment as a will-o'-the-wisp? Did the representatives of this epoch or their posterity perhaps gild this period with false brilliance? We would like to demonstrate that if the seventeenth century was the century of genius, then the eighteenth century is in and of itself the great genius (Table 4.1, Figure 4.49).

Let us begin by quoting Kant's own definition of the word "enlightenment":

> Enlightenment is mankind's exit from its self-incurred immaturity. Immaturity is the inability to make use of one's own understanding without the guidance of another. Self-incurred is this ability if its cause lies not in the lack of understanding, but rather in the lack of resolution and the courage to use it without the guidance of another. Sapere aude! Have the courage to use your own understanding! is thus the motto of enlightenment.

Dare to think! In this century, the thinkers did indeed dare to trust their reason; they believed in common sense, in nature, and in its eternal laws. Man, nature, reason, and experience are the crucial concepts that made a new worldview possible. This worldview is universal, and it explains all phenomena with immanent causes within its closed circle (Section 4.3.2).

They believed in reason, though with the awareness that reason is not infallible. They knew that most trouble is caused by fanatics, by those who "wish to place their own chains on the whole world"; therefore tolerance is needed for those who think differently.

If we could say no more but that, in this century, the achievements of the previous century were systematized, that the scope of application of the methods already worked out was enlarged, and that scientific progress was made accessible to a wide circle of intellectually active people, in other words, if the eighteenth century had served only a pedagogically informing and educational function, even then it would have been deserving of our admiration. "If I have seen farther, it is by standing on the shoulders of giants" holds true to be sure, but someone had to lift us up onto those shoulders, and that is exactly what their pedagogy did.

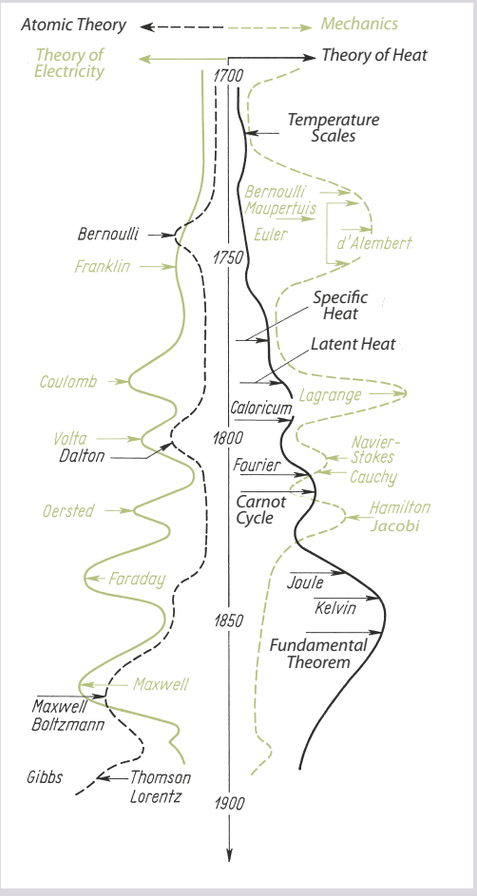
However, in order to temper a possible one-sidedness in our evaluation of the eighteenth century resulting from our admiration, let us recall Goethe's words of warning:

> That the history of the world must be rewritten from time to time is a fact about which there is no doubt in our day. However, such a necessity arises not because much that has happened has been finally discovered, but because new points of view have been given by those of an advanced time who have been led to points of observation where that past can be viewed and evaluated in a new way. It is the same with the sciences. Not only the discovery of previously unknown relationships and objects in nature, but also the advancing beliefs and opinions change much and are worth taking note of from time to time. I would single out the passing eighteenth century as a period that must be examined in this regard. In its great accomplishments it revealed many shortcomings and did manifold injustice to the previous centuries, particularly the less educated. In this sense one can call it self-wise in that it is quite proud of a certain clear consistency and became accustomed to judging everything according to a very clear standard. Excessive doubt and decisive denial alternated with each other to bring forth one and the same effect: an arrogant self-satisfaction and a rejection of all that cannot be immediately obtained or understood.
>
> Where is respect for lofty unreachable goals to be found? Where the feeling for a seriousness that sinks itself into boundless depths? How rare is forbearance for bold unsuccessful effort! How rare patience for what takes a long time! This is not the place to investigate whether the lively Frenchman or the dry German is more at fault and to what extent both mutually contribute to this widespread tone. One may look through the works, notebooks, papers in which shorter or longer reports of the lives of learned men, their character and works, are given; one may look through dictionaries, libraries, necrologies, and seldom find that a problematic nature is represented with thoroughness and fairness. Indeed, one comes to the aid of decent persons of earlier times by attempting to free them of the suspicion of magic; but now it would again be necessary that one look after them in some other way and free them again from the hands of such exorcists who, to drive away the ghosts, make it a sacred duty to drive away the spirit itself.
>
> —Goethe, *Geschichte der Farbenlehre* (History of Color Theory)

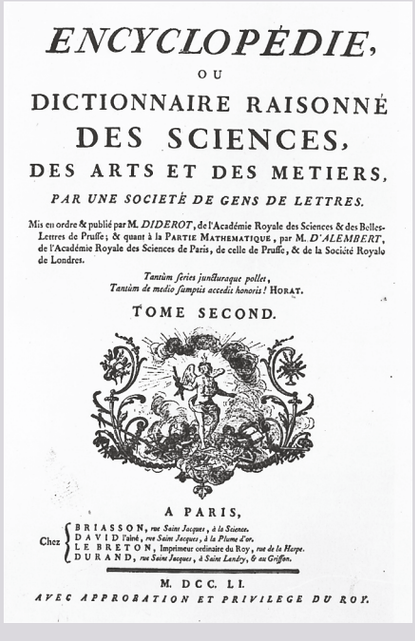
4.3.2 *The Great Encyclopedia*

History warns of two pitfalls in philosophical thought: belief in authority and the passion to construct philosophical systems. The encyclopedists therefore allowed no authority to hold sway and constructed no system; instead, they analyzed all current knowledge and presented the results one after another. Thus, there arose out of details gathered from many sources the monumental work of the century, the *Encyclopédie*. The *Encyclopédie* represents no philosophical system in the old sense of the term, yet it does present a unity that is expressed in the comportment of a newly self-confident citizenry.

If you were to pick up the *Encyclopédie* originally published by Diderot and d'Alembert (Figures 4.50, 4.51), you would find topics and a tone in every article that even today would be accepted as normal among intellectuals throughout the world. The encyclopedists would be our equals in debating epistemology, morality, religion, human rights, equality, and humanity. Most of the time, we even share their somewhat naive belief in the omnipotence of reason (Figure 4.52).

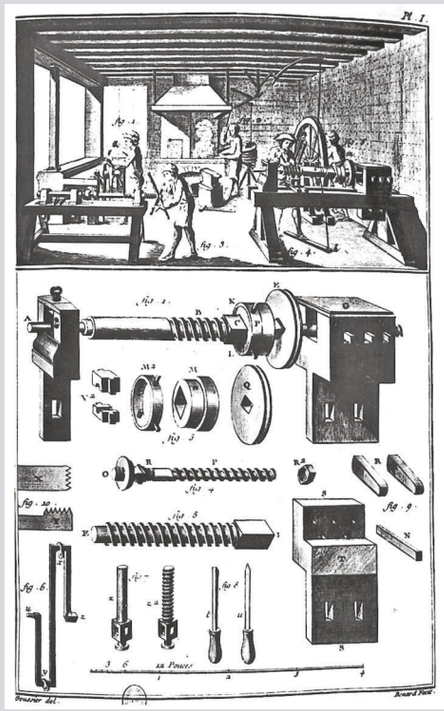


▲ **Figure 4.49** The relative intensity of development in four branches of classical physics.


ENCYCLOPÉDIE,
OU
DICTIONNAIRE RAISONNÉ
DES SCIENCES,
DES ARTS ET DES METIERS,
PAR UNE SOCIETÉ DE GENS DE LETTRES.
Mis en ordre & publié par M. DIDEROT, de l'Académie Royale des Sciences & des Belles-Lettres de Prusse; & quant à la PARTIE MATHEMATIQUE, par M. D'ALEMBERT, de l'Académie Royale des Sciences de Paris, de celle de Prusse, & de la Société Royale de Londres.
Tantùm series juncturaque pollet,
Tantùm de medio sumptis accedit honoris! HORAT.
TOME SECOND.
A PARIS,
Chez BRIASSON, rue Saint Jacques, à la Science.
DAVID l'aîné, rue Saint Jacques, à la Plume d'or.
LE BRETON, Imprimeur ordinaire du Roy, rue de la Harpe.
DURAND, rue Saint Jacques, à Saint Landry, & au Griffon.
M. DCC. LI.
AVEC APPROBATION ET PRIVILEGE DU ROY.

◀ **Figure 4.50** The great French *Encyclopédie* is one of the most influential series of books in history. It is above all the fruit of French genius in the eighteenth century and a symbol of the Age of Enlightenment. People from a great variety of social positions and points of view were engaged in its creation, even the royal censor de Malesherbes. Among the collaborators we find such names as Voltaire, Rousseau, d'Alembert, Holbach, Helvétius,

continued on next page



▲ **Figure 4.50** *continued*

Montesquieu, Quesnay, and Buffon. However, the *Encyclopédie* never would have appeared without the enthusiasm, patience, organizational ability, and courage of Denis Diderot (Figure 4.52).

The publisher, le Breton, had originally planned on translating a popular English work, Ephraim Chambers's *Cyclopaedia*. However, Diderot advised against this idea. He managed to secure the cooperation first of d'Alembert, then of the entire French intellectual elite, some of whose names are listed above. The first and second volumes of the work, projected to run to eight volumes, appeared in 1751 and 1752. After attacks, prohibition, and subsequent approval, by 1757 the seventh volume had already appeared. After an assassination attempt on the king, printing of further volumes and the distribution of the work were prohibited. In the somewhat defused atmosphere after the expulsion of the Jesuits, the concluding text-only volumes 8 through 17, which had been printed by 1762, were distributed—in secrecy—to the 4250 subscribers. By 1772, the 11 additional volumes of illustrations were also completed. All in all, with its five supplementary and two index volumes, the work consists of 35 volumes.

Engineers and craftspeople can be delighted in how the *Encyclopédie* appreciates their activities. On the very title page, the arts and crafts are listed on equal terms with the sciences (Figures 4.53, 4.54).

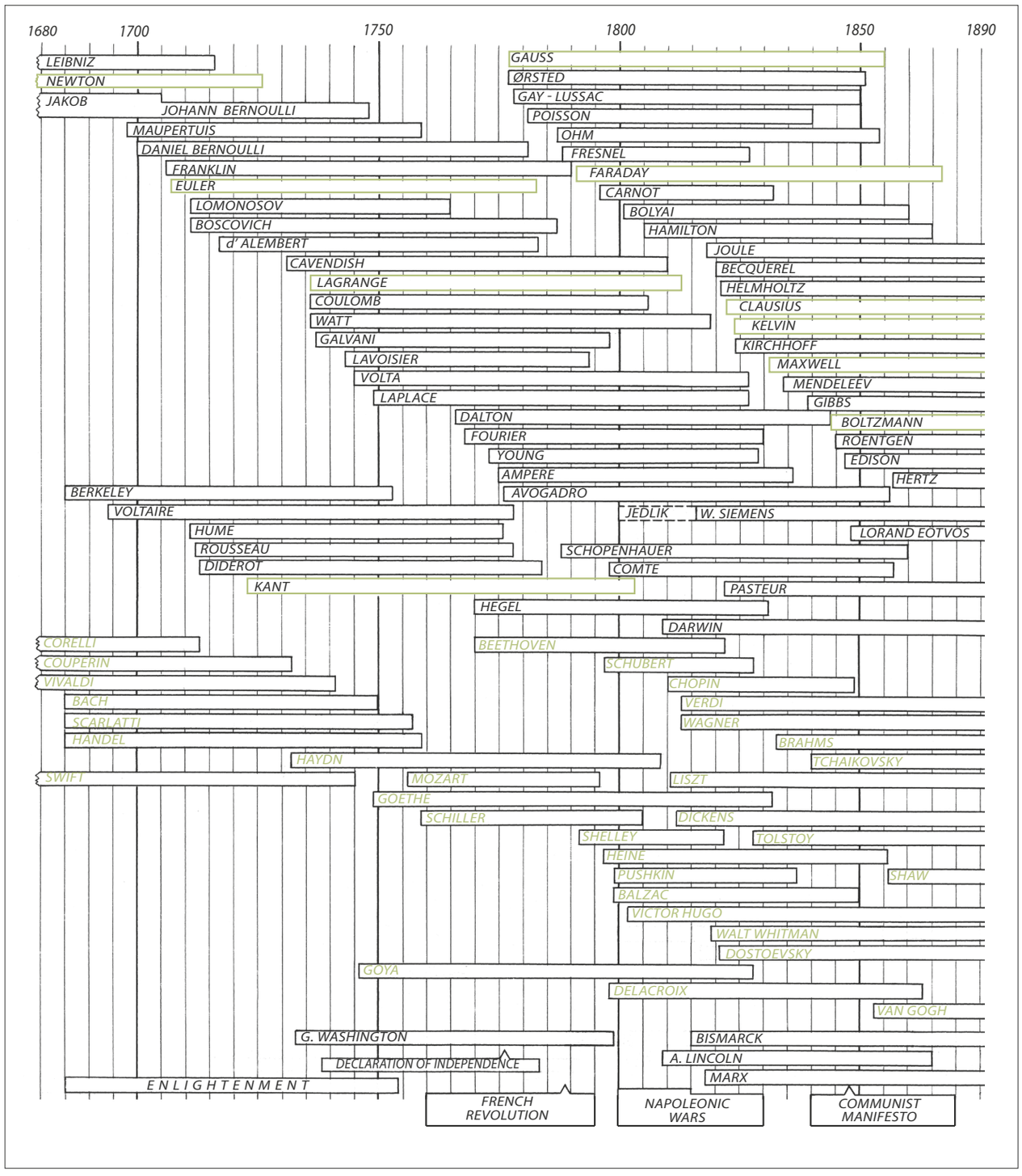
> But while justly respecting great geniuses for their enlightenment, society ought not to degrade the hands by which it is served. The discovery of the compass is no less advantageous to the human race than the explication of its properties would be to physical science. Finally, considering in itself the principle of the distinction about which we are speaking, how many alleged scholars are there for whom science is in truth only a mechanical art? What real difference is there between a head stuffed with facts without order, without utility, and without connection, and the instinct of an artisan reduced to mechanical operation?
> The contempt in which the mechanical arts are held seems to have affected to some degree even their inventors. The names of these benefactors of humankind are almost all unknown, whereas the history of its destroyers, that is to say, of conquerors, is known to everyone. However, it is perhaps in the artisan that one must seek the most admirable evidences of the sagacity, the patience, and the resources of the mind.

D'Alembert delights physicists by giving a place to Descartes and Newton in the history of philosophy, a place that even today we would like to see restored.

The most important innovation of the *Encyclopédie* is perhaps that in it science and philosophy appear for the first time in a new role: The world needs to be explained, but it needs to be changed as well. For the time being, such change should proceed in the direction guided by the unchangeable laws that are discovered by reason. A direction extended to history and class consciousness will be heralded one hundred years later by the dialectic and historical materialism postulating historical development as universal law.

4.3.3 D'Alembert: Preface to the Encyclopedia

> While poorly instructed or badly intentioned adversaries made open war on it, philosophy sought refuge, so to speak, in the works of a few great men. They had not the dangerous ambition of removing the blindfolds from their contemporaries' eyes; yet silently in the shadows they prepared from afar the light which gradually, by imperceptible degrees, would illuminate the world.
>
> The immortal Chancellor of England, Francis Bacon [1561–1626], ought to be placed at the head of these illustrious personages. His works, so justly esteemed (and more esteemed, indeed, than they are known), merit our reading even more than our praises. One would be tempted to regard him as the greatest, the most universal, and the most eloquent of the philosophers, considering his sound and broad views, the multitude of objects to which his mind turned itself, and the boldness of his style, which everywhere joined the most sublime images with the most rigorous precision. Born in the depths of the most profound night, Bacon was aware that philosophy did not yet exist, although many men doubtless flattered themselves that they excelled in it (for the cruder a century is, the more it believes itself to be educated in all that can be known). Therefore, he began by considering generally the various objects of all the natural sciences. He divided these sciences into different branches, of which he made the most exact enumeration that was possible for him. He examined what was already known concerning each of these objects and made the immense catalogue of what remained to be discovered. This is the aim of his admirable book *The Advancement of Learning*. In his *Novum Organum*, he perfects the views that he had presented in the first book, carries them further, and makes known the necessity of experimental physics, of which no one was yet aware. Hostile to systems, he conceives of philosophy as being only that part of our knowledge which should contribute to making us better or happier, thus apparently confining it within the limits of the science of useful things, and everywhere he recommends the study of Nature. His other writings were produced on the same pattern. Everything, even their titles, proclaims the man of genius, the mind that sees things in the large view. He collects facts, he compares experiments and



◀ Table 4.1
The builders of classical physics.

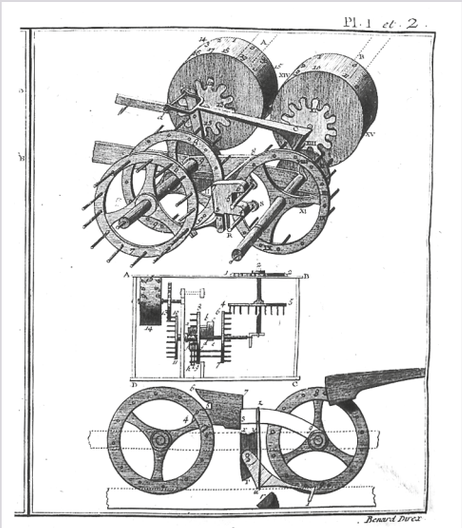
In the seventeenth century, physics and philosophy were closely intertwined. In the eighteenth century, in the spirit of the *Encyclopédie*, a place was also made for "handicrafts" and for technology, and finally, in the nineteenth century, the connection between physics and technology (Plate XX) was made even tighter in the attempt to further decouple physics from philosophy. The nineteenth century is the century of the great inventors. In the table, we encounter the extreme types: The Hungarian Ányos Jedlik (1800–1895) had constructed an electric motor as early as 1828, and in 1861, he discovered the principle of the self-energizing dynamo; he was modest and worked in relative seclusion. Today we celebrate Werner von Siemens (1816–1892) as the inventor of the dynamo, for it was he who recognized the economic significance of his apparatus, constructed in 1867. Thomas Alva Edison (1847–1931) was also not only an inventor, but at the same time, a research director and entrepreneur, paying particular attention to the economic value of his discoveries. We have Edison to thank for the electric light, electrical distribution networks, the phonograph, and the nickel–iron battery, as well as many other devices.

Over the course of the eighteenth century, the static worldview of the seventeenth century was rejected, and in the nineteenth century, attention began to be paid to dynamics, motion, and change, that is, to continuing development and evolution. Thus the book *Systema Naturae* (1735–1788) by Linnaeus (1707–1778) was merely a taking stock of static forms, while Charles Darwin (1809–1882), with his *On the Origin of Species* (1859) and *The Descent of Man* (1871), assigned to mankind its proper place in the developmental hierarchy, building on work of famous precursors such as his grandfather Erasmus Darwin, Jean Baptiste de Lamarck (1744–1829), and even Kant.

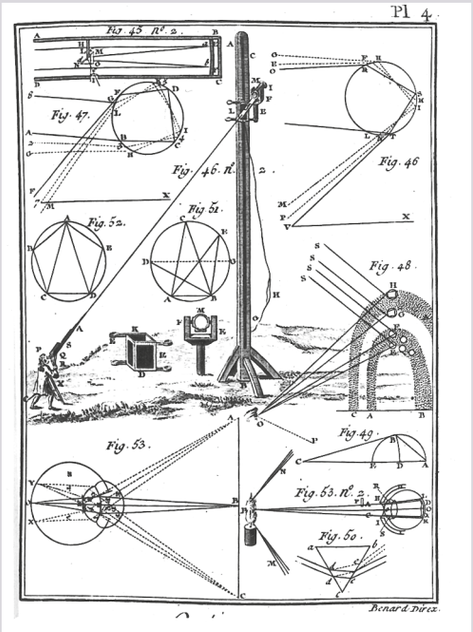
In economics, the classicists such as Adam Smith (1723–1790) (*An Inquiry into the Nature and Causes of the Wealth of Nations*, 1776) were still seeking immutable laws, but others such as Karl Marx (1818–1883) were searching for the driving forces of change and progress.

The work of Louis Pasteur (1822–1895), as well as that of his younger contemporary and rival Robert Koch (1843–1910), is of epochal importance in demonstrating the role of microbes in disease and the development of cures for such diseases.

Although his name does not appear in the table, we mention here Gregor Mendel (1822–1884), who formulated the quantitative laws of inheritance, thus laying the groundwork for one of the signature sciences of the twentieth century. His forgotten work was rediscovered around 1900. Thus arose a new science—genetics. In a significant development, in 1953, James Watson and Francis Crick described the spatial structure of the carrier of genetic information, DNA (deoxyribonucleic acid), as a double helical spiral.



(a)



(b)

▲ **Figure 4.51** Two typical illustrations from the great French *Encyclopédie*: **(a)** construction diagram for Pascal's calculating engine and **(b)** explanatory figures under the rubric "Optics."

points out a large number to be made; he invites scholars to study and perfect the arts, which he regards as the most exalted and most essential part of human science; he sets forth with a noble simplicity his *Conjectures and Thoughts* on the different objects worthy of men's interest; and he would have been able to say, like that old man in Terence, that nothing which touches humanity was alien to him. Natural science, ethics, politics, economics, all seem to have been within the competence of that brilliant and profound mind. And we do not know which we ought to admire more, the riches he lavishes upon all the subjects he treats or the dignity with which he speaks of them. His writings can best be compared to those of Hippocrates on medicine; and they would be no less admired, nor less read, if the culture of the mind were as dear to mankind as the conservation of health. But in every area only the works of those who head a school of disciples make a brilliant impression. Bacon was not of that number, and the form of his philosophy prevented it; it was too wise to astonish anyone. Scholasticism, which continued to dominate, could not be overthrown except by bold and new opinions. And apparently circumstances are not such that a philosopher who is content to say to men: "Here is the little that you have learned, there is what remains for you to find," is destined to cause much stir among his contemporaries. …

Chancellor Bacon was followed by the illustrious Descartes [1596–1650]. That exceptional man, whose fortune has varied so much in less than a century, possessed all the qualities necessary for changing the face of philosophy: a strong imagination, a most logical mind, knowledge drawn from himself more than from books, great courage in battling the most generally accepted prejudices, and no form of dependence which forced him to spare them. …

As a philosopher he [Descartes] was perhaps equally great, but he was not so fortunate. Geometry, which by the nature of its object always gains without losing ground, could not fail to make a progress that was most sensible and apparent for everyone when plied by so great a genius. Philosophy found itself in quite a different state. There everything remained to be done, and what do not the first steps in any branch of knowledge cost? One is excused from making larger steps by the merit of taking any at all. If Descartes, who opened the way for us, did not progress as far along it as his sectaries [apologists] believe, nevertheless the sciences are far more indebted to him than his adversaries will allow; his method alone would have sufficed to render him immortal. His *Dioptrics* is the greatest and the most excellent application that has yet been made of geometry to physics. In other words, we see his inventive genius shining forth everywhere, even in those works which are least read now. If one judges impartially those vortices which today seem almost ridiculous, it will be agreed, I daresay, that at that time nothing better could be imagined. The astronomical observations which served to destroy them were still imperfect or hardly established. Nothing was more natural than to postulate a fluid which carried the planets. Only a long sequence of phenomena, of reasonings, and of calculations, and consequently a long sequence of years, could cause such an attractive theory to be renounced. Moreover, it had the singular advantage of explaining gravitation of bodies by the centrifugal force of the vortex itself, and I am not afraid to assert that this explanation of weight is one of the finest and most ingenious hypotheses that philosophy has ever imagined. Thus, physicists had to be carried forward almost in spite of themselves by the theory of central forces and by experiments made much later before they would abandon it. Let us recognize, therefore, that Descartes, who was forced to create a completely new physics, could not have created it better; that it was necessary, so to speak, to pass by way of the vortices in order to arrive at the true system of the world; and that if he was mistaken concerning the laws of movement, he was the first, at least, to see that they must exist.

His metaphysics, as ingenious and new as his physics, suffered virtually the same fate. And it too can be justified by more or less the same reasons; for such is the fortune of that great man today, that after having had innumerable disciples, he is reduced to a handful of apologists. No doubt he was mistaken in admitting the existence of innate ideas. But had he retained that single truth taught by the Aristotelians concerning the origin of ideas through the senses, perhaps it would have been difficult to uproot the errors that debased this truth by being alloyed with it. Descartes dared at least to show intelligent minds how to throw off the yoke of scholasticism, of opinion, of authority—in other words, of prejudices and barbarism. And by that revolt whose fruits we are reaping today, he rendered a service to philosophy perhaps more difficult to perform than all those contributed thereafter by his illustrious successors. He can be thought of as a leader of conspirators who, before anyone else, had the courage to arise against a despotic and arbitrary power and who, in preparing a resounding revolution, laid the foundations of a more just and happier government, which he himself was not able to see established. If he concluded by believing he could explain everything, he at least began by doubting everything, and the arms which we use to combat him belong to him no less because we turn them against him. Moreover, when absurd opinions are of long duration, one is sometimes forced to replace them by other errors, if one cannot do better, in order to disabuse the human



race. The uncertainty and vanity of the mind are such that it always needs an opinion to which to cleave. It is a child to whom one must offer a toy in order to take a dangerous weapon away from him. He will quit his toy by himself when he reaches the age of reason. In thus deceiving the philosophers, or those who believe themselves such, one teaches them at least to distrust their intelligence, and that frame of mind is the first step toward truth. Thus was Descartes persecuted in his own lifetime, as if he had come to bring truth to men.

Newton [1642–1727], whose way had been prepared by Huygens, appeared at last, and gave philosophy a form which apparently it is to keep. That great genius saw that it was time to banish conjectures and vague hypotheses from physics, and that this science was uniquely susceptible to the experiments of geometry. It was perhaps with this aim that he began by inventing calculus and the method of series, whose applications are so extensive in geometry itself and still more so in explaining the complicated effects that one observes in Nature, where everything seems to take place by various kinds of infinite progressions. The experiments on weight and the observations of Kepler led the English philosopher to discover the force which holds the planets in their orbits. Simultaneously he showed how to distinguish the causes of their movements and how to calculate them with a precision such as one might reasonably expect only after several centuries of labor. Creator of an entirely new optics, he made the properties of light known to men by breaking it up into its constituent parts. Anything we could add to the praise of the great philosopher would fall far short of the universal testimonial that is given today to his almost innumerable discoveries and to his genius, which was at the same time far-reaching, exact, and profound. He has doubtless deserved all the recognition that has been given him for enriching philosophy with a large quantity of real assets. But perhaps he has done more by teaching philosophy to be judicious and to restrict within reasonable limits the sort of audacity which Descartes had been forced by circumstances to bestow upon it. His Theory of the World (for I do not mean his System) is today so generally accepted that men are beginning to dispute the author's claim to the honor of inventing it (because at the beginning great men are accused of being mistaken, and at the end they are treated as plagiarists). To the people who find everything in the work of the ancients I leave the pleasure of discovering gravitation of planets in those works, even if it is not there. But even supposing that the Greeks had had the idea of it, what was only a rash and romantic

▲ **Figure 4.52** The "board of directors" of the encyclopedists. Sitting across from the viewer with raised arm is Voltaire; to his right, Diderot, and to his left, Helvétius. In the right half of the figure, we see Baron Holbach (wearing a hat); with his back to the viewer sits Condorcet; to the left, next to him, d'Alembert, and then Turgot.

We shall speak elsewhere about Voltaire and Diderot. Claude-Adrien Helvétius did not author a single entry in the encyclopedia; however, his intellectual weight qualified him to be counted among the encyclopedists. On his work *De l'Esprit* (Concerning the Spirit), Russell quotes the contemporary Bentham:

> What Bacon was to the physical world, Helvétius was to the moral. The moral world had its Bacon but its Newton is yet to come.

According to Locke, our minds are at birth a *tabula rasa*, and education determines what will be written on this blank page. Helvétius believed in the omnipotence of education, to which all the influences of the environment contribute. His ethics is based on an egoism, interpreted in a positive ("equipped with far-sightedness") sense. Marquis de Condorcet (1743–1794), a protégé of d'Alembert, was, as a mathematician, elected to the *Académie* in his youth. The second edition of his *Éléments du calcul des probabilités et son application aux jeux de hasard, à la loterie et aux*

continued on next page

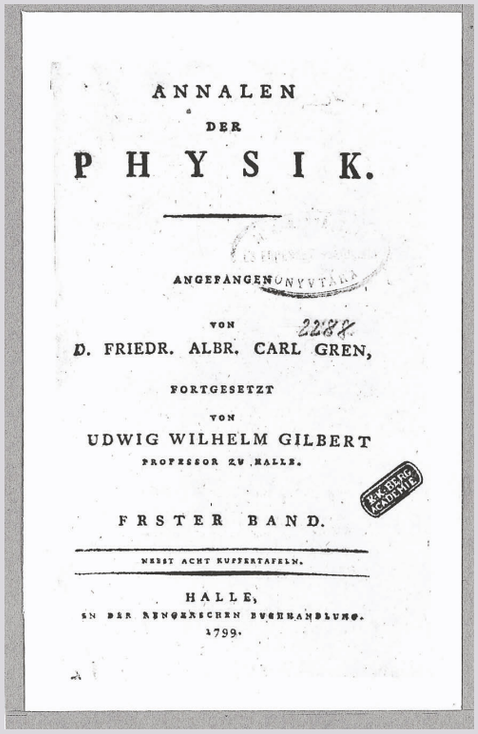
Figure 4.52 *continued*

jugements des hommes, which appeared after his death, in 1805, is of significance not only for mathematics, but also for sociology. Condorcet was an active participant in the French Revolution. In 1792, he presented a far-reaching plan for a reform of the educational system: elementary schools in every village, and universities in every département; all schools under state jurisdiction, obligatory and tuition-free; selection according to ability; educational opportunities for girls and continuing education for adults. His ideas were in their essence adopted in France 100 years later. To escape the terror of the revolution, he went into hiding, and during that time, he wrote his book *Sketches for a Historical Picture of the Progress of the Human Mind*. He believed that the inequalities between nations and classes would disappear and that there was no limit—intellectual, moral, or physical—to the degree to which mankind can be perfected and that mankind would indeed someday become perfect.

Condorcet anticipated Malthus's theory on population growth, but believed that with regulation of births, a solution could be found. Condorcet did not live to see the appearance of his book, suffused with a belief in the power of reason and predicting a bright future. In tragic counterpoint to his writings, he died in prison—probably committed suicide—after his arrest.

Baron Holbach (1723–1789) came from a family of the German nobility; however, France was his true home. The audacity of his 1770 work *Le Système de la Nature* distressed even his philosopher colleagues.

Turgot (1727–1781) was briefly finance minister under Louis XVI; he is best known for his economic theories and his attempts at reform.


ANNALEN
DER
PHYSIK.
ANGEFANGEN
VON
D. FRIEDR. ALBR. CARL GREN,
FORTGESETZT
VON
UDWIG WILHELM GILBERT
PROFESSOR ZU HALLE.
ERSTER BAND.
MEBST ACHT KUPFERTAFELN.
HALLE,
IN DER RENGERSCHEN BUCHHANDLUNG.
1799.

▲ **Figure 4.53** At the end of the eighteenth century, the first technical dictionaries and encyclopedias appeared, such as Fischer's *Physikalische Wörterbuch* shown here. The work

continued on next page

system with them became a demonstration in the hands of Newton. That demonstration, which belongs exclusively to him, constitutes the real merit of his discovery. Without such support, the theory of attraction would only be a hypothesis like so many others. If today some celebrated writer takes it into his head to predict without any proof that one day we will succeed in making gold, would our descendants be justified in trying on this pretext to deprive a chemist who actually achieved it of the glory of his great work? And does the invention of telescopes belong any less to its authors, even though some ancient may have thought it possible that we would one day extend the range of our vision? …

Locke [1632–1704] undertook and successfully carried through what Newton had not dared to do, or perhaps would have found impossible. It can be said that he created metaphysics, almost as Newton had created physics. He understood that the abstractions and ridiculous questions which had been debated up to that time and which had seemed to constitute the substance of philosophy were the very part most necessary to proscribe. He sought the principal causes of our errors in those abstractions and in the abuse of signs, and that is where he found them. In order to know our soul, its ideas, and its affections, he did not study books, because they would only have instructed him badly; he was content with probing deeply into himself, and after having contemplated himself, so to speak, for a long time, he did nothing more in his treatise, *Essay Concerning Human Understanding* [1690], than to present mankind with the mirror in which he had looked at himself. In other words, he reduced metaphysics to what it really ought to be: the experimental physics of the soul—a very different kind of physics from that of bodies, not only in its object, but in its way of viewing that object. In the latter study we can, and often do, discover unknown phenomena. In the former, facts as ancient as the world exist equally in all men; so much the worse for whoever believes he is seeing something new. Reasonable metaphysics can only consist, as does experimental physics, in the careful assembling of all these facts, in reducing them to a corpus of information, in explaining some by others, and in distinguishing those which ought to hold the first rank and serve as the foundation. In brief, the principles of metaphysics, which are as simple as axioms, are the same for the philosophers as for the general run of people. But the meager progress that this science has made for such a long time shows how rarely these principles are applied felicitously, whether because of the difficulty that surrounds such an application, or perhaps also because of the natural temptations that prevent us from holding ourselves within bounds when we engage in metaphysical speculations. Nevertheless, the title of metaphysician is still rather common in our century, for we love to squander everything. But how few persons are really worthy of that name! How many are there who earn it only by the unfortunate talent of obscuring clear ideas with a great deal of subtlety and of preferring their own extraordinary notions to true ones which are always simple? One should not be astonished, therefore, if the majority of those who are called metaphysicians have so little regard for one another. I have no doubt that this title will soon become an insult for our men of intelligence, as the name "sophist," degraded by those who bore it in Greece, was rejected by true philosophers, even though it means "a sage."

We may conclude from all this history that England is indebted to us for the origins of that philosophy which we have since received back from her. It is perhaps a greater distance from substantial forms to vortices than from vortices to universal gravitation; just as perhaps a greater interval exists between pure algebra and the idea of applying it to geometry than between the small triangle of Barrow and differential calculus.

Such are the principal geniuses that the human mind ought to regard as its masters. Greece would have raised statues to them, even if she had been obliged to tear down those of a few conquerors in order to give them room.

—d'Alembert, *Preliminary Discourse to the Encyclopedia of Diderot* [pp. 74–76, 77–85]

4.3.4 *Belief in the Solid Foundation of Physics: Kant*

We come into this world with no inborn ideas: All our knowledge comes to us mediated by sensory experience from the outside world—thus we are told by Locke. And what guarantees the reliability of this knowledge, the inevitability of the truths distilled from experience? Nothing, answers Hume, since all of our claims—for example, those regarding causes and effects—are nothing but beliefs resulting from our having become accustomed to the world, and these beliefs reflect more on our senses than on our intellect. The skeptical Scotsman Hume wakened the Königsberg recluse Immanuel Kant (Figure 4.55) from his dogmatic slumber. Indeed, even today we read Hume's stern words with a slight pang of conscience:

> When we run over libraries, persuaded of these principles, what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: For it can contain nothing but sophistry and illusion.

The reliability of our scientific discoveries is guaranteed by the inner structure of our thought; thus might one briefly summarize KANT's counterargument.

KANT's extensive presentation of this idea fills almost one thousand pages of his *Critique of Pure Reason*. "We can call pure reason the principles whereby we know anything *a priori*; the critique of pure reason is therefore the investigation of the possibilities and limits of the application of pure reason."

It is, of course, fitting that we give the question to be answered in Kant's own words:

> There can be no doubt that all our knowledge begins with experience. ... In the order of time, therefore, we have no knowledge antecedent to experience, and with experience all our knowledge begins.
>
> But though all our knowledge begins with experience, it does not follow that it all arises out of experience. ...
>
> This, then, is a question which at least calls for closer examination and does not allow of any off-hand answer—whether there is any knowledge that is thus independent of experience and even of all impressions of the senses. Such knowledge is entitled *a priori*, and distinguished from the empirical, which has its sources *a posteriori*, that is, in experience. [pp. 41–42]

To approach the fundamental problem, we must learn to distinguish analytic and synthetic judgments.

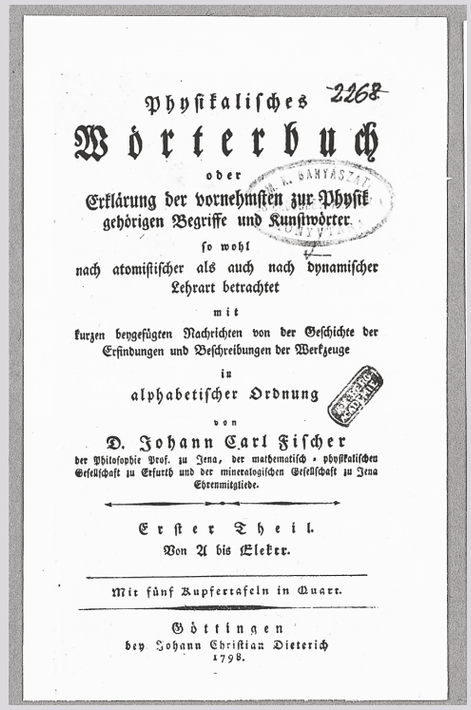
> *Analytic judgments* (affirmative) are ... those in which the connection of the predicate with the subject is thought through identity; those in which this connection is thought without identity should be entitled *synthetic* [*judgements*. We may call the former explanatory (explicative) and the latter extending (ampliative) ...]. The former, as adding nothing through the predicate to the concept of the subject, but merely breaking it up into those constituent concepts that have all along been thought in it, although confusedly, can also be entitled explicative. The latter, on the other hand, add to the concept of the subject a predicate which has not been in any wise thought in it, and which no analysis could possibly extract from it; and they may therefore be entitled ampliative. [p. 48]

And here we see the key statement that will become the physicists' article of faith during the nineteenth century—especially for German physicists: *The tenets of natural sciences are a priori synthetic judgments.*

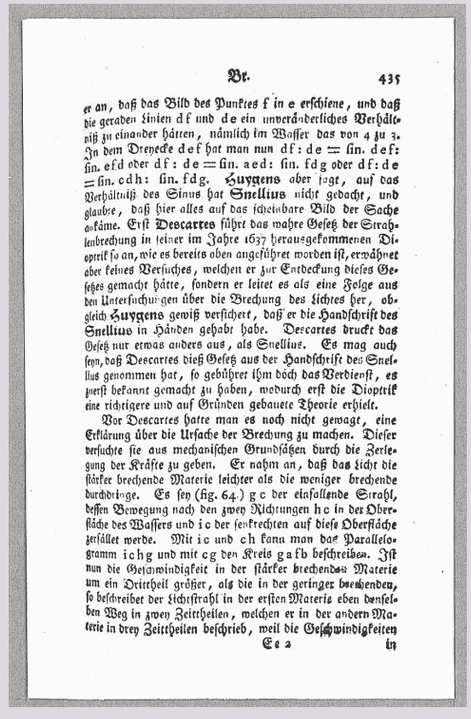
Now we are in a position to express the specific problem of pure reason:

> Much is already gained if we can bring a number of investigations under the formula of a single problem. For we not only lighten our own task, by defining it accurately, but make it easier for others, who would test our results, to judge whether or not we have succeeded in what we set out to do. Now the proper problem of pure reason is contained in the question: *How are a priori synthetic judgments possible?* ...
>
> In the solution of the above problem, we are at the same time deciding as to the possibility of the employment of pure reason in establishing and developing all those sciences which contain a theoretical a priori knowledge of objects, and have therefore to answer the questions:

▼ Figure 4.53 *continued*
is up-to-date and written on a high level; the historical summaries are worth reading even today. In the same period, in 1799, the first volume of the journal *Annalen der Physik* appeared, a journal that would see the publication of many important articles. (Library of the University for Heavy Industry, Miskolc.)



Physikalisches 2268
Wörterbuch
oder
Erklärung der vornehmsten zur Physik gehörigen Begriffe und Kunstwörter
so wohl
nach atomistischer als auch nach dynamischer Lehrart betrachtet
mit
kurzen beygefügten Nachrichten von der Geschichte der Erfindungen und Beschreibungen der Werkzeuge
in
alphabetischer Ordnung
von
D. Johann Carl Fischer
der Philosophie Prof. zu Jena, der mathematisch-physikalischen Gesellschaft zu Erfurth und der mineralogischen Gesellschaft zu Jena Ehrenmitgliede.
Erster Theil.
Von A bis Ekker.
Mit fünf Kupfertafeln in Quart.
Göttingen
bey Johann Christian Dieterich
1798.

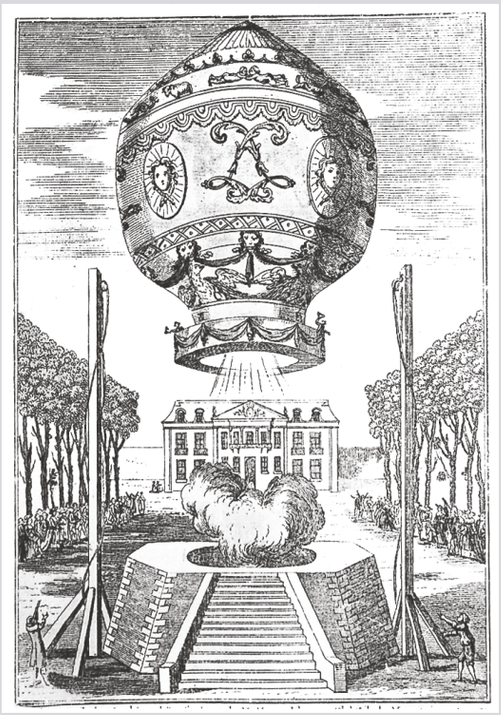


Br. 435

er an, daß das Bild des Punktes f in e erschiene, und daß die geraden Linien df und de ein unveränderliches Verhältniß zu einander hätten, nämlich im Wasser das von 4 zu 3. In dem Dreyecke def hat man nun df : de = sin. def : sin. efd oder df : de = sin. aed : sin. fdg oder df : de = sin. cdh : sin. fdg. Huygens aber sagt, auf das Verhältniß des Sinus hat Snellius nicht gedacht, und glaubte, daß hier alles auf das scheinbare Bild der Sache ankäme. Erst Descartes führt das wahre Gesetz der Strahlenbrechung in seiner im Jahre 1637 herausgekommenen Dioptrik so an, wie es bereits oben angeführet worden ist, erwähnet aber keines Versuches, welchen er zur Entdeckung dieses Gesetzes gemacht hätte, sondern er leitet es als eine Folge aus den Untersuchungen über die Brechung des Lichtes her, obgleich Huygens gewiß versichert, daß er die Handschrift des Snellius in Händen gehabt habe. Descartes druckt das Gesetz nur etwas anders aus, als Snellius. Es mag auch seyn, daß Descartes dieß Gesetz aus der Handschrift des Snellius genommen hat, so gebührt ihm doch das Verdienst, es zuerst bekannt gemacht zu haben, wodurch erst die Dioptrik eine richtigere und auf Gründen gebauete Theorie erhielt.

Vor Descartes hatte man es noch nicht gewagt, eine Erklärung über die Ursache der Brechung zu machen. Dieser versuchte sie aus mechanischen Grundsätzen durch die Zerlegung der Kräfte zu geben. Er nahm an, daß das Licht die stärker brechende Materie leichter als die weniger brechende durchdringe. Es sey (fig. 64.) gc der einfallende Strahl, dessen Bewegung nach den zwey Richtungen hc in der Oberfläche des Wassers und ic der senkrechten auf diese Oberfläche zerfället werde. Mit ic und ch kann man das Parallelogramm ichg und mit cg den Kreis gafb beschreiben. Ist nun die Geschwindigkeit in der stärker brechenden Materie um ein Drittheil größer, als die in der geringer brechenden, so beschreibet der Lichtstrahl in der ersten Materie eben denselben Weg in zwey Zeittheilen, welchen er in der andern Materie in drey Zeittheilen beschrieb, weil die Geschwindigkeiten

Ee 2 in



▲ **Figure 4.54** Symbol of technology in the Age of Enlightenment: In the Montgolfier brothers' hot-air balloon, man rises into the air for the first time in 1783.

Our eighteenth century will certainly not have to be ashamed if it sooner or later passes on to the nineteenth its inventory of newly gained knowledge and acquired things, even if the transmittal had to take place tomorrow. We would like to cast a brief glance at what it could answer its successor if tomorrow it were asked by it, What have you delivered and what new have you seen? It could boldly answer, I have determined the shape of the Earth; I have taught to defy thunder; I have drawn lightning like champagne out of a bottle; I have discovered animals that in their marvels surpass the fable of the hydra; discovered fish that—what even Jupiter could not do—kill weaker fish, even under water, with invisible lightning; through Linnaeus I have had created the first usable inventory of the works of nature; I have seen the return of a comet when the vacation granted it by Halley came to a close, and in my eighty-ninth year, I await the second return; instead of the single air that my forefathers knew, I count thirteen varieties; I have transformed solids into air and air into solids; I have forged quicksilver; lifted huge weights with fire; shot them with water as with gunpowder; I have tricked plants into creating illegitimate children; made steel to flow like butter with burning tinder; I have melted glass under water; cast down gold from the throne as the heaviest solid it had usurped for millennia and replaced it with a white metal; I have provided a new type of outstanding telescope that even Newton held to be impossible; I have switched the poles of a natural magnet in seconds and switched it again; I have hatched eggs without hens and without the warmth of brooding. I have made it so that there is now a single bishop at Rome, just as at Hildesheim. I have stamped on the head of a powerful and dangerous

continued on next page

How is pure mathematics possible?
How is pure science of nature possible?

Since these sciences actually exist, it is quite proper to ask *how* they are possible; for that they must be possible is proved by the fact that they exist. [pp. 55–56]

Kant, in opposition to extreme empiricism, but also to extreme rationalism, wishes to show that in cognition both factors play a role. He states categorically that while all knowledge begins with perception, a significant role is played by the perceiving human mind: Raw perception becomes knowledge only once the mind organizes perception within certain a priori structures, by which is meant thought structures that existed prior to any perception. Among these structures are *viewpoints* (*Anschauungsformen*)—with space as the external viewpoint and time as the internal viewpoint—and *categories*, which include, for example, relational categories such as substance, accident, and furthermore cause and effect. The categories are thus "a priori conditions of experience because only with their help can our reasoning sum up the multiple layers of perception into the unity of transcendent apperception." The categories therefore are those concepts that a priori prescribe the laws for phenomena and hence also for nature, which is the setting for all phenomena.

Kant himself called this new point of view—that in cognition the objects must adapt themselves to the recognizing mind—a Copernican Revolution in philosophy. In Kant's words, we can summarize the role of sensory perception and reason, that is, of *empeiria* and *ratio*, as follows: Without sensory perception, objects would not exist for us, and without reason, they could not be fathomed. A concept without sensory perception is empty, while perception without concept is blind. Therefore, it is just as necessary to make the concepts perceivable (that is, to bring the objects into view) as to process the perception rationally (that is, to subordinate it to concepts). Reason is incapable of sensory perception, and the senses cannot imagine anything.

In the history of philosophy, Kantian epistemology represents an important nexus into which the philosophical currents of the past merge and out of which many later philosophical developments and systems will burst forth. For physicists, this is of special importance; they feel that this was the last attempt to satisfy the universal desire of physicists to find some secure knowledge that is "eternally valid," that is, not to be refuted by any future experience, no matter how modest this secure knowledge might be. This desire of physicists is expressed most poignantly in such remarks made by Einstein as, "The most incomprehensible thing about the universe is that it is comprehensible," and, "What I am really interested in is whether God could have made the world in a different way; that is, whether the necessity of logical simplicity leaves any freedom at all."

The fact is that physics has transcended Kant's conception; which is to say, it no longer recognizes either space or time as a priori structures, nor are causality and substance recognized as organizing principles for reasoning. The last possibility of assuming the existence of truths of this kind, or for seeking them, seems to have been lost. Most of the physicists of the nineteenth century, including Helmholtz and Hertz, but also even Planck, accepted Kant's philosophy, which gave them assurance that their physical knowledge was in fact true. At the turn of the twentieth century, Mach and Einstein consciously set themselves in opposition to Kant. Kant's influence on modern natural philosophy is typified by the fact that in dealing with the problems of space and time and of causality and substance, one cannot ignore the arguments that Kant gave for their a priori existence.

Any physicist attempting to find an ally in philosophers for a unified and firm foundation for science will read with dismay the following—somewhat

simplified—list of the "solid" conclusions reached by philosophy in Wisdom's book *Philosophy and Its Place in Our Culture*:

> Since there is a widespread sense among non-philosophers and also among philosophers that in the course of its two-and-a-half-thousand years of existence it has made no progress, I venture to give the following list of achievements (even if strikingly negative):
>
> Knowledge in general is not obtainable by Reason alone (though it is in certain narrowly defined fields). ...
>
> No knowledge is based on pure experience or pure observation.
>
> There are no synthetic a priori truths.
>
> No well-formed statements (other than self-contradictions) are false or meaningless a priori. ...
>
> All empirical knowledge is fallible.
>
> [pp. 105–106]

4.4 From Effluvium to the Electromagnetic Field

4.4.1 *Peter of Maricourt and Gilbert*

In the previous discussions, we have repeatedly encountered a pattern of development in the various branches of physics, primarily in mechanics and astronomy: The Greeks reached a relatively high level, the Middle Ages gradually rediscovered the ancient legacy, and building on this foundation, later centuries surpassed those achievements. This schema fails to hold for electromagnetic phenomena, for here, European science received nothing more from the Greeks than nomenclature. Thus "electricity" and "magnetism" are derived from the Greek words for amber (ήλεκτρον) and magnetite (ήλίδος Μαγνήτις); see Quotation 4.18. Medieval scholars found nothing in Aristotle about magnets and were therefore compelled to develop new methods on their own. In 1269, Peter of Maricourt (Petrus Peregrinus) carried out a number of investigations into the properties of magnets and even made use of experimental methods, which was unusual at that time. Using a small iron needle, he determined the forces on the surface of a spherical magnet, drawing the needle's direction point by point and thereby obtaining—as we would put it today—the directions of the magnetic field lines. These experiments revealed that the magnetic field lines run like the meridians on the globe, meeting at two opposite poles (Figure 4.56). Indeed, it is Maricourt who gave us the term "pole."

Figure 4.57 reproduces one of Maricourt's sketches, which shows an idea for a perpetual motion machine based on magnetism.

The extent to which interest in magnetic phenomena was outside the mainstream is shown by the fact that neither Thomas Aquinas nor Albertus Magnus, both of whom were contemporaries of Maricourt, mentions him or his work.

More than three centuries passed before William Gilbert (1544–1603), court physician to the English queen, Elizabeth I, continued the work of Peter of Maricourt in a remarkably similar spirit and with similar methods. Gilbert's book *De magnete, magneticisque corporibus et de magno magnete tellure* appeared in 1600, the same year in which Giordano Bruno was burned at the stake. This work is also significant from the point of view of natural philosophy in that it stresses the importance of direct experimentation before Galileo (Quotation 4.19). Gilbert's most important contribution can already be seen in the book's title: The Earth is a large magnet. This is how Gilbert presents the theory of the compass. He recog-

Figure 4.54 *continued*

> hydra of [religious] orders; And what have I seen? Oh, enough. I have seen Peter I and Catherine and Friedrich and Joseph and Leibniz and Newton and Euler and Winckelmann and Mengs and Harrison and Cook and Garrick. Are you satisfied with that? Good. But look here at a few trifles: Here I have a new enormous country, here a fifth part of the world, there a new planet, and a small convincing little proof that our Sun is a satellite, and look here, finally I have made, in my eighty-third year, an airship.
>
> —Lichtenberg, *Gelehrte und gemeinnützige Aufsätze*, 1783



Figure 4.55 Immanuel Kant (1724–1804): Studies in mathematics, philosophy, and theology at the University of Königsberg; there he became a lecturer in 1755 and professor in 1770.

Even in his youth, Kant made notable contributions to the development of the natural sciences. For example, he interpreted the forces acting among particles of matter in the Newtonian sense and postulated the existence of both attractive and repulsive forces, which, however, did not balance each other because the first was to vary with distance as $1/r^2$, whereas the second varied as $1/r^3$. His name is enshrined in physics with the Kant–Laplace hypothesis on the origin of the solar system, which was presented in the 1755 treatise *Allgemeine Naturgeschichte und Theorie des Himmels*. (Laplace's treatise appeared in 1796 with the title *Exposition du système du monde*.) The presentations of the two authors do not agree on every point; however, they have in common the bold intention of explaining the present state of the solar system as the result of development from a primordial cloud, thereby opposing the dogma that the state of the cosmos had remained unchanged since its creation. With Laplace, the planets break off from the already condensed rotating Sun, while Kant has the planets condensing on their own, which is closer to today's consensus.

Kant's most important philosophical works are *Kritik der reinen Vernunft* (*Critique of Pure Reason,* 1781), *Kritik der praktischen Vernunft* (*Critique of Practical Reason,* 1788), and *Kritik der Urteilskraft* (*Critique of Judgment,* 1790).

> *The Copernican Revolution in Metaphysics*
>
> Hitherto it has been assumed that all our knowledge must conform to objects. But all attempts to extend our knowledge of objects by establishing something in regard to them a priori, by means of concepts, have, on this assumption, ended in failure. We must therefore make trial whether we may not have more success in the tasks of

continued on next page

Figure 4.55 *continued*

metaphysics, if we suppose that objects must conform to our knowledge. This would agree better with what is desired, namely, that it should be possible to have knowledge of objects a priori, determining something in regard to them prior to their being given. We should then be proceeding precisely on the lines of Copernicus' primary hypothesis. Failing of satisfactory progress in explaining the movements of the heavenly bodies on the supposition that they all revolved round the spectator, he tried whether he might not have better success if he made the spectator revolve and the stars to remain at rest. A similar experiment can be tried in metaphysics, as regards the *intuition* of objects. If intuition must conform to the constitution of the objects, I do not see how we could know anything of the latter a priori; but if the object (as object of the senses) must conform to the constitutions of our faculty of intuition, I have no difficulty in conceiving such a possibility.

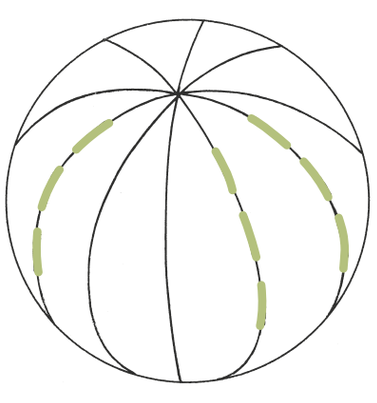
—Kant, *Critique of Pure Reason*, preface to the second edition [p. 22]

Quotation 4.18

To pass on, I will begin to discuss by what law of nature it comes about that iron can be attracted by that stone which the Greeks call magnet from the name of its home, because it is found within the national boundaries of the Magnetes. This stone astonishes men, because it often makes a chain out of little rings hanging from it. For you may sometimes see five or more hanging in a string and swayed by a light breeze, where one hangs from another beneath it, and one from another learns the stone's power and attraction: to such a distance does its power hold force, oozing through and through.

In matters of this sort many principles have to be established before you can give a reason for the thing itself, and you must approach by exceedingly long and roundabout ways; accordingly I crave all the greater attention of ears and mind.

—Lucretius, *On the Nature of Things*, Book VI, ll.905–920



◀ **Figure 4.56** Peter of Maricourt had already maintained in 1269 that the "magnetic lines of force" of a spherical magnet, like the Earth's meridians, meet at the poles.

nized the character of the forces acting between magnetic poles, discovered both attractive and repulsive forces, and determined, moreover, that breaking a magnet into two cannot separate the two poles because each fragment again becomes a magnet with two poles. He observes the magnetic inclination, that is, the deviation of the compass needle from the horizontal, and surmises that it must therefore be possible to use this information in determining geographic latitude without the need for astronomical observation.

Gilbert also made detailed investigations of electrical phenomena. He knew, for example, that not only amber, but also a whole host of other materials (glass, wax, sulfur, and some gemstones, for example) could be electrified through friction. He described some significant differences between electrical and magnetic phenomena, his most important contribution relating perhaps to the character of the forces: Magnets elicit a turning response (*verticitas*), whereas electrical force is expressed through attraction (*attractio*). Gilbert did not yet mention the two kinds of electrical charges and did not yet know that among electrical phenomena, repulsion might also occur (Figure 4.58).

4.4.2 *The Chronology of Progress*

With Gilbert's work, the seventeenth century made its contribution to the investigation of electromagnetic phenomena in its first year, and essentially made no progress thereafter. Descartes, of course, included all known natural phenomena in his philosophy, so in this uniform cosmology there was a place for magnetic phenomena as well (Figure 4.59). We also have to bring up for special mention the multifaceted and ingenious experimentalist, Otto von Guericke, the mayor of Magdeburg, who constructed the first "triboelectric" machine, which produced static electricity by friction (Figure 4.60), thereby laying the experimental foundations for intensive electrostatic investigation. However, his scientifically minded contemporaries were busy with and to some extent blinded by the success in mechanics, in the formulation of natural-philosophical foundations, and in the derivations of quantitative answers using the most advanced mathematical methods of the era. As a result, interest in electricity diminished to such an extent that von Guericke's electricity-producing machine was forgotten for two generations.

As we saw in Section 3.2, Kepler attributed an important role to magnetism. He surmised that the motion of the planets is due to the magnetic attraction of the Sun.

In Newton, we find statements about magnetism and electricity in his *Opticks*, though always in the form of questions (Quotation 4.20). In the two "questions" related to these topics, Newton supports two different conceptions concerning the nature of electric and magnetic forces. In one, he essentially adopts Gilbert's idea that magnetic and electric substances fill the surrounding space with a fluid (*effluvium*). This fluid was assumed to be capable of penetrating ordinary heavy bodies. In the other question, Newton interprets magnetic and electric interactions as action at a distance. Here we meet the brilliant conjecture that material particles interact at very small distances via their electrical forces, with this interaction being independent of whether they are put into an electrified state with the aid of friction.

The first half of the eighteenth century witnessed a renewed interest in electricity. By the middle of the century, it had become as fashionable to perform electrical experiments in the salons of the nobility as it was to read the *Encyclopédie* (Figure 4.61). The craze for experimentation increased even more after the invention of the Leyden jar; now one could perform experiments that were so visually

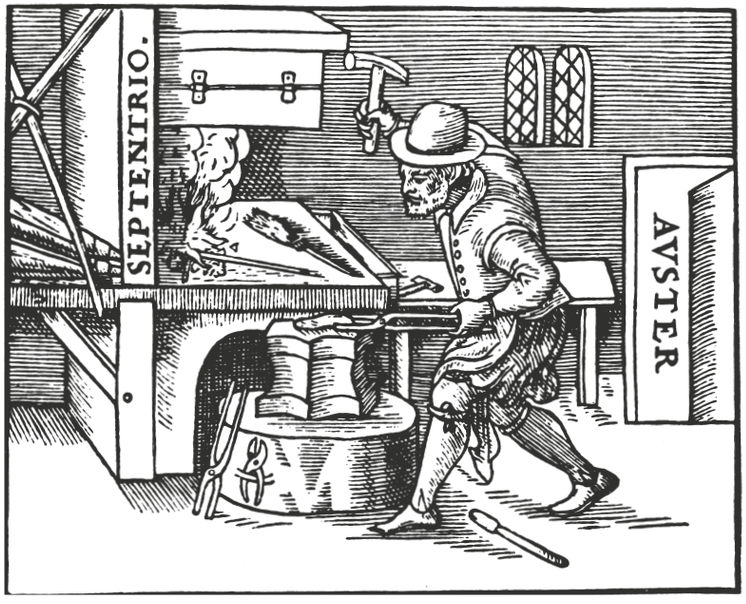
impressive that they not only amused salon guests in polite society, but could serve as attractions at public shows and fairs.

Table 4.2 offers a schematic representation of the problems in electricity and magnetism that were the focus of investigation from the beginning of the eighteenth century together with the names of the researchers who played a significant role in this work and what they achieved.

We see that in the first three-quarters of the eighteenth century, it was essentially qualitative research in electrostatics that was carried out. This work served, of course, as a necessary precondition for the quantitative investigations that would begin in the last quarter of the century. In the first decades of the nineteenth century, then, the fundamental laws of electrostatics and—we should add—magnetostatics were brought into the form in which they are used today.

Experimental aids to the discovery of relationships between electrical and magnetic phenomena became possible with the invention of a source of strong and steady electrical current: the galvanic cell. Interestingly, its discovery by VOLTA occurred precisely in the year 1800.

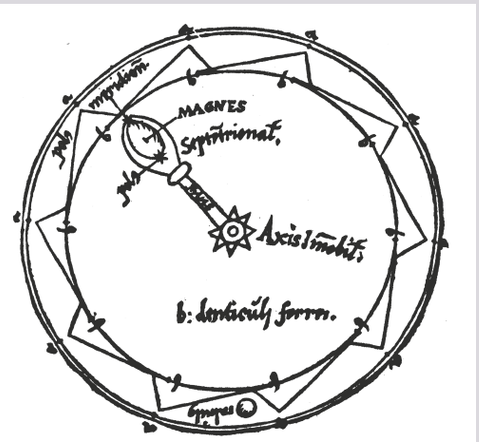
It is somewhat surprising that it was not until 1820 that it came to be noticed that a magnetic needle could be influenced by an electrical current, but then, in only a few years, all the laws relating to the connection between electrical current and magnetism were written down in the form to which we are now accustomed.



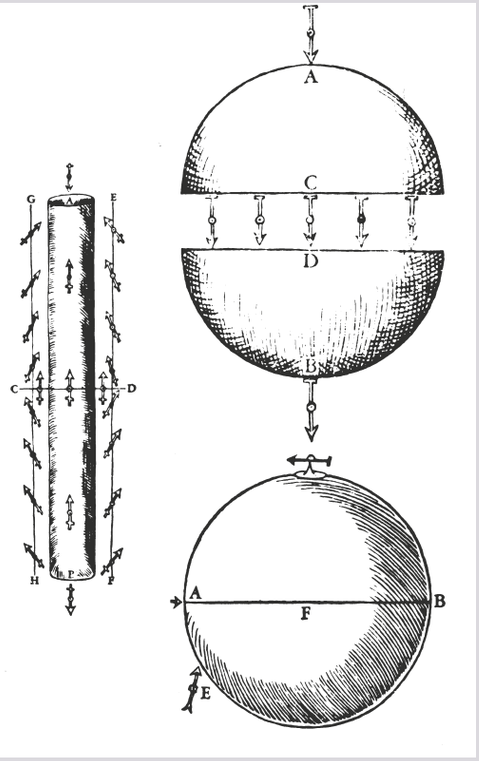
◀ ▶ **Figure 4.58**
In GILBERT's book *De magnete …*, we find many familiar pictures.

After this, only two more discoveries were needed in order to complete the entire edifice of classical electrodynamics. Both of these are of fundamental importance, and they are connected to the two greatest personalities of electrodynamics: FARADAY, who found the relationship between the changes in a magnetic field over time and the associated electric field (in simpler terms, the law of induction), and MAXWELL, who recognized the converse phenomenon, thereby bringing to completion the field of classical electrodynamics with the system of equations that bears his name.

Behind the increasing precision and refinement of the quantitative description of electromagnetic phenomena lurks a change of view of crucial significance; namely, by the end of the eighteenth century, the idea, going back to NEWTON, of action at a distance had become completely acceptable, not only in science but also as an intellectual concept. Thus not only did Coulomb's laws reflect this new attitude and



▲ **Figure 4.57** PETER OF MARICOURT hoped to use the magnetic force to construct a perpetual motion machine.



Quotation 4.19

As handed down by [PLATO and ARISTOTLE], the loadstone merely attracted iron, the rest of its virtues were all undiscovered. But that the story of the loadstone might not appear too bare and too brief, to this singular and sole known quality there

continued on next page

Quotation 4.19, ***continued***

were added certain known figments and falsehoods, which in the earliest times, no less than nowadays, used to be put forth by raw smatterers and copyists to be swallowed by men. As for instance, that if a loadstone be anointed with garlick, or if a diamond be near, it does not attract iron. Tales of this sort occur in Pliny, and in Ptolemy's *Quadripartitum*; and the errors have been sedulously propagated, and have gained ground (like ill weeds that grow apace) coming down even to our own day, through the writings of a host of men, who, to fill out their volumes to a proper bulk, write and copy out pages upon pages on this, that, and the other subject, of which they knew almost nothing for certain of their own experience. Such fables of the loadstone even Georgius Agricola himself, most distinguished in letters, relying on the writings of others, has embodied as actual history in his books *De Natura Fossilium*. Galen noted its medicinal power. ... But his commentator ... repeats the story of the garlick and the diamond, and moreover introduces Mahomet's shrine vaulted with loadstones, and writes that, by the exhibition of this (with the iron coffin hanging in the air) as a divine miracle, the public were imposed upon. But this is known by travelers to be false. Yet Pliny relates that Chinocrates the architect had commenced to roof over the temple of Arsinoe at Alexandria with a magnet-stone, that her statue of iron placed therein might appear to hang in space. His own death, however, intervened, and also that of Ptolemy, who had ordered it to be made in honour of his sister. ... And as yet we have not set ourselves to overthrow by argument those errors and impotent reasonings of theirs, nor many other fables told about the loadstone, nor the superstitions of impostors and fabulists: for instance ... that a white loadstone may be procured as a love potion: or as Hali Abbas thoughtlessly reports, that if held in the hand it will cure gout and spasms: Or that it makes one acceptable and in favour with princes, or eloquent, as Pictorio has sung ...: Or that by day it has a certain power of attracting iron, but by night the power is feeble, or rather null: Or that when weak and dulled the virtue is renewed by goats' blood, as Ruellius writes: Or that Goats' blood sets a loadstone free from the venom of a diamond: ... Or that it removed sorcery from women, and put to flight demons, as Arnaldus de Villanova dreams. ... With such idle tales and trumpery do plebeian philosophers delight themselves and satiate readers greedy for hidden things, and unlearned devourers of absurdities.

—William Gilbert, *On the Magnet* [pp. 1–2, 6, 7]

become new supports for Newton's ideas, but Ampère's laws of electrodynamic were also born in accord with the fundamental postulates of Newtonian mechanics. Slowly, step by step, to some extent under the coercion of experimental facts and based on the work of Faraday and Maxwell, the concept of the *electromagnetic field as physical reality* began to gain acceptance.

In the following sections we pursue the developments sketched above in greater detail.

4.4.3 *Qualitative Electrostatics*

Joseph Priestley's 1767 book *The History and Present State of Electricity, with Original Experiments* offers an interesting summary of the qualitative investigations into electrostatics that had been carried out during the first three-quarters of the eighteenth century and even previously. Priestley himself was actively involved in several different branches of science and is known above all for his work in chemistry, in particular for the discovery of oxygen. His book is of interest for a variety of reasons. As the title informs us, it is a summary, or a status report as we would call it today. However, what is also of interest is that Priestley went to the trouble to reproduce the experiments himself and was thus able to contribute a number of original insights into electrostatics. Quotation 0.5, taken from that book, should be read by all those who are under the impression that it is only a phenomenon of our times that the proliferation of research papers has become so great that it is impossible for any one individual to keep up with new developments, or that the rapid acceleration in the pace of scientific research began only recently.

Table 4.2 shows the principal participants in qualitative electrostatics: Stephen Gray, Dufay, Franklin, Aepinus, and Priestley. In Figure 4.62 can be seen two pages from the table of contents of Priestley's book, and we would now like to spend a bit of time discussing the researchers who are linked, according to Priestley, with the beginning and end of this phase of development.

Stephen Gray (1666–1736) worked with such simple equipment as a glass rod. His most important observation is that certain materials, then called nonelectric substances, conduct electricity. This nomenclature goes back to Gray himself and his pupil Jean Théophile Desaguliers, a Huguenot who had fled France. Gray had determined that a test component made of a "nonelectric" material, when hung from an isolating silken thread or laid on an isolating plate, could be brought into a state of electrification. Figure 4.63 presents the principle of the experiment that Gray carried out at his country estate with the help of a friend. He first determined that an electrical effect could be detected even on a metal nail driven into the stopper of a rubbed glass cylinder. Then he attached a string to this nail, and was successful in transmitting the electrical effect. He conducted the experiments first in the castle gallery and then continued them out of doors in the fine autumn weather. He succeeded in transmitting the electrical state over a distance of 886 feet. We must point out, however, that this remained an electrostatic experiment and not one involving the conduction of electric current. It is characteristic of Gray's inventiveness that he successfully employed this method to transmit signals. Gray here also made an observation that was to be later of great importance in treating the questions of the similarities and differences between electric and thermal phenomena: The electrical status of a body is a function of its surface area and not its volume.

Stephen Gray carried on a lively correspondence with the overseer of the French royal gardens, Charles François de Cisternay Dufay (1698–1739). Dufay's

great accomplishment was the discovery that there exist two kinds of electricity. Previously, it had been known only that a body that had been electrified by means of friction first attracts light objects in its vicinity and then, after they are touched, repels them, and that electrified bodies repel one another. It then was observed, however, that in rubbing a glass rod and a piece of resin (Quotation 4.21), these two objects do not repel, but attract. DUFAY therefore distinguished two types of electricity: glass electricity (*électricité vitreuse*) and resinous electricity (*électricité résineuse*). As a consequence, alongside the single-fluid theory of electricity there arose a two-fluid theory, which was worked out in detail by the official naturalist of the French royal court, JEAN ANTOINE NOLLET (1700–1770), who conceived of the two types of electricity in the form of an *effluvium* and an *affluvium* surrounding electrified bodies.

A new impetus to experimentation was provided by the Leyden jar, the invention of the pastor VON KLEIST, born in Vicewo (Vietzow), Poland, and by MUSSCHENBROEK, who was a professor in Leyden. VON KLEIST came upon the invention by chance, MUSSCHENBROEK in the process of systematic experimentation, though also with the assistance of chance. MUSSCHENBROEK's goal was to prevent the well-known phenomenon that a conductor suspended in air by isolating threads, when brought into an electrified state, will sooner or later lose that state. To this end, he "electrified" water inside a glass flask by connecting a metallic rod leading into the flask through the stopper to an electrification machine utilizing friction. In one hand he held the flask, and when he touched the metal rod with the other (Figure 4.64), he experienced a powerful electric shock. MUSSCHENBROEK's panicked fright is colorfully described in Quotation 4.22.

Incidentally, the second part of this quotation is worthy of note: Every scientist must display a certain *furor heroicus*, that is, a readiness to put his or her life on the line for the sake of scientific investigation, with renown as the reward for such sacrifice. RICHMAN, the man mentioned in the quotation, lost his life in St. Petersburg through his experiments with electricity.

As we have mentioned, from this time on, it had become fashionable to experiment with electricity. Larger and ever more elaborate apparatuses were built and new experimental tools were employed. Figure 4.65 comes from a 1770 book of J. B. HORVÁTH; Figure 4.66 shows some measuring instruments from MUSSCHENBROEK's book. Even physicians saw in electrical effects the promise of hidden possibilities and introduced electroshock therapy, although not with the purposes for which it has been used more recently.

Perhaps the most interesting personality and most successful experimentalist of this epoch was BENJAMIN FRANKLIN (1706–1790, Plate XVIII). He was the first American to make significant contributions to the development of European science. FATHER NOLLET, already mentioned, had completely different ideas from those of FRANKLIN, and on becoming aware of FRANKLIN's discoveries after the translation of his works into French, simply could not believe that a physicist by the name of FRANKLIN existed and that moreover he lived in Philadelphia. He protested that the entire book must have been written by his detractors in order to besmirch his reputation.

FRANKLIN learned about electricity from an itinerant showman. Fortunately for FRANKLIN, he did not read the books of his European colleagues; his intellectual arsenal consisted of NEWTON's *Principia* and *Opticks*. Thus he was able to interpret his experimental findings more freely. FRANKLIN is best known for his invention of the lightning rod. Even before FRANKLIN, it had been apparent—because of the light and crackling sound of an electrical spark—that lightning is a similar


De Quatuor Corporibus. 599

fumere debeant ; quæ veluti exiguas columnas concipere poſſumus, tribus ſtriis, in modum cochlearum intortis excavatas. Idque magis vel minus, prout axi Vorticis

P.599.

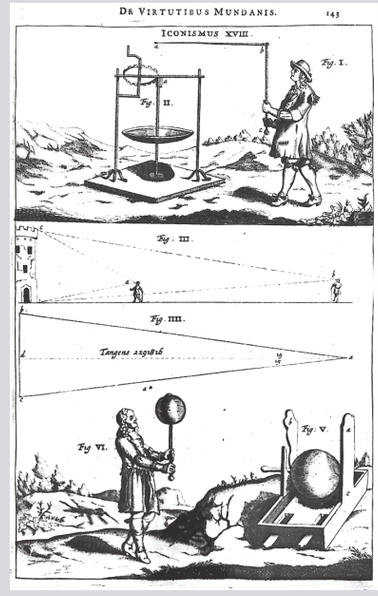
remotiores, aut viciniores tranſeunt; quia globuli ſecundi elementi celerius in illis, quàm in iſtis circumvolvuntur.

Notandum etiam ex eo, quod ipſæ veniant verſus medium cœli, ex partibus contrariis : Unæ ſcilicet ab Auſtrali, aliæ à Boreali polo, dum interim totus Vortex circa ſuum axem, in unas & easdem partes movetur,

3. Particulæ ſtriatæ variis modis ſunt inſertæ.

Pp 4

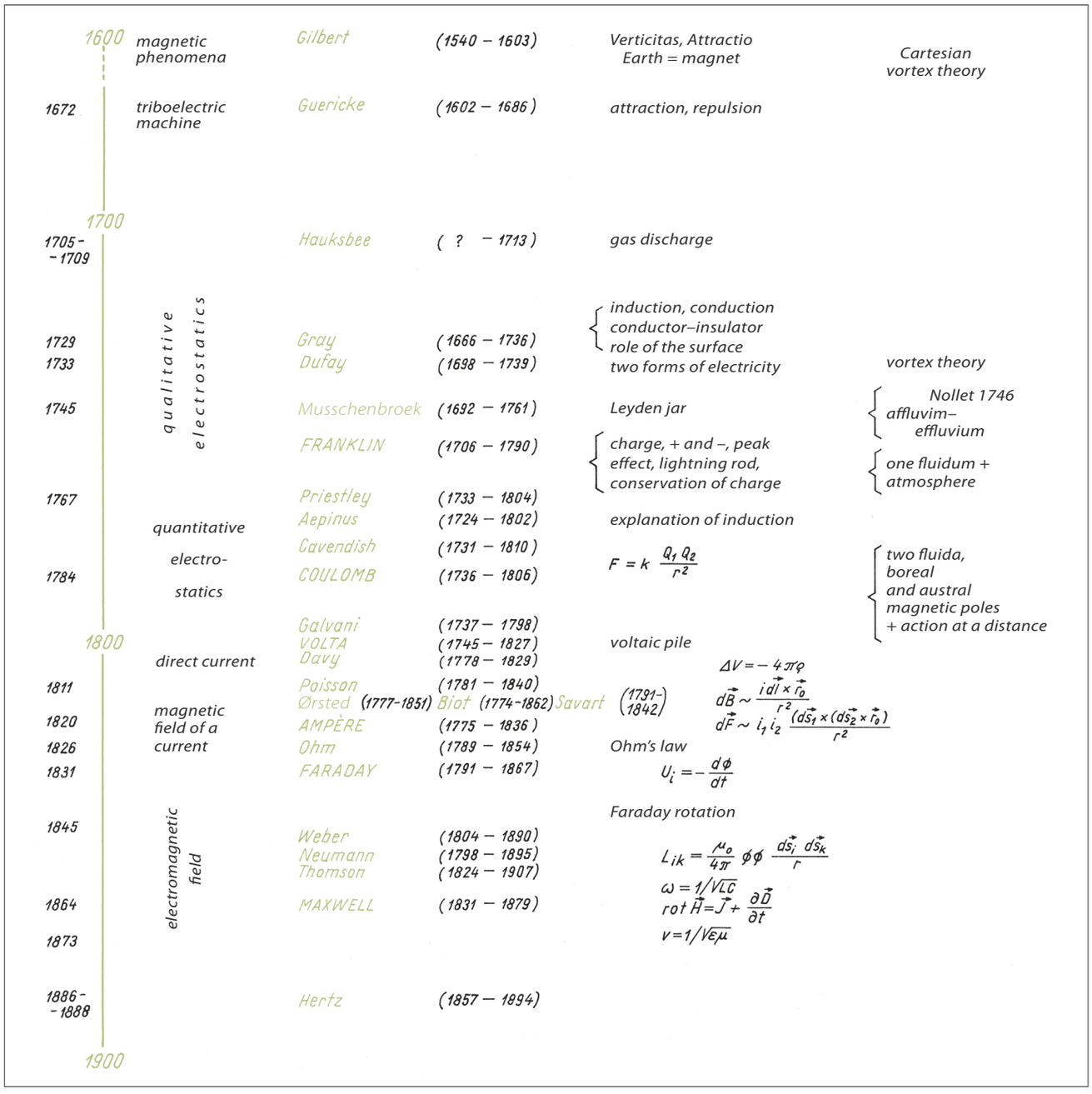
◀ Figure 4.59 DESCARTES's interpretation of magnetism.



◀ Figure 4.60 GUERICKE's triboelectric machine (Fig. V). (Library of the Hungarian Academy of Sciences.)



◀ Figure 4.61 In the Age of Reason, electricity was a source of amusement in educated circles [Taton 1957].



◀ **Table 4.2** Chronology of discoveries in the field of electrodynamics.

Quotation 4.20

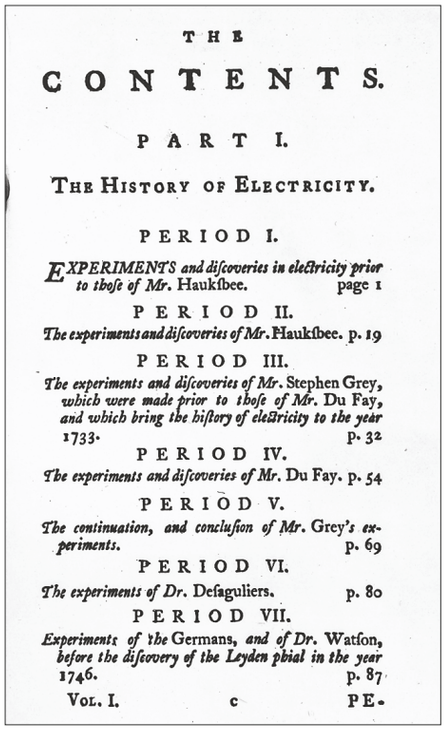
[If any one ask how a Medium can be so rare] ... Let him also tell me how an electrick Body can by Friction emit an Exhalation so rare and subtile, and yet so potent, as by its Emission to cause no sensible Diminution of the weight of the electrick Body, and to be expanded through a Sphere, whose Diameter is above two Feet, and yet to be able to agitate and carry up Leaf Copper or Leaf Gold, at the distance of above a Foot from the electrick Body? And how the

continued on next page

electrical phenomenon. However, Franklin was the first to provide a complete proof of this assumption (Figure 4.67). He stretched an electrical conductor between a kite flying at a great height and a Leyden jar and showed that atmospheric electricity is just as capable of charging a Leyden jar as an electrification machine.

Franklin further observed that a metallic point could conduct charge to a body and then conduct the charge away from it (Figure 4.68).

The reader will perhaps have noticed that we have been avoiding the word "charge," for this word is attached to a specific idea relating to the nature of electricity. The word "charge" was introduced by Franklin, and so from now on, we will no longer require such circumlocutions as "bodies that have been brought into a state of electrification"; instead, we will speak of "electrically charged" bodies or bodies that have been given an electrical charge.



THE

CONTENTS.

PART I.

THE HISTORY OF ELECTRICITY.

PERIOD I.

EXPERIMENTS and discoveries in electricity prior to those of Mr. Haukſbee. page 1

PERIOD II.

The experiments and discoveries of Mr. Haukſbee. p. 19

PERIOD III.

The experiments and discoveries of Mr. Stephen Grey, *which were made prior to those of Mr.* Du Fay, *and which bring the history of electricity to the year* 1733. p. 32

PERIOD IV.

The experiments and discoveries of Mr. Du Fay. p. 54

PERIOD V.

The continuation, and concluſion of Mr. Grey's *experiments.* p. 69

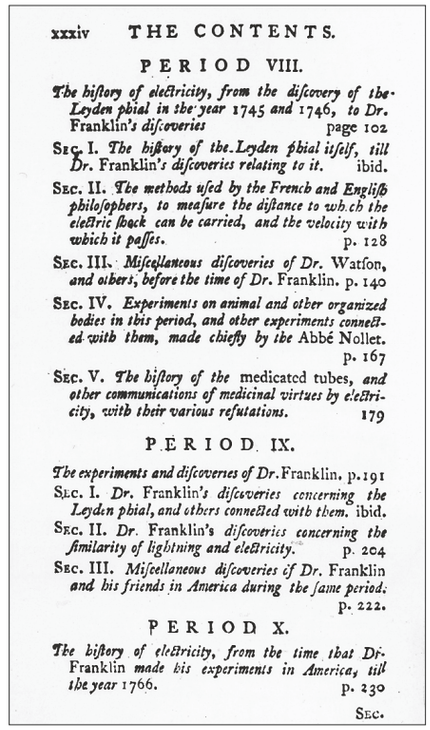
PERIOD VI.

The experiments of Dr. Deſaguliers. p. 80

PERIOD VII.

Experiments of the Germans, *and of Dr.* Watſon, *before the diſcovery of the Leyden phial in the year* 1746. p. 87

VOL. I. c PE-



xxxiv THE CONTENTS.

PERIOD VIII.

The hiſtory of electricity, from the diſcovery of the Leyden phial in the year 1745 *and* 1746, *to Dr.* Franklin's *diſcoveries* page 102

SEC. I. *The hiſtory of the Leyden phial itſelf, till Dr.* Franklin's *diſcoveries relating to it.* ibid.

SEC. II. *The methods uſed by the French and Engliſh philoſophers, to meaſure the diſtance to which the electric ſhock can be carried, and the velocity with which it paſſes.* p. 128

SEC. III. *Miſcellaneous diſcoveries of Dr.* Watſon, *and others, before the time of Dr.* Franklin. p. 140

SEC. IV. *Experiments on animal and other organized bodies in this period, and other experiments connected with them, made chiefly by the* Abbé Nollet. p. 167

SEC. V. *The hiſtory of the* medicated tubes, *and other communications of medicinal virtues by electricity, with their various refutations.* 179

PERIOD IX.

The experiments and diſcoveries of Dr. Franklin, p. 191

SEC. I. *Dr.* Franklin's *diſcoveries concerning the Leyden phial, and others connected with them.* ibid.

SEC. II. *Dr.* Franklin's *diſcoveries concerning the ſimilarity of lightning and electricity.* p. 204

SEC. III. *Miſcellaneous diſcoveries of Dr.* Franklin *and his friends in America during the ſame period.* p. 222.

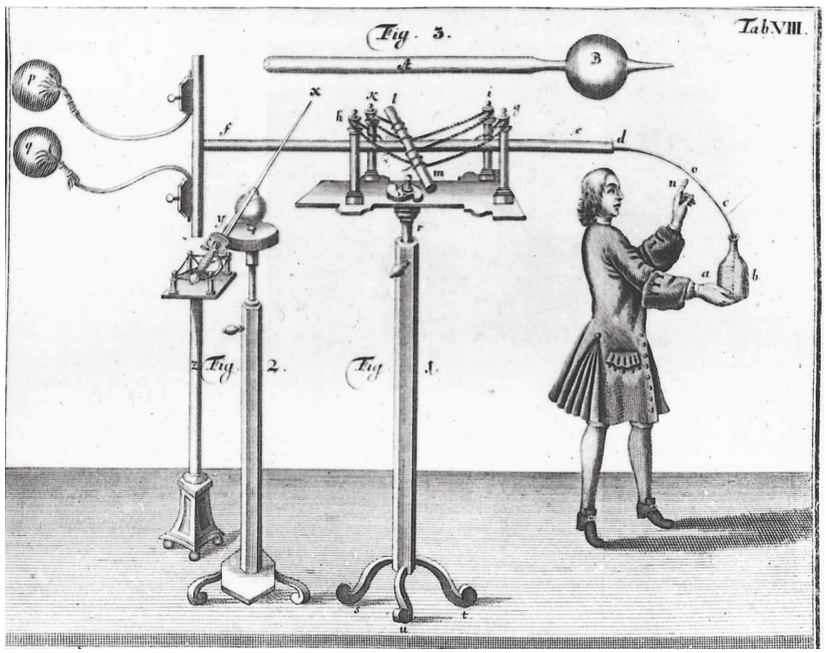
PERIOD X.

The hiſtory of electricity, from the time that Dr. Franklin *made his experiments in America, till the year* 1766. p. 230

SEC.

▲ **Figure 4.62** Two pages from the table of contents of PRIESTLEY's 1767 book *History and Present State of Electricity with Original Experiments*.

For FRANKLIN, there was only one type of charge, which he took to be glass electricity as described by DUFAY. The electrical state of a body depends on whether the body has more or less of these charges than normal. If the former, then we can say the body is charged with glass electricity, or—after EULER—that it is positively charged. If there is a shortage of electrical charges, then the body possesses—in the old terminology—resinous charge, or in the newer form, negative charge. Thus negative electric charge represents a shortage of charge; that is, the electrical charge has been removed from the material in question (Figure 4.69).



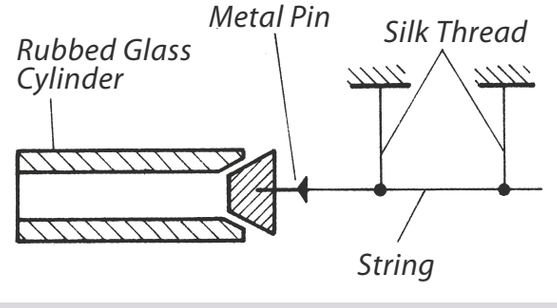
◀ **Figure 4.64** MUSSCHENBROEK discovered the "amplifying effect" of the Leyden jar in this manner [Winkler 1746]. (Library of the Hungarian Academy of Sciences.)

Quotation 4.20, *continued*

Effluvia of a Magnet can be so rare and subtile, as to pass through a Plate of Glass without any Resistance or Diminution of their Force, and yet so potent as to turn a magnetick Needle beyond the Glass? ...

Have not the small Particles of Bodies certain Powers, Virtues, or Forces, by which they act at a distance, ... upon one another for producing a great Part of the Phaenomena of Nature? For it's well known, that Bodies act one upon another by the Attractions of Gravity, Magnetism, and Electricity; and these Instances shew the Tenor and Course of Nature, and make it not improbable but that there may be more attractive Powers than these. For Nature is very consonant and conformable to her self. How these Attractions may be perform'd, I do not here consider. What I call Attraction may be perform'd by impulse, or by some other means unknown to me. I use that Word here to signify only in general any Force by which Bodies tend towards one another, whatsoever be the Cause. For we must learn from the Phaenomena of Nature what Bodies attract one another, and what are the Laws and Properties of the Attraction, before we enquire the Cause by which the Attraction is perform'd. The Attractions of Gravity, Magnetism, and Electricity, reach to very sensible distances, and so have been observed by vulgar Eyes, and there may be others which reach to so small distances as hitherto escape Observation; and perhaps electrical Attraction may reach to such small distances, even without being excited by Friction.

—NEWTON, *Opticks*, Queries 22, 31



▲ **Figure 4.63** Schematic representation of GRAY's experiment.

Quotation 4.21

It is then certain ... that bodies which have become electric by contact are repelled by those which have rendered them electric; but are they repelled likewise by other electrified bodies of all kinds? And do electrified bodies differ from each other in no respect save their intensity of electrification? An

continued on next page

Quotation 4.21, ***continued***

examination of this matter has led me to a discovery which I should never have foreseen, and of which I believe no-one hitherto has had the least idea. ...

We see, then ... that there are two electricities of a totally different nature—namely, that of transparent solids, such as glass, crystal, &c. and that of bituminous or resinous bodies, such as amber, copal, sealing-wax, &c. Each of them repels bodies which have contracted an electricity of the same nature as its own, and attracts those whose electricity is of the contrary nature. We see even that bodies which are not themselves electrics can acquire either of these electricities, and that then their effects are similar to those of the bodies which have communicated it to them.

—Dufay, 1733 [Whitaker 1958, pp. 43–44]

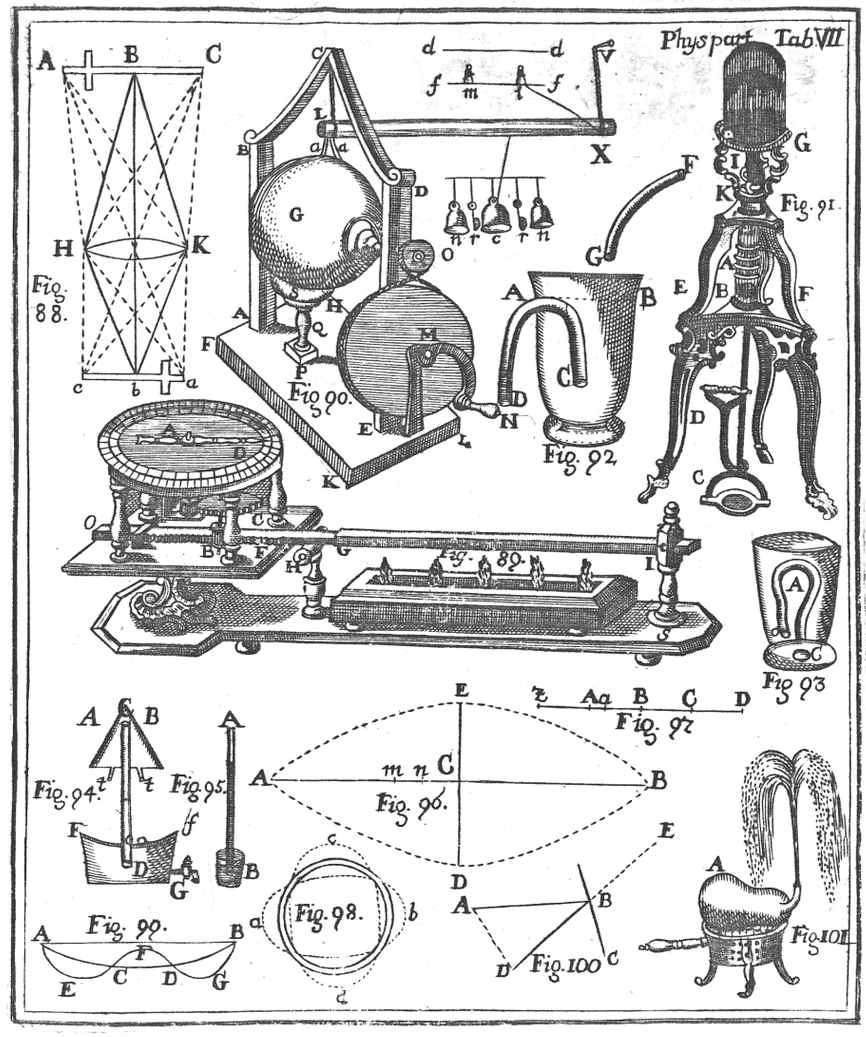
Quotation 4.22

Mr. Muschenbroek, who tried the experiment with a very thin glass bowl, says, in a letter to Mr. Réaumur, which he wrote soon after the experiment, that he felt himself struck in his arms, shoulder and breast, so that he lost his breath, and was two days before he recovered from the effects of the blow and the terror. He adds, that he would not take a second shock for the kingdom of France. ...

We are not, however, to infer from these instances, that all the electricians were struck with this panic. Few, I believe, would have joined with the cowardly professor, who said that he would not take a second for the kingdom of France. Far different from these were the sentiments of the magnanimous Mr. Boze, who with a truly philosophical heroism, worthy of the renowned Empedocles, said he wished he might die by the electric shock, that the account of his death might furnish an article for the memoirs of the French Academy of Sciences. But it is not given to every electrician to die in so glorious a manner as the justly envied Richman.

In France as well as in Germany experiments were made to try how many persons might feel the shock of the same phial. The Abbé Nollet, whose name is famous in electricity, gave it to one hundred and eighty of the guards, in the King's presence; and at the grand convent of the Carthusians in Paris, the whole community formed a line of nine hundred toises, by means of iron wires between every two persons (which far exceeded the line of one hundred and eighty of the guards) and the whole company upon the discharge of the phial, gave a sudden spring, at the same instant of time, and all felt the shock equally.

—J. B. Priestley, *The History and Present State of Electricity* [Vol. I, pp. 106–108, 125–126]

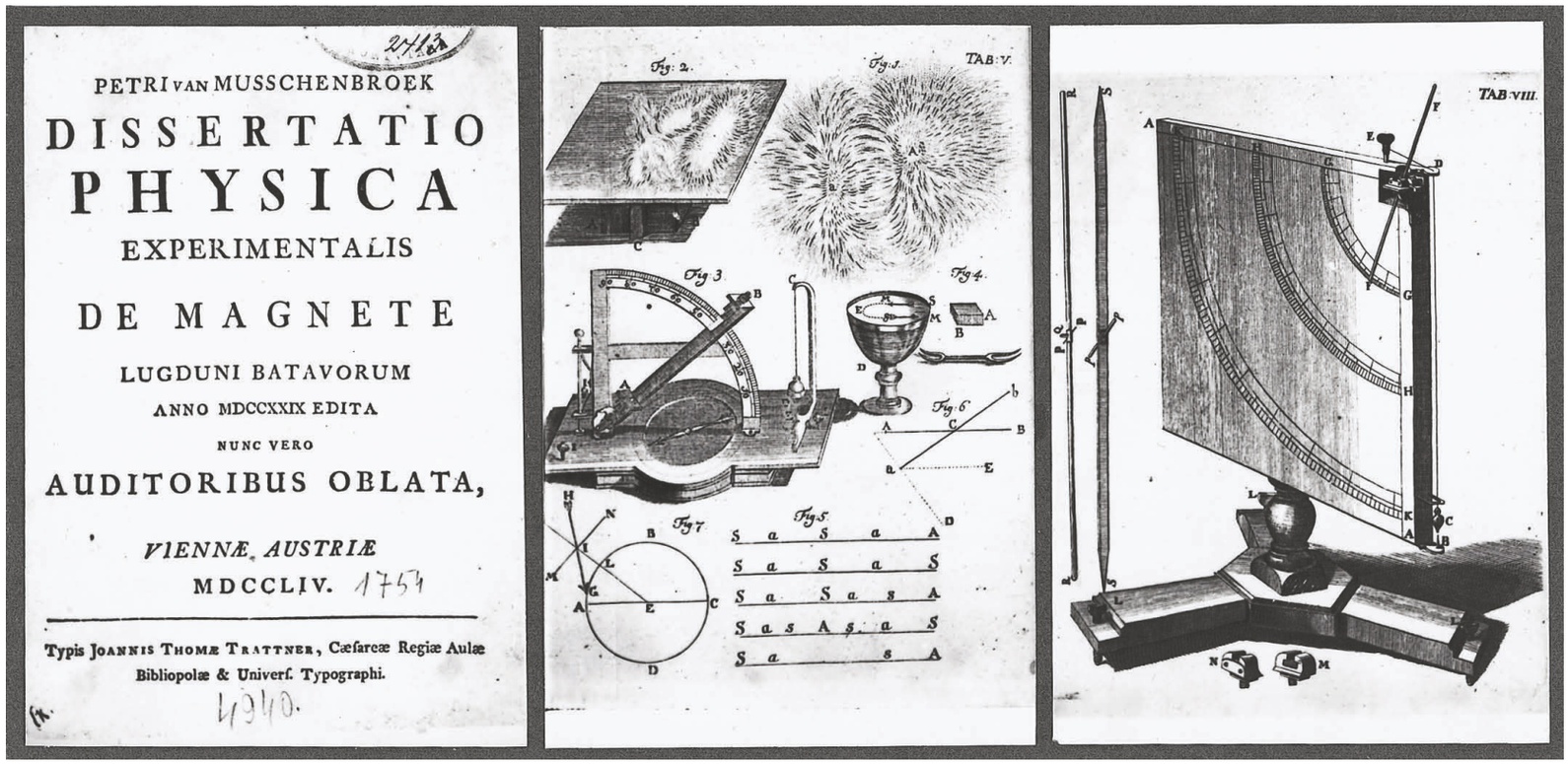


▶ **Figure 4.65** In the middle of the eighteenth century, experiments in electrostatics were carried out with instruments like these [Horvath 1770].

To explain electrical phenomena, Franklin assumed that electrical charges repel one another, whereas electrical charge and matter are mutually attractive. However, he could not thus explain why two bodies with a shortage of charge, consisting, so to speak, of "naked" material, repel each other. In our current terminology, the two bodies each carry a negative charge, and it seems obvious to us that there should be a repulsive force between them.

Franz Ulrich Theodor Aepinus (1724–1802), born in Rostock and active in St. Petersburg, attempted to find an answer to these open questions. According to his view, the particles of matter deprived of electricity repel one another just like particles of matter that constitute the electrical charge itself. One may interpret this statement, if we may use today's terminology, to say that the particles of matter remain after the removal of charges in a kind of ionized state.

We cannot determine exactly what motivated Franklin to view the electricity of glass as something present and to speak of a lack of something in regard to the electricity of resin. The terms "positive" and "negative" are logical consequences of such a viewpoint. Franklin apparently took the point of view that when a charged and uncharged body touched, the charges always flowed in one direction only. This is also the case, according to our present knowledge, when we consider

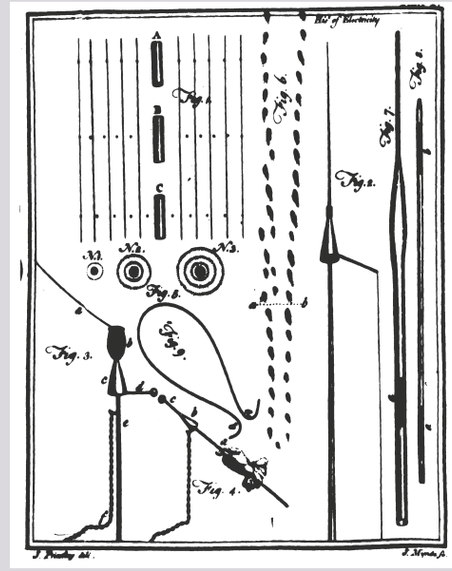


▲ **Figure 4.66** Title page and two typical pages of Musschenbroek's book. (Library of the University for Heavy Industry, Miskolc.)

the flow of electric charges in metallic bodies. We know that, contrary to Franklin's assumption, it is the negative charges that flow in metals while the positive charges remain in place. However, in Franklin's experiments in the transmission of charge to a body or the removal of charge from a body with the help of a metallic point, charge currents move in both directions in the air space between the metallic point and the body. It is quite likely that Franklin was led to the assumption that only glass electricity flows by the shape of the light phenomenon that faintly glows during such experiments, which is visible in the dark and indeed resembles a jet of water coming from a watering can.

Franklin's experiments and his interpretation of the results, which were then refined by Aepinus, finalized the charge conservation law. Charges were not produced by friction, but merely separated. By way of illustrating the separation of charges, Franklin compared bodies in an uncharged state to a wet sponge and the removal of charge by rubbing the body to the dripping of water when the sponge is compressed.

Alongside this monistic theory of electricity, however, the dualistic theory gained in importance. Even Coulomb embraced this model. Both models are able to explain the fact of conservation of charge, which follows automatically from the single-fluid model because a charge deficit necessarily agrees with the notion of removed charge. The action of force, however, can be more simply explained with the two-fluid model. To wit, if one assumes the existence of two forms of electricity, then it may simply be assumed that the force is proportional to both charges, that is to say, to their product, or as we would put it today, is proportional to $Q_1 \times Q_2$. With the single-fluid model, one must postulate a repulsive force between the excess charges, an attractive force between matter and charges, and finally, a repulsive force between matter and matter. The following argument shows how the two models lead to the same conclusion.



▲ **Figure 4.67** Franklin's experimental contraptions.

> As every circumstance relating to so capital a discovery as this (the greatest, perhaps, that has been made in the whole compass of philosophy, since the time of Isaac Newton) cannot but give pleasure to all my readers, I shall endeavour to gratify them with the communication of a few particulars which I have from the best authority.

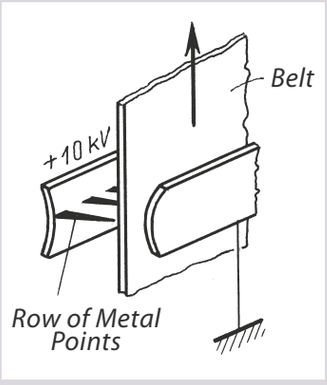
continued on next page

Figure 4.67 *continued*

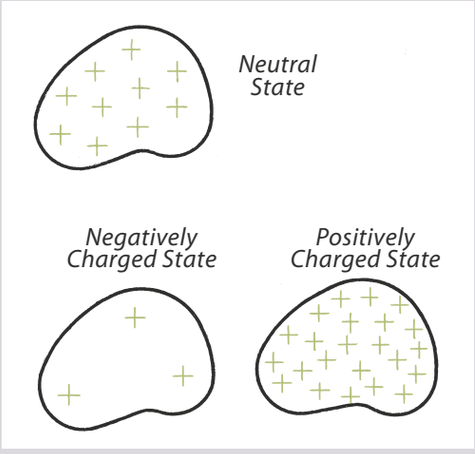
To demonstrate, in the completest manner possible, the sameness of the electric fluid with the matter of lightning, Dr. Franklin, astonishing as it must have appeared, contrived actually to bring lightning from the heavens, by means of an electrical kite, which he raised when a storm of thunder was perceived to be coming on. This kite had a pointed wire fixed upon it, by which it drew the lightning from the clouds. This lightning descended by a hempen string, and was received by a key tied to an extremity of it; that part of the string which was held in his hand being of silk, that the electric virtue might stop when it came to the key. He found that the string would conduct electricity even when nearly dry, but that when it was wet, it would conduct it quite freely; so that it would stream out plentifully from the key, at the approach of a person's finger.

At this key he charged phials, and from electric fire thus obtained, he kindled spirits [ignited alcohol], and performed all other electrical experiments which are usually exhibited by an excited globe or tube.

—J. B. Priestley, *The History and Present State of Electricity*, 1775 [pp. 216, 215]



◀ **Figure 4.68** The moving belt in a Van de Graff generator is charged with the "wonderful Effect of Points" discovered by Franklin. These generators have been used in accelerators for nuclear research to create several million volts of electric potential.



▲ **Figure 4.69** According to Franklin, there was only one type of charge.

We set $Q_1 = E_1 - M_1$ and $Q_2 = E_2 - M_2$, where M is the quantity of matter that in the uncharged state neutralizes the effect of the electrical charge E. The product $Q_1 \times Q_2$ then becomes

$$\begin{aligned} Q_1 \times Q_2 &= (E_1 - M_1) \times (E_2 - M_2) \\ &= E_1 \times E_2 - E_1 \times M_2 - E_2 \times M_1 + M_1 \times M_2, \end{aligned}$$

the right-hand side of which contains precisely the mathematical expressions for the above-mentioned interactions.

On the occasion of a farewell banquet for his scientifically minded friends, Franklin projected a vision of a future age of electricity. Before the dinner, a spirit stove on the far side of a river was lit with the aid of an electric spark, the turkey destined for the festive meal was killed with an electric shock, and the guests drank to the health of all the famous "electricians" in the world from electrified champagne glasses to the crackling of the discharge of electrified jars.

By the middle of the eighteenth century, additional phenomena had been described—even if only in qualitative form—whose significance would only later become clear (Quotations 4.23, 4.24 and Figure 4.70).

4.4.4 *Quantitative Electrostatics*

By now, the time had become ripe for establishing quantitative laws to describe electrical phenomena, and indeed, several researchers succeeded in quantifying the electrical attractive force. The "natural-philosophical" basis for this was the same for all of them: Following Newton, they assumed the existence of forces acting at a distance between charged bodies and looked for the laws that described those forces. The law was found independently by four researchers: Priestley, Cavendish, Robison, and Coulomb, after the latter of whom the law was named. Priestley formulated the law precisely in his 1767 book and provided a reasoned basis for it. It was known from experiments that charge accumulates on the surface of electrical conductors and that there is no perceptible electric charge in the interior of hollow bodies made of electrically conducting material (Quotation 4.25). The forces emanating from the surface charges could compensate for an arbitrary point in the interior of a closed surface only if the force were inversely proportional to the square of the distance from the charge (Figures 4.71, 4.72). Cavendish arrived at the law via a similar train of thought; moreover, he confirmed the law with the aid of a torsion balance. It is of historical interest that the idea of using a torsion balance to measure small forces was mentioned independently by several researchers. Cavendish credits a Reverend Michell for being the source of the idea of the torsion balance as well as for providing him with the first exemplar.

Robison made his measurements in 1769, earlier than Coulomb, but did not produce a consistent formulation of the law. He obtained different values for repulsive and attractive forces: in one case he got slightly higher values than the square law, and in the other, slightly lower ones.

As an engineer and military officer, Charles Coulomb (1736–1806, Figure 4.73) spent a number of years overseeing fortifications, and it was not until 1776 that he was able to devote all of his energies to scientific work. He made notable contributions to a number of scientific fields; for example, he was awarded a competition prize from the French Academy of Sciences for the construction of the best compass, and on the basis of that work was elected to the academy. In a prize-winning paper, Coulomb describes a torsion balance and shows that the torque necessary to rotate the torsion thread through a given angle is proportional to the fourth power of the

thread's diameter and inversely proportional to its length, with the proportionality factor depending on the material of which the thread is constructed. We also have Coulomb to thank for an extensive study of sliding friction in the case of rigid bodies. He was the first to investigate in depth the question of the force required to hold a ship on an inclined plane before launch.

Figure 4.74 shows Coulomb's torsion balance as it appeared in his original article.

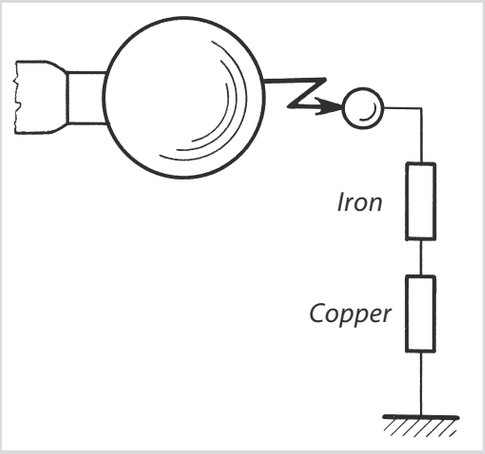
Cavendish would today occupy a higher place in the history of the study of electricity had he published his results promptly. As it turned out, they were only published 100 years after the fact, in 1879. The reason for publication then was that Lord Kelvin had discovered in Cavendish's manuscripts the formula for the relationship between the charge on a sphere and the charge on a flat circular plate at the same potential. While this was only an experimental result, it suggested that Cavendish must have surmised some concepts and their qualitative roles that were to become very important later on. A treatment of this problem is even today possible only with university-level knowledge and would be considered a difficult question on an examination in theoretical electrodynamics.

Of course, Cavendish did not write about potential but used the term "degree of electrification" instead. Thus for him, two metallic conductors have the same degree of electrification when they are joined by an electrical conductor. Today we would say that both conductors are on the same potential. Cavendish raised the question of the relationship between the charges of both bodies in this case. When determining this charge relationship, Cavendish in effect determined the ratios of the electric capacitances of the two metallic bodies. The carrying out of these measurements required a clear understanding of electrical phenomena. Cavendish did not report on the details of how these measurements were made, but it is likely that he touched the conductors many times with a small test sphere, each time discharging the test sphere by grounding it.

Cavendish also investigated the properties of dielectrics and determined that the ability of a conductor to store a charge is altered if various insulating materials are placed in the vicinity of the conductor. This fact was rediscovered by Faraday almost two generations later.

Cavendish also measured the conductivity of various substances and introduced the concept of resistance, preceding Ohm by half a century. What is astonishing here is that he determined the ratio of conductivities of seawater and iron to be $1 : 4 \times 10^6$, an excellent result. Cavendish remarked that he obtained this value by means of a simple measuring procedure. Today, one would measure electrical resistance also with very simple instruments, namely a voltmeter and an ammeter. In Cavendish's time, the creation of such devices was still far in the future, and one learns from his manuscript how he took these measurements: Cavendish compared the strengths of the electric shocks he felt when he touched—through the various materials—the electrodes of a charged Leyden jar. From the subjective sensation of the shock, he estimated numerical values for the conductivities. It is practically a miracle that he was able to obtain such good results with such a method.

Once the laws for the interactions of electrically charged bodies had been represented mathematically, there was nothing to stand in the way of the application of the mathematical apparatus that had been worked out for treating gravitation from being applied to electrostatic phenomena. This was done by Poisson in an 1811 publication. With this work, electrostatics had finally arrived at the same level of mathematical perfection as mechanics. In addition, Poisson also contributed to the coming of age of magnetostatics in 1824.



▲ **Figure 4.70** Priestley, on advice from Franklin, was the first to investigate the effect of the same current passing through different conductors.

Quotation 4.23

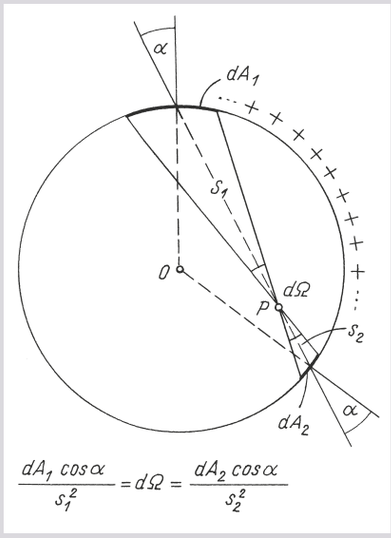
In a conversation I once had with Dr. Franklin, Mr. Canton, and Dr. Price, I remember asking whether it was probable that there was any difference in the conducting power of different metals. ... I have since endeavoured to carry into execution a scheme proposed by Dr. Franklin, viz. transmitting the same explosion of the battery through two wires at the same time, of two different metals, and of the same thickness. ... I first joined a piece of iron wire, and a piece of copper wire. The explosion totally dispersed the iron, and left the copper untouched. ... From these experiments it is easy to settle the order in which the metals above mentioned are to be ranked, with respect to the power of electricity to melt them. It is as follows. *Iron, brass, copper, silver, gold*. ... I make no doubt but that an explosion which melts a copper wire of any given diameter would disperse an iron wire of twice the diameter.

—J. B. Priestley, *The History and Present State of Electricity* [Vol. II, pp. 368–371]

Quotation 4.24

Magnets have been observed to lose their virtue, or to have their poles reversed by lightning. The same did Dr. Franklin by electricity. By electricity he frequently gave polarity to needles, and reversed them at pleasure.

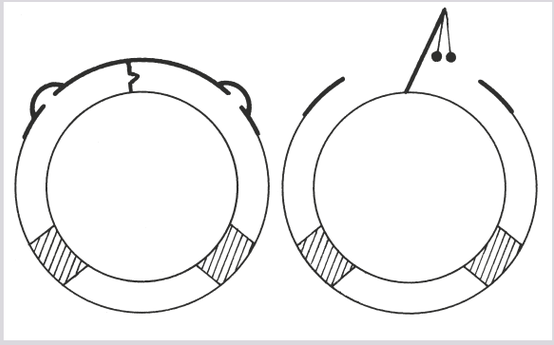
—J. B. Priestley, *The History and Present State of Electricity* [Vol. I, p. 214]



▲ **Figure 4.71** If electrical force falls off with distance as $1/r^2$, then the forces from dA_1 and dA_2 acting on the point P cancel each other out, since we have

$$F_1 \propto \frac{dA_1}{s_1^2} = \frac{d\Omega}{\cos\alpha} = \frac{dA_2}{s_2^2} \propto F_2.$$

The entire surface of the sphere can be decomposed into pairs of surface elements whose effects balance each other out.



▲ **Figure 4.72** An experimental arrangement for proving that inside a metallic electrode there is no electrical effect. At the same time, this will also prove the $1/r^2$ law, at least to the extent of the accuracy of the measurement. Let us denote any deviation from the inverse-square law by the quantity e in the formula $1/r^{2+\varepsilon}$. The precision available to Cavendish allowed the limit $|\varepsilon| < 1/50$; Maxwell could improve this to $|\varepsilon| < 1/21{,}600$. Today, it has been further reduced to $|\varepsilon| < 3 \times 10^{-16}$ (Williams, Faller, Hill, 1971). These investigations have a fundamental importance for modern physics, being connected with the question of whether the photon has a nonzero rest mass. In the case of a nonzero mass, according to Yukawa (see Figure 5.140), the potential would have, instead of $1/r$, the following dependence on distance: $e^{-r/r_0}/r$, where $r_0 = h/2\pi mc$ and h is Planck's constant, m the rest mass, and c the velocity of light. Using this formula, $|\varepsilon| < 3 \times 10^{-16}$ implies an upper bound on the rest mass of the photon of $m < 2 \times 10^{-50}$ kg. This mass is many orders of magnitude less than the rest mass of the neutrino. Thus, at least according to our current knowledge, we may consider the rest mass of the photon to be equal to zero.

Laplace (Figure 4.75) had earlier determined that for a system of point masses whose individual particles act on one another analogously to those in Newton's law of gravitation, the force on a given point mass can be represented by partial derivatives of a certain quantity. Laplace did not give this quantity the name "potential," by which we know it today, but he did specify the value

$$V = G\sum_i \frac{m_i}{r_i}.$$

Laplace also showed that this quantity satisfies the equation

$$\Delta V \equiv \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0,$$

today known as Poisson's equation. Poisson also determined that at locations where continuously distributed electrical charges are located, this equation must be modified to read

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -4\pi\rho.$$

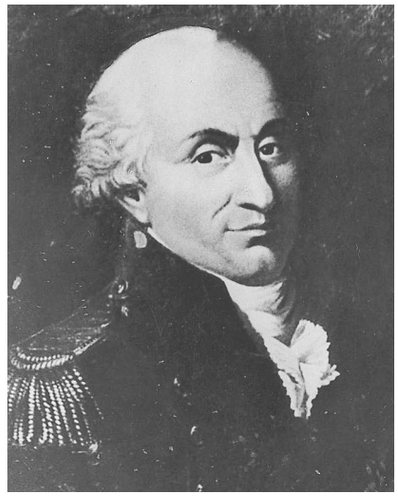
In this form, the equation is known as the Laplace–Poisson equation. Here ρ denotes the charge density.

Poisson was able to solve electrostatic problems that even by today's standards are quite complicated. For example, he calculated the charge distribution on the surface of two opposing metallic spheres.

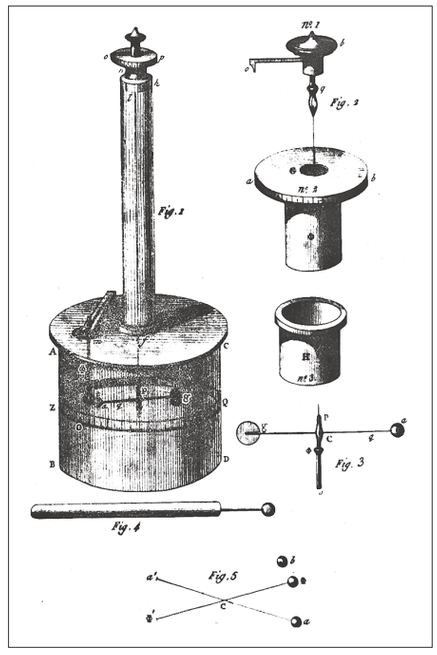
Finally, electrostatics was brought into its present-day form by Green and Gauss (Figure 4.76).

4.4.5 *Flow of Electric Charge*

Even in the electrostatic experiments it was obvious that during charging or discharging, a flow of electrical charges takes place; in other words, one can talk about an *electrical current*. The laws of electrical currents and, in particular, their magnetic effects could be successfully investigated only when it had become possible to generate sustained flows of electric charge of a strength suitable for experimentation. Various effects of the flowing charges have been previously



▲ **Figure 4.73** Charles Auguste de Coulomb (1736–1806): After nine years service as an engineering officer, began to work on scientific problems. He confirmed the law named for him with measurements carried out between 1784 and 1789. In his 1785 book *Théorie des machines*, he considered problems in friction and elasticity, including torsion. These investigations led to the construction of the torsion balance.



▲ **Figure 4.74** Coulomb's torsion balance. The balance and the measurements obtained with it were published beginning in 1784.



▲ **Figure 4.76** Carl Friedrich Gauss (1777–1855): Son of a mason; studied at the University of Göttingen, then instructor in Brunswick, and from 1807 until his death, director of the newly founded observatory and professor of mathematics and astronomy in Göttingen.

In his early years he worked on problems in number theory. The results are collected in his 1801 book *Disquisitiones Arithmeticae*. He shifted the focus of his work about every ten years: 1800–1820, astronomy; 1820–1830, geometry; 1830–1840, theoretical physics. His most important summary works are *Theoria motus corporum coelestium* (1809), *Disquisitiones circa superficies curvas* (1827), *Intensitas vis magneticae terrestris ad mensuram absolutam revocata* (1833), and *Dioptrische Untersuchungen* (1840).

Of his contributions to mathematics that immediately influenced the development of physics or have continued to have application in physics, we mention the following: the method of least squares, the Gaussian error function, the internal geometry of surfaces, the hypergeometric series, the Gaussian complex plane, and the gamma function (complex factorial).

His most important contributions to physics are these: the principle of least constraint, the general theory of optical mapping with lenses, Gauss's theorem on electrostatics, the Gaussian system of units, a measuring procedure for the magnetic moment using his method of "principal positions."

Gauss is rightly referred to as the prince of mathematicians. His mathematical talent was revealed in his earliest years. He was nine years old when his elementary school class was given the exercise of adding up the numbers from 1 to 60, which Gauss solved in a matter of seconds, having noted that 61 = 1 + 60 = 2 + 59 = 3 + 58, and so on, and that there are 30 such pairs, and therefore $30 \times 61 = 1830$.

In searching for solutions to the equation $x_n - 1 = 0$, Gauss realized which regular polygons could be constructed with straightedge and compass; he was thus able to construct a 17-sided regular polygon. He was so proud of this result that he ordered the figure to be carved on his tombstone. Although it can be seen from Gauss's diary that by 1818—thus before Bolyai and Lobachevsky—he had discovered non-Euclidean geometry, he did not dare publish his results, because he feared that the world was not yet mature enough to accept such ideas.



◀ **Figure 4.75** Pierre-Simon Laplace (1749–1827): Of peasant stock, he first considered a religious vocation. D'Alembert secured a position for the 22-year-old Laplace as mathematics teacher in Paris. Laplace actively took part in social and scientific-organizational life. Together with the chemist Claude-Louis Berthollet (1748–1822), founded the Society of Arcueil for the promotion of young scientists. Member of the *Académie*; named minister of the interior for a brief period by Napoleon during the period of the Consulate.

His more important works are these: *Exposition du système du monde* (1796); in this work written in a high literary popular style, we find a theory of the origin of the solar system (Kant–Laplace hypothesis).

Mécanique céleste (1799–1825), five volumes; here, alongside a number of concrete astronomical problems, the Laplace equation $\Delta U = 0$ is considered. Laplace's contributions to probability theory and statistics are found in the two treatises *Théorie analytique des probabilités* (1812) and *Essai philosophique sur les probabilités* (1814). In the foreword to the last-named work, we meet the frequently cited Laplace's demon.

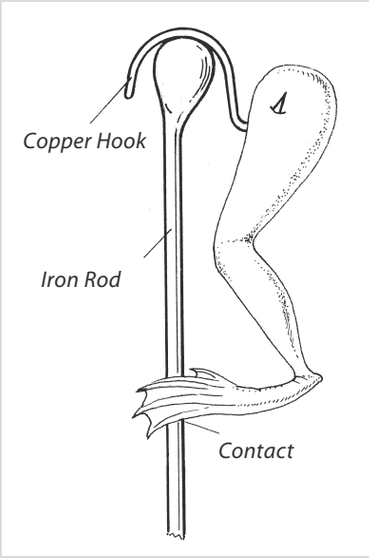
We shall have more to say in the text about Laplace's investigations in the theory of heat, still based on the *caloricum* theory. Here we would like to mention only the results on the theoretical determination of the speed of sound. Mersenne and Gassendi had measured it earlier, while Newton determined it theoretically ($v = \sqrt{p / \rho}$, where p is the pressure, ρ the density). The measurements carried out with greater precision by William Derham (1708) brought to light a discrepancy with the Newtonian theory. Laplace suggested that the adiabatic changes that occur as a result of compression and rarefaction of the air arising from wavelike propagation also need to be taken into account ($v = \sqrt{\kappa p / \rho}$, where $\kappa = c_p/c_v$, with c_p the specific heat at constant pressure, c_v the specific heat at constant volume).

After the publication of the first volume of the *Mécanique céleste*, Laplace was asked by Napoleon why the heavens are discussed over hundreds of pages, but God is never mentioned. Laplace is said to have answered, *Je n´ai pas eu besoin de cette hypothèse là, Sire* (I had no need of that hypothesis, your Majesty).

Quotation 4.25

May we not infer from this experiment, that the attraction of electricity is subject to the same laws with that of gravitation, and is therefore according to the squares of the distances; since it is easily demonstrated, that were the earth in the form of a shell, a body in the inside of it would not be attracted to one side more than another.

—J. B. Priestley, *The History and Present State of Electricity* [Vol. II, p. 374]



◀ **Figure 4.77** Galvani's most important observation: To induce a twitch in a frog's leg, the arrangement shown suffices, and neither a spark nor lightning nor atmospheric electricity is needed.

Quotation 4.26

But I took the animal into a closed room, and placed it on an iron plate; and when I pressed the hook which was fixed in the spinal marrow against the plate, behold! the same spasmodic contractions as before. I tried other metals at different hours on various days, in several places, and always with the same result, except that the contractions were more violent with some metals than with others. After this I tried various bodies which are not wood; but nothing happened. This was somewhat surprising, and led me to suspect that electricity is inherent in the animal itself. This suspicion was strengthened by the observation that a kind of circuit of subtle nervous fluid (resembling the electric circuit which is manifested in the Leyden jar experiment) is completed from the nerves to the muscles when the contractions are produced.

—Galvani [Whitaker 1958, pp. 68–69]

Quotation 4.27

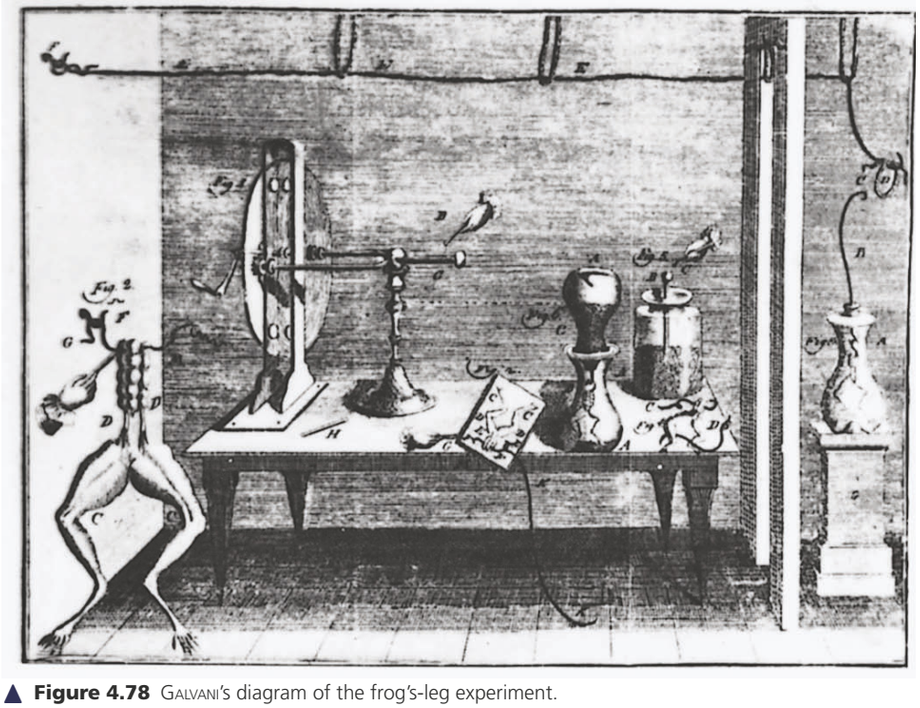
Yes, the apparatus of which I speak, and which will doubtless astonish you, is nothing but the assemblage of a number of good conductors of different types, arranged in a certain fashion. 30, 40, 60, or more pieces of copper—or, better, silver—each applied to a piece of tin—or, what is much better, zinc—and an equal number of levels of water, or some other liquid that is a better conductor than pure water, such as salt water, lye, &c., or pieces of cardboard, skin, &c., that have thoroughly absorbed these liquids; such layers interposed between each pair or combination of the two different metals, such an alternating sequence, and always in the same order, of these three types of conductors: this is all

continued on next page

observed. The heat and light from electrical sparks was evident; further, it was also known that in the proximity of a very strong discharge iron may become magnetized or its magnetic polarity can be reversed (Quotation 4.24).

A breakthrough in this field came in the year 1800. In that year, Alessandro Volta (1745–1827) wrote a letter to the president of the Royal Society in which he described a discovery involving constant direct current. The prehistory of this discovery goes back to the year 1780. Luigi Galvani (1737–1798), professor of anatomy at the University of Bologna, had learned from one of his assistants that while making an anatomical preparation of a nerve from a frog's leg, it had been observed that if one touched the nerve with a dissecting knife at a time when a discharge spark was being emitted from a—then ubiquitous—triboelectric machine elsewhere in the laboratory, the nerve twitched. It was only in 1791 that Galvani described how this phenomenon came to his attention and the various experiments that he conducted. From today's perspective, we can say that if Galvani's colleagues had actually observed a simultaneity between the electrical discharge and the twitching of the frog's leg, then they, in fact, could have been registering an electromagnetic wave emanating from the sparks. In the course of his own experiments, Galvani discovered that frogs' legs hung with copper hooks on an iron window grating would twitch if they accidentally touched the iron grating (Figure 4.77). This fact convinced Galvani that it was not something like a storm or atmospheric electricity that was responsible for the phenomenon. He finally carried out a variety of laboratory experiments (Figure 4.78) from which he concluded (Quotation 4.26) that this electrical phenomenon had its origin in the frog's leg itself, and with this he introduced the notion of animal electricity.

At about this time, Volta was professor of physics at the rival University of Pavia, where he had made a name for himself with a very effective electrostatic machine, called an electrophorus, and—more importantly—a very sensitive electroscope (Figure 4.79). Volta started out by reproducing one of Galvani's



▲ **Figure 4.78** Galvani's diagram of the frog's-leg experiment.

experiments, and he accepted Galvani's interpretation that the electrical phenomenon originated in the frog's leg itself. But in the following year, in 1793, Volta performed very carefully a series of experiments already referenced in 1754 by the Swiss researcher Sulzer: If we connect two different metal pieces together at one end then touch one of the pieces with our tongue, we perceive either a mild acid or a mild alkali taste, depending on the type of metal, just as when we touch with our tongue the positive or negative pole of a triboelectric machine. Volta immediately saw that the frog's leg in Galvani's experiments served no function other than to make the presence of electricity visible and that the crucial factor was the contact between the two kinds of metal. He showed that no twitching occurs if only there is only one type of metal involved in the experiment. Finally, using an electroscope, he was able to prove directly that two connected pieces of different metals will each have a charge after being separated. Volta's great discovery was that this effect could be enhanced by stacking several zinc and copper plates alternately with a third material between them, such as damp cardboard, which Volta called a conductor of the second kind (Quotation 4.27 and Figures 4.80, 4.81).

Volta did not understand how his galvanic battery worked. He did, however, correctly state that the electricity thus obtained had the same properties as the electricity obtained through friction. He was of the opinion that the two metals played passive roles and thus the current flow that began when the circuit was closed would continue indefinitely without any change occurring in the battery. Although during Volta's lifetime the impossibility of a mechanical perpetual motion machine was already a well-accepted fact, with the mysterious and weightless electrical fluidums, everything still seemed possible.

Already in 1800, the same year in which Volta's discoveries were announced, intensive work with galvanic batteries was begun in England. A leading role here was played by Humphry Davy (1778–1829). He determined that chemical processes occurring in the galvanic batteries have a significant role in the electrical phenomena. His work can be seen as the starting point for the field of electrochemistry and indeed, as the starting point for all theories about the electrical nature of chemical processes.

The characteristic quantities in the description of the flow of electrical charges are the current intensity and the "electromotive force" that creates the flow. They are linked by the simplest formula in electrical engineering: Ohm's law.

"Let's take a galvanic battery and measure its voltage U; then we connect the poles of the battery with conductors of various lengths and various diameters made of the same material, and we measure the current intensity I in each case; finally, we repeat the experiment with various materials. We will immediately obtain Ohm's law in the following form:

$$U = I \times R.$$

Here, $R = R_{\text{internal}} + R_{\text{external}}$ and $R_{\text{external}} = \rho(l/A)$, where ρ is the specific resistance, l is length, and A is the cross-sectional area of the conductor."

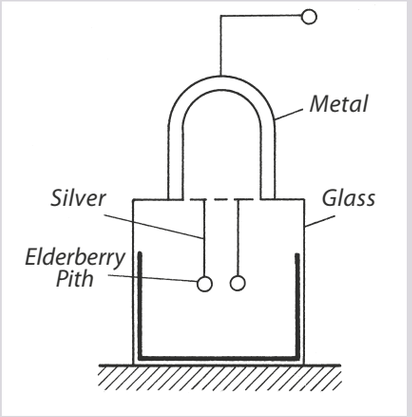
It is all so simple, and we do not understand why it was discovered only after the magnetic effect of electrical currents. But, of course, we are looking at this from today's point of view, which gives us a false picture.

Georg Simon Ohm (Figures 4.82, 4.83) was the first to publish the relationship that now bears his name; he did so in two articles from the years 1826 and 1827. [The relationship in Figure 4.83 that reads $X = a/(b + x)$ would today be written $I = U/(R_{\text{ext}} + R_{\text{int}})$.] However, Ohm's theory was not accepted until 1841, and it was not until a quarter century after his discovery that he attained his professorial status.

Quotation 4.27, ***continued***

that constitutes my new instrument, which imitates, as I have said, the effects of Leyden bottles, or electric batteries, by producing the same shocks as these; [and] whose activity, in truth, remains far below that of the aforesaid batteries, even when these have been charged to a high level, with respect to the force and noise of its explosions, to its spark, to the distance at which the discharge can occur, &c., equaling only the effects of a battery charged to a very weak degree, though a battery with an immense capacity; in other respects, however, it infinitely surpasses the virtue and power of these very same batteries, insofar as it does not need, as they do, to be charged in advance by means of foreign electricity; and insofar as it is capable of producing shocks every time one touches it in the correct manner, no matter how frequent these contacts are.

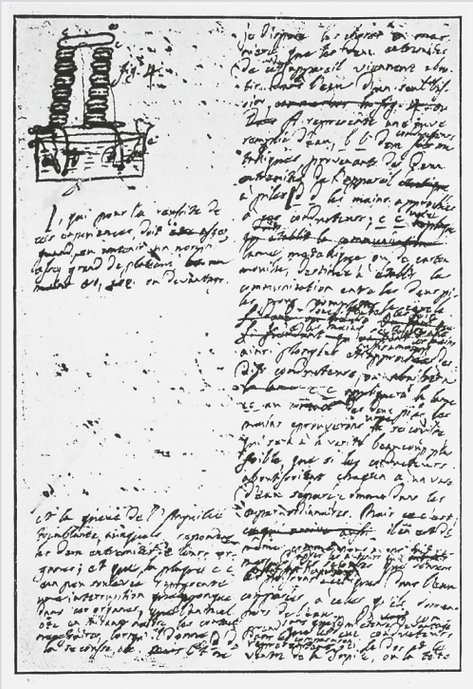
—Alessandro Volta, "On the Electricity Excited by the Mere Contact of Conducting Substances of Different Kinds," 1800



◀ **Figure 4.79** Volta's electroscope.



▲ **Figure 4.80** Volta demonstrates the voltaic pile to General Napoleon Bonaparte.



▲ **Figure 4.81** The construction of the voltaic pile (from one of Volta's notebooks).

Quotation 4.28

You certainly have a right to ask why it is inconceivable that no one tried the action of the voltaic pile on a magnet for twenty years. However, I believe that a cause of this is easily discovered: it simply existed in Coulomb's hypothesis on the nature of magnetic action; everyone believed this hypothesis as though it were a fact; it simply discarded every possibility of the action between electricity and so-called magnetic wires; the restriction was such that when M. Arago spoke of these new phenomena [of electromagnetism] at the Institute his remarks were rejected just as the ideas of stones falling from the heavens were rejected when M. Pictet read a memoir to the Institute on these stones. Everyone had decided that all this was impossible. ... Everyone resists changing ideas to which he is accustomed.

—Ampère, Letter to a friend, 1820 [Williams 1966, p. 60]

In establishing his law, Ohm could draw on two significant developments. Thomas Johann Seebeck (1770–1831) discovered thermoelectricity in 1821, and so Ohm was able to obtain a current source with constant voltage. On the theoretical front, Fourier's 1822 work on thermal conduction helped Ohm formulate analogous laws for electrical conduction. The problems that Ohm had to overcome were more conceptual than technical: for example, it was not clear whether a current along a conductor is constant or whether it may be "used up" in the process, and it was also not clear what was the relationship between the potential, as known from electrostatics, and the quantity measurable in the electric circuit that was somehow analogous to the concept of temperature. It was also unknown whether current flows on the surface or through the interior of a conductor.

Ohm's simple law was extended by Gustav Kirchhoff to more complicated circuits. In 1845, Ohm worked out the two Kirchhoff laws for general circuits. Kirchhoff also made crucial progress in explaining the subject conceptually; for example, he pointed to the shared nature of the potential in the Poisson equation and the "electroscopic force" in Ohm's law.

Kirchhoff's first law, or current law, states that the sum of the currents meeting at a node in a circuit is equal to zero. In forming this sum, the currents flowing into the nodal point are considered to be negative, and those flowing out are considered positive.

Kirchhoff's second law concerns voltage: if one considers any closed loop within an electrical circuit, the sum of the "electromotive forces" (the internal voltages) is equal to the drop in voltage across the resistors. Again, the appropriate positive and negative signs must be used.

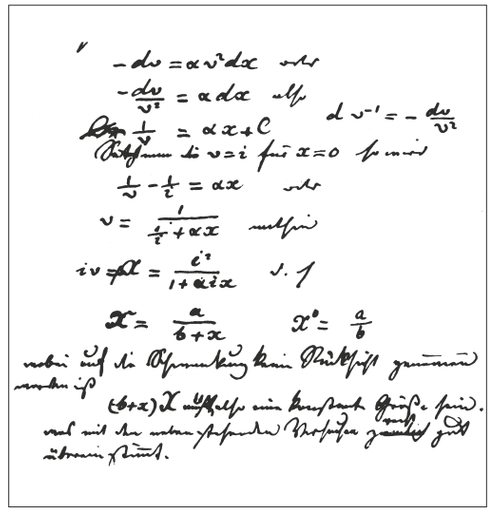
The introduction in 1894 of complex resistances, or impedances, in treating alternating-current circuits is the work of the American engineer Charles Steinmetz (1865–1923). For the quantitative treatment of phenomena in networks that are powered by generators with complex temporal voltage curves, a most original, almost magical, method was given by Oliver Heaviside (1850–1925), a method that was able to be justified mathematically only with difficulty via the methods of the Laplace transform and the theory of distributions.



▶ **Figure 4.82** Georg Simon Ohm (1789–1854): Was senior teacher of mathematics and physics at the gymnasium in Cologne (1817–1828); from 1833 directed the polytechnic school in Nuremburg, and was appointed professor at the University of Munich in 1849.

In addition to his well-known discoveries in basic electrical circuits, Ohm did significant work in acoustics. He investigated the role of overtones in human hearing (1843). Helmholtz seized on Ohm's work on hearing in working out his resonance theory.

▼ **Figure 4.83** Ohm's law in its original form.



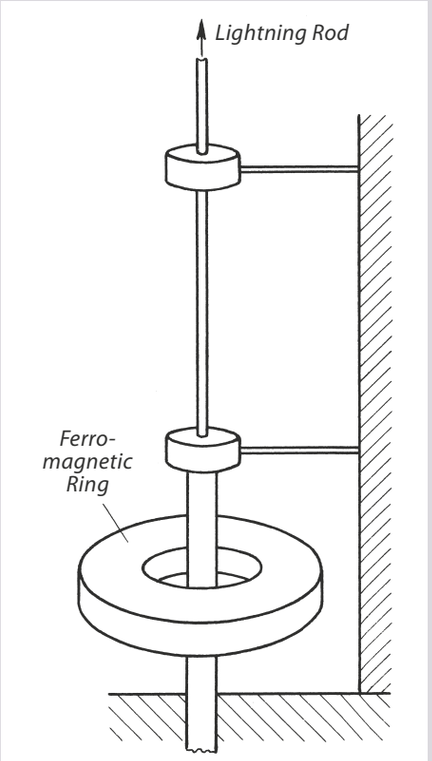
4.4.6 *The Magnetic Field of Electric Currents: Cross-Fertilization from Natural Philosophy*

Throughout the first two decades of the nineteenth century, experimenters were able to make use of equipment that could produce constant currents of suitable strength to bring conductors to incandescence and to carry out electrochemical investigations. Therefore, it seems surprising that the magnetic effects of current were only discovered in 1820.

At the beginning of the nineteenth century, a number of observations should have suggested to investigators that there is some kind of connection between magnetic fields and electrical current or, as one would have said at the time, the flow of the electric *fluidum* produces a magnetic effect in the surroundings. For instance, it was already known that in a house stuck by lightning, steel objects—knives, for example—that were close to the lightning strike became magnetized. Even today it is still a common practice to measure the very high currents (on the order of 100,000 amperes) that occur in lightning by using the magnetic effect (Figure 4.84). At that time, however, this was given little attention because no one was looking for a connection between electricity and magnetism. COULOMB's writings, as AMPÈRE remarked in a letter (Quotation 4.28), excluded all such possibilities. Strangely, the impulse to seek such a relationship came from philosophy: The extreme mechanical materialism emanating from the rationalism of the eighteenth century was protested against by Romanticism in art, literature, and philosophy. This movement emphasized a more unified and dynamic description of nature and mankind. In SCHELLING's natural philosophy, all natural phenomena are represented as diverse manifestations of a single fundamental principle, in constant battle with one another, but eventually reaching equilibrium.

ØRSTED embraced this philosophy; he spent years searching for a connection between electricity and magnetism. In this regard, natural philosophy exerted an immediate positive influence on the development of physics, and we have to rate this influence even more highly when we consider that FARADAY was also thinking along such lines. As is clear from ØRSTED's memoirs (Quotation 4.29), such a unifying natural-philosophical point of view can have disadvantages as well. ØRSTED assumed at the outset that the magnetic effect should emanate from an electrical conductor like light or heat and together with light and heat. For this reason, he started by looking for the magnetic effect around conductors that glowed with current. He chose a very thin platinum wire as the conductor because that could be made to glow readily. In fact, a weak current was sufficient for heating the thin wire, but this worked against the success of the experiment.

ØRSTED's discovery was of a purely qualitative character (Figure 4.85), and the theory that he proposed contributed neither an explanation of the phenomenon nor useful suggestions for further experimentation. Nevertheless, it was so completely unexpected that it received great attention in Europe. ØRSTED sent his article, which was written in Latin, to all the relevant scientific societies in Europe. It is already clear from the letter of AMPÈRE cited above (Quotation 4.28) that there was a general reluctance to believe in the correctness of the observation. However, the speed with which further theoretical and experimental results were achieved in this area proves that the leading intellects of the time had soon completely accepted this idea. Both the necessary experimental equipment and the requisite mathematical apparatus were at hand, so that within a few years, the associated theoretical description as we know it today had been completed.



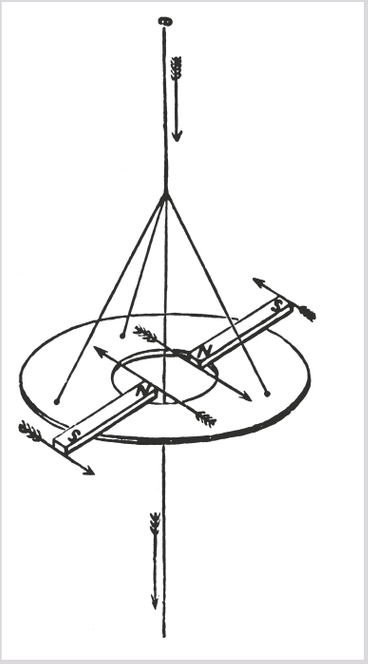
▲ **Figure 4.84** Even today, the magnitude of the current in a lightning strike (up to hundreds of thousands amperes) is determined from the magnetization of a ring made of ferrous material slipped onto the grounding stake.

Quotation 4.29

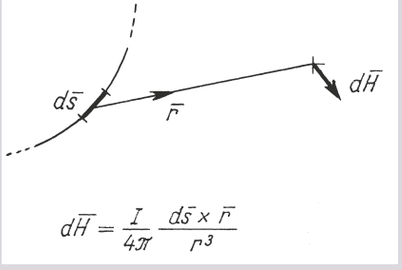
Electromagnetism itself was discovered in the year 1820, by Professor HANS CHRISTIAN OERSTED, of the University of Copenhagen. Throughout his literary career, he adhered to the opinion, that the magnetic effects are produced by the same powers as electrical. He was not so much led to this, by the reason commonly alleged for this opinion, as by the philosophical principle, that all phenomena are produced by the same original power. ...

In the month of July 1820, he again resumed the experiment, making use of a much more considerable galvanical apparatus. The success was now evident, yet effects were still feeble in the first repetitions of the experiment, because he employed only very thin wires, supposing that the magnetical effect would not take place, when heat and light were not produced by the galvanical current; but he soon found that conductors of a greater diameter give much more effect; and he then discovered, by continued experiments during a few days, the fundamental law of electromagnetism, viz. *that the magnetical effect of the electrical current has a circular motion round it.*

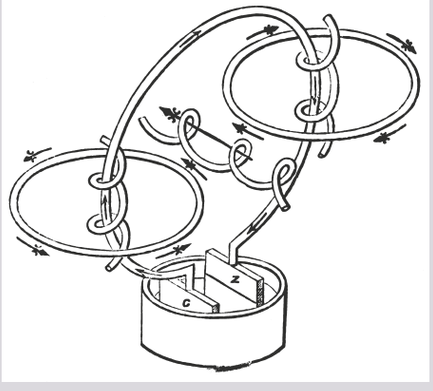
—ØRSTED, Article about his own discovery in *The Edinburgh Encyclopaedia* [Williams 1966, pp. 56, 58]



◀ **Figure 4.85** Ørsted's experiment as illustrated by Maxwell half a century later.



◀ **Figure 4.86** The Biot–Savart law as expressed today.



◀ **Figure 4.87** Maxwell explained the relationships between a current and a magnetic field with such drawings.

In 1820, the year of Ørsted's discovery, Biot and Savart gave a quantitative description of the magnetic effect generated by a current flowing through a conductor at any point in space. To measure the force, they employed a magnetized needle and used its period of oscillation as a measure of the intensity of the magnetic effect. Biot and Savart were primarily experimentalists, and their measurements were limited to two simple setups. Laplace then helped them to work out the precise expression of the law that followed. He pointed out that the measured effect could be understood as the sum of the effects of short segments of the conductor; the effect of each segment—a differential *current element*—being inversely proportional to the square of the distance to the point of observation and proportional to the current, to the length of the segment, and to a trigonometric factor expressing the apparent shortening of the segment due to its angle as "seen" from the observation point (Figures 4.86, 4.87).

Biot investigated the dependence of Earth's magnetic field on altitude. To this end, he ascended in a balloon with Gay-Lussac, reaching an altitude of more than 6000 feet (2000 meters) after overcoming great difficulties (Figure 4.88). (We note in passing that Gay-Lussac made a lone ascent in a balloon up to an altitude of 23,000 feet.) Through their measurements, they determined that the strength of Earth's magnetic field remains essentially unchanged up to the altitudes that they reached.

4.4.7 *The Interaction of Currents: An Extension of Newton's Ideas*

Also in the year 1820, André Marie Ampère (Figure 4.89) investigated the interaction of currents experimentally and proposed a mathematical theory of this interaction (Figure 4.90). Maxwell considered Ampère's work in this area to be among the most significant contributions ever in the history of science and called its author the "Newton of electrodynamics." He viewed Ampère's work as exemplary in the rigor of its logic, although it was clear to him that Ampère did not arrive at his results in the way suggested by his publications (Quotation 4.30).

On the basis of experimental observations and the application of Newton's natural philosophy, Ampère sought to derive a law for determining the strength of the interaction between two current elements. He based his derivation on the following experimental observations (Figures 4.91, 4.92):

1. The force is proportional to the product of the two currents.
2. The force is unchanged if, in the case of constant current intensities, we multiply all lengths—that is, the distances of the current elements from one another and the lengths of the current elements themselves—by a constant factor.
3. The net force exerted by a circuit on a given current element of another circuit is always perpendicular to that current element.
4. The forces between two current elements satisfy Newton's third law: they are equal to each other and are directed along the line joining them.

Beginning with these axioms, Ampère arrived at the following law of forces:

$$\mathrm{d}^2 F_A = \frac{ii' \,\mathrm{d}s\,\mathrm{d}s'}{r^2}\left(\cos\varepsilon - \frac{3}{2}\cos\theta\cos\theta'\right).$$

The meaning of the symbols in the formula is clarified in Figures 4.91 and 4.92. Using today's vector notation, this relationship can be expressed as

$$\mathrm{d}^2 \mathbf{F}_A = ii'\mathbf{r}\left(\frac{\mathrm{d}\mathbf{s}\,\mathrm{d}\mathbf{s}'}{r^3} - \frac{3}{2}\frac{(\mathrm{d}\mathbf{s}\cdot\mathbf{r})(\mathrm{d}\mathbf{s}'\cdot\mathbf{r})}{r^5}\right). \qquad (1)$$



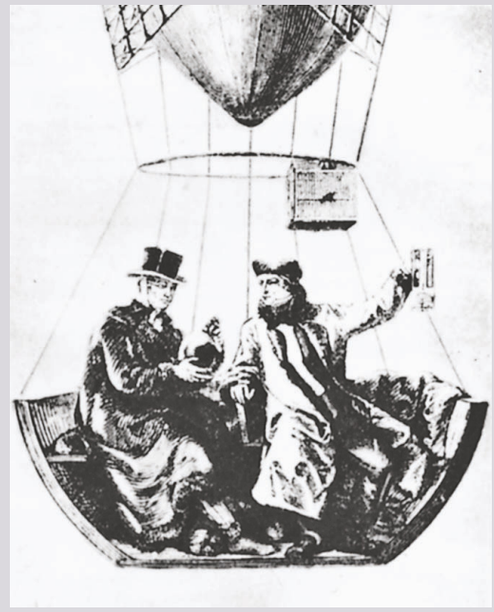
▲ **Figure 4.89** André Marie Ampère (1775–1836): Taught physics in Bourg and Lyon, then professor at the École Polytechnique; in 1820, he derived the law describing the electrodynamic interaction of currents. Today, the influence of substances on a magnetic field can still be best understood with the model of Ampère's molecular current.

In today's physics texts, one finds a different relationship for the interaction of two current elements, because in the form given above, the force law cannot easily be brought into agreement with the Biot–Savart law for the magnetic effect of a current element. The reason is that today we do not require that the principle *action = reaction* be satisfied for every current element. However, because Ampère's formula does a good job of describing all measurements relating to a closed circuit, the discrepancy that we mentioned indicates only that the decomposition of the resultant force into a sum of forces that can be associated with the individual current elements is not unique. In fact, the relationship for the resultant force does not change if on the right-hand side of Ampère's formula (1) a term is added that disappears upon integration over a closed circuit and therefore has no effect on the result. A simple expression of this sort is

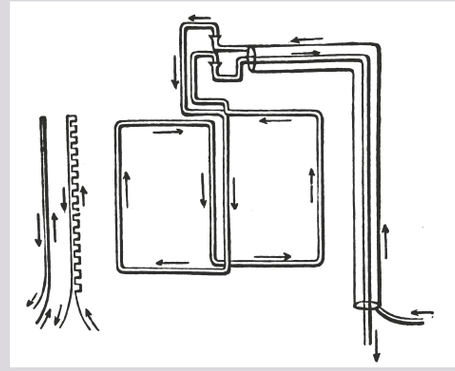
$$(\mathrm{d}\mathbf{s}\cdot\mathbf{r})\,\mathrm{d}\mathbf{s}'.$$

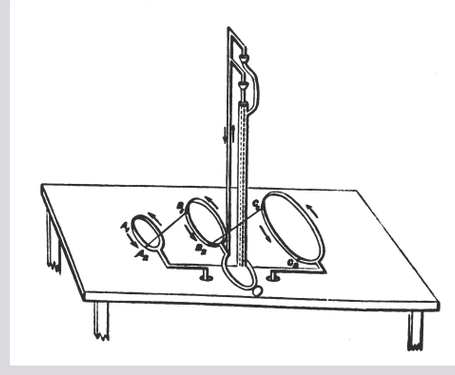
This expression may be summed, for fixed d$\mathbf{s}'$, over all d$\mathbf{s}$ (that is, over a closed circuit S with elements d$\mathbf{s}$). We see, then, from Figure 4.92 that d$\mathbf{s}$ = –d$\mathbf{r}$, and consequently,

$$\mathrm{d}\mathbf{s}\cdot\mathbf{r} = -(\mathbf{r}\cdot\mathrm{d}\mathbf{r}) = -\frac{1}{2}\mathrm{d}\mathbf{r}^2.$$



▲ **Figure 4.88** Biot and Gay-Lussac ascending in a balloon to measure the magnetic field of Earth [Millikan 1965].





▲ **Figure 4.90** Ampère's experiment as presented by Maxwell. The illustrations in Ampère's original publications are filled with so many details that they tend to obscure the essential ideas.

Quotation 4.30

The experimental investigation by which Ampère established the laws of the mechanical action between electric currents is one of the most brilliant achievements in science.

The whole, theory and experiment, seems as if it had leaped, full grown and full armed, from the brain of the "Newton of electricity." It is perfect in form, and unassailable in accuracy, and it is summed up in a formula from which all the phenomena may be deduced, and which must always remain the cardinal formula of electrodynamics.

The method of Ampère, however, though cast into an inductive form, does not allow us to trace the formation of the ideas which guided it. We can scarcely believe that Ampère really discovered the law of action by means of the experiments which he describes. We are led to suspect, what, indeed, he tells us himself, that he discovered the law by some process which he has not shewn us, and that when he had afterwards built up a perfect demonstration, he removed all traces of the scaffolding by which he had raised it.

Faraday, on the other hand, shews us his unsuccessful as well as his successful experiments, and his crude ideas as well as his developed ones, and the reader, however inferior to him in inductive power, feels sympathy even more than admiration, and is tempted to believe that, if he had the opportunity, he too would be a discoverer. Every student should therefore read Ampère's research as a splendid example of scientific style in the statement of a discovery, but he should also study Faraday for the cultivation of a scientific spirit, by means of the action and reaction which will take place between the newly discovered facts as introduced to him by Faraday and the nascent ideas in his own mind.

—James Clerk Maxwell, *A Treatise on Electricity and Magnetism* [pp. 175–176]

We have thus to integrate a total differential over a closed curve, and therefore, as is well known, the integral vanishes.

Of course, we may also investigate total differentials of a more complicated form. For example, if we add the total differential

$$\mathrm{d}\left[\mathbf{r}(\mathrm{d}\mathbf{s}'\cdot\mathbf{r})\frac{ii'}{r^3}\right], \quad -\mathrm{d}\mathbf{r}=\mathrm{d}\mathbf{s},$$

to the force $\mathrm{d}^2\mathbf{F}_A$, we obtain

$$\mathrm{d}^2\mathbf{F}=\mathrm{d}^2\mathbf{F}_A+\mathrm{d}\left[\mathbf{r}(\mathrm{d}\mathbf{s}'\cdot\mathbf{r})\frac{ii'}{r^3}\right]=\frac{ii'}{r^3}\left[(\mathrm{d}\mathbf{s}\cdot\mathbf{r})\mathrm{d}\mathbf{s}'+(\mathrm{d}\mathbf{s}'\cdot\mathbf{r})\mathrm{d}\mathbf{s}-(\mathrm{d}\mathbf{s}\cdot\mathrm{d}\mathbf{s}')\mathbf{r}\right].$$

This expression is completely symmetric with respect to the current elements, so both forces are equal. However, they are no longer parallel to the line joining the current elements. If we integrate $\mathrm{d}^2\mathbf{F}$ over the circuit S and take into account the vectorial relationship

$$(\mathrm{d}\mathbf{s}'\cdot\mathbf{r})\mathrm{d}\mathbf{s}-(\mathrm{d}\mathbf{s}\cdot\mathrm{d}\mathbf{s}')\mathbf{r}=\mathrm{d}\mathbf{s}'\times(\mathrm{d}\mathbf{s}\times\mathbf{r})$$

as well as the fact that the integral of the quantity $(\mathrm{d}\mathbf{s}\cdot\mathbf{r})$ over a closed path is zero, we then obtain

$$\mathrm{d}\mathbf{F}=ii'\oint\left[\mathrm{d}\mathbf{s}'\times\frac{\mathrm{d}\mathbf{s}\times\mathbf{r}}{r^3}\right]=i'\mathrm{d}\mathbf{s}'\times\oint_S i\frac{\mathrm{d}\mathbf{s}\times\mathbf{r}}{r^3}=i'\mathrm{d}\mathbf{s}'\times\mathbf{B}$$

for the force exerted by the circuit S on the current element $i'\mathrm{d}\mathbf{s}'$. This completely corresponds to the notion that the circuit S produces a magnetic field $\mathbf{B}$ at the location of the current element $i'\mathrm{d}\mathbf{s}'$ according to the Biot-Savart law:

$$\mathbf{B}=\oint_S i\frac{\mathrm{d}\mathbf{s}\times\mathbf{r}}{r^3}$$

and that this field, in turn, exerts the following force to the current element $i'\mathrm{d}\mathbf{s}'$ placed there:

$$\mathrm{d}\mathbf{F}=i'\mathrm{d}\mathbf{s}'\times\mathbf{B}.$$

Looking at it in this manner, it is no longer the case that the principle *action = reaction* is satisfied for the force between two current elements.

Ampère also gave a microphysical interpretation of the magnetic properties of matter—one that is still a useful model for understanding such properties today.

Ampère noticed and then proved theoretically that a current loop and a very flat magnet (magnetic double layer) generate exactly the same fields in the surrounding space (Figure 4.93). On the basis of this observation, he then associated small currents—the Ampère molecular circular currents—with the volume element of a larger body; the fields produced by these currents will then combine in the surrounding space. The total outward effect will depend on the orientation of these circuits with respect to one another and with respect to the external magnetic field (Figure 4.94).

Ampère's law of interaction was revised 20 years later by Wilhelm Eduard Weber (1804–1891). Although this revision led to a dead end, it is still worth mentioning here because in it we can see an approach to the later classical theory of electrons. Furthermore, the limits placed on a precise description of electromagnetic phenomena in the framework of Newton's theory of action at a distance become clear. Weber hypothesized that in an electrical current within a conductor, equal quantities of charge flow in each direction at the same velocity. According to Weber's theory, one therefore has

$$i=2\lambda u, \quad i'=2\lambda' u',$$

where λ is the charge per unit length and u the charge velocity. If we now consider Ampère's law, equation (1), together with the relations

$$\frac{\mathrm{d}r}{\mathrm{d}t}=\frac{\mathrm{d}r}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t}=\frac{\mathrm{d}r}{\mathrm{d}s}u, \quad \mathrm{d}\mathbf{s}\cdot\mathbf{r}=-r\frac{\mathrm{d}r}{\mathrm{d}s}\mathrm{d}s,$$

then after some rearrangement, we may express the force between the moving charges as

$$F = \frac{Q_1 Q_2}{r^2}\left[1 - \frac{1}{c^2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + \frac{2r}{c^2}\frac{\mathrm{d}^2 r}{\mathrm{d}t^2}\right]. \qquad (2)$$

The first term in this formula clearly describes the electrostatic force, while the other two terms describe electrodynamic forces. There is no need to criticize this formula in detail; that was already done by Weber's contemporaries. They noted, for example, that in a very simple arrangement, one could show that a moving charge would have to exert a magnetic field even on a stationary charge. Nevertheless, it is worth pointing out a conclusion that is not so obvious: Weber's theory relates magnetic interactions exclusively to the motion of the electrical charge, with the charge velocity and acceleration appearing as characteristic quantities in the formula. (In this respect, the theory agrees with the later theory of the electron.) However, if we recall the analogy between gravitational and electrostatic forces, which was the model for the derivation of Coulomb's law, then perhaps we should not be surprised that the analogy may now be considered in the opposite direction (Figure 4.95). Thus, astronomers attempted to take the velocities of the heavenly bodies into account in their calculations of the gravitational force, and they obtained, on the basis of Weber's formula (2), a relationship of the form

$$F \sim \frac{m_1 m_2}{r^2}\left[1 - \frac{1}{h^2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + \frac{2r}{h^2}\frac{\mathrm{d}^2 r}{\mathrm{d}t^2}\right],$$

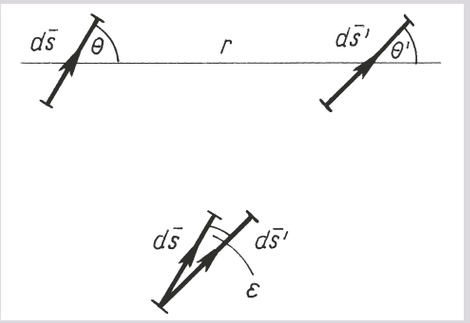
where h denotes the speed of propagation of the gravitational effect (Tisserand, 1872). We meet here for the first time the perihelion motion of Mercury as an effect that would later become the touchstone for the validity of new theories of gravitation. Even if one takes into account the perturbing actions of the other planets on Mercury's orbit, when using the Newtonian theory of gravitation, the degree of agreement between the observed and calculated values of the advance of the perihelion is less than satisfactory. A remainder of 38″ cannot be accounted for. The above formula provides an additional contribution of 14″ if the speed of propagation of gravity is set at the velocity of light. It was Riemann who later succeeded in finding a force law that could explain the complete 38″. The perihelion motion of Mercury will appear again in this book as an experimental proof of the correctness of Einstein's general theory of relativity (Section 5.2).

4.4.8 *Faraday: The Greatest of the Experimentalists*

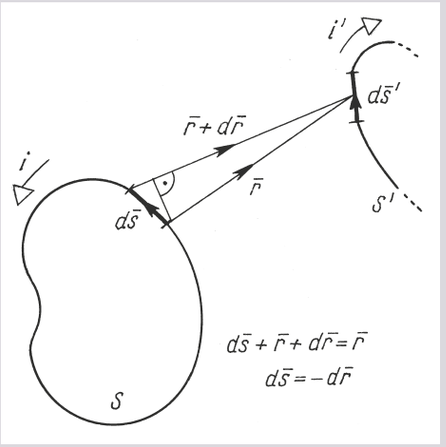
Let us now return to the problems in electrodynamics that were being investigated during the first years of the 1820s. At this time, Faraday (Figures 4.96, 4.98), who is often cited as the most important experimental physicist ever, began his investigations of electromagnetic phenomena. In 1821, he was asked to prepare a paper summarizing what was known in the field up to then and, like Priestley before him, he reproduced all the earlier experiments he was reporting on. In this report, Faraday already superseded his predecessors; for example (cf. Figure 4.97), he proved that a current flowing through a conductor exerts a force on a pole of a bar magnet and showed that this force acts along a circle, incidentally giving the principle of the first electric motor.

Figure 4.98 collects all the areas in which Faraday made significant contributions over the course of his life. Here we review briefly the most important concrete results:

1. The induction theorem (now named after Faraday) is the first of these. Influenced by Romantic natural philosophy, a number of researchers expressed the opinion that, analogously to the phenomenon of the influence of electrostatics on the current in a circuit, there should be some sort of influence of the current in one circuit on the current in a second circuit (Figure 4.99). Based on this analogy, an effect, for example, should appear whereby a current would generate a current in a neighboring conductor simply because such a current is flowing in the first circuit. Even Ampère conjectured something of this kind, although in a letter in 1822, he mentioned that such an effect does not exist and there is merely something to be observed when the current is switched on or off. However, Ampère did not investigate this

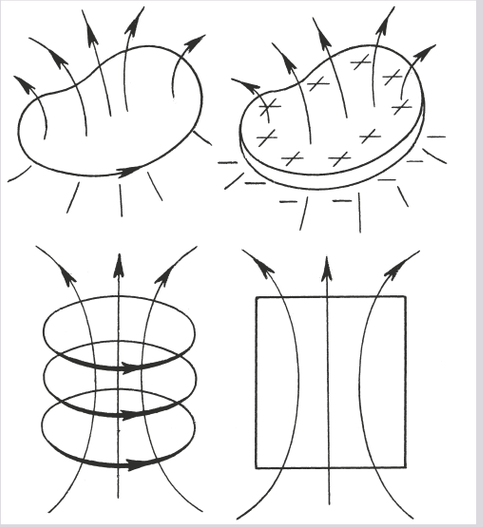


▲ **Figure 4.91** Definitions of the notation used in Ampère's force law.

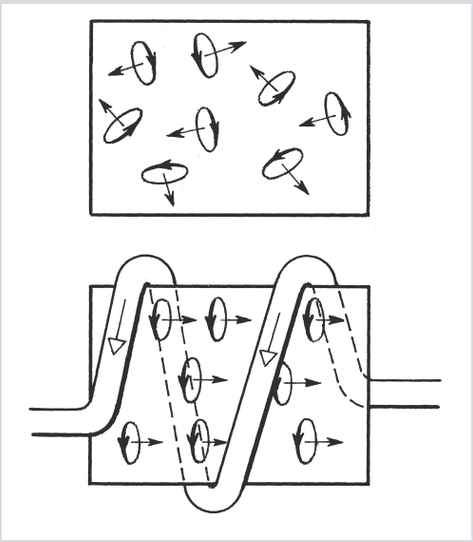


▲ **Figure 4.92** If we fix the path element d**s′**, then for the differential path element d**s** on the circuit S, we have d**s** = –d**r**, so that –(**r** · d**r**) can be rewritten as follows:

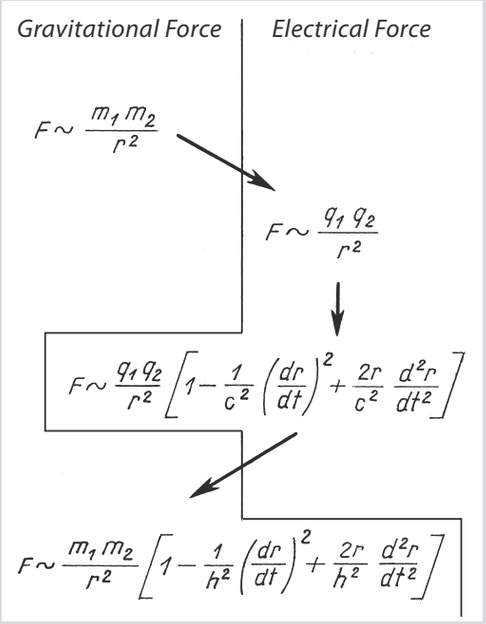
$$-(\mathbf{r}\cdot\mathrm{d}\mathbf{r}) = -\mathbf{r}\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}\mathbf{s}}\mathrm{d}\mathbf{s}.$$



▲ **Figure 4.93** Ampère realized that the magnetic field of a current is equal to the field of a magnetic double layer. It follows that the magnetic field of a coil with current flowing through it and the magnetic field of a bar magnet (**B**-field) have the same form.



▲ **Figure 4.94** In a magnetic material, the Ampère circular currents align under the influence of an external magnetic field.

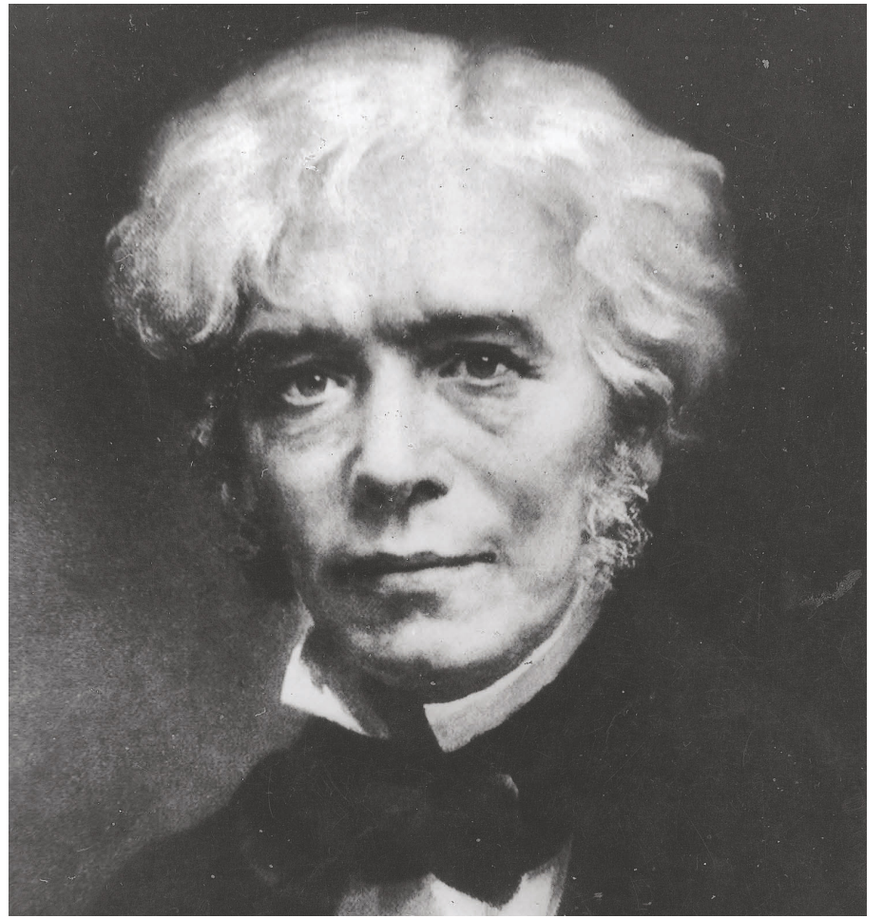


▲ **Figure 4.95** Coulomb's law was formulated in analogy to Newton's law of gravitation, and Weber's law served as the starting point for a new law of gravitation that was supposed to be more precise, but turned out to be incorrect.

latter effect, thereby missing out on—as he himself later bitterly remarked—a very significant discovery. Even in the lists of experiments that Faraday kept, we find over the course of many years the entry "no effect." Eventually, Faraday noticed the same phenomenon that Ampère had seen but ignored: To obtain an effect, something in the circuit had to change or—to put it more generally—the magnetic state of the entire system had to change. Faraday also found that the current arising in the second circuit is inversely proportional to the total resistance of this circuit, in other words, that the voltage created by a change in the magnetic state is independent of the type of conductor, for only then can the current be exactly proportional to the reciprocal of the resistance of the circuit. Faraday provided a quantitative formulation of this law using the simple and expressive expedient of the lines of force (Figure 4.100):

In a resting circuit loop, the induced voltage is proportional to the change in the number of lines of force per unit time ($U = -d\Phi/dt$).

For a moving conductor, the induced voltage is proportional to the number of force lines intersected per unit time [$U = \mathbf{I}(\mathbf{v} \times \mathbf{B})$].



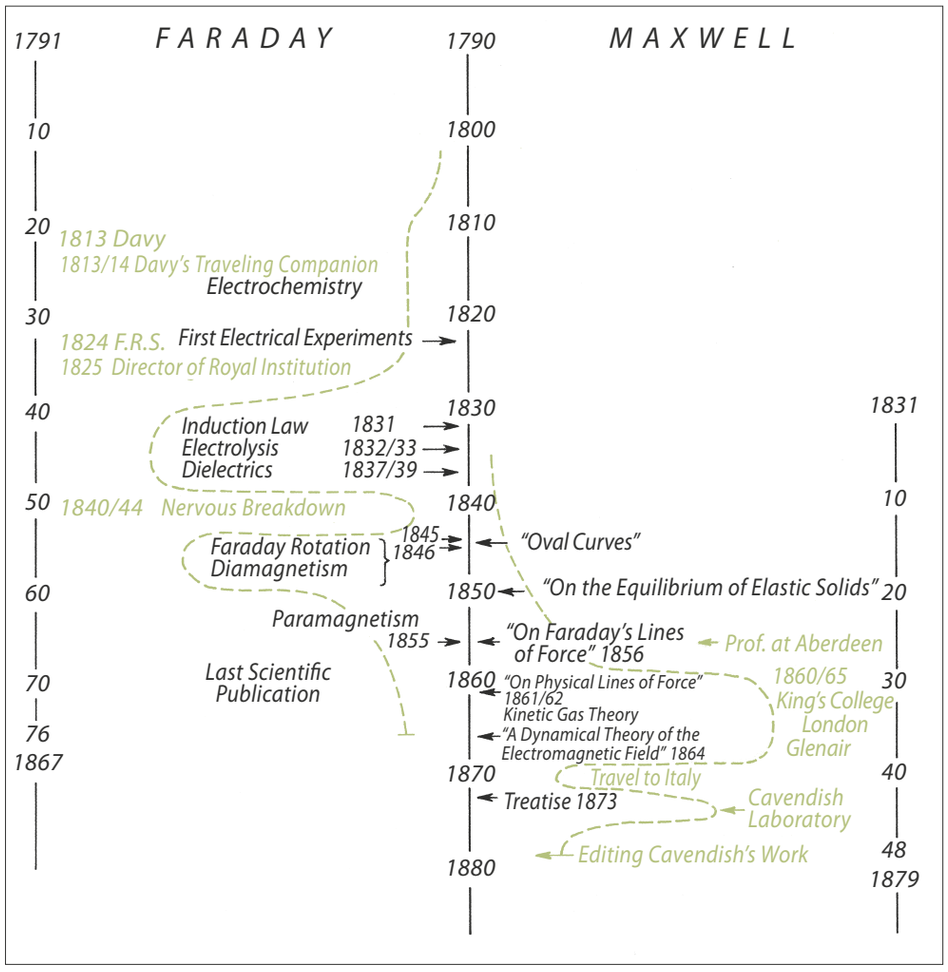
▲ **Figure 4.96** Michael Faraday (1791–1867): Apprenticed to a bookbinder, then laboratory assistant, and finally secretary to Davy. 1824, member of the Royal Society; from 1825, director of the Royal Institution. His first works are devoted to problems in chemistry (1823: *On Fluid Chlorine*). He also worked on technological problems: production of stainless steel and glass with particular optical properties. After 1820, he began to study electricity. 1821: construction of the "rotation apparatus" (Figure 4.97); August 29, 1831: law of induction; 1833: laws of electrolysis; 1845: Faraday effect. In the years 1832–1856, his work was published as *Experimental Researches in Electricity* in successive paragraphs numbered 1 to 3340.

Faraday is seen as the most important experimental physicist of all time, with whom, at most, Rutherford can be compared.

Today, it is not really necessary to mention the practical significance of the induction law: It serves as the theoretical basis for the operation and the engineering design of all electrical generators and transformers.

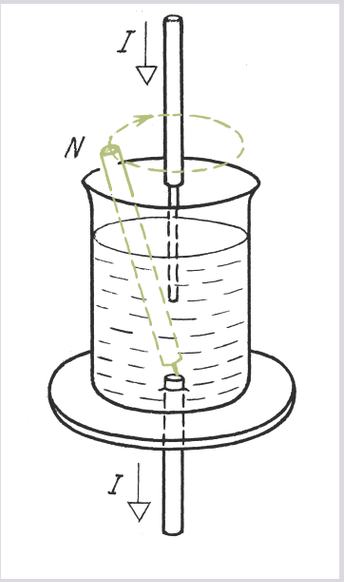
2. The laws of electrolysis derived by FARADAY in the next two years are also of great significance. These are now known as Faraday's laws.

The first law states that the quantity of separated substances produced during electrolysis is proportional to the charge flow. By the second law, well-defined chemically meaningful values are assigned to the proportionality

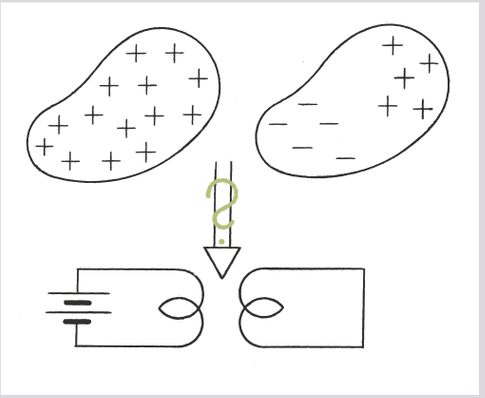


▲ **Figure 4.98** Life and work of the two greatest figures in the field of electrodynamics who also maintained scientific contact with each other and greatly respected each other.

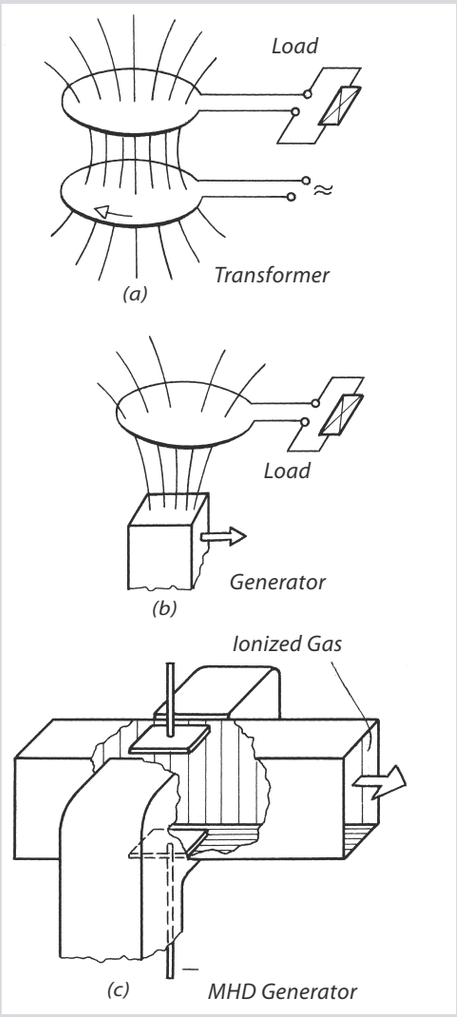
> … before I began the study of electricity I resolved to read no mathematics on the subject till I had first read through FARADAY's *Experimental Researches on Electricity.* …
>
> As I proceeded with the study of FARADAY, I perceived that his method of conceiving the phenomena was also a mathematical one, though not exhibited in the conventional form of mathematical symbols. I also found that these methods were capable of being expressed in the ordinary mathematical forms, and thus compared with those of the professed mathematicians. …
>
> It is of great advantage to the student of any subject to read the original memoirs on that subject, for science is always most completely assimilated when it is in the nascent state, and in the case of FARADAY's *Researches* this is comparatively easy, as they are published in a separate form, and may be read consecutively. If by anything I have here written I may assist any student in understanding FARADAY's modes of thought and expression, I shall regard it as the accomplishment of one of my principal aims—to communicate to others the same delight which I have found myself in reading FARADAY's *Researches*.
>
> —JAMES CLERK MAXWELL, Preface to *A Treatise on Electricity and Magnetism* [Vol. I, pp. viii, ix, xi]



◀ **Figure 4.97** With this setup, FARADAY clearly demonstrated the circular magnetic lines of force and gave us the first electric motor as well.



▲ **Figure 4.99** An electrically charged conductor causes the charges to separate in a nearby conductor (electrostatic induction). Could there be a similar phenomenon between two circuits?



▲ **Figure 4.100 (a)** Faraday's law of induction: If the number of field lines surrounded by a circuit changes, then a voltage is induced in it. The emphasis here is on change, on movement, on the dynamic aspects.

> Have had an iron ring made (soft iron), iron round and 7/8 inches thick and ring 6 inches in external diameter. Wound many coils of copper wire round one half, the coils being separated by twine and calico – there were three lengths of wire each about 24 feet long and they could be connected as one length or used as separate lengths. By trial with a trough each was insulated from the other. Will call this side of the ring *A*. On the other side but separated by an interval was wound wire in two pieces together amounting to about 60 feet in length, the direction being as with the former coils; this side call *B*.
>
> Charged a battery of 10 pr. plates 4 inches square. Made the coil on *B* side one coil and connected its extremities by a copper wire passing to a distance and just over a magnetic needle (3 feet from iron ring). Then connected the ends of one of the pieces on *A* side with battery; immediately a sensible effect on needle. It oscillated and settled as last in original position. On *breaking* connection of *A* side with battery again a disturbance of the needle.

continued on next page

factor that appears in the first law; to wit, it is stated that with equal quantities of charge flow, the quantities of separated material are directly proportional to the element's equivalent weight (Figure 4.101). In addition to its practical ramifications, this law has tremendous theoretical significance, because it makes it possible to descend into the world of atoms and demonstrate relationships between microphysical quantities: $M = (1/9.65 \times 10^7) \times (A/z)It$, giving the mass produced in terms of the current I, the time t, the atomic weight A, and the valence z.

3. Faraday also did important work on dielectrics; indeed, the very word "dielectric" came from Faraday. He introduced the notion of dielectric constant into physics and specified the methods for measuring it (Figure 4.102).
4. Faraday conjectured that a magnetic field should have an effect beyond that which it has on so-called magnetic materials. He thought that all matter without exception must exhibit some magnetic properties. He categorized materials as paramagnetic and diamagnetic, and then made a thorough investigation of diamagnetic materials.
5. The assumption that there must be some interrelation among various physical phenomena led Faraday to the discovery of yet another effect, having both practical and theoretical importance, which also bears his name. Faraday was searching for an interaction between magnetic fields and light, and following a number of fruitless experiments, he finally discovered that the plane of oscillation of linearly polarized light is rotated in certain media if they are located within a magnetic field (Figure 4.103). This rotation of the polarization plane is called the Faraday rotation.
6. We note as an aside that in his later years, Faraday searched without success for an interaction between gravitation and electromagnetism (Quotation 4.31). Even to this day, there has still been no successful demonstration of such a relationship.

We have yet to mention another extremely important contribution by Faraday. We will refrain from speaking about it in superlatives only because we would then have to do so for all of Faraday's work. This contribution has to do with the introduction of a completely new way of looking at electromagnetic phenomena. Moreover, here we encounter a strange phenomenon in the history of science, which calls for a brief explanation. The Cartesian vortex theory was abandoned in France toward the end of the seventeenth century and the beginning of the eighteenth century in favor of Newton's theory of action at a distance. The generation of mathematicians and theoretical physicists active at the end of the eighteenth century—whose leading representatives were Lagrange, Laplace, Poisson, and Ampère—had made Newtonian principles so much a part of their own thinking that they dealt with the new phenomenon of electromagnetism in terms of these principles. We have seen a concrete example of this in the case of Ampère, who in his derivation of the fundamental law of electrodynamics began with the Newtonian worldview. Because of the understandably great influence of the French mathematicians, no one dared to take a different approach. Faraday, however, was for the most part self-taught and never received a formal education. Maxwell, who was an admirer of Faraday and who brought Faraday's discoveries into a mathematical form to develop them further, continues the story thus in *A Treatise on Electricity and Magnetism*:

> For instance, Faraday, in his mind's eye, saw lines of force traversing all space where the mathematicians saw centers of force attracting at a distance; Faraday saw a medium where they saw nothing but distance; Faraday saw the

seat of the phenomena in real actions going on in the medium, they were satisfied that they had found it in a power of action at a distance impressed on the electric fluids. ...

It was perhaps for the advantage of science that Faraday, though thoroughly conscious of the fundamental forms of space, time, and force, was not a professed mathematician. He was not tempted to enter into the many interesting researches in pure mathematics which his discoveries would have suggested if they had been exhibited in a mathematical form, and he did not feel called upon either to force his results into a shape acceptable to the mathematical taste of the time, or to express them in a form which mathematicians might attack. He was thus left at leisure to do his proper work, to coordinate his ideas with his facts, and to express them in natural, untechnical language. [Vol. I, p. ix; Vol. II, p. 176]

As already mentioned, Faraday was influenced by Romantic natural philosophy. He was also influenced by Boscovich, who, as we have seen in Section 4.2, viewed the particles of matter as centers of force with an effect that spreads out over all of space and assumed that this effect was somehow inherent in all matter. Faraday, who was educated in the sober outlook of the empirical tradition, now sought to make these qualitative ideas more realistic with the help of concrete experiments and then attempted to use the ideas made more realistic in this way to interpret the phenomena. He made tangible the imagined lines of force with the help of iron filings. He produced images (Figure 4.104) like those that can be found today in any introductory physics textbook. As we have seen with Ampère, as well as with Weber, who developed Ampère's theory further, the theory of action at a distance is capable of explaining the interaction of circuits without having to postulate a medium in the space between them. Faraday, in contrast, assumed that in the regions surrounding a conductor with current flowing through it an electromagnetic field emerges with its lines of force. Thus, at the place of the second conductor, completely independently of whether we placed a conductor there or if there is current flowing in that conductor, that is in *space* itself, even in a vacuum, there will be a particular state distinguished from the state that would have been there in the absence of the current-carrying conductor in the vicinity. The first circuit therefore does not act directly on the second circuit, but rather by way of this "state of readiness." What is the historical significance of such a separation of the interaction process into two steps? Can this particular state of space be viewed as something that actually exists? One can show mathematically that as long as the changes are not too rapid, both formulations lead to the same results. However, if we concede the experimental fact that the effect of one circuit needs some time to reach the other, then the hypothesis of a medium of transmission is necessary, and its function is obvious. When Faraday was working out his idea of the electromagnetic field with the aid of electric and magnetic lines of force, the existence of electromagnetic waves had not yet been either theoretically founded or experimentally proven. However, the picture that Faraday had worked out for himself suggested to him the idea that electromagnetic effects could be transmitted by the lines of force in the form of transversal oscillations.

But Faraday's view is more than just the starting point for the future field theory, for his analogies between electric current and the "current" of magnetic field lines have proven very useful in power engineering and pedagogically in the visualization of complex magnetic phenomena.

Figure 4.100 *continued*

Made all the wires on *A* side one coil and sent current from battery through the whole. Effect on needle much stronger than before.

—Michael Faraday, Diary entry, August 29, 1831 [Elliott 1966, p. 257]

(b) Emil Lenz (1804–1865): Member of the academy at St. Petersburg, in 1833 formulated Lenz's law, according to which the direction of the induced current is such that its effect will hinder the cause that created it, so when a conductor is moved in a magnetic field, the induced current will flow so that the force exerted on it by the magnetic field will oppose its movement.

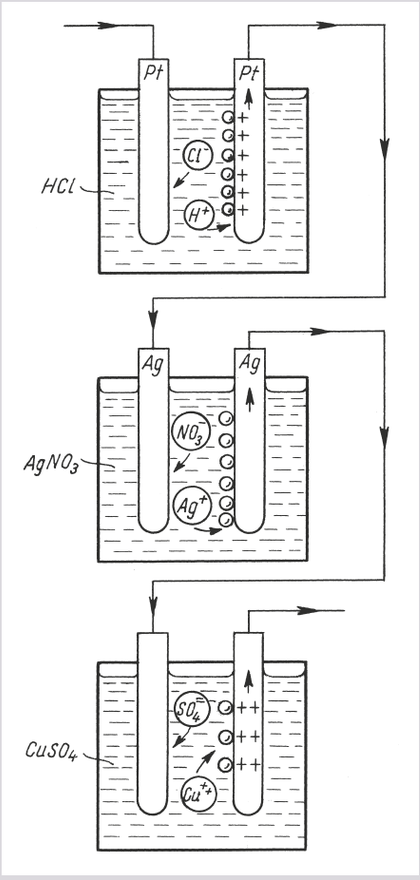
Joseph Henry (1797–1878): American physicist who came close to discovering the induction law. In 1828 he produced a very strong electromagnet by winding isolated wire into a multilayered coil. In 1832 he observed self-induction, and in 1842 determined experimentally that oscillations can result from the discharge of a capacitor.

Although in Faraday's experiments the dynamo principle (in the form of a copper disk rotating between the poles of a strong permanent magnet) and the transformer principle (in two coils wound around the same iron core) already played a role, it was a long way to a practical implementation of these ideas. Taking part in the development of the electric generator were Pixii (primitive alternating current generator, 1832), Clarke (rectification using a commutator, 1836), Pacinotti (1860), Gramme (ring-shaped armature, 1868), Ányos Jedlik (self-exciting dynamo, 1861), Siemens (recognition of the significance and practicality of the autoexcitation principle, 1866).

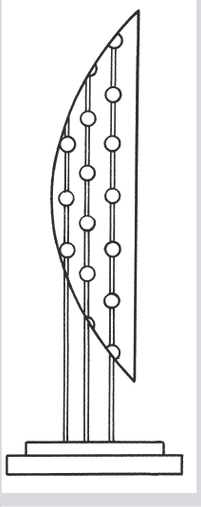
A number of researchers also took part in the perfection of the transformer; here we mention only Yablochkov (1876) and the triumvirate Déri–Bláthy–Zipernowsky (1885).

In the further development of generators, motors, transformers, and distribution systems, Nikola Tesla (1856–1943), who began his career in the Austro-Hungarian Empire (Budapest, 1880–1882), played an outstanding role: magnetic rotational field, multiphase system, asynchronous motor, Tesla coil.

(c) A modern magnetohydrodynamic (MHD) generator, transforming heat energy or kinetic energy directly into electricity, also employs the induction principle.



▲ **Figure 4.101** Demonstration of Faraday's electrolysis law. The details of the process, such as the indicated ions, were not yet known in his time.



◀ **Figure 4.102** Staring out from the model given by Faraday for dielectrics, antennas in modern microwave technology may be constructed of artificial dielectrics.

> If the space round a charged globe were filled with a mixture of an insulating dielectric, as oil of turpentine or air, and small globular conductors as shot, the latter being at a little distance from each other so as to be insulated, then these would, in their condition and action resemble what I consider to be the condition and action of the particles of the insulating dielectric itself. If the globe were charged, these little conductors would all be polar; if the globe were discharged they would all return to their normal state, to be polarized again upon the recharging of the globe.
>
> —Faraday, *Experimental Researches* [Tricker 1966, p. 82]

In his explanation of induction, Faraday speaks of an "electrotonic" state of a magnetic field—a designation that both Maxwell and Boltzmann took from him—but it is only the temporal change in this state that generates a measurable effect in the form of an electric field, or more precisely, according to Faraday, an induced electric voltage in a conductor.

4.4.9 *Maxwell: The Fundamental Laws of Electromagnetic Fields*

Faraday's ideas about the lines of force and the key role of the medium were so foreign to the leading physicists of the time that despite the recognition that they accorded the brilliant experimentalist, they could not see a sufficient foundation there for a new theory. There was but one exception, who, however, was to have a decisive influence on the further development of the theory of electromagnetism: James Clerk Maxwell (Figure 4.105).

Maxwell was 40 years younger than Faraday, born in the year in which Faraday discovered the law of induction. Despite their great difference in age, the two of them had an opportunity to exchange ideas, at first through correspondence, but later also in person. In their personal qualities Maxwell and Faraday are both among the most attractive personalities in the history of physics. In his modesty, Maxwell is perhaps surpassed by few among the great physicists. He presented his findings as though they were no more than a mathematical formulation of Faraday's ideas. For his part, Faraday spoke and wrote highly of Maxwell's unique ability to express the most difficult theoretical ideas and results in a way that is understandable even by mathematically less well educated experimental physicists and that serves as a source of ideas for further research (Quotation 4.32).

The right side of Figure 4.98, which we introduced in the previous section, summarizes the outward details of Maxwell's life and work. We see that his most productive years were those that he spent in London. In this period, he worked out the general theory of electromagnetic fields and the kinetic theory of gases. At the moment, we are of course only interested in the former of these.

As we have mentioned several times already, Maxwell made it his mission to clothe Faraday's ideas in mathematical garb, thereby making them more accessible to theoretical physicists. It is most instructive to follow the development of Maxwell's ideas from his first article, written in 1855, "On Faraday's Lines of Force," to his 1873 summation, *A Treatise on Electricity and Magnetism*. At the beginning of his research, Maxwell still considered it necessary to make use of mechanical models of electromagnetic phenomena, using them to derive quantitative relationships. In his early works, Maxwell appears to have taken the model too literally. However, the model gradually disappeared, and what remained were only the electromagnetic field equations. Figure 4.106 shows one of the mechanical models as well as drawings representing such models. One should note how the electromagnetic phenomena were explained with the help of the models. We may imagine—thus Maxwell began his argument—that a magnetic field consists of vortices circling about the magnetic field lines. For homogeneous magnetic fields, the model consists of an array of vortices that have the same direction of rotation. It is easily seen that vortices with the same direction of rotation would disturb each other if they touched. This gave Maxwell the idea to arrange for "idler wheels" between the vortices that rotate in the opposite direction and remain fixed in place if all the vortices turn at the same speed. In Figure 4.106(b)—following Maxwell—the vortices are displayed as hexagons for the sake of simplicity, so that

the role of the idler wheels can be more easily discerned. Using these, Maxwell illustrated the electric field according to the model. He assumed that the motion of the idler wheels carries with it an electric field or current. With the model, many experimental observations can be interpreted. For example, if a current flows in a conductor, then due to the movement of the idler wheels, the vortices, which represent the magnetic field, start to turn, meaning that a magnetic field is produced. More complex relationships can also be derived from the model. For example, let



▲ **Figure 4.105** James Clerk Maxwell (1831–1879): The most important theoretical physicist of the nineteenth century. He brought classical physics to a close. Studies in mathematics and physics in Edinburgh, then in Cambridge. 1856: appointed professor at Aberdeen, and from 1860 to 1865, taught at King's College, London. In 1865, he returned to his estate in Scotland to devote himself exclusively to scientific work. In 1871, he was named director of the newly organized Cavendish Laboratory at Cambridge, which would play an important role later in British scientific life.

As a young researcher, Maxwell worked on the problem of color vision; however, his crucial work was in the areas of electrodynamics and the kinetic theory of gases. He first formulated the Maxwell equations in his 1862 article *On Physical Lines of Force*. In 1873, his two-volume work *A Treatise on Electricity and Magnetism* appeared.

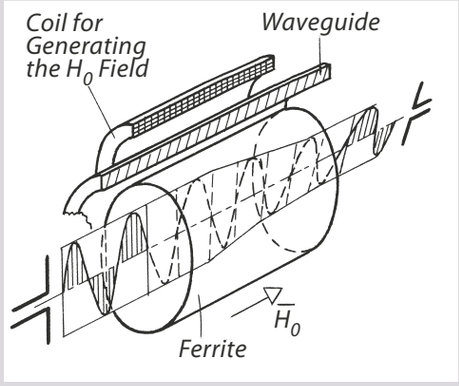
His contributions to kinetic gas theory are contained in *The Theory of Heat* (1871). Here, Maxwell was the first to emphasize that it is necessary and important to trace the fundamental units back to atomic constants.

> … it was his task to build and complete the classical theory, and in so doing he achieved greatness unequalled. His name stands magnificently over the portal of classical physics, and we can say this of him: by his birth, James Clerk Maxwell belongs to Edinburgh, by his personality he belongs to Cambridge, by his work he belongs to the whole world.
>
> —Max Planck, 1931 [Elliott 1966, p. 263]

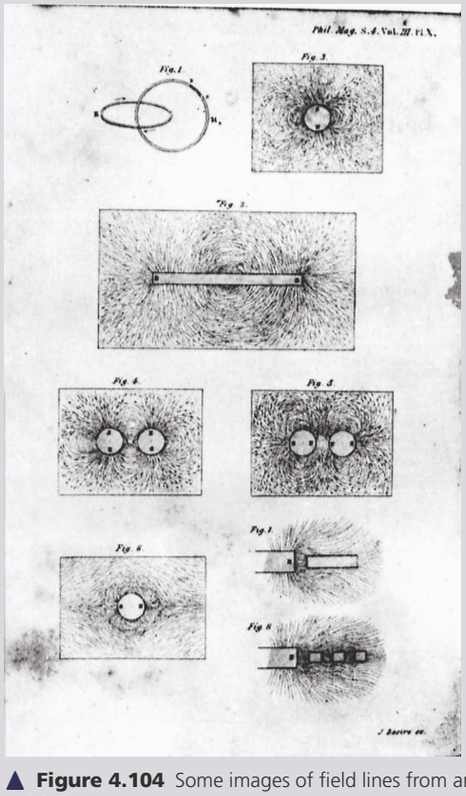
Quotation 4.31

Under the full conviction that the force of gravity is related to other forms of natural power, and is a fit subject for experiment, I endeavoured on a former occasion to discover its relation with electricity, but unsuccessfully. Under the same deep conviction, I have recently striven to procure evidence of its connection with either electricity or heat.

—Faraday [Tricker 1966, p. 26]



▲ **Figure 4.103** Faraday effect: In certain media, the polarization plane of linearly polarized light, which propagates parallel to the field lines of a magnetic field, is rotated. This effect is essentially the basis for the functioning of microwave and optical isolators in modern communication technology.



▲ **Figure 4.104** Some images of field lines from an article by Faraday.

Quotation 4.32

There is one thing I would be glad to ask you. When a mathematician engaged in investigating physical actions and results has arrived at his own conclusions, may they not be expressed in common language as fully, clearly, and definitely as in mathematical formulae? If so, would it not be a great boon to such as we to express them so—translating them out of their hieroglyphics that we might also work upon them by experiment. I think it must be so, because I have always found that you could convey to me a perfectly clear idea of your conclusions which, though they may give me no full understanding of the steps of your process, gave me the results neither above nor below the truth, and so clear in character that I can think and work from them.

—Faraday, Letter to Maxwell, November 13, 1857 [Tricker 1966, p. 25]

us say that the magnetic field in space is not constant and its intensity changes in the direction above to below, then in the model, the speed of rotation of the magnetic vortex also decreases in the direction above to below. In this case, due to the difference in the rotational speeds of the vortices on their two sides, the idler wheels cannot remain in the same place and will start to move; furthermore, they will move perpendicular to the direction in which the magnetic field is changing.

In this way, the spatial change in the magnetic field and the electric current are in close connection with each other.

Formulating the connection between the spatial change of the magnetic field in a particular direction and the electrical current mathematically yields, in fact, the law (for stationary current) known today as the first Maxwell equation,

$$\operatorname{curl}\mathbf{H} = \mathbf{J}.$$

Maxwell then sought a mathematically ascertainable quantity that would allow for the representation of Faraday's electrotonic state. It is in no way surprising that it is precisely this quantity that played a crucial role in Maxwell's theory (Quotation 4.33). However, in today's didactically revised version of Maxwell's theory, this quantity, called the vector potential and denoted by **A**, no longer is considered a fundamental quantity, but rather a derived one. The temporal change in this quantity is intimately connected to the induced electric field strength and the induced voltage, for Maxwell showed that the time derivative of its line integral over a closed path is equal to the induced voltage. But this is equivalent to the statement that the time derivative of the quantity itself provides the electric field for every spatial point of the electric field.

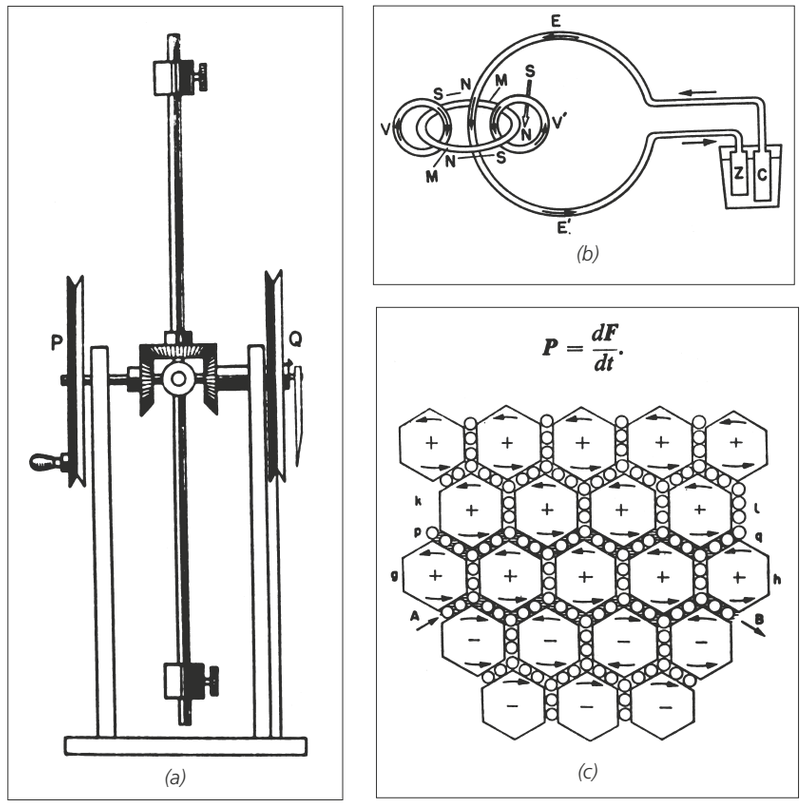
Somewhat more generally, we may also write

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \operatorname{grad}\varphi,$$

since the integral over a closed path does not change if we add a term to the integrand that can be represented as the gradient of a scalar quantity. It is worth noting that Maxwell also established his more general relation for a moving conductor with the help of a hydrodynamic analogy.

Using Stokes' theorem, the line integral of the vector potential can be transformed into

$$\oint_s \mathbf{A}\,d\mathbf{s} = \oint_a \operatorname{curl}\mathbf{A}\,d\mathbf{a}.$$



▶ **Figure 4.106** **(a)** Maxwell's mechanical model for visualizing phenomena that occur in inductively coupled circuits. **(b)** The current in the conductor creates the magnetic field. **(c)** The mechanical model of the electromagnetic field.

It appears therefore that certain phenomena in electricity and magnetism lead to the same conclusion as those of optics, namely, that there is an ethereal medium pervading all bodies, and modified only in degree by their presence; that the parts of this medium are capable of being set in motion by electric currents and magnets; that this motion is communicated from one part of the medium to another by forces arising from the connections of those parts; that under the action of these forces there is a certain yielding depending on the elasticity of these connections; and that therefore energy in two different forms may exist in the medium, the one form being the actual energy of motion of its parts, and the other being the potential energy stored up in the connections, in virtue of their elasticity.

—James Clerk Maxwell, *A Dynamical Theory of the Electrical Field* [p. 39]

It turns out to be useful here to introduce the magnetic induction **B** via the relationship

$$\operatorname{curl} \mathbf{A} = \mathbf{B},$$

whose surface integral gives the magnetic flux.

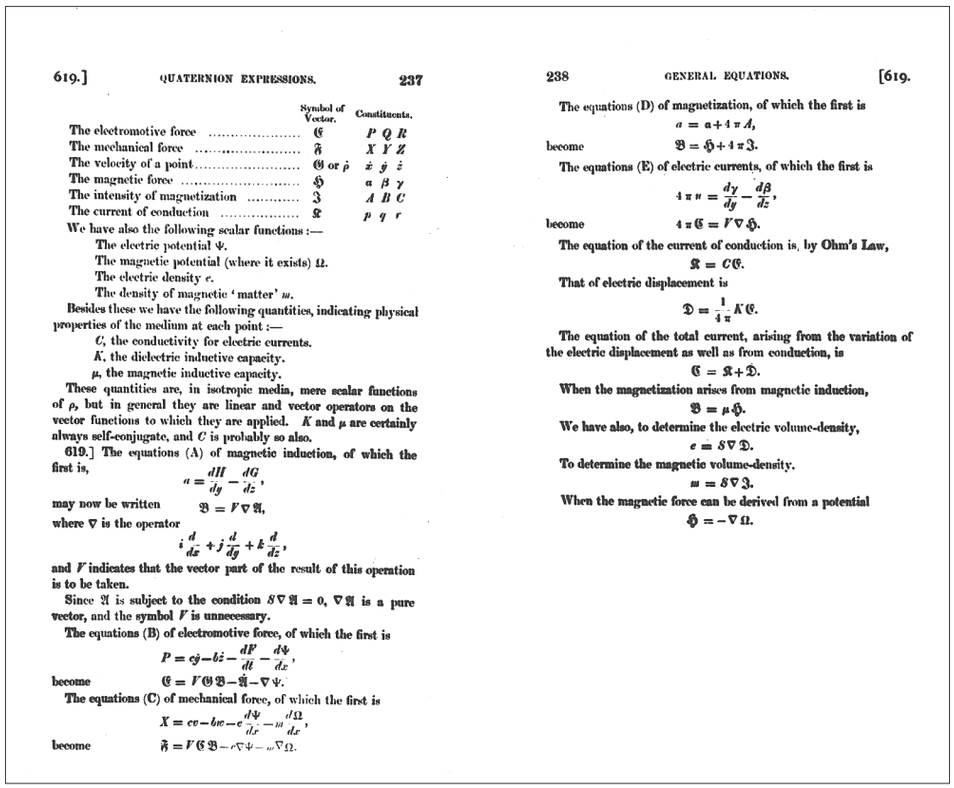
In what follows, we present Maxwell's fundamental equations in the form that Maxwell used them, together with the form generally in use today:

$$\mu \mathbf{H} = \operatorname{curl} \mathbf{A} \rightarrow \operatorname{div} \mathbf{B} = 0, \quad \mathbf{B} = \mu \mathbf{H};$$
$$E = -\frac{\partial \mathbf{A}}{\partial t} - \operatorname{grad} \varphi \rightarrow \operatorname{curl} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}.$$

The mathematically expressed condition for electromagnetic waves to develop is the existence of a *displacement current*. For Maxwell, following Faraday, it was obvious that the electrical charges in a dielectric are displaced from a resting position under the influence of an electric field, so that a temporal change in an electric field induces a current. This current is called a displacement current. Faraday also found it natural that since empty space, the vacuum, is a transmitter of electromagnetic force effects just like any solid body, in a like manner the vacuum should also have the properties that make the transmission necessary, including the displacement of charges. Here, Maxwell also attempted to clarify the effect with a model.

Figure 4.107 shows the entire system of Maxwell's equations as they are given in his *Treatise*.

Figure 4.108 indicates Maxwell's major step forward in comparison with the laws of Ørsted, Ampère, and Faraday. The first pair of diagrams asserts that the flow of electric charges induces a magnetic field; according to the second picture, every temporal change in the magnetic field (more precisely, the magnetic induction) leads to an electric field. All of today's high-current technology is based on these two laws. The second pair of diagrams asserts additionally that every electric



619.] QUATERNION EXPRESSIONS. 237

	Symbol of Vector.	Constituents.
The electromotive force	$\mathfrak{E}$	$P\ Q\ R$
The mechanical force	$\mathfrak{F}$	$X\ Y\ Z$
The velocity of a point	$\mathfrak{C}$ or $\dot{\rho}$	$\dot{x}\ \dot{y}\ \dot{z}$
The magnetic force	$\mathfrak{H}$	$\alpha\ \beta\ \gamma$
The intensity of magnetization	$\mathfrak{I}$	$A\ B\ C$
The current of conduction	$\mathfrak{K}$	$p\ q\ r$

We have also the following scalar functions :—

The electric potential Ψ.

The magnetic potential (where it exists) Ω.

The electric density e.

The density of magnetic 'matter' m.

Besides these we have the following quantities, indicating physical properties of the medium at each point :—

C, the conductivity for electric currents.

K, the dielectric inductive capacity.

μ, the magnetic inductive capacity.

These quantities are, in isotropic media, mere scalar functions of ρ, but in general they are linear and vector operators on the vector functions to which they are applied. K and μ are certainly always self-conjugate, and C is probably so also.

619.] The equations (A) of magnetic induction, of which the first is,

$$a = \frac{dH}{dy} - \frac{dG}{dz},$$

may now be written

$$\mathfrak{B} = V \nabla \mathfrak{A},$$

where ∇ is the operator

$$i\frac{d}{dx} + j\frac{d}{dy} + k\frac{d}{dz},$$

and V indicates that the vector part of the result of this operation is to be taken.

Since $\mathfrak{A}$ is subject to the condition $S \nabla \mathfrak{A} = 0$, $\nabla \mathfrak{A}$ is a pure vector, and the symbol V is unnecessary.

The equations (B) of electromotive force, of which the first is

$$P = cy - bz - \frac{dF}{dt} - \frac{d\Psi}{dx},$$

become

$$\mathfrak{E} = V \mathfrak{C} \mathfrak{B} - \mathfrak{A} - \nabla \Psi.$$

The equations (C) of mechanical force, of which the first is

$$X = ev - bm - e\frac{d\Psi}{dx} - m\frac{d\Omega}{dx},$$

become

$$\mathfrak{F} = V \mathfrak{C} \mathfrak{B} - e\nabla\Psi - m\nabla\Omega.$$

238 GENERAL EQUATIONS. [619.

The equations (D) of magnetization, of which the first is

$$a = \alpha + 4\pi A,$$

become

$$\mathfrak{B} = \mathfrak{H} + 4\pi \mathfrak{I}.$$

The equations (E) of electric currents, of which the first is

$$4\pi p = \frac{d\gamma}{dy} - \frac{d\beta}{dz},$$

become

$$4\pi \mathfrak{C} = V \nabla \mathfrak{H}.$$

The equation of the current of conduction is, by Ohm's Law,

$$\mathfrak{K} = C \mathfrak{E}.$$

That of electric displacement is

$$\mathfrak{D} = \frac{1}{4\pi} K \mathfrak{E}.$$

The equation of the total current, arising from the variation of the electric displacement as well as from conduction, is

$$\mathfrak{C} = \mathfrak{K} + \dot{\mathfrak{D}}.$$

When the magnetization arises from magnetic induction,

$$\mathfrak{B} = \mu \mathfrak{H}.$$

We have also, to determine the electric volume-density,

$$e = S \nabla \mathfrak{D}.$$

To determine the magnetic volume-density.

$$m = S \nabla \mathfrak{I}.$$

When the magnetic force can be derived from a potential

$$\mathfrak{H} = -\nabla \Omega.$$

Quotation 4.33

The conception of such a quantity, on the changes of which, and not on its absolute magnitude, the induction current depends, occurred to Faraday at an early stage of his *Researches*. He observed that the secondary circuit, when at rest in an electromagnetic field which remains of constant intensity, does not shew any electrical effect, whereas, if the same state of the field had been suddenly produced, there would have been a current. Again, if the primary circuit is removed from the field, or the magnetic forces abolished, there is a current of the opposite kind. He therefore recognized in the secondary circuit, when in the electromagnetic field, a "peculiar electrical condition of matter," to which he gave the name of the Electrotonic State. He afterwards found that he could dispense with this idea by means of considerations founded on the lines of magnetic force, but even in his latest *Researches*, he says, "Again and again the idea of an *electrotonic* state has been forced on my mind."

The whole history of this idea in the mind of Faraday, as shewn in his published *Researches*, is well worthy of study. By a course of experiments, guided by intense application of thought, but without the aid of mathematical calculations, he was led to recognize the existence of something which we now know to be a mathematical quantity, and which may even be called the fundamental quantity in the theory of electromagnetism. But as he was led up to this conception by a purely experimental path, he ascribed to it a physical existence, and supposed it to be a peculiar condition of matter. ...

The scientific value of Faraday's conception of an electrotonic state consists in its directing the mind to lay hold of a certain quantity, on the changes of which the actual phenomena depend.

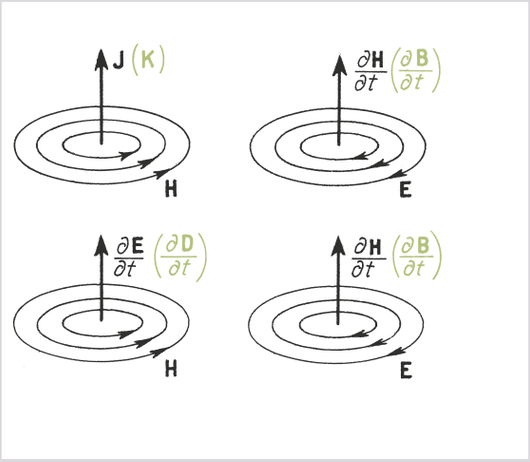
—James Clerk Maxwell, *A Treatise on Electricity and Magnetism* [pp. 187–188]

◀ **Figure 4.107** The system of Maxwell's equations as they appear in Maxwell's book.

War es ein Gott, der diese Zeichen schrieb,
Die mit geheimnißvoll verborg'nem Trieb
Die Kräfte der Natur um mich enthüllen
Und mir das Herz mit stiller Freude füllen.

Was it a god that fashioned this design,
Which with a secret thrust divine
Makes Nature's powers about me manifest
And fills my heart with quiet happiness.

—Ludwig Boltzmann, paraphrasing *Faust* in praising Maxwell [Solymar 1999, p. 141]



▲ **Figure 4.108** Maxwell's fundamentally new idea: the displacement current.

> One of the chief peculiarities of this treatise is the doctrine which it asserts, that the true electric current $\mathfrak{C}$, that on which the electromagnetic phenomena depend, is not the same thing as $\mathfrak{K}$, the current of conduction, but that the time-variation of $\mathfrak{D}$, the electric displacement, must be taken into account in estimating the total movement of electricity, so that we must write,
>
> $$\mathfrak{C} = \mathfrak{K} + \dot{\mathfrak{D}}.$$
>
> —Maxwell, *A Treatise on Electricity and Magnetism*, Vol. II, article 610

These diagrams can be seen as pictographs of the first two equations,

$$\text{curl}\mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad \text{and} \quad \text{curl}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}.$$

The two additional equations

$$\text{div}\mathbf{D} = \rho \quad \text{and} \quad \text{div}\mathbf{B} = 0$$

express the physical fact that the **D**-lines come from the charges, that is, that the **B**-field is source-free. There is no "real" magnetic charge.

The group of equations

$$\mathbf{D} = \varepsilon\mathbf{E}, \quad \mathbf{B} = \mu\mathbf{H}, \quad \mathbf{J} = \gamma\mathbf{E} + \rho\mathbf{v}$$

creates a relationship among the field data through the material constants. The equation for the energy density,

$$w = \frac{1}{2}\mathbf{ED} + \frac{1}{2}\mathbf{HB},$$

and the formula derivable from this for the power density, $\mathbf{P} = \mathbf{E} \times \mathbf{H}$, and for the force density,

$$\mathbf{f} = \rho\mathbf{E} + (\mathbf{J} + \rho\mathbf{v}) \times \mathbf{B},$$

finally build the bridges to mechanics, thermodynamics, and so on.

We thus have the complete system of Maxwell's equations. The various subfields of electrodynamics turn out to be special cases of these all-encompassing formulas.

continued on next page

field that changes over time (more precisely, every changing displacement) induces a magnetic field. Finally, this relationship makes possible the existence of electromagnetic waves. Therefore, an electromagnetic field can exist even in empty space—far from any current-carrying conductor—because electric and magnetic fields mutually induce each other: An electric field arises wherever a magnetic field changes over time, and a magnetic field is created wherever a time-varying electric field exists. This close relationship between these two fields brings about the interplay that results in electromagnetic waves (Figure 4.109).

Long before the theoretical and experimental investigation of electromagnetic waves in open air, the question was raised of how quickly electromagnetic effects are propagated. This question achieved practical significance with the development of wire-based telecommunications. In 1854, William Thomson gave an incomplete telegraph equation in the form

$$\frac{1}{C}\frac{\partial^2 U}{\partial x^2} = R\frac{\partial U}{\partial t},$$

whose solution is a progressive damped wave with a speed of propagation proportional to the square root of the frequency (C and R denote the capacity and resistance, respectively, of the pair of conductors with respect to the unit of length). In 1857, Kirchhoff derived the equation

$$\frac{\partial^2 U}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} + K\frac{\partial U}{\partial t},$$

which in the ideal case, $K = 0$, has a solution with a speed of propagation independent of the frequency. The telegraph equation in the form usually used today was given by Heaviside in 1876, and Poincaré subsequently established the connection between telegraphy theory and Maxwell's equations.

As for the other Maxwell equations, we will mention here only that William Thomson introduced the field vectors **B** and **H** to characterize a magnetic field and investigated the relationships between these fields (including for crystalline substances) in the years 1849–1851. He discovered the relationship between the energy density and the fields (1851–1853). The electric field vectors **D** and **E** were essentially first introduced by Maxwell.

It was only with difficulty that the physicists of the era began to use Maxwell's equations as part of their theoretical toolkit, although the greatest of them realized that, with these equations, classical electromagnetic theory had been brought to a final synthesis.

The main reason for the difficulties was the lack of mechanical perspective. Brought up in the spirit of Descartes and Huygens, scientists had become accustomed to think in terms of mechanical concepts. Faraday remained true to this worldview, and Maxwell, too, could break away from it only with difficulty. One can see the intensity of this striving for a mechanical interpretation of electromagnetic phenomena from the large number of different display models, devised with great ingenuity and requiring considerable skill from the craftsmen to construct, that were created at the time for the purpose of classroom demonstrations. In Figure 4.106(c), we already introduced a simple model of this sort, devised by Maxwell himself, illustrating the flow of electric currents in two inductively coupled circuit loops.

An illustration of the general attitude is given by the epigraph of Boltzmann's book on Maxwell's theory of electromagnetism:

> "Thus should I then with sweaty brow
> Teach you that which I do not know."
>
> These words of Faust have never been more accurately applied than to him who desires to hold forth on the true nature of electricity. One should therefore not take with too much skepticism this bit of poetry as an introduction to a collection of dry formulas.

Among the few illustrations in that book, the representation of a more complex mechanical model occupies a prominent place (Figure 4.110).

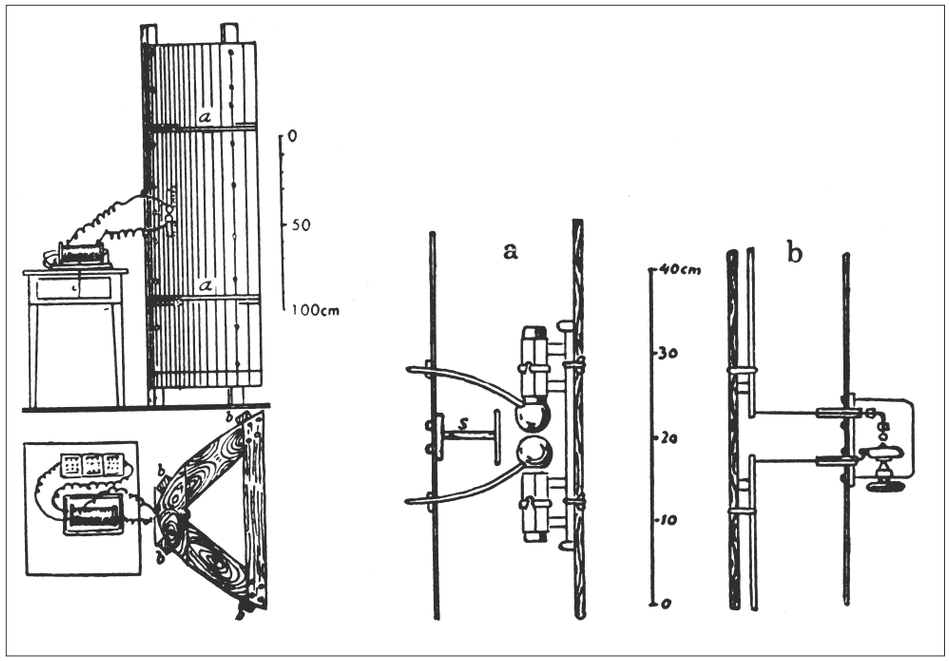
Toward the end of the century, the famous physicist Lord Rayleigh still saw it as necessary to buttress Maxwell's equations with the help of a complex model of the ether.

Not only did Maxwell's theory of electromagnetic fields have enormous practical importance because it began a revolution in the development of telecommunications, but it exerted a lasting influence on natural philosophy as well (Figures 4.111, 4.112). It forced physicists to view natural phenomena more abstractly and to grant the quantities arising in an abstract description a real existence, even if they could not be explained with the help of a mechanical model whose properties would correspond directly to our sensory experience (Quotations 4.34, 4.35).

4.4.10 *The Electromagnetic Theory of Light*

There is another characteristic of electrodynamic theory that justifies calling it an important juncture in the history of physics: in this theory two previously independently developed branches of physics meet: electrodynamics and optics (see Figure 0.10).

In order to gain a proper appreciation of the significance of this synthesis, we begin by looking backward to the development of the physics of light after Newton and Huygens. In the eighteenth century, no significant progress was made in this area. Indeed, we might even say that there was a certain degree of regression; the scholars of that period adopted only a simplified variant of Newton's theory, namely, a simplified corpuscular theory. The fermentation began in 1800, with the publication of an article by Thomas Young (Figure 4.113), in which he emphasized the advantages of Huygens' theory (Quotation 4.36). As a result of his investigations in the ensuing years (1801–1803), Young developed a wave theory of light that agrees almost perfectly with what is taught in today's high schools and colleges. He was the first to postulate that light consists of periodic waves. Using this assumption, he correctly described the phenomenon of interference and



▲ **Figure 4.111** Experimental setup by Hertz to show the existence of electromagnetic waves.

Figure 4.108 *continued*

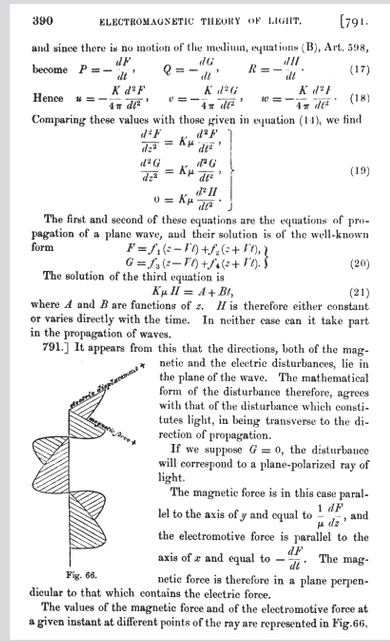
Waves result as solutions to the wave equations

$$\Delta\mathbf{E} = \varepsilon_0\mu_0 \frac{\partial^2\mathbf{E}}{\partial t^2}, \quad \Delta\mathbf{H} = \varepsilon_0\mu_0 \frac{\partial^2\mathbf{H}}{\partial t^2}$$

in vacuum, and

$$\Delta\mathbf{A} - \varepsilon\mu \frac{\partial^2\mathbf{A}}{\partial t^2} = -\mathbf{J}, \quad \Delta\varphi - \varepsilon\mu \frac{\partial^2\varphi}{\partial t^2} = -\frac{\rho}{\varepsilon}$$

in the presence of matter.


390 ELECTROMAGNETIC THEORY OF LIGHT. [791.

and since there is no motion of the medium, equations (B), Art. 598, become $P = -\frac{dF}{dt}$, $Q = -\frac{dG}{dt}$, $R = -\frac{dH}{dt}$. (17)

Hence $u = -\frac{K}{4\pi}\frac{d^2F}{dt^2}$, $v = -\frac{K}{4\pi}\frac{d^2G}{dt^2}$, $w = -\frac{K}{4\pi}\frac{d^2H}{dt^2}$. (18)

Comparing these values with those given in equation (14), we find

$$\left.\begin{aligned} \frac{d^2F}{dz^2} &= K\mu\frac{d^2F}{dt^2}, \\ \frac{d^2G}{dz^2} &= K\mu\frac{d^2G}{dt^2}, \\ 0 &= K\mu\frac{d^2H}{dt^2}. \end{aligned}\right\} \quad (19)$$

The first and second of these equations are the equations of propagation of a plane wave, and their solution is of the well-known form

$$\left.\begin{aligned} F &= f_1(z - Vt) + f_2(z + Vt), \\ G &= f_3(z - Vt) + f_4(z + Vt). \end{aligned}\right\} \quad (20)$$

The solution of the third equation is

$$K\mu H = A + Bt, \quad (21)$$

where A and B are functions of z. H is therefore either constant or varies directly with the time. In neither case can it take part in the propagation of waves.

791.] It appears from this that the directions, both of the magnetic and the electric disturbances, lie in the plane of the wave. The mathematical form of the disturbance therefore, agrees with that of the disturbance which constitutes light, in being transverse to the direction of propagation.

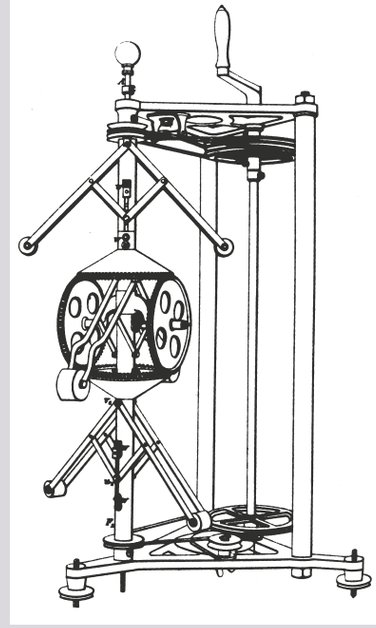
If we suppose $G = 0$, the disturbance will correspond to a plane-polarized ray of light.

The magnetic force is in this case parallel to the axis of y and equal to $\frac{1}{\mu}\frac{dF}{dz}$, and the electromotive force is parallel to the axis of x and equal to $-\frac{dF}{dt}$. The magnetic force is therefore in a plane perpendicular to that which contains the electric force.

Fig. 66.

The values of the magnetic force and of the electromotive force at a given instant at different points of the ray are represented in Fig. 66.

◀ **Figure 4.109** Representation of the plane electromagnetic wave in Maxwell's book.



◀ **Figure 4.110** Boltzmann's model for illustrating the inductive coupling of two electric circuits.



▲ **Figure 4.112** Heinrich Hertz (1857–1894): From 1878, assistant to Helmholtz in Berlin, although he had intended to become a civil engineer. 1880: dissertation on electromagnetic induction in rotating bodies. During his time as assistant for Helmholtz, Hertz worked on gas discharge. 1886, professor at the Technical University, Karlsruhe.

1889, as successor to Clausius, professor of physics at Bonn University. We have Hertz to thank for a large number of discoveries, and even his unsuccessful cathode-ray experiments goaded his student Lenard along the correct path. Best known are his investigations in which he showed the existence of electromagnetic waves and that they behave the same way as light waves, thus proving that Maxwell's theory is correct. Through the publication of his book *Über die Grundgleichungen der Elektrodynamik für ruhende Körper* (*On the Fundamental Equations of Electrodynamics for Bodies at Rest*), Maxwell's theory was finally able to triumph in continental Europe.

> The most direct, and in a sense the most important, problem which our conscious knowledge of nature should enable us to solve is the anticipation of future events, so that we may arrange our present affairs in accordance with such anticipation. As a basis for the solution of this problem we always make use of our knowledge of events which have already occurred, obtained by chance observation or by prearranged experiment. In endeavoring thus to draw inferences as to the future from the past, we always adopt the following process.
>
> We form for ourselves images or symbols of external objects; and the form which we give them is such that the necessary consequents of the images in thought are always the images of the necessary consequents in nature of the things pictured. In order that this requirement may be satisfied, there must be a certain conformity between nature and our thought. Experience teaches us that the requirement can be satisfied, and hence that such a conformity does exist.

continued on next page

interpreted with its help the results of a number of experiments that had been performed by Newton. For example, he offered a theory of Newton's rings. Young was also aware of the significance of coherence. However, his ideas met with resistance from the leading physicists of the time. Biot, for instance, believed in the corpuscular theory until his death in 1862, which was shortly before the appearance of the electromagnetic theory of light. In England, they spoke of a sacrilege against the great and true teachings of Newton. In France, however, Augustin Fresnel (Figure 4.114) promulgated views similar to those of Young; he even used the interference principle to determine wavelengths. Fresnel, together with Arago, who brought Fresnel and Young into scientific contact, constructed an interferometer. In this apparatus, two beams of light from the same source interfere and then are directed along different paths and finally superimposed. A more developed variant is the interferometer invented by Michelson (1883), which he used for his seminal experiment relating to the theory of relativity.

The creators of the wave theory of light were expected to answer the question of the carrier medium of light waves. This question arises when we compare light waves with waves on the surface of water or with sound waves propagating in air. Thus did the ether appear yet again as a hypothetical material filling all of space; for example, we quote here an observation of Fresnel, written in 1822: *La lumière n'est qu'un certain mode de vibration d'un fluide universel* (Light is nothing more than a certain form of vibration of a universal fluid).

Experiments in the polarization of light brought new information regarding the form of vibration of the ether. We recall that both Huygens and Newton attributed particular significance to the double refraction of light in Iceland spar. Newton also found that light exhibits a certain "sidedness": certain planes are favored by the light waves exiting the spar. Étienne Louis Malus (1775–1812) determined that natural light, too, when reflected through a certain angle, behaves like one of the rays exiting a double-refracting crystal. Malus called this phenomenon *polarization*. Then Brewster determined the angle at which light is completely polarized. Both saw in these effects the Newtonian corpuscles. Young then proved a relationship between polarization and the transversality of waves: the particles of ether vibrate perpendicularly to the direction of light's propagation.

In 1820, Fresnel offered a completely theoretical description of the phenomena described above. With him, we meet for the first time a wave representation of the form

$$a \sin\left[\omega\left(t - \frac{x}{c}\right) + \alpha\right],$$

and he also described both linear and circular polarized waves in mathematical form. Although he worked with a mechanical ether model, in other words, he assumed that the ether obeys the same mechanical laws as real fluids or solid bodies, he still managed to obtain the correct relationships for reflection and refraction as well as for the propagation of waves in crystals. This success, however, was due entirely to Fresnel's brilliant intuition: he prescribed boundary conditions on the interface between two media that do not follow from mechanics but lead to the correct result.

In the first half of the nineteenth century, researchers focused on making the mechanical ether theory more precise and to provide a better theoretical basis for it. Among those who worked in this area, we find such names as Cauchy and Franz Ernst Neumann (1798–1895). Although toward the end of the century, there were again repeated efforts to improve this theory, they invoked ever more forced

assumptions about the properties of the ether. For example, to explain transversality, the ether had to be viewed as an ideal solid, yet, it is evident that the celestial bodies pass through it unhindered in their motion. To explain this, the ether would have to be instead a very fine gaseous substance. For these reasons, at the time of the birth of Maxwell's theory, many physicists suspected that a fundamental revision was going to be necessary.

There were several interesting attempts at such a revision before these contradictions were resolved by the theory of relativity. We have already encountered the name G. G. Stokes (1819–1903) in the context of the flow of viscous fluids; Stokes' theorem—which relates the surface integral to the line integral along the boundary curve of the surface—has multiple applications in a variety of areas of physics. Stokes assumed that the ether was a substance like wax, or more precisely, a substance with the following waxlike properties: It reacts to a sudden action like a solid; it is shattered by the blow of a hammer like a piece of glass. However, under the effect of a constant, steady force, it flows like a liquid.

Stokes also proposed that the planets draw a part of the ether along with them.

By the middle of the nineteenth century, it was clear that there existed some sort of relationship between light and electricity. In 1849, Fizeau measured the speed of light by using an apparatus, now named for him, consisting of a rapidly rotating toothed wheel that interrupts a light beam. Since the experiment is entirely terrestrial, its conditions can be easily controlled. In 1851, Fizeau also determined the speed of light in flowing liquids. The results of his experiment will play a role in the theory of relativity. At the same time, precision measurements in the investigation of electrical phenomena assumed ever greater significance. It became an important project to compare the fundamental units of the various electrical and magnetic systems of measurement. In 1855, Wilhelm Weber (1804–1891) and Rudolf Kohlrausch (1809–1858) determined the relationship between the electrostatic and electromagnetic units of charge.

As we would say today, Weber and Kohlrausch compared the electrical charge determined by the equation

$$F = \frac{e_1 e_2}{r^2}$$

with the charge determined by

$$F = \frac{i_1 \, \mathrm{d}l_1 \, i_2 \, \mathrm{d}l_2}{r^2}.$$

The measurement of the former quantity can be carried out, with the help of a Thomson balance, for example, and the other with a Helmholtz balance. Weber and Kohlrausch found that the two units, as given in the equations above, are related by a dimension of speed; its numerical value, in turn, is equal to the speed of light in a vacuum.

This relationship, which at the time came as an enormous surprise, follows immediately from Maxwell's equations. The first and second of Maxwell's equations (written for the case of a vacuum) have as their solution a plane wave that exhibits all the properties that we expect of light. The speed of propagation of this wave is given by the above-mentioned quotient; that is, electromagnetic waves propagate at the velocity of light. The waves are transversal—both the electric and magnetic fields oscillate perpendicularly to the direction of propagation. The properties of

Figure 4.112 *continued*

When from our accumulated previous experience we have once succeeded in deducing images of the desired nature, we can then in a short time develop by means of them, as by means of models, the consequences which in the external world only arise in a comparatively long time, or as the result of our own interposition. We are thus enabled to be in advance of the facts, and to decide as to present affairs in accordance with the insight so obtained. The images which we here speak of are our conceptions of things. With the things themselves they are in conformity in *one* important respect, namely, in satisfying the above-mentioned requirement. For our purpose it is not necessary that they should be in conformity with the things in any other respect whatever. As a matter of fact, we do not know, nor have we any means of knowing, whether our conceptions of things are in conformity with them in any other than this *one* fundamental respect.

—Heinrich Hertz, *The Principles of Mechanics* [pp. 1–2]

Hertz had not yet realized the significance of radio waves for practical applications. The following researchers are most closely associated with the development of today's radio technology:

Aleksandr Stepanovich Popov (1859–1906): 1895, construction of a receiver; 1896, creation of a radio link between two buildings.

Guglielmo Marconi (1874–1937): 1896, experimental demonstration in London; 1899, radio link across the English Channel; 1901, radio link across the Atlantic.

Carl Ferdinand Braun (1850–1918): 1874, rectifying effect using semiconductors; 1897, cathode ray tubes; 1898, Braun transmitter with coupled resonators and crystal detector.

Quotation 4.34

The electromagnetic field is that part of space which contains and surrounds bodies in electric or magnetic conditions.

It may be filled with any kind of matter, or we may endeavour to render it empty of all gross matter, as in the case of Geissler's tubes and other so-called vacua.

There is always, however, enough of matter left to receive and transmit the undulations of light and heat, and it is because the transmission of these radiations is not greatly altered when transparent bodies of measurable density are substituted for the so-called vacuum, that we are obliged to admit that the undulations are those of an ethereal substance, and not of the gross matter, the presence of which merely modifies in some way the motion of the ether.

We have therefore some reason to believe, from the phenomena of light and heat, that there is an ethereal medium filling space and permeating bodies, capable of being set in motion and of transmitting that motion from one part to another, and of communicating that motion to gross matter so as to heat it and affect it in various ways.

Now the energy communicated to the body in heating it must have formerly existed in the moving

continued on next page

Quotation 4.34, ***continued***

medium, for the undulations had left the source of heat some time before they reached the body, and during that time the energy must have been half in the form of motion of the medium and half in the form of elastic resilience. From these considerations Professor W. Thompson has argued, that the medium must have a density capable of comparison with that of gross matter, and has even assigned an inferior limit to that density.

We may therefore receive, as a datum derived from a branch of science independent of that with which we have to deal, the existence of a pervading medium, of small but real density, capable of being set in motion, and of transmitting motion from one part to another with great, but not infinite, velocity.

—James Clerk Maxwell, *A Dynamical Theory of the Electrical Field* [pp. 34–35]



▲ **Figure 4.113** Thomas Young (1773–1829): A child prodigy; could read by age 2. His many talents are revealed in the fact that he was a painter, musician, philologist, physician, and physicist. From 1800, he had a medical practice in London, but from 1801 to 1804, he was professor of physics at the Royal Institution. In 1807, his lectures were published in two volumes (*A Course of Lectures on Natural Philosophy and the Mechanical Arts*). As a physicist, he is known primarily for his proof of the wave nature of light. In 1793, he worked on problems of visual acuity and color vision. In 1801, discovered the phenomenon of interference as the best proof of the wave nature of light. In 1817, demonstrated the transversality of light waves. Among other significant contributions to physics are studies of cohesion and the consequent estimation of the size of molecules.

Young also made a name in archaeology. It was his idea that the framed hieroglyphics on the Rosetta stone, discovered in 1799, represented proper names and that these signs were to be interpreted as phonetic. Using this idea, I. F. Champollion succeeded in interpreting the texts on the stone, written in hieroglyphics, demotic script, and Greek.

light in a plane (linear) and circularly polarized light can be more precisely formulated with the help of electric and magnetic field vectors (Quotation 4.37).

Of course, the triumphal march of the electromagnetic theory of light was not free of obstacles. For example, if we calculate the velocity of light in a dielectric, then Maxwell's theory yields the relationship

$$v = \frac{c}{\sqrt{\varepsilon_r \mu_r}}.$$

Here ε_r is the relative dielectric constant, μ_r the relative magnetic permeability, c the velocity of light in vacuum, and v the velocity of propagation of light in the dielectric.

In optics, the relationship

$$v = \frac{c}{n}$$

was established at that time based on Fizeau's measurements. The electromagnetic theory of light also correctly gives the velocity of light in a medium when the Maxwellian relationship

$$n^2 = \varepsilon_r \mu_r$$

is satisfied between the electromagnetic material constants ε_r, μ_r and the optical index of refraction. If we note that the index of refraction of water is $n = 1.33$, the dielectric constant is $\varepsilon_r = 80$, and the magnetic permeability is $\mu_r = 1$, then we see at once that the given relationship is not satisfied. However, from this we cannot conclude that the electromagnetic theory of light is false, for in specifying the numerical values, we have completely overlooked dispersion (dependence on frequency), and that both the index of refraction and the electromagnetic material constants depend on frequency. It is clear that, in the above formulas, we should use values that relate to a single frequency. However, the index of refraction is generally measured for light of frequency of order $10^{15} s^{-1}$, whereas the electromagnetic material constants are measured in a direct-current field, that is, at a frequency of zero or else in a low-frequency alternating field, so that we should not be surprised to find a deviation from the expected values. In the framework of classical electrodynamics, a successful theory of dispersion was given only toward the end of the century.

Meanwhile, in 1875, Lorentz solved Maxwell's equations in the presence of material media, using only macroscopic material constants. By making adroit choices for the boundary conditions at the interface between the different materials, where the tangential components of the electric and magnetic fields and the normal components of the electrical displacement and the magnetic induction change continuously, he was able to derive all the laws of optics—for example Fresnel's formulas and the Brewster angle—from the electromagnetic theory of light.

After the success of the theory, a decade later, Hertz demonstrated the identical nature of electromagnetic and light waves experimentally, as well.

The equivalence of electromagnetic waves and light waves is not obvious from the formal mathematical viewpoint. In the following, we outline briefly, emphasizing only the critical steps, the complex route by which we get from the Maxwell equations describing the behavior of electromagnetic waves to the Fermat principle that determines the path of a ray of light, that is, the transition

$$\left.\begin{aligned} \operatorname{curl} \mathbf{H} &= \varepsilon \frac{\partial \mathbf{E}}{\partial t} \\ \operatorname{curl} \mathbf{E} &= -\mu \frac{\partial \mathbf{H}}{\partial t} \end{aligned}\right\} \rightarrow \delta \int_{P_1}^{P_2} n \, \mathrm{d}s = 0.$$

The material characteristic values ε and μ are "slowly" varying functions of the positional coordinates: $\varepsilon = \varepsilon(x,y,z)$, $\mu = \mu(x,y,z)$, where "slowly" varying means that ε and μ are considered to be constant in a region of space with greatest linear dimension λ_0. The time dependence is taken to be of the form $e^{i\omega t}$, and λ_0 is defined by the equation $2\pi c/\omega = \lambda_0$, where $c = 1/\sqrt{\varepsilon_0 \mu_0}$ is the speed of light in a vacuum. We further assume that ε and μ do not differ greatly from ε_0 and μ_0. Let us attempt to solve the Maxwell equations with functions of the form

$$\mathbf{E} = \mathbf{E}_0(x,y,z) e^{-\mathrm{j}k\mathfrak{E}(x,y,z)} e^{\mathrm{j}\omega t}, \quad \mathbf{H} = \mathbf{H}_0(x,y,z) e^{-\mathrm{j}k\mathfrak{E}(x,y,z)} e^{\mathrm{j}\omega t}, \tag{3}$$

where $k = 2\pi/\lambda_0$. If we insert these functions into the Maxwell equations and neglect all terms in which (after suitable rearrangement) ω appears in the denominator (quasioptic approximation for high frequen-

cies), after laborious computation, we obtain the following equation for $\mathbf{E}_0(x,y,z)$ (and a similar equation for $\mathbf{H}_0(x,y,z)$):

$$\left[\frac{\varepsilon\mu}{\varepsilon_0\mu_0} - (\operatorname{grad}\mathfrak{E})^2\right]\mathbf{E}_0(x,y,z) = 0.$$

A nontrivial solution is possible only if

$$\frac{\varepsilon\mu}{\varepsilon_0\mu_0} - (\operatorname{grad}\mathfrak{E})^2 = 0,$$

in particular, if

$$\left(\frac{\partial\mathfrak{E}}{\partial x}\right)^2 + \left(\frac{\partial\mathfrak{E}}{\partial y}\right)^2 + \left(\frac{\partial\mathfrak{E}}{\partial z}\right)^2 = n^2, \tag{4}$$

where the index of refraction $n = n(x,y,z)$ is introduced via $n^2 = (\varepsilon/\varepsilon_0)(\mu/\mu_0) = \varepsilon_r\mu_r$.

From the system of equations (3), it follows that the surfaces of the form $\mathfrak{E}(x,y,z) = \text{constant}$ represent the equiphase surfaces of the wave; the tangent of the path of the light ray is perpendicular to these surfaces at each point. Let us represent the path parametrically as

$$x = x(\tau), \quad y = y(\tau), \quad z = z(\tau).$$

Since, as we have just stated, the normal of the equiphase surface is parallel to the tangent of the path curve, we have the following equations:

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = a\frac{\partial\mathfrak{E}}{\partial x}, \quad \frac{\mathrm{d}y}{\mathrm{d}\tau} = a\frac{\partial\mathfrak{E}}{\partial y}, \quad \frac{\mathrm{d}z}{\mathrm{d}\tau} = a\frac{\partial\mathfrak{E}}{\partial z}.$$

Equation (4) can therefore be rewritten as

$$\frac{1}{a^2}\left[\left(\frac{\mathrm{d}x}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)^2\right] = n^2. \tag{5a}$$

This equation has the same form as the energy equation for the motion of a mass point in a potential field:

$$\frac{m}{2}\left[\left(\frac{\mathrm{d}x}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)^2\right] = U(x,y,z), \tag{5b}$$

where U is the negative potential energy. The corresponding quantities are

$$\frac{1}{a^2} \leftrightarrow \frac{m}{2}, \quad \tau \leftrightarrow t, \quad n^2 \leftrightarrow U(x,y,z).$$

However, we know that for equation (5b), the Maupertuis principle of least action holds:

$$\delta\int_{P_1}^{P_2} mv\,\mathrm{d}s = 0,$$

where $v = \sqrt{2/m}\sqrt{U}$. Since $n \leftrightarrow \sqrt{U}$ and the constant factors in the determination of the extreme values are insignificant, we arrive at Fermat's principle:

$$\delta\int_{P_1}^{P_2} n\,\mathrm{d}s = 0.$$

4.4.11 *Lorentz's Theory of the Electron*

Classical electrodynamics reached its culmination in 1891with the electron theory of Hendrik Lorentz (Figure 4.115). Lorentz assumed that the immaterial ether was the carrier of electromagnetic fields, created by resting or moving electric charges. Resting or uniformly moving charges create static, or stationary, fields, whereas electromagnetic waves are given off by accelerated moving charges. In the presence of matter, this picture is modified to the extent that the electric charges

Quotation 4.35

[B]efore Maxwell people conceived of physical reality—in so far as it is supposed to represent events in nature—as material points, whose changes consist exclusively of motions, which are subject to total differential equations. After Maxwell they conceived physical reality as represented by continuous fields, not mechanically explicable, which are subject to partial differential equations. This change in the perception of reality is the most profound and fruitful one that has come to physics since Newton.

—Einstein, *Ideas and Opinions* [p. 269]



◀ Figure 4.114 Jean Augustin Fresnel (1788–1827): One of the most important representatives of the French school, which was concerned with proving the wave nature of light. It is of interest that Fresnel was born in a place called Broglie (in Normandy). He worked as a highway engineer, working on scientific problems in his free time. From 1823: member of the French Academy; from 1825: member of the Royal Society. Fresnel became known for his precise formulation of Huygens' principle, the mathematical description of waves, and the discovery of the relationship between polarization and transversality. Fresnel correctly interpreted the rotation of the polarization plane as a double diffraction of circularly polarized waves. The formulas he derived correctly describing the intensity of reflected and refracted waves today bear his name.

Dominique François Jean Arago (1786–1853): Highly respected scholar of the French school mentioned above. In 1806, he and Biot carried out geodetic measurements in Spain necessary for the precise definition of the meter. He introduced the scientific world to Ørsted's experiment and was the first to observe that a magnetic needle placed over a rotating copper disk will be set in motion (Arago's experiment). From 1830, Arago was secretary of the *Académie*. He took an active part in political life—he was prime minister in 1848—and was also known for his activities in popularizing science.

Hippolyte Fizeau (1819–1896): Encouraged by Arago, in 1848 measured the velocity of light under terrestrial conditions (using a rotating strobe disk), first in air and then, in 1851 and 1859, in flowing water. Jean Bernard Léon Foucault (1819–1868) worked on the same problem using a rotating mirror, obtaining in 1849 the excellent value 2.98×10^8 m/sec for the velocity of light. In measuring the velocity of light in various media, he determined that it is less in a liquid or solid medium than in air, thus further bolster

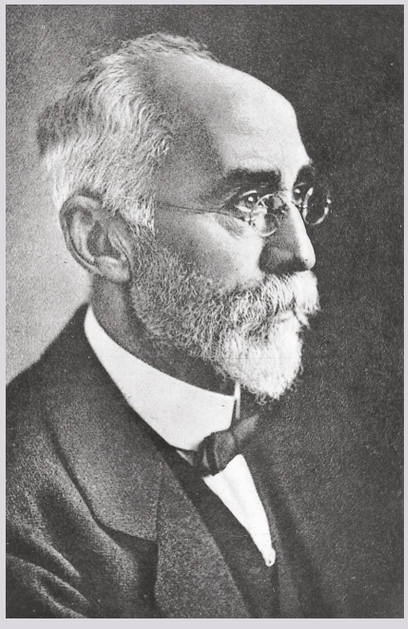
continued on next page

Figure 4.114 *continued*

ing the case for the wave theory of light. From FOUCAULT, we also have a theory and practical realization of the gyroscope. Worthy of mention is also the very long pendulum (Foucault pendulum) hanging from the dome of the Pantheon, which gives evidence, accessible even to the layman, of the rotation of Earth. FOUCAULT also carried out a more precise version of ARAGO's experiment and determined that eddy currents are induced in electric conductors moving in the presence of a strong magnetic field; these currents are called Foucault currents.

In understanding the wave nature of light, in addition to the French school and the Englishman THOMAS YOUNG, the Munich optician JOSEPH FRAUNHOFER (1787–1826) also played an important role. FRAUNHOFER united extraordinary tool-making skills with mathematical ability. He produced the first optical grating, a glass plate with etched grooves, 300 lines/mm (7500 lines/inch), which he used to measure the wavelengths of light. The black lines that were discovered in the Sun's spectrum are called Fraunhofer lines in his honor.

CHRISTIAN DOPPLER (1803–1853): Teacher in Prague, then in Vienna. In 1842, his work *On the Colored Light of the Binary Stars* was published, with the description of the effect that now bears his name: The wavelength of light will change depending on whether the source approaches or recedes from the observer.



◀ **Figure 4.115** HENDRIK ANTOON LORENTZ (1853–1928): Studies in mathematics and physics at the University of Leyden; there from 1878, professor of theoretical physics. His 1875 treatise *Réflexion et réfraction de la lumière dans la théorie électromagnétique* contains the derivation of Fresnel's formulas from the fundamental equations of electromagnetism. Among his multifaceted achievements (for example, he is responsible for the plan to drain the Zuider Zee), we should mention in particular the electron theory of metals named for him (1892), the theoretical interpretation of the Zeeman effect (1896), and his preliminary work leading to the theory of relativity (Lorentz–Fitzgerald contraction, 1892; Lorentz transformation, 1899).

LORENTZ was rightly one of the most popular and internationally renowned physicists at the turn of the century, although later he had great reluctance to abandon his reservations about the relativity theory as formulated by EINSTEIN and about PLANCK's quantum theory.

present within it, either freely moving or coupled elastically, also create an electromagnetic field in the ether. The interaction of light—or, more generally speaking, electromagnetic waves—with matter takes the form of the electromagnetic field of the incident wave exerting a force on the charges, which have mass, and such an interaction changes their state of motion. The incident wave and the secondary wave emitted by the accelerated particles together constitute the resultant wave. Thus, a major role in the Lorentz theory is played by the Lorentz force that describes the interaction

$$\mathbf{F} = q\mathbf{E} + q(\mathbf{v} \times \mathbf{B}).$$

In the phenomenological form of the Maxwell equations, the material properties are described by material constants such as the dielectric constant, magnetic permeability, and electrical conductivity. In the Lorentzian theory, however, the microscopic contributions resulting from the motion of the individual particles in the material are each taken into account separately. Employing a suitable spatial and temporal averaging of these microscopic contributions, LORENTZ succeeded in calculating the macroscopic material constants, thereby giving a microphysical explanation for dispersion, emission, and absorption. He was even able to find an interpretation for the behavior of an atom in a magnetic field, that is, for the splitting of the spectral lines (Zeeman effect) that was in agreement with experimental observation. For their work, LORENTZ and ZEEMAN were awarded the Nobel Prize in 1902.

Lorentz's electron theory is in use even today, as it is easy to understand, and it correctly describes the electromagnetic behavior of macroscopic bodies qualitatively, or, so to speak, to zeroth quantitative approximation.

The Lorentzian theory expresses conspicuously the original asymmetry of Maxwell's equations: There is no real magnetic charge (magnetic monopole); the magnetic property of matter appears only in dipoles. In the practical (engineering) discussion of electromagnetic waves, it is customary to introduce magnetic charge and magnetic current, which facilitates the calculations, so that we obtain symmetric expressions for E and H and we can find useful heuristic analogies.

More recently, particle physicists (DIRAC, 1948; SCHWINGER, 1969) took the search for symmetry seriously. Perhaps the reason we were unable to find monopoles in nature is that they latch onto each other with a "superstrong" interaction. Perhaps very powerful cosmic rays are able to break such strong bonds, and thus, given favorable conditions, monopoles might accumulate. This could be the case on the surface of the Moon. Moon rocks were therefore examined to see whether they contained magnetic monopoles. Moving magnetic monopoles—magnetic currents—would create a circular closed electric field around themselves, just as moving electric charges—ordinary electric current—create a circular closed magnetic field around themselves. No such effect has been found thus far.

Not just classical electrodynamics but all of prerelativistic physics culminated in Lorentz's theory. Its worldview was as follows: The fundamental task of physics is the examination of the two basic substances, the ether and tangible matter. The ether takes the role of Newtonian absolute space, the proper frame of reference and the carrier of electromagnetic phenomena. Matter moves in this space, and, relative to this movement, it interacts with the ether's oscillating state—in other words, with the electromagnetic waves.

Before this theory could achieve something resembling a final state, the results of a number of experiments, inspired by the theory itself, pointed to the incorrectness of the entire conception and finally led to the theory of relativity.

4.5 Heat and Energy

4.5.1 *The Thermometer*

On the one side, philosophy; on the other, technology—with such support, the nineteenth century discovered the most basic law of physics, one that was only strengthened by the later "one–two punch" of quantum mechanics and relativity—the law of conservation of energy.

Romantic natural philosophy aroused in the hearts of physicists the wish to discover what was common and consistent in the relationships among the variety of phenomena and transformations while technological advances led to questions about the efficiency of engines. It is not surprising, then, that progress in knowledge about the nature of heat was closely related to the development of the concept of energy. The importance of "living force," which today is called kinetic energy, was surmised already at the end of the seventeenth century, except that its disappearance during an inelastic collision or due to friction was a very confusing fact. Leibniz already recognized that in these cases the living force is transformed into the living forces of the particles that constitute the given object. The main argument leading to the law of conservation of energy is in this way connected with questions about the nature of heat. Naturally, as shown in Figure 4.116, other fragments of physical knowledge also had important roles, for example, the connection between electrical phenomena and heat, or the heat generated by chemical reactions, as well as the relationship between electrical and chemical phenomena.

One would expect that ideas concerning the nature of heat would have been based on a conception that was already widespread at the end of the seventeenth century, namely, that heat has its origin in the motion of the particles that make up a material body. This would then have led directly to the connection between the two forms of energy: heat and kinetic energy. However, that is not what happened. In a seemingly superfluous and at first glance surprising detour in the history of physics, the kinetic theory of heat was abandoned, and in its place a theory of heat substance (*caloricum*) was adopted. It is only the result of our hasty judgment about what should have happened that names such as Joseph Black have fallen into obscurity, although we owe our thanks to him for such quantitative concepts as heat quantity, specific heat, latent heat, melting point, and boiling point. It has also been forgotten that findings in Carnot's and Fourier's theory of heat that are seen today as fundamental and used in teaching are based on the theory of the *caloricum*.

We must immediately add that these results, arrived at by using the most up-to-date mathematical apparatus of the time, actually hindered the development of the kinetic theory of heat, which was mired down in its search for quantitative conclusions from its qualitative ideas. For this reason the detour was necessary; only in this roundabout way could the qualitative notions be introduced whose measurements then would allow for the formulation of the simplest laws.

It was already clear in the Middle Ages that in connection with heat, two fundamentally different quantities had to be considered: one is an *intensity* and the other is a *quantity*. Naturally, all this was expressed only vaguely. But they could sense that, for example, a flame somehow possesses a greater heat intensity than a piece of warm iron, whereas in the piece of iron the quantity of heat is greater than in the flame. Today, we call these temperature and heat quantity.

The measurements were made possible by the development in the seventeenth century of the most important instrument for the theory of heat: the thermometer.

Quotation 4.36

Much as I venerate the name of Newton, I am not therefore obliged to believe that he was infallible. I see ... with regret that he was liable to err, and that his authority has, perhaps, sometimes even retarded the progress of science.

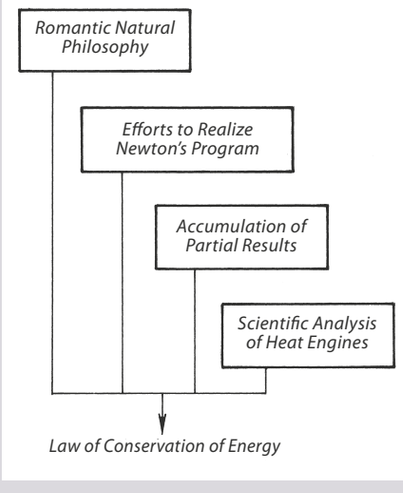
—Thomas Young [Mason 1953, p. 379]

Suppose a certain number of equal waves of water to move upon the surface of a stagnant lake, with a certain constant velocity, and to enter a narrow channel leading out of the lake. Suppose then another similar cause to have excited another equal series of waves, which arrive at the same channel, with the same velocity, and at the same time with the first. Neither series of waves will destroy the other, but their effects will be combined: if they enter the channel in such a manner that the elevations of one series coincide with those of the other, they must together produce a series of greater joint elevations; but if the elevations of one series are so situated as to correspond to the depressions of the other, they must exactly fill up those depressions, and the surface of the water must remain smooth; at least I can discover no alternative, either from theory or from experiment.

—Thomas Young [Elliott 1966, p. 10]

Whenever two portions of the same light arrive at the eye by different routes, either exactly or very nearly in the same direction, the light becomes most intense when the differences of the routes is any multiple of a certain length, and least intense in the intermediate state of the interfering portions; and this length is different for light of different colours.

—Thomas Young, "An Account of Some Cases of the Production of Colours," 1802 [Elliott 1966, p. 11]



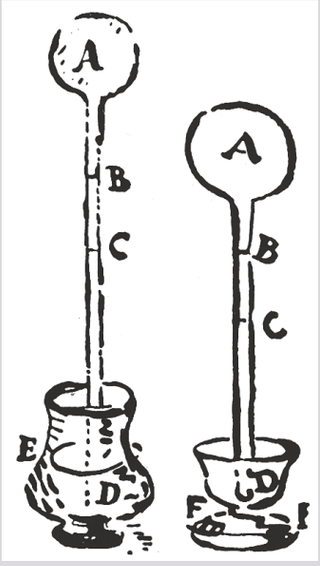
◀ **Figure 4.116** Roadmap leading to the law of conservation of energy.

Quotation 4.37

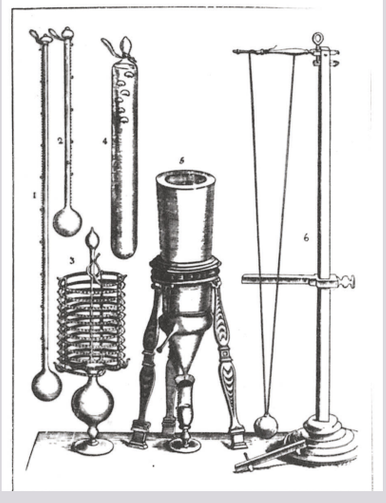
The general equations are next applied to the case of a magnetic disturbance propagated through a non-conducting field, and it is shewn that the only disturbances which can be so propagated are those which are transverse to the direction of propagation, and that the velocity of propagation is the velocity *v*, found from experiments such as those of Weber, which expresses the number of electrostatic units of electricity which are contained in one electromagnetic unit.

This velocity is so nearly that of light, that it seems we have strong reason to conclude that light itself (including radiant heat, and other radiations if any) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws.

—Maxwell, *A Dynamical Theory of the Electromagnetic Field* [Magie 1935, p. 537]



◀ **Figure 4.117** The barothermoscope, attributed to Galileo, whose readings depend on temperature and also on barometric pressure.



◀ **Figure 4.118** Thermometers and other scientific instruments from the second half of the seventeenth century.

As mentioned in Section 1.4, the fact that air expands when heated was known and used already in Alexandria. However, this phenomenon was not employed for the measurement of temperature until the beginning of the seventeenth century. Figure 4.117 shows an apparatus whose construction is attributed to Galileo, although he himself did not describe it in detail. This measuring instrument is called the barothermoscope because its reading depends on the air pressure as well as on the temperature. The first enclosed alcohol thermometer (Figure 4.118) is ascribed to Ferdinando II de' Medici, grand duke of Tuscany. In Florence, under the auspices of the Accademia del Cimento (Academy of Experiment), 1657–1667, a variety of glass masterpieces were produced with scales formed of colored glass beads. It is worthwhile to especially draw attention to thermometer #4 in the illustration. An alcohol-filled ampoule contains a number of small, hollow glass spheres. The weights of the spheres are adjusted in such a way that as the density of alcohol decreases with rising temperature, one sphere after the other—at different given temperatures—sinks to the bottom.

The eighteenth century brought nothing essentially new to the construction of the thermometer. However, reproducibility of temperature measurements was achieved by temperature scales between two set scale values corresponding to two easily achievable phenomena. The temperature scale that was to become most widely used was introduced by the Swedish astronomer Anders Celsius (1701–1744). The zero point of this scale corresponds to the temperature of melting ice, and the boiling point of water is designated the second fixed point on the scale with the value of 100 °C. In Figure 4.119, alongside the Celsius scale we can see the Fahrenheit scale, which was introduced around the same time, and also the "absolute" (Kelvin) temperature scale, which was established in the second half of the nineteenth century.

4.5.2 *Progressive in Its Day: The Caloricum Theory of Joseph Black*

With the introduction of the thermometer, a quantitative formulation of the fundamental laws of thermodynamics was made possible. This was accomplished by Joseph Black (1728–1799).

Black was a professor of chemistry and medical sciences, first in Glasgow (from 1756 until 1766) and then in Edinburgh (1766 up to his death). His most important discoveries in the field of thermodynamics date from his time in Glasgow. Black himself wrote no books; we know of his theory primarily because of a book written in 1803 by his student Robison, in which Black's lectures are reproduced (Figure 4.120). Intriguingly, Robison dedicated this book to James Watt, as the most talented and famous of Black's students (Quotation 4.38).

A quotation from this book (Quotation 4.39) makes it clear how little the concepts of temperature and heat quantity were understood in Black's time. By taking measurements, Black confirmed the fact, previously known, that bodies in contact tend to assume the same temperature. This phenomenon was interpreted by Black as the result of equal distribution of heat among the various bodies. Black accurately called this state heat *equilibrium* of the bodies and concluded that anyone talking about heat *equality* was confusing two quantities, namely temperature and heat quantity.

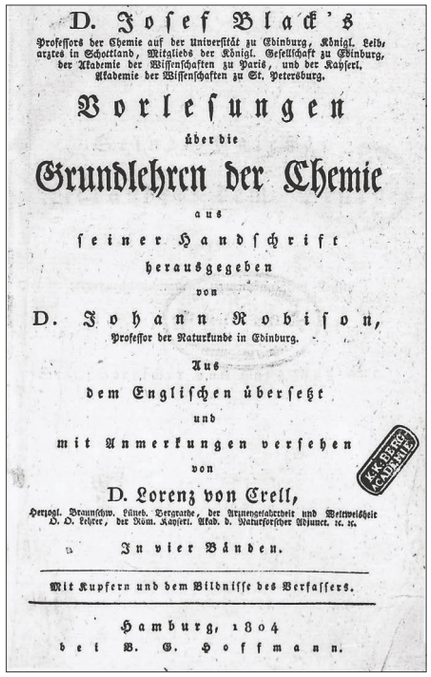
Through his investigations into the heat quantities needed to achieve equal increase in temperature in bodies of equal mass but different chemical composition, Black introduced the notion of specific heat. He refuted the generally accepted opinion that this quantity should be proportional to the mass for bodies of equal

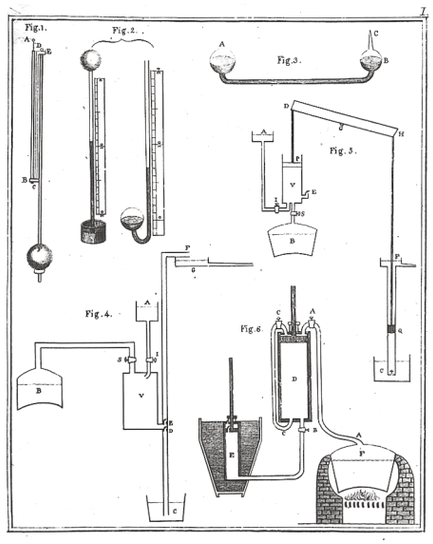
volume, in other words, to their density. Black believed that his measurements contradicted the kinetic theory of heat: From the kinetic theory, it would follow—at least at first glance—that when more particles or heavier particles move, their living force will be greater. Black added, "I see little possibility of refuting this counterargument." As things turned out, the answer was only provided 100 years later, with the aid of the equipartition principle (Quotation 4.40).

Black's train of thought with regard to latent heat is so clear and convincing that it could be transcribed verbatim into our high school physics texts (Quotation 4.41). His final conclusion—specifically that heat is a substance—is no longer accepted. It is remarkable, however, how cautiously he set forth his hypotheses. In the last lines of the quotation, we hear from the true experimental physicist who suggests the limited practical use for hypotheses arrived at through an overabundance of imagination.

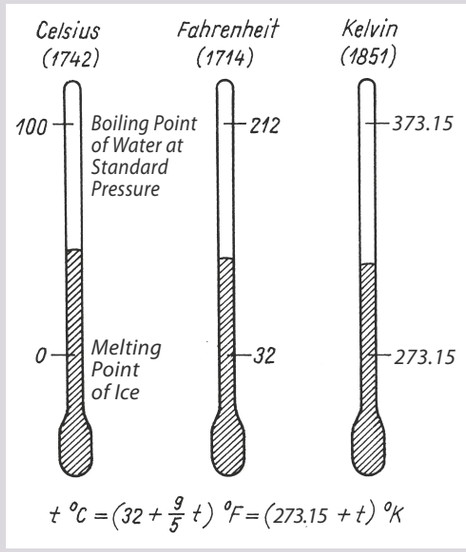
What did Black or his contemporary, the important chemist Lavoisier, understand by *caloricum* or "heat substance"? Black himself refers to Cleghorn, who in 1779 summarized the properties of heat substance as follows: The *caloricum* is an elastic liquid, a *fluidum*, whose particles are mutually repellent, being simultaneously attracted by the particles of ordinary ponderable (weighable) matter, where the attractive force depends on the quality of the material and its aggregate state. This *fluidum* can be neither created nor destroyed; like ponderable matter, it satisfies the conservation law. The *caloricum* can be present both measurably and latently in ponderable matter. In the latter case, the heat substance forms to some extent a chemical bond with the ponderable matter.

There was no universal agreement as to the weight of the heat substance. Many researchers attempted to decide this question on the basis of measurements. Yet we may easily imagine the difficulties and the many sources of error that arose in measurements taken when the material present in one pan of the balance had to be heated after the pans had been balanced. Measurements seemed to indicate


D. Josef Black's
Professors der Chemie auf der Universität zu Edinburg, Königl. Leibarztes in Schottland, Mitgliedes der Königl. Gesellschaft zu Edinburg, der Akademie der Wissenschaften zu Paris, und der Kayserl. Akademie der Wissenschaften zu St. Petersburg.
Vorlesungen
über die
Grundlehren der Chemie
aus
seiner Handschrift
herausgegeben
von
D. Johann Robison,
Professor der Naturkunde in Edinburg.
Aus
dem Englischen übersetzt
und
mit Anmerkungen versehen
von
D. Lorenz von Crell,
Herzogl. Braunschw. Lüneb. Bergrathe, der Arzneygelahrtheit und Weltweisheit D. D. Lehrer, der Röm. Kayserl. Akad. d. Naturforscher Adjunct. ꝛc. ꝛc.
In vier Bänden.
Mit Kupfern und dem Bildnisse des Verfassers.
Hamburg, 1804
bei B. G. Hoffmann.



▲ **Figure 4.120** Black's lectures in German. Title page and a typical page of drawings. (Library of the University for Heavy Industry, Miskolc.)



▲ **Figure 4.119** Three important temperature scales.

Quotation 4.38

Fortunately for Dr. Black, and for the world, he had now gotten a pupil who was as keenly interested in the scientific question as the Professor. This was Mr. James Watt, then employed in fitting up the instruments in the McFarlane Observatory of the University; a philosopher in the most exalted sense of the word, who never could be satisfied with a conjectural knowledge of any subject, and who grudged no labor or study to acquire certainty in his researches. He chanced to have in his hands, for repairs, a model of Newcomen's steam engine, belonging to the Natural Philosophy Class, and was delighted with the opportunity which this small machine gave him for trying experiments connected with the theory of ebullition, which he had just learned from Dr. Black. These he prosecuted in a most happy train of success and did not stop, till his steam engine was rendered more like the most docile of animals, than a frame of lifeless matter; so that, while its power is competent to the lifting a house from its place, a child of ten years old shall, with a touch of his hand, make it go fast or slow, forwards or backwards, and act either forcefully or feebly. This gentleman attached to Dr. Black by every tie of respect, esteem, and affection, supplied him with proofs and illustrations in abundance, of all the points on which the professor wanted information. These were always recited in the class, with the most cordial acknowledgment of obligation to Mr. Watt.

—John Robison's editor's preface to Joseph Black's *Lectures on the Elements of Chemistry*

Quotation 4.39

Any person who reflects on the ideas which we annex to the word heat will perceive that this word is used for two meanings, or to express two different things. It either means a sensation excited in our organs, or a certain quality, affection, or condition of the bodies around us, by which they excite in us that sensation. The word is used in the first sense when we say, we feel heat; in the second when we say, there is heat in the fire, or in a hot stove. ...

We must therefore adopt, as one of the most general laws of heat, that "all bodies communicating freely with each other, and exposed to no inequality of external action, acquire the same temperature, as indicated by a thermometer." All acquire the temperature of the surrounding medium. ...

This is what has been commonly called an equal heat, or the equality of heat among different bodies; I call it the *equilibrium of heat*. The nature of this equilibrium was not well understood, until I pointed out a method of investigating it. Dr. Boerhaave imagined, that when it obtains, there is an equal quantity of heat in every equal measure of space, however filled up with different bodies, and professor Muschenbroek expresses his opinion to the same purpose: "*Est enim ignis aequaliter per omnia, non admodum magna, distributus, ita ut in pede cubico auri et aeris ut plumarum, par ignis sit quantitas*." The reason they give for this opinion is, that to whichever of these bodies the thermometer be applied, it points to the same degree.

But this is taking a very hasty view of the subject. It is confounding the quantity of heat in different bodies with its general strength or intensity, though it is plain that these are two different things, and should always be distinguished, when we are thinking of the distribution of heat.

—Joseph Black, *Lectures on the Elements of Chemistry* [pp. 21–22, 74, 75]

Quotation 4.40

It was formerly a common supposition, that the quantities of heat required to increase the heat of different bodies by the same number of degrees, were directly in proportion to the quantity of matter in each; and therefore, when the bodies were of equal size, the quantities of heat were in proportion to their density. But very soon after I began to think on this subject (anno 1760), I perceived that this opinion was a mistake, and that the quantities of heat which different kinds of matter must receive, to reduce them to an equilibrium with one another, or to raise their temperature by an equal number of degrees, are not in proportion to the quantity of

continued on next page

that the *caloricum* had some weight, though very small in comparison to the total weight of the object containing it. There were also measurements according to which a negative weight would have had to be assigned to the *caloricum*. The idea of a weightless substance—imponderable matter—would have caused no particular difficulties; physicists had already become accustomed to it in the case of light as well as that of the electric *fluidum*.

4.5.3 *Rumford: But Heat Is Still a Form of Motion!*

Declaring that heat is a form of motion, Count Rumford (Benjamin Thompson, 1753–1814) took up the fight against the substance theory of heat. In the course of an adventurous life, Rumford worked in a wide variety of fields and left behind contributions not only to science but also to other realms of human endeavor. In 1800, for example, he founded the English Royal Institution, whose first director was Davy and which later became famous through the work of Faraday. Rumford's scientific work was done in Munich, where he served as advisor to the Bavarian king and later as director of the military arsenal. He also founded social institutions and a state employment agency; and to this day, his name lives on in Rumford soup and the Rumford fireplace.

Rumford first set himself the task of verifying and completing the measurements relating to the weight of the heat substance. He noticed that latent heat was a suitable object of these measurements because, for example, a relatively large amount of heat is required for the melting of ice, and the same quantity of heat is liberated when water freezes without a change in its temperature. This method allowed for the elimination of the sources of error resulting from differences in temperature. With carefully executed experiments, in which he investigated the influence of even the smallest possible differences in temperature on the balance arm, Rumford was able to prove unambiguously that the heat substance, if indeed it existed, must have a vanishingly small weight. According to Rumford, he achieved such a degree of precision in his weighing that he would have been able to detect a deviation of one part in a million in the weight of an object. Rumford's description of his measurements can be found in Quotation 4.42, together with a cautious note that his results are almost self-evident if one considers heat not as a substance, but as motion.

The weakest point of the substance theory was its inability to provide a plausible explanation for the creation of heat through friction. Within the framework of substance theory, it was necessary to assume that friction so alters the state of a body that its heat capacity is reduced, with the result that the unchanged quantity of heat substance can raise its temperature. By examining the heat generated by the process of boring cannon barrels in great detail, Rumford intended to deal a death-blow to this theory. He was first able to prove that the specific heat of the shavings from the barrel remained unchanged. Then he was able to establish that from a given body kept warm by friction we can extract a heat quantity proportional to time; in other words, we can extract as much heat from it as we want. From these observations, Rumford drew confidently his final conclusions: Heat cannot be a substance, because if it were, then one would not be able to remove an unlimited quantity from a body. Heat can be nothing other than motion that can be continually recreated by mechanical friction, so that one can draw off heat from a body as long as this heat is created by mechanical work (Quotation 4.43). Quotation 4.44, from Rumford's article, shows that he came quite close to recogniz-

ing the equivalence of mechanical energy and heat energy. From the results of his measurements, with hindsight, we can even determine a value for the equivalence, which will play an important role later.

In saying farewell to Rumford, we must mention that he saw his investigations as basic research without immediate practical application and therefore had no expectation of material support. He writes in self-justification that the cannon barrels used in his experiments did not go to waste, but were put to use for their intended purpose.

From today's vantage point of the complete triumph of the kinetic theory, we would be inclined to see Rumford's conclusions as definitive. However, such was not the case at the beginning of the nineteenth century. Rumford's experimental results were accepted, to be sure, but the attempt was made to interpret them on the basis of heat substance. The fact, for example, that one can remove an inexhaustible quantity of heat from a body was interpreted to mean that in such cases the body serves only as a conduit; in reality, the heat substance flows into the body from the environment, which constitutes a practically inexhaustible reservoir.

To be sure, the kinetic theory also had its difficulties in the quantitative—and also qualitative—explanation of certain phenomena.

The following table summarizes the most important phenomena as well as commentary as to which of the two theories—heat-substance theory and kinetic theory—can more convincingly explain them.

	Heat conduction	Thermal radiation	Latent heat	Frictional heat	Quantitative conclusion possible?
Heat substance theory	yes	yes	yes	no	yes
Kinetic theory	yes	no	no	yes	no

To this table it must be added that the flow of heat substance (thermal conduction) could be presented as a very simple analogy to fluid flow, whereas with the kinetic theory, conduction was very difficult to describe because knowledge of the statistical nature of collision processes would have been needed. Nevertheless, in the table we state that the kinetic theory would eventually be able to accomplish this task. The *caloricum* theory can easily explain thermal radiation, that is, transmission of heat through a vacuum, as a flow of imponderable heat substance through a vacuum. The kinetic theory is helpless in the face of this phenomenon. Today we say that, in this situation, heat is transmitted from one body to another as electromagnetic radiation, that is, in a form of motion different from the usual motion of matter.

Looking at the above table, it is not difficult to understand why the great majority of scholars in the first decades of the nineteenth century still accepted the substance theory—not with full conviction, but as a useful working hypothesis.

4.5.4 *Fourier's Theory of Heat Conduction*

One of the most important successes of heat substance theory was achieved by Fourier with his mathematical theory of heat conduction.

Jean Baptiste Joseph Fourier (Figure 4.121) came from a poor family. It was thanks to the French Revolution—and, following it, Napoleon—that he had

Quotation 4.40, ***continued***

matter in each, but in proportions widely different from this, and for which no general principle or reason can yet be assigned. ...

Quicksilver, therefore, has less *capacity* for the matter of heat than water (if I may be allowed to use this expression) has; it requires a smaller quantity of it to raise its temperature by the same number of degrees. ... We must, therefore, conclude that different bodies, although they be of the same size, or even of the same weight, when they are reduced to the same temperature or degree of heat, whatever that be, may contain very different quantities of the matter of heat; which different quantities are necessary to bring them to this level, or equilibrium, with one another.

It may have been remarked that the discoveries which have been made in this way are very unfavorable to one of the opinions which have been formed of the nature of heat. Many have supposed that heat is a tremulous, or other, motion of the particles of matter, which tremendous motion they imagined to be communicated from one body to another. But, if this were true, we must admit that the communication would be in conformity with our general experience of the communication of tremulous motion. We are not at liberty to feign laws of motion different from those already admitted, otherwise we can make any supposition account for any phenomena that we please. The denser substances ought surely to be the most powerful in communicating heat to others, or exciting it in them. The fact, however, in a great many examples, and yet not in all, is just the reverse. Such an opinion is therefore totally inconsistent with the phenomena. I do not see how this objection can be evaded.

—Joseph Black, *Lectures on the Elements of Chemistry* [pp. 76, 77, 80]

Quotation 4.41

Fluidity was universally considered as produced by a small addition to the quantity of heat which a body contains, when it is once heated up to its melting point; and the return of such a body to a solid state as depending on a very small diminution of the quantity of heat after it is cooled to the same degree; that a solid body, when it is changed into a fluid, receives no greater addition to the heat within it than what is measured by the elevation of temperature indicated after fusion by the thermometer; and that, when the melted body is again made to congeal, by a diminution of its heat, it suffers no greater loss of heat than what is indicated also by the simple application to it of the same instrument.

continued on next page

Quotation 4.41, ***continued***

This was the universal opinion on this subject, so far as I know, when I began to read my lectures in the University of Glasgow, in the year 1757. But I soon found reason to object to it, as inconsistent with many remarkable facts, when attentively considered; and I endeavored to shew, that these facts are convincing proofs that fluidity is produced by heat in a very different manner.

I shall now describe the manner in which fluidity appeared to me to be produced by heat, and we shall then compare the former and my view of the subject with the phenomena.

The opinion I formed from attentive observation of the facts and phenomena, is as follows. When ice, for example, or any other solid substance, is changing into a fluid by heat, I am of opinion that it receives a much greater quantity of heat than what is perceptible in it immediately after by the thermometer. A great quantity of heat enters into it, on this occasion, without making it apparently warmer, when tried by that instrument. This heat, however, must be thrown into it, in order to give it the form of a fluid; and I affirm that this great addition of heat is the principal, and most immediate cause of the fluidity induced.

And, on the other hand, when we deprive such a body of its fluidity again, by a diminution of its heat, a very great quantity of heat comes out of it, while it is assuming a solid form, the loss of which heat is not to be perceived by the common manner of using the thermometer. ...

When we perceive that what we call heat disappears in the liquefaction of ice, and reappears in the congelation of water, and a number of analogous phenomena, we can hardly avoid thinking it a substance, which may be united with the particles of water, in the same manner as the particles of Glauber's salt are united with them in solution, and may be separated as these are. But, since heat has never been observed by us in a separate state, all our notions of this union must be hypothetical. ...

Many have been speculations and views of ingenious men about this union of bodies with heat. But, as they are all hypothetical, and as the hypothesis is of the most complicated nature, being in fact a hypothetical application of another hypothesis, I cannot hope for much useful information by attending to it. A nice adaption of conditions will make almost any hypothesis agree with the phenomena. This will please our imagination, but does not advance our knowledge.

—Joseph Black, *Lectures on the Elements of Chemistry* [pp. 111–112, 184, 185]

the opportunity to develop his talents. He had begun his investigations related to thermal conduction as early as 1807, but their publication was at first rejected by the strict referees Lagrange, Laplace, and Legendre. Fourier's results finally appeared in 1822, in the book *Théorie analytique de la chaleur*, in which he starts out with a very clear physical description of the problem and proceeds to offer an illuminative presentation of the mathematical issues involved.

For the following discussion, let us imagine a body with an arbitrary initial temperature distribution. We could picture this by using tiny thermometers applied to various locations on the body. Each thermometer would at the outset show a particular value, perhaps all identical. From its surface, the body gives off heat, which passes into the environment, while heat flows from the interior of the body toward the surface. We can therefore expect that the thermometers located at various positions around the body, at any given time, will show different temperatures (Figure 4.122). The goal of the mathematical investigation is then to determine the temperatures displayed by the thermometers as a function of position and time.

Fourier's research was of fundamental importance for the development of both mathematics and physics. Fourier found that the solution to the mathematical form of the problem, at least in its simplest case where the temperature is sought as a function of position, can be represented by combinations of sine functions with various arguments. Since the initial condition, that is, the temperature distribution at the beginning of the experiment, can be prescribed arbitrarily, it is natural to ask, *how can an arbitrary function be built up just from sine functions?* In the language of mathematics, this question relates to the representation of arbitrary functions by what are now called Fourier series. Thus, Fourier's approach opened up an entire new branch of mathematics, the theory of Fourier series, and physicists also found the Fourier series an extraordinarily effective method for solving many of their problems.

In the following, we stick to Fourier's description of the problem and, for the sake of simplicity, confine our attention to the one-dimensional case.

As shown in Figure 4.123, suppose we are given, at time $t = 0$, the temperature distribution $T(x,0)$ in a rod that runs along the x-axis. Furthermore, suppose that for all future times $t > 0$, the two ends of the rod are held at a fixed temperature, for example 0° C. This can be realized in practice by placing the ends of the rod in thermal contact with a very large container filled with water at 0° C. We assume that the heat capacity of this container is so great that the heat flowing from the rod into the heat sink causes no change in the container's temperature. We assume that the cylindrical rod is covered by a heat-insulating layer, so that heat can only flow along the axis of the rod. Fourier was already aware, or, to put it more accurately, correctly assumed, that the heat flow J is proportional to the difference in temperature per unit length, or equivalently, the temperature gradient $\partial T/\partial x$, and flows in the direction of decreasing temperature. With this assumption, Fourier arrived at the relationship

$$J = -\kappa \frac{\partial T}{\partial x},$$

where κ is the so-called thermal conductivity constant. Moreover, Fourier took into account that the heat content of a segment of the rod of length ∂x can vary over time, since the quantities of heat flowing in or out of the two lateral surfaces do not have to be equal. This yields the second equation,

$$\frac{\partial J}{\partial x} = -\rho c \frac{\partial T}{\partial t},$$

where c is the specific heat and ρ is the thickness of the rod. (For the sake of simplicity, we have set the cross-sectional area of the rod equal to 1.) Combining the two equations, we immediately obtain the following differential equation for thermal conduction:

$$\frac{\partial^2 T}{\partial x^2} = \frac{\rho c}{\kappa} \frac{\partial T}{\partial t} = k^2 \frac{\partial T}{\partial t}. \qquad (1)$$

Even today, differential equations of this type are called heat equations, even though we meet them in other branches of physics as well, such as the theory of long-distance electrical transmission and the theory of diffusion processes. We therefore seek the function

$$T = T(x,t)$$

that satisfies the above differential equation, with the boundary conditions

$$T(0,t) = 0, \quad T(l,t) = 0, \quad t > 0,$$

and the initial condition

$$T(x,0) = f(x),$$

where $f(x)$ is some prescribed function.

Fourier solved equation (1) by applying the usual method of separation of variables. He assumed that the desired function could be represented as the product of two functions, one of them a function of x alone and the other a function of t alone:

$$T(x,t) = \Phi(x)\Psi(t).$$

If we substitute this product into the initial differential equation,

$$\Psi(t)\frac{\mathrm{d}^2\Phi(x)}{\mathrm{d}x^2} = k^2\Phi(x)\frac{\mathrm{d}\Psi(t)}{\mathrm{d}t},$$

we obtain, after a bit of manipulation,

$$\frac{1}{k^2\Phi(x)}\frac{\mathrm{d}^2\Phi(x)}{\mathrm{d}x^2} = \frac{1}{\Psi(t)}\frac{\mathrm{d}\Psi(t)}{\mathrm{d}t} = -\lambda.$$

One side of this equation depends only on the variable x, whereas the other side depends only on t, so the equation can only be satisfied if both sides are equal to some constant. (This constant, denoted by λ, is called the separation constant.) If this is the case, then the initial differential equation, which is a partial differential equation, decomposes into two ordinary differential equations

$$\frac{\mathrm{d}^2\Phi(x)}{\mathrm{d}x^2} + \lambda k\Phi(x) = 0, \quad \frac{\mathrm{d}\Psi(t)}{\mathrm{d}t} + \lambda\Psi(t) = 0,$$

whose solutions can be written at once:

$$\Phi(x) = b\sin\left(\sqrt{\lambda}kx + c\right), \quad \Psi(t) = a\,\mathrm{e}^{-\lambda t}.$$

The boundary conditions are satisfied when we require of the function $\Phi(x)$ that

$$\Phi(0) = \Phi(l) = 0.$$

These conditions are satisfied if we have

$$c = 0, \quad \sqrt{\lambda}kl = \pi v, \quad v = 1,2,\ldots,$$

and so the separation constant assumes the value

$$\lambda_v = \left(\frac{\pi v}{kl}\right)^2.$$

Taking into account the time-dependent term, we obtain

$$T(x,t) = \sum_v T_v(x,t) = \sum_v c_v \exp\left[-\left(\frac{\pi v}{kl}\right)^2 t\right]\sin\frac{v\pi x}{l} \quad (c_v = ab_v),$$

where we have taken into consideration the fact that for the linear differential equation (1), the sum of solutions is again a solution.

Quotation 4.42

Having determined that water does not acquire or lose any weight upon being changed from a state of liquid to that of ice, and vice versa, I shall now take my final leave of a subject which has long occupied me, and which has cost me much pains and trouble; for I am fully convinced, from the results of the afore-mentioned experiments, that if heat be in fact a *substance*, or matter—a fluid *sui generis*, as has been supposed—which, passing from one body to another, and being accumulated, is the immediate cause of the phenomena we observe in heated bodies, it must be something so infinitely rare, even in its most condensed state, as to baffle all our attempts to discover its weight. And if the opinion which has been adopted by many of our ablest philosophers, that heat is nothing more than an intestine [internal] vibratory motion of the constituent parts of heated bodies, should be well founded, it is clear that the weights of bodies can in no wise be affected by such motion.

—Count Rumford, "An Inquiry Concerning the Weight Ascribed to Heat," 1799 [pp. 58–59]

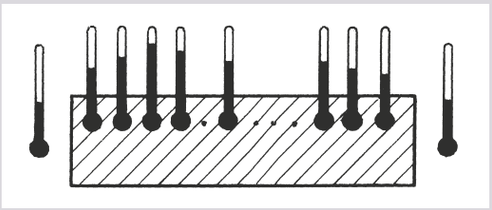
Quotation 4.43

It is hardly necessary to add, that any thing which any insulated body, or system of bodies, can contribute to furnish without limitation, cannot possibly be a material substance: and it appears to me to be extremely difficult, if not quite impossible, to form any distinct idea of any thing, capable of being excited, and communicated, in the manner the heat was excited and communicated in these experiments, except it be *motion*.

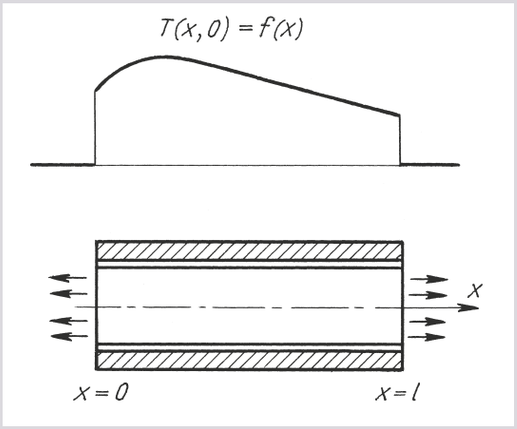
—Count Rumford, "An Inquiry Concerning the Source of the Heat Which Is Excited by Friction" [p. 99]



◀ Figure 4.121 Joseph Fourier (1768–1830): 1796: professor at the École Polytechnique. 1816: member of the French Academy. His *Théorie analytique de la chaleur*, in which the theory of the Fourier series also appears, was published in 1822. He also investigated problems of measurement errors.



▲ **Figure 4.122** FOURIER's problem: At time $t = 0$, a certain temperature distribution in a body is given. How will the temperature change at various points in the body as a function of time?



▲ **Figure 4.123** The one-dimensional heat-conduction problem: We are investigating the cooling of the rod given the initial temperature distribution and given that the rod can transmit heat to the environment only at its end surfaces.



◀ **Figure 4.124** JOSEPH-LOUIS GAY-LUSSAC (1778–1850): After his studies at the École Polytechnique, worked as a chemist in BERTHOLLET's laboratory (in Arcueil), then was associated for a time with HUMBOLDT (who also belonged to the Arcueil circle; the director of the physics department of the laboratory was LAPLACE). 1809–1832: professor of chemistry at the École Polytechnique and professor of physics at the Sorbonne.

Two laws are associated with his name: 1802, investigation of the expansion of gases under constant pressure, where he obtained the relation

$$V = V_0\left[1+\alpha(t-t_0)\right].$$

The coefficient of expansion a in this formula is the same for all gases to very good approximation. GAY-LUSSAC also measured its value, and for $t = 0$°C he obtained the excellent result a = (1/267) K^{-1} (the exact value being 1/273.15 K^{-1}).

continued on next page

The mathematical problem of representing an arbitrary function by a Fourier series arises in the attempt to satisfy the initial conditions. If we set $t = 0$ in the general solution procedure, we obtain the relationship

$$T(x,0) = f(x) = \sum_{v=1}^{\infty} c_v \sin\frac{v\pi x}{l}.$$

On the left-hand side of this equation, we have an arbitrary function, and on the right-hand side, we have an infinite series of sine functions.

In order to determine the unknown coefficients c_v associated with the function $f(x)$, FOURIER at first offered a very complicated method that was extremely questionable with respect to mathematical rigor. He developed the sine functions into the power series

$$\sin vx = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} v^{2n-1}}{(2n-1)!} x^{2n-1}.$$

He then interchanged the order of summation, thereby obtaining

$$f(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)!} \left(\sum_{v=1}^{\infty} v^{2n-1} c_v \right) x^{2n-1}.$$

FOURIER then compared the resulting power series with the Maclaurin series of the function $f(x)$:

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(0) x^k,$$

finally obtaining the following algebraic system of infinitely many equations in an infinite number of unknowns c_v:

$$\sum_{v=1}^{\infty} v^{2n-1} c_v = (-1)^{n-1} f^{(2n-1)}(0), \quad n = 1,2,3\ldots.$$

Later, FOURIER also found the relationship, familiar to us today,

$$c_v = \frac{2}{\pi}\int_0^{\pi} f(s)\sin vs \,\mathrm{d}s.$$

In studying this relationship, FOURIER determined that the value of the integral can be interpreted geometrically as the area under the graph of the function $f(s)\sin vs$ over the interval from 0 to π, and therefore one can carry out the transformation as well for the case of "arbitrary" functions that are not analytic in the sense of LAGRANGE. This way, whether he intended or not, FOURIER contributed to the extension of the notion of function.

4.5.5 *Caloricum and the State Equation*

The theory of the *caloricum* was able to produce useful results even beyond the connection with heat flow. LAPLACE attempted also to derive the state equations for gases on the basis of this theory. Reading his works today, one cannot help but be amazed by the ability of this brilliant mathematician to juggle hypotheses and formulas. LAPLACE begins with the assumption that due to the heat substance that surrounds the particles of mater like a cloud, a repulsive force emerges between the particles that can be expressed by the relationship

$$F = Hc^2\varphi(r),$$

where H is a constant, c the density of the *caloricum*, and $\varphi(r)$ a function that decreases rapidly with the distance between the particles. (We note, in this discussion, that we are assuming for the first time the action of short-range forces.) By considering the sum of such forces, LAPLACE derives the formula

$$P = 2\pi HK\rho^2 c^2$$

for the pressure of the gas. The constant K describes the resultant of all forces acting on a single particle. This constant has a finite value even in the case of infinitely many particles because of the above-mentioned short range of interaction between the particles. However, the derived formula is obviously incorrect because the gas density ρ appears to the second power instead of to the correct first. This issue was explained away effortlessly by LAPLACE by arguing that the density of the material and the density of the

caloricum are not independent of each other—indeed, that as a consequence of the radiative equilibrium between them, there must exist a dependence of the form

$$\rho c^2 = \Pi(u),$$

in which the right-hand side depends solely on the temperature u. With this assumption, Laplace finally obtains

$$P \sim \rho\Pi(u),$$

which is the correct result: The equation has the same form of Gay-Lussac's law if, like Laplace, we use the quantity $\Pi(u)$ that depends only on the temperature instead of the temperature itself (Figure 4.124).

The greatest accomplishment of Laplace's theory is the derivation, in which he uses the method described above, of the state equation for the adiabatic change of state:

$$pV^{\kappa} = \text{constant}.$$

4.5.6 *The Carnot Cycle*

Efforts to increase the efficiency of heat engines led to the recognition—especially by James Watt (Figure 4.125)—of the necessity to merge theory with practical application. The experiments of Sadi Carnot (Figure 4.126) were also dedicated to this task. Carnot envisioned the mode of operation of a heat engine as being similar to that of a water-powered engine, whereby water (in itself incapable of being transformed into work) cascades from a higher to a lower level, thereby producing mechanical work. In the case of heat engines, one can think of the water level as being replaced by the temperature, and the flowing mass of water by the heat substance that, like the water, is neither created nor destroyed. If one accepts this analogy in calculating the efficiency of a steam engine, then the result is

$$\eta = \frac{Q(T_1 - T_2)}{QT_1} = \frac{T_1 - T_2}{T_1} = 1 - \frac{T_2}{T_1}.$$

This example demonstrates both the heuristic value of analogical thinking and the dangers of its uncritical application. If we compare, following Figure 4.127, the operating principles of a water-powered engine, a heat engine, and a direct-current motor, we can see that in all three cases, work is performed by a flow from a higher level (a higher potential) to a lower, assuming, however, that the quantities of electricity and heat remain unchanged. The analogy between the water-wheel and the electric motor works, in the sense that in both cases the inflow per unit time of substance (water or electric charge) at the higher level is equal to the outflow per unit time at the lower level. However, Carnot erroneously believed the heat substance to be conserved, when, in fact, the amount of heat at a lower temperature flowing out of a heat engine is less than the amount at a higher temperature flowing in. Part of the heat is transformed in the machine into mechanical work. Such cannot be said about either water or electric charge, that some part of the water or some of the charge were somehow turned into mechanical work.

The surprising fact is that the Carnot formula for efficiency is still correct and is still in use today. But there is a simple explanation for this: if in a given time interval we supply Q_1 heat at temperature T_1 to the engine, and at the same time remove Q_2 heat and temperature T_2, then, in the ideal case, the difference $Q_1 - Q_2$ is turned into mechanical work, so the efficiency will be

$$\eta = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1}.$$

Figure 4.124 *continued*

1808: publication of the results of several years of research, according to which the volumes of chemically reacting gases and the gaseous reaction products are in the proportions of whole numbers that tend to be small. For example, if a unit volume of hydrogen reacts with a unit volume of chlorine, then after the reaction, there will be two unit volumes of hydrochloric acid: $H_2 + Cl_2 = 2HCl$. Dalton propounded the law of simple and multiple proportions around this time. Dalton and Gay-Lussac did not get along, so it was only later that Amadeo Avogadro was able to unite their results, by pointing out the difference between atoms and molecules and thereby formulating the law that bears his name.

Of interest are also the balloon ascents that Gay-Lussac made in 1804 with Biot for measuring Earth's magnetic field (Figure 4.88). Later, Gay-Lussac ascended alone to a height of 7000 meters to measure air pressure and temperature and take air samples at various altitudes.

An important result of Gay-Lussac's multifaceted work in the area of chemistry is the discovery (together with Thénard) of the element boron.

We should mention as well that Jacques-Alexandre-César Charles, in part before Gay-Lussac and in part at the same time, but independently, also discovered the law of the thermal expansion of gases.

Putting together the laws of Boyle–Mariotte and Gay-Lussac yields the state equation for ideal gases, and after the introduction of absolute temperature $T = t + 273.15$, the formula $pV = RT$ for this law becomes obvious. Clapeyron gave the state equation in this form, but it was a long journey before W. Thomson (Lord Kelvin) introduced the thermodynamic temperature scale named for him (1848), which is independent of the properties of the ideal gas.

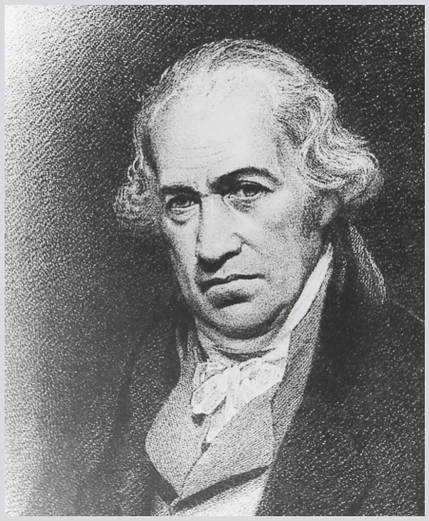
Quotation 4.44

As the machinery used in this experiment could easily be carried round by the force of one horse (though, to render the work lighter, two horses were actually employed in doing it), these computations show further how large a quantity of heat might be produced, by proper mechanical contrivance, merely by the strength of a horse, without either fire, light, combustion, or chemical decomposition; and, in a case of necessity, the heat thus produced might be used in cooking victuals.

But no circumstances can be imagined, in which this method of procuring heat would not be disadvantageous; for, more heat might be obtained by using the fodder necessary for the support of a horse, as fuel ...

For fear I should be suspected of prodigality in the prosecution of my philosophical researches, I think it is necessary to inform the Society, that the cannon I made use of in this experiment was not sacrificed to it. The short hollow cylinder which was formed at the end of it ... called in the German language the *verlorner Kopf* ... the head of the cannon to be thrown away.

—Count Rumford, "An Inquiry Concerning the Source of the Heat Which Is Excited by Friction" [p. 95]



◀ Figure 4.125
James Watt (1736–1819): Son of a well-to-do shipwright. Unable to attend school regularly because of his health. In 1757, opened in Glasgow a mechanical workshop affiliated with the university, and here, in addition to his skills as a toolmaker, acquired the theoretical knowledge that turned him from a technician into a scientist. In 1763, he began the repairs to a model of a Newcomen engine that belonged to the university. 1765: the condensing steam engine is born. The centrifugal regulator and the indicator for measuring steam pressure in the cylinder are also his inventions. Watt's significant contributions are twofold: He brought the steam engine into a practical form and helped it thereby to achieve the role that it would play in the course of the Industrial Revolution, while also giving impetus to scientifically based investigation into increasing the engine's efficiency, on which the foundations of thermodynamics were laid.



◀ Figure 4.126
Nicolas Léonard Sadi Carnot (1796–1832): Engineering officer. Wrote his famous treatise *Réflexion sur la puissance motrice de feu et sur les machines propres à développer cette puissance* (published in Paris in 1824) in 1821 while fleeing the Bourbons. In this work, Carnot still supports the existence of a heat substance (*caloricum*); however, from notes that he wrote shortly before his death, it is clear that he had recognized the kinetic theory as a possible alternative. One can even find in his ideas hints of the principle of conservation of energy.

Today we know that, in the case of the process investigated by Carnot, the so called *Carnot cycle* (Figure 4.128), the following relationship holds:

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2},$$

In other words,

$$\frac{T_2}{T_1} = \frac{Q_2}{Q_1}.$$

This means that Carnot's formula, even though it was derived on the basis of incorrect assumptions, happens to be completely correct. In his later years—as it is evident from his correspondence and notes—Carnot saw the relationship between heat and mechanical work more clearly. His untimely death prevented him from taking the decisive step, but the following quotation shows just how close he had come to the principle of conservation of energy.

> It may be objected at this point that perpetual motion, which has been shown to be impossible for purely mechanical processes, may not be so where the effects of heat or electricity are concerned. But is it conceivable that the phenomena of heat or electricity are [not] due to ... some kind of motion and hence that they are not governed by the general laws of mechanics? Moreover, do we not know from experience that every attempt to produce perpetual motion has been futile, whatever method has been used, and that no one has ever managed to generate motion that is truly perpetual, in other words, motion that would continue without any effect on the materials being used to create it?
>
> It has sometimes been suggested that the electrical generating apparatus (Volta's pile) could be a source of perpetual motion, and attempts have been made to achieve this by constructing dry piles which are claimed to undergo no changes. But, no matter what procedure has been adopted, the devices have always deteriorated quite markedly, whenever they have been used at all vigorously and for any length of time.
>
> In a broad philosophical sense, the words "perpetual motion" should not be taken simply to mean motion that will continue indefinitely, following some initial impetus. They refer to the effect produced by a device, or any apparatus, that is capable of yielding unlimited quantities of motive power, capable of successively disturbing the state of rest of all those objects in nature that are not in motion and of overcoming their principle of inertia. Such a device would even possess the force necessary to move the entire universe and to sustain and constantly accelerate its motion. This is what we mean by the true creation of motive power. If such creation was possible, we would have no need to seek motive power in air currents, water, or combustible materials. We would have at our disposal a limitless supply on which we could draw infinitely.
>
> —Sadi Carnot, *Reflexions on the Motive Power of Fire* [pp. 69–70]

4.5.7 *The Kinetic Theory of Heat: First Steps*

With Carnot's results, the theory of heat substance reached the limits of its explanatory power. While in the first half of the nineteenth century this theory continued to serve as the background for investigations into the physics of heat, even its most successful proponents, such as Carnot and Laplace, were already considering the possibility that heat is, in fact, a form of motion of the particles of

matter. The only problem was that proceeding from this point of view, they were not able to derive quantitative laws.

As we mentioned before, the kinetic theory already had a long history. Indeed, Bacon (Section 3.4) had categorically declared that there must exist a close relationship between heat and motion. This idea was generally accepted throughout the seventeenth century. However, the quantitative formulation of the theory was first taken up seriously by the "Basel school." Jakob Hermann (1678–1733) had claimed in 1716, although without any justification, that heat is proportional to the density of a body and the square of its motion. As in so many other areas, here, too, Euler contributed pioneering work: He arrived at a numerical value for the speed of the particles: $v \approx 477\ \mathrm{ms}^{-1}$. In the year 1738, Daniel Bernoulli, in his work on hydrodynamics, derived the relationship

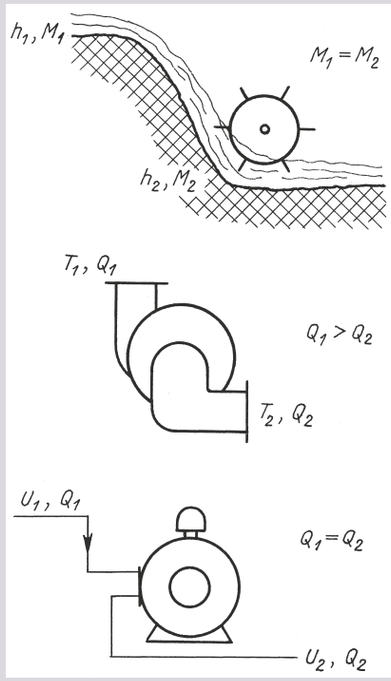
$$p \sim nmv^2,$$

where he began with assumptions that are very similar to those underlying today's kinetic theory of gases. Figure 4.129 is taken from his book. Bernoulli surmised that the pressure exerted by a gas is the result of collisions of particles bouncing about as they try to move on a straight path. As can be seen in the figure, Bernoulli pictured the particles of matter neither as identical shapes nor as geometrically regular. Bernoulli's quantitative theories were not immediately taken note of, and they were forgotten for almost a century. The first time we encounter Bernoulli's ideas again is in 1816, in the work of John Herapath (1790–1868), who, however, made no further progress and indeed, contrary to Bernoulli, incorrectly associated temperature with the *impulse* of the particles. Thus it was easy for his opponents, among them no less an authority than Davy, to silence him.

Significant progress was made at the beginning of the 1840s in the theory derived by John James Waterston (1811–1883). The fate of Waterston's investigations is most instructive. An unknown neophyte among professional physicists, he sent his work from Bombay to the Royal Society. His article was turned over to two referees who were recognized specialists. Only one of the two found anything remotely positive to say in an otherwise negative review, while the other declared the entire work to be nothing but nonsense (Quotation 4.45).

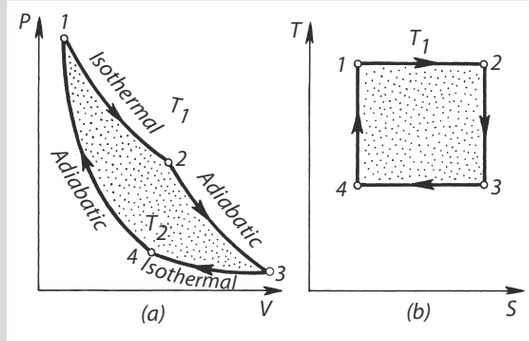
Waterston's work, and indeed his very name, would have faded into obscurity had it not been possible for him finally to publish a short extract from his work and had not Lord Rayleigh discovered, in 1891, the original manuscript in the archives of the Royal Society. As a result, Waterston's article eventually appeared in the journal of the Royal Society after a delay of half a century. In connection with this story, Lord Rayleigh's advice to young physicists is particularly noteworthy: "A young author who believes himself capable of great things would usually do well to secure the favourable recognition of the scientific world by work whose scope is limited, and whose value is easily judged, before embarking on higher flights" (see Quotation 4.45). And Lord Rayleigh meant this completely seriously.

Quotation 4.46 reproduces Waterston's article almost completely, summarizing his theory. The author's "peculiar theory" has an entirely modern tone. We may judge Waterston's most important accomplishment to be that he succeeded in formulating with greater precision the findings of the earlier kinetic theories. Furthermore, he was the first to formulate a special case of the equipartition principle, namely that every atom of two gases at temperature equilibrium has the same average kinetic energy. We shall have more to say about the role of the equipartition theorem when we treat classical statistical theory. Here we mention only that

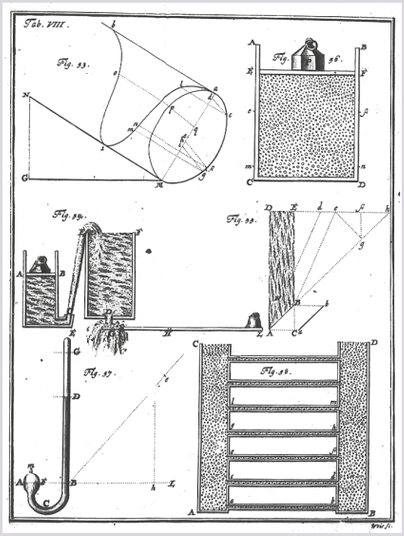


▲ **Figure 4.127** How work is extracted by a water wheel, a heat engine, and a direct-current motor. The water wheel and the electric motor work analogously, but the pattern breaks down with the heat engine because though the heat flows from a higher level to a lower one, it changes in its quantity. This was first noticed by Joule. Carnot did not yet realize the problem in his *Réflexion*… .

> We may rightly compare the moving force of heat with that of a waterfall. The moving force of the latter depends on the height and on the quantity of fluid: the moving force of heat depends on the applied quantity of *caloricum* and—what could be called the height of fall—on the temperature difference that exists between the heat-exchanging bodies.



▲ **Figure 4.128** The Carnot cycle, represented in **(a)** a *V-P* and in **(b)** an *S-T* coordinate system.



◀ **Figure 4.129** An illustration in Bernoulli's work on the kinetic theory of gases (Fig. 96). (Library of the University for Heavy Industry, Miskolc.)

Quotation 4.45

It is difficult to put oneself in imagination into the position of the reader of 1845, and one can understand that the substance of the memoir should have appeared speculative and that its mathematical style should have failed to attract. But it is startling to find a referee expressing the opinion that "the paper is nothing but nonsense, unfit even for reading before the Society." Another remarks "that the whole investigation is confessedly founded on a principle entirely hypothetical, from which it is the object to deduce a mathematical representation of the phenomena of elastic media. It exhibits much skill and many remarkable accordances with the general facts, as well as numerical values furnished by observation.... The original principle itself involves an assumption which seems to me very difficult to admit, and by no means a satisfactory basis for a mathematical theory, viz., that the elasticity of a medium is to be measured by supposing its molecules in vertical motion, and making a succession of impacts against an elastic gravitating plane." ... [The history of this paper suggests that] highly speculative investigations, especially by an unknown author, are best brought before the world through some other channel than a scientific society, which naturally hesitates to admit into its printed records matter of uncertain value. ... [Perhaps one may go further and say that] a young author who believes himself capable of great things would usually do well to secure the favourable recognition of the scientific world by work whose scope is limited, and whose value is easily judged, before embarking on higher flights.

—Lord Rayleigh, Introduction to J. J. Waterston, "On the Physics of Media that Are Composed of Free and Perfectly Elastic Molecules in a State of Motion"

Waterston was the first to succeed in answering Black's question as to why the kinetic theory did not necessarily imply a proportionality between the specific heat of a body and its density.

Given all this, we can well understand why Waterston had absolutely no influence on the further development of the kinetic theory. It was not in England, but in Germany a decade later, that the theory was advanced, or more precisely, restarted from a somewhat more basic level than where Waterston started.

4.5.8 *The Law of Conservation of Energy*

Meanwhile, by the beginning of the 1840s, the time had become ripe for a formulation of the law of conservation of energy.

The discovery of this law is usually linked to three names: Julius Robert Mayer (1814–1878, Figure 4.130), James Prescott Joule (1818–1889, Figure 4.131), and Hermann von Helmholtz (1821–1894, Figure 4.132). However, we should add that similar ideas had occurred to many other researchers. Thus, for example, Justus Liebig wrote about the connection between heat energy and chemistry as well as the role of energy in biological processes.

Robert Mayer, too, was led to his own research by certain physiological phenomena. As a naval physician, he noticed that the blood drawn from the veins of sailors was of a brighter red in the tropics than what was usual in their chillier home climate. He correctly concluded that in the tropics, the oxidation processes of an organism proceed with lesser intensity because part of the heat required by life functions is supplied by the environment. His 1842 paper entitled *Bemerkungen über die Kräfte der unbelebten Natur* (Remarks on the Forces of the Inorganic Nature) was strongly influenced by natural philosophy, and it is therefore no great surprise that it was rejected for publication by the leading physics journal of the time, *Annalen der Physik*. Eventually, Liebig helped to place the article in the journal *Annalen der Chemie*. In reading the introduction (Quotation 4.47), we can understand the unease of the physicists with a work in which every claim is derived from the basic axiom that *causa aequat effectum*, meaning the cause is to be equated with its effect. Of a similar nature is the strange argument with which Robert Mayer tries to make it plausible that falling bodies increase in temperature. He starts from the experience that when we compress a body, that is, when we decrease its volume, heat is created. Let's consider a stone falling from a high place: as it falls, the volume of Earth to which the stone also, after all, belongs will thereby be reduced; hence, the temperature will rise …

At the end of his article, almost unexpectedly, we find the correct conclusion that when a body falls from a given height, its temperature must rise by a corresponding definite amount. This is posed in the form of a question: from what height must a body fall in order to achieve an increase in temperature of one degree? Although Mayer, as we would expect, did not carry out the experiment, he nevertheless gives an answer: Using the relationship of the specific heat of gases under constant pressure and volume, he estimates the mechanical equivalent for heat and obtains a numerical value of about 360 kpm. This value is far from the 425 kpm measured by Joule, which is today considered valid. Nonetheless, the order of magnitude of the result and the theoretical approach were both entirely correct.

In 1841, Joule published the law that is today named for him regarding the heat effect of electric current, whereby the heat generated by current flowing through

a resistor is proportional to the square of the current, the resistance, and the time:

$$Q \sim I^2 Rt.$$

His seminal work on the law of conservation of energy was published in 1845 with the title *On the Existence of an Equivalent Relation between Heat and the Ordinary Forms of Mechanical Power*. The method he used for taking measurements and the measuring device employed are today part and parcel of the standard high school physics curriculum (Figure 4.133). The potential energy of a body of a given weight is transformed into heat as the descent of the weight drives stirring paddles in some liquid. Of course, the measurements must be made with great care. Joule's first measurements did not yield the value that is considered correct today. However, with a refinement of his techniques, elimination of sources of error, or compensating for them, he arrived at a value for the mechanical equivalent of heat that is very close to what is in use today: 1 kcal ≈ 425 kpm.

Helmholtz's work on the conservation of energy was published in 1847 under the title *Über die Erhaltung der Kraft* (*On the Conservation of Force*). Although Helmholtz's style also exhibits the influence of Romantic natural philosophy, his arguments are presented in a form that would completely satisfy the tastes of today's physicists.

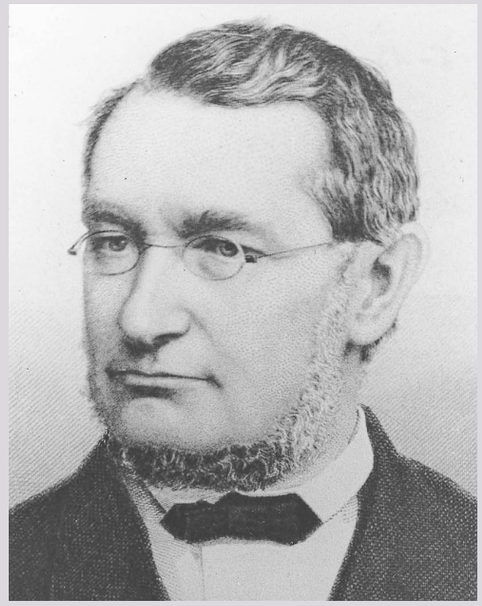
Because of the importance of the law of conservation of energy, it is worth taking the time to read some of the reminiscences of those who participated in its discovery, and here we are thinking primarily of the sober, realistic Joule, who was raised in the tradition of English empiricism. We learn from the memoirs he wrote in 1850, that he, too, was of the opinion that the time was ripe for an articulation of the law of conservation of energy. Nevertheless, this was accepted only with great hesitation by the scientists of the time. Joule's first report met with complete indifference except for one enthusiastic young man who did pay attention to it—with skepticism, to be sure, but nonetheless sensing the importance of the discovery. This was William Thomson, the future Lord Kelvin (Quotation 4.48).

4.5.9 *The Kinetic Theory of Gases*

After the formulation of the law of conservation of energy, the kinetic theory of matter, at first in the form of a kinetic theory of gases, came back into the limelight. More to the point, we should say that this theory was discovered anew, and now nothing stood in the way of its general acknowledgment. Referring to Herapath, in 1848 Joule himself had attempted to calculate the velocity of a hydrogen molecule and the specific heat of a gas under constant volume. His work was published in 1851 but did not arouse any particular attention.

The breakthrough in this approach came thanks to August Karl Krönig (1822–1879). While his brief 1856 article contained no new ideas, because it was the opinion of a noted professor, it gave the long-overdue impetus to the field. From this time on, macroscopic thermodynamics, the kinetic theory of gases, and classical statistics mutually reinforced their parallel development. Among those associated with this era were the likes of Rudolf Clausius (Figure 4.134), J. C. Maxwell, William Thomson (Figure 4.135), and Ludwig Boltzmann (Figure 4.136), to mention only the most important names.

It was Clausius who laid the cornerstone of both the macroscopic and microscopic theories. In his 1857 *Annalen der Physik* article *Über die Art der Bewegung, welche wir Wärme nennen* (On the form of motion that we call heat), he derives



▲ **Figure 4.130** Julius Robert Mayer (1814–1878): His observations on the color of blood in Surabaya (Java), while serving as a ship's surgeon in 1840, provided an impetus to the formulation of the energy conservation law. In 1841, the journal *Annalen der Physik* declined to publish his work, and it appeared a year later in the *Annalen der Chemie* with the title *Bemerkungen über die Kräfte der unbelebten Natur*. In 1845, Mayer thoroughly investigated, in a way that is convincing even to today's physicists, the relationship between the work of expansion of a gas and the participating heat quantity. In 1848, he conjectured that the heat of the Sun stems from the kinetic and gravitational energy of the meteors that strike it. Physicists were slowly becoming aware of the law of conservation of energy, but they considered Joule and Helmholtz as its discoverers. Mayer fought long in vain for recognition of his contribution, suffered a nervous breakdown, and in 1850 attempted suicide. Tyndall, in an 1862 lecture to the Royal Institution, finally conceded that Robert Mayer should be granted first place among the discoverers of the energy conservation law.

Quotation 4.46

The author deduces the properties of gases, with respect to heat and elasticity, from a peculiar form of the theory which regards heat as consisting in small but rapid motions of the particles of matter. He conceives that the atoms of a gas, being perfectly elastic, are in continual motion in all directions, being restrained within a limited space by their collisions with each other, and with the particles of surrounding bodies. The *vis viva* of those motions in a given portion of gas constitutes the quantity of heat contained in it.

He shows that the result of this state of motion must be to give the gas an elasticity proportional to the mean square of the velocity of the molecular motions, and to the total mass of the atoms contained in unity of bulk; that is to say, to the density of the medium.

continued on next page

DOI: 10.1201/9781032697697-6
This chapter has been made available under a CC-BY-NC-ND license.

Quotation 4.46, ***continued***

This elasticity, in a given gas, is the measure of temperature. Equilibrium of pressure and heat between two gases takes place when the number of atoms in unity of volume is equal, and the *vis viva* of each atom equal. Temperature, therefore, in all gases, is proportional to the mass of one atom multiplied by the mean square of the velocity of the molecular motions, being measured from an *absolute zero* 491° below the zero of Fahrenheit's thermometer. [Probably a misprint instead of 461°, the correct value being 459°.]

If a gas be compressed, the mechanical power expended in the compression is transferred to the molecules of the gas increasing their *vis viva*; and conversely, when the gas expands, the mechanical power given out during the expansion is obtained at the expense of the *vis viva* of the atoms.

—J. J. Waterston, Abstract for "On a General Theory of Gases," 1851 [Truesdell 1968, p. 294]



◀ **Figure 4.131** James Prescott Joule (1818–1899): Owner of a brewery, perhaps the last autodidact who made a significant contribution to science. His strength was a finely honed measurement technique. In recognition of his scientific accomplishments, he was named president of the Literary and Philosophical Society of Manchester. In 1840, he worked out the laws for the heat generated by an electric current. 1843: measurement of the heat equivalent of mechanical work. Over three decades, he determined the numerical conversion factors for energy transformation with various techniques; these experiments contributed decisively to the general recognition of the concept of energy and the energy conservation law. His friendship with Thomson (later Lord Kelvin) had a great influence on his work: The Joule–Thomson effect is the change in temperature that occurs in the expansion of a gas without doing work and while kept insulated so that no heat exchange with the environment occurs. At room temperature, all gases—with the exception of hydrogen, helium, and neon—will cool under these conditions, so that this effect can be used for the liquefaction of gases. At very low temperatures, hydrogen and helium can also be cooled this way.

the state equation

$$pv = \frac{nmu^2}{3},$$

where u^2 is mean square velocity of the molecules. His derivation is similar to what appears in today's textbooks on theoretical physics. Clausius determined that the pressure of a gas is a function of the energy of the translational motion of the molecules. The translational energy is a portion of the internal energy of the gas, to which other energies may also contribute. Thus, with increasing complexity of the molecules, the energy contribution resulting from the vibrational motion of the atoms within the molecules increases as well. Clausius derived the following relationship between the kinetic energy of the translational motion E_k and the total heat q:

$$\frac{E_k}{q} = \frac{3}{2}\frac{c_v - c_p}{c_p},$$

where c_v is the specific (related to the volume unit) heat of the gas at constant volume, and c_p is the specific heat under constant pressure.

Clausius also introduced the notion of free path length into the theory and explained the apparent contradiction between the extraordinarily high velocities of the gas particles (for hydrogen, it is 1800 ms^{-1}) and the small speed of diffusion.

For Maxwell, the kinetic gas theory was an interesting exercise in mechanics. In a letter, Maxwell writes that it would be worthwhile to investigate, independently of the kinetic theory of gases, the behavior of a large number of elastic spheres interacting exclusively by collision processes. The main result of his 1860 work is the derivation of the velocity distribution function that now bears his name. Using it, one can determine the number of particles with velocities in the interval from v to $v + \mathrm{d}v$ (Figure 4.137) via

$$\frac{\mathrm{d}N}{N} = \frac{4}{\alpha^3\sqrt{\pi}} v^2 \, \mathrm{e}^{-v^2/\alpha^2} \, \mathrm{d}v,$$

where α^2, defined by $\overline{v^2} = \frac{3}{2}\alpha^2$, is a quantity proportional to the mean square velocity.

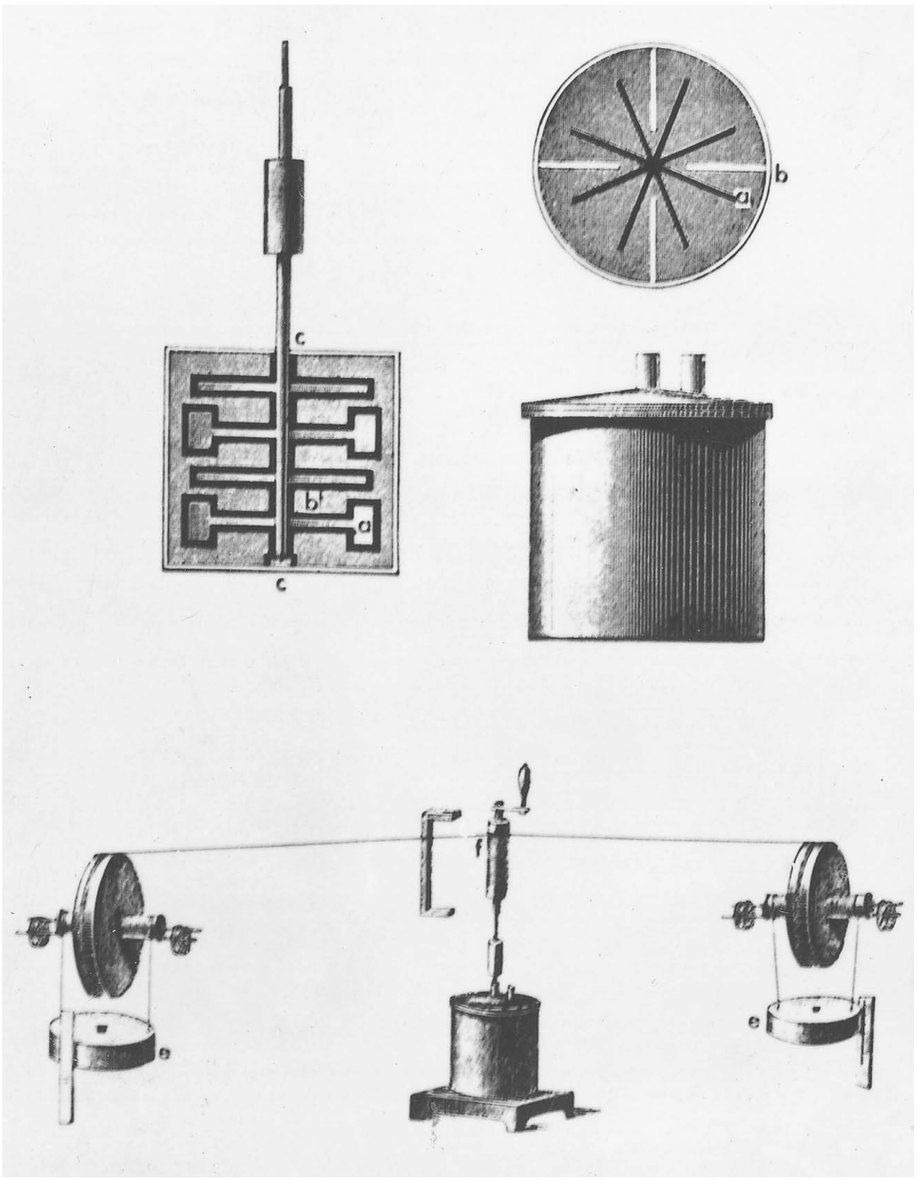
Also of importance is Maxwell's conclusion that for a mixture of gases in thermodynamic equilibrium, every molecule possesses the same average kinetic energy (equipartition theorem). This principle was generalized by Maxwell in the same article: the same energy is associated with each degree of freedom of the gas particles. But equal average energy is also additionally allotted to every degree of freedom that is bound to a potential energy. This average energy is equal to $\frac{1}{2}kT$, where T is the absolute temperature and

$$k = 1.38\times10^{-23}\ \mathrm{JK^{-1}} = 1.38\times10^{-16}\ \mathrm{erg\,K^{-1}}$$

is one of the universal constants of physics (the Boltzmann constant).

Maxwell then calls attention to the fact that the theoretical values thus derived for the specific heats differ from the experimental results by more than the margin of error in the experiments (Figure 4.138).

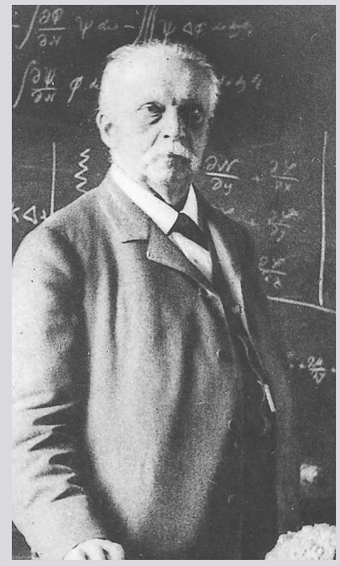
In 1866, Maxwell derived the following general relationship that would later become so significant: $c_p / c_v = (f + 2) / f$, where f is the number of degrees of freedom of the gas molecules.



▲ **Figure 4.133** Joule's experimental apparatus for determining the mechanical equivalent of heat.

4.5.10 *The Second Law of Thermodynamics*

In macroscopic thermodynamics, Clausius again takes center stage. Having accepted the equivalence of heat and mechanical work, or, more generally, the principle of conservation of energy, he noticed that for an explanation of thermodynamic phenomena, yet another concept was also needed. He began with the 1854 axiom that heat cannot of its own accord pass from a colder body to a warmer one. This axiom is a generally understandable way of stating the second law of thermodynamics (Figure 4.139(a)). An axiom equivalent to this is the following: it is impossible to construct a perpetual motion machine of the second type, by which is meant an apparatus that without violating the law of conservation of energy, nevertheless completely (with 100% efficiency) transforms heat into mechanical work (Figure 4.139(b)).



◀ **Figure 4.132** Hermann von Helmholtz (1821–1894): Began his career as a military physician. Part of his scientific work was in the area of physiology: In 1851 he determined the velocity of nerve excitation. 1871: professor at the University of Berlin. His work extended over almost the entire range of physics. 1847: formulation of the law of conservation of energy in a form that is closest to our current view. 1859: working out of the hydrodynamic vortex theorem that bears his name (J. J. Thomson based his vortex atomic model on these theorems). Although himself not an advocate of the atomic theory, in 1881 he emphasized that an atomic structure of matter implies an atomic structure of electricity. In the second half of the nineteenth century, Helmholtz played an outstanding role in the system of higher education and the development of scientific life in Germany. Many German scientists at the beginning of the twentieth century saw Helmholtz as their intellectual father.

Quotation 4.47

Forces are causes; therefore, the following fundamental law applies: *causa aequat effectum*. If cause *c* has effect *e*, then $c = e$. If *e* is itself the cause of another effect *f*, then $e = f$, etc., $c = e = f = \ldots = c$. In a chain of causes and effects, as is illuminated from the nature of an equation, a term or a part of a term can never become zero. This first property of all causes we call *indestructibility*.

If the given cause *c* has caused the like effect *e*, then *c* has ceased to be; *c* has become *e*; if after the creation of *e*, *c* remains in whole or in part, then a further effect would have to correspond to this remaining cause, and the effect of *c* must therefore be $> e$, which contradicts the assumption $c = e$. Since therefore *c* changes into *e*, *e* into *f*, etc., these quantities must be seen as different manifestations of one and the same object. The capacity of being able to assume different forms is the second significant property of all causes. Summarizing both properties, we say, causes are (quantitatively) *indestructible* and (qualitatively) *changeable objects*.

Two categories of causes are to be found in nature between which, according to experience, no transformations take place. One category consists of the causes to which are associated the property of ponderability and impenetrability—matter. The others are the causes that do not have these properties—forces that we may call imponderable in recognition of a property that they lack. Forces

continued on next page

Quotation 4.47, ***continued***

are therefore these: *indestructible*, *changeable*, *imponderable objects*.

It is now agreed that for vanishing motion in many cases (*exceptio confirmat regulam*), no other effect can be found than heat, for the created heat no other cause than motion; we therefore prefer the assumption that heat results from motion to the assumption of a cause without effect and an effect without cause, like a chemist establishing a connection between H and O on the one hand and water on the other instead of unquestioningly allowing H and O to vanish and water to appear in an unexplained manner.

We may clarify the natural relationship that arises among gravitational force, motion, and heat in the following way. We know that heat appears when the individual mass elements of a body move closer together; compression creates heat; now, what holds for the smallest mass elements and their smallest interstices must hold as well for large masses and measurable volumes. The falling of a weight is an actual reduction in the volume of the Earth, and therefore must be related to the resulting heat; this heat must be precisely proportional to the size of the weight and its (original) distance. ...

We close our theses, which arise necessarily from the fundamental law *causa aequat effectum* and stand in full accord with all natural phenomena with a practical consequence. ...

We must find how high a given weight must be raised above the Earth's surface so that its gravitational force is equivalent to the heating of an equal weight of water from 0° to 1° C. That such an equation can actually be established in nature can be considered the summary of what we have said above.

Using the established theorems on the heat and volume relationships of gases, one finds the sinking of a column of mercury compressing a gas to be equal to the heat quantity released by the compression, and it turns out—the relative powers of the capacities of atmospheric air under equal pressure and volume = 1.421—that the falling of a weight from a height of about 365 m corresponds to a warming of an equal weight of water from 0° to 1°. If one compares the power of our best steam engines with this result, one sees how only a small portion of the heat generated in the boiler is actually converted into motion or lifting power, and this could serve as justification for searching for other ways to represent motion fruitfully than the sacrifice of the chemical difference of C and O, namely through transformation of electricity created by chemical means into motion.

—Robert Mayer, "Bemerkungen über die Kräfte der unbelebten Natur," 1842

We arrive at the notion of entropy and thereby a mathematical formulation of the second law of thermodynamics in the following way: In an ideal reversible Carnot cycle, we have $Q_1 / T_1 = Q_2 / T_2$, or equivalently, $Q_1 / T_1 - Q_2 / T_2 = 0$. More generally, we may write $\sum Q_i / T_i = 0$, where the quantities of heat flowing into the system are taken as positive, and those flowing out as negative. Given a general ideal reversible cycle, it can be broken down into small "elementary" cycles. For each of these, the above equation holds with the incoming or outgoing heat quantity $\mathrm{d}Q$ and with the (variable) temperature T. Thus, in 1865, it was established by Clausius that for an arbitrary reversible cycle that is run to completion, one has the equation $\oint \mathrm{d}Q / T = 0$.

From this, it follows that the quantity $\mathrm{d}Q / T$ is a total differential. The value of the line integral for a process that does not return to its original state therefore depends only on the initial and final states:

$$\frac{\mathrm{d}Q}{T} = \mathrm{d}S, \quad \int_A^B \frac{\mathrm{d}Q}{T} = S_B - S_A.$$

If the point A is held fixed, then one has a quantity that depends only on the state B. This new quantity of state is the entropy of the material. For irreversible cyclical processes. Clausius was able to prove the equation

$$\oint \frac{\mathrm{d}Q}{T} < 0,$$

where heat flowing into the system must be taken as positive.

From this, it follows that the entropy of a closed system cannot decrease (Figure 4.139(c)).

The two laws of thermodynamics can now be written as

$$\mathrm{d}U = \mathrm{d}Q + \mathrm{d}A$$

and

$$\mathrm{d}Q \leq T\,\mathrm{d}S,$$

or equivalently:

1. The energy of a closed system is constant.
2. The entropy of a closed system can only increase or remain constant.

The work of the chemist Walther Nernst (1864–1941) and the physicist Max Planck then led (in 1904 and 1911, respectively) to the formulation of the third law of thermodynamics:

3. The entropy of a system in thermodynamic equilibrium approaches zero as the temperature approaches absolute zero.

The further development of phenomenological thermodynamics and its extension to heterogeneous systems in the 1880s is due primarily to Helmholtz and Gibbs. Then in the twentieth century, the focus of interest moved to irreversible processes, with particular emphasis on biological phenomena (Prigogine, Onsager, Gyarmati).

At first glance, entropy appears to be a totally abstract quantity, of interest solely in theoretical physics. However, entropy is a quantity of state just like pressure, volume, and temperature, so it has practical engineering applications, for example, in the design of steam engines with temperature/entropy diagrams (see Figure 4.128).

Many of the phenomena in daily life can be related through the concept of entropy. It makes it possible to describe quantitatively simple laws: for example, if we have in a closed system of bodies at different temperatures, sooner or later their temperatures will come into equilibrium.

This conclusion was projected to the whole universe in the so-called heat death theory, according to which the universe at present is in a state of relatively low entropy, which means that we can find bodies in it with widely differing temperatures. But every process that takes place in the universe leads to an increase in its total entropy, so the universe must eventually reach a state of maximal entropy in which all temperature differences have been equalized, and thus one in which the possibility for life no longer exists. From the standpoint of physics, however, this conclusion is incorrect for two reasons. The first is that the law of increase in

entropy holds true only for closed, finite systems; the second objection relates to the probabilistic nature of the second law, which describes the probable course of events, from which, even if only with a small probability, deviations are possible.

It was Kelvin who, in 1852, first emphasized that in natural processes, there is a general tendency for all forms of energy to be transformed eventually into heat, equalizing any temperature differences. The pessimism of Schopenhauer, which came into fashion at this time, provided the philosophical background for these fresh scientific theories.

The vision of the French poet Jules Laforgue (1860–1887) in *Funeral March for the Death of the Earth* (Quotation 4.49) is just as viable today as it was then, but at the time the catastrophe was going to be due to the blind laws of nature; today, however, it is human nature (or perhaps the blind laws of human nature) that is responsible for what will come.

Furthermore, from what we now know, it seems that the laws of nature are more likely to incinerate us than to freeze us: After a few billion years, our Sun will be transformed into a red giant and will eventually become so large as to encompass Earth's orbit.

Quotation 4.48

[T]he conversion of heat [...] into mechanical effect is probably impossible,* certainly undiscovered. In actual engines for obtaining mechanical effect through the agency of heat, we must consequently look for the source of power, not in any absorption and conversion, but merely in a transmission of heat. ...

*[In the footnote] This opinion seems to be nearly universally held by those who have written on the subject. A contrary opinion however has been advocated by Mr. Joule of Manchester; some very remarkable discoveries which he made with reference to the *generation* of heat by the friction of fluids in motion, and some known experiments with magneto-electric machines, seeming to indicate an actual conversion of mechanical effect into caloric.

—Sir William Thomson (Lord Kelvin), *Mathematical and Physical Papers* [pp. 102–103]

4.5.11 *Entropy and Probability*

By the mid-1860s, the two laws of macroscopic (phenomenological) thermodynamics had firmly taken root; following the interpretation of the state equation, the kinetic theory of gases successfully dealt with the problems of heat conduction and viscosity (Maxwell, *On the Dynamical Theory of Gases*, 1866–1868).

The state equation and even the first law of thermodynamics had been derived from simple mechanical laws (conservation of momentum and energy), so it is no wonder that the need immediately arose to interpret the second law of thermodynamics in terms of the kinetic theory. Entropy plays a central role in the mathematical formulation of the second law, so the first serious efforts were directed toward proving within the framework of the kinetic theory—by investigating reversible processes—that the expression $\mathrm{d}S = \mathrm{d}Q / T$ is a complete differential, in other words, that the entropy S exists as a quantity of state (Rankine 1865, Boltzmann 1866, Clausius 1871).

Their starting point was the principle of least action in a somewhat more general form; but to reach the desired goal, in the beginning the movement of individual particles had to be restricted in extremely unnatural ways, for example, with each particle having to move in a closed path with identical periods. Although these restrictions could be moderated later, by the beginning of the 1870s it had become clear that particularly in the case of irreversible processes, an interpretation of the second law using solely mechanical principles was not possible.

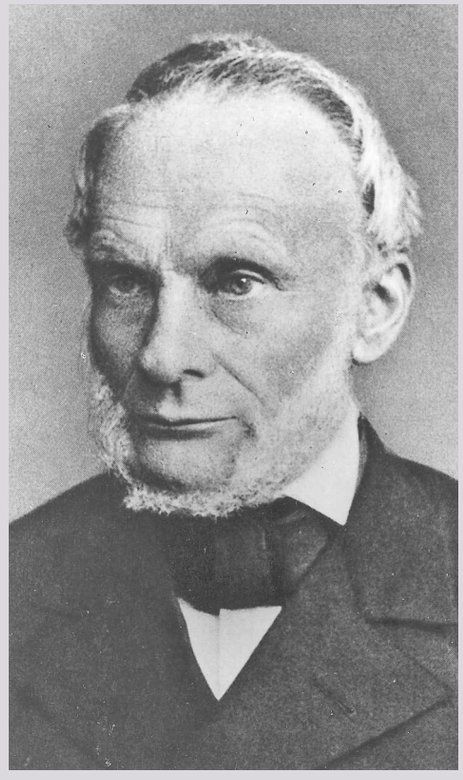
The trains of thought followed by Clausius and Boltzmann can be summarized as follows [Spassky 1977]: We write the action function of an individual particle:

$$\int_{t_1}^{t_2} (\mathfrak{E} - U)\,\mathrm{d}t = \int_{t_1}^{t_2} [\mathfrak{E} - (E_0 - \mathfrak{E})]\,\mathrm{d}t = \int_{t_1}^{t_2} 2\mathfrak{E}\,\mathrm{d}t - \int_{t_1}^{t_2} E_0\,\mathrm{d}t.$$

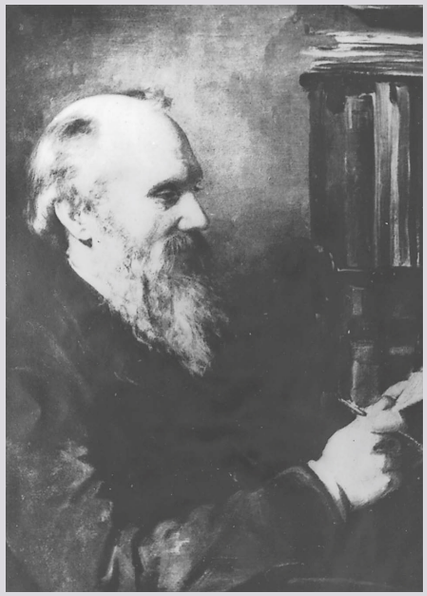
Here $\mathfrak{E}$ is kinetic energy, U is potential energy, and E_0 is total energy. For the variation, the constant E_0 is irrelevant, and so the last term may be omitted. We therefore have

$$\int_{t_1}^{t_2} 2\mathfrak{E}\,\mathrm{d}t = \int_{t_1}^{t_2} mv^2\,\mathrm{d}t = \int_{t_1}^{t_2} mvv\,\mathrm{d}t = \int_{P_1}^{P_2} mv\,\mathrm{d}s.$$

Suppose our system consists of n particles each with three degrees of freedom. And suppose that the state



▲ **Figure 4.134** Rudolf Clausius (1822–1888): After studies at the University of Berlin, taught at a number of German universities as well as the Swiss Federal Institute of Technology, in Zurich. 1865: formulation of the second law of thermodynamics and introduction of the notion of entropy. His work on the kinetic gas theory, which began in 1857, is also of significance.



◀ Figure 4.135 WILLIAM THOMSON (LORD KELVIN) (1824–1907): From 1846 to his retirement, professor at the University of Glasgow. Known primarily for his experiments in the theory of heat (Joule–Thomson effect, introduction of the absolute temperature scale, 1848) and the formula $\omega = 1/\sqrt{LC}$ for the resonance frequency of an oscillating electrical circuit, but he was also instrumental in designing and laying the transatlantic telegraph cable (1854–1866). His thermodynamic experiments were collected in the 1851 publication *On the Dynamical Theory of Heat*.



◀ Figure 4.136 LUDWIG BOLTZMANN (1844–1906): After studies in Vienna, assistant to J. STEFAN. Taught theoretical physics in Graz, then in Vienna, Munich, and Leipzig. Was also an outstanding experimentalist; for example, he verified the relationship $n^2 = \varepsilon_r \mu_r$ required by the Maxwell electromagnetic theory of light for the case of sulfur. His main results: relationship between entropy and thermodynamic probability, MAXWELL–BOLTZMANN distribution function, theoretical foundation of the STEFAN–BOLTZMANN law for blackbody radiation.

BOLTZMANN was a multifaceted personality. He studied music with BRUCKNER and was also active as a writer. In 1906, probably less on account of the attacks of the "energists" than from a deterioration of his mental faculties, he took his own life.

of the system is characterized by specifying the $3n$ positional coordinates in $3n$-dimensional space:

$$x^{(1)}, y^{(1)}, z^{(1)}; x^{(2)}, y^{(2)}, z^{(2)}; \ldots; x^{(n)}, y^{(n)}, z^{(n)}.$$

Suppose that at time t_1, the system is at the point

$$P_1\left(x_1^{(1)}, y_1^{(1)}, z_1^{(1)}; x_1^{(2)}, y_1^{(2)}, z_1^{(2)}; \ldots; x_1^{(n)}, y_1^{(n)}, z_1^{(n)}\right)$$

in configuration space and is moving along a curved path in this space, arriving at time t_2 at the space point

$$P_2\left(x_2^{(1)}, y_2^{(1)}, z_2^{(1)}; x_2^{(2)}, y_2^{(2)}, z_2^{(2)}; \ldots; x_2^{(n)}, y_2^{(n)}, z_2^{(n)}\right).$$

If we supply the system with heat δQ, then the system describes a different, neighboring path between P_1 and P_2. The variation in the action function for the entire system is given, after some not very complicated intermediate calculations, by

$$\delta \int_{t_1}^{t_2} 2\mathfrak{E}\,\mathrm{d}t = \int_{t_1}^{t_2} (\delta\mathfrak{E} + \delta U + \delta L)\,\mathrm{d}t + \sum_i m^{(i)}\left(\dot{x}^{(i)}\delta x^{(i)} + \dot{y}^{(i)}\delta y^{(i)} + \dot{z}^{(i)}\delta z^{(i)}\right)\Big|_{t_1}^{t_2}.$$

Here $\mathfrak{E}$ is now the sum of the kinetic energies of all the particles, U is the potential energy, L is the work of the system against external forces, and $\mathfrak{E} + U = E$ is the total energy of the system. If we now take the average values for all the quantities that appear, for example, in the equation

$$\bar{\mathfrak{E}} = \frac{1}{\tau}\int_{t_1}^{t_2} \mathfrak{E}\,\mathrm{d}t, \quad \tau = t_2 - t_1,$$

for the kinetic energy, we obtain the equation

$$2\delta\left(\tau\bar{\mathfrak{E}}\right) = \tau\left(\delta\bar{E} + \delta\bar{L}\right) + \sum_i m^{(i)}\left(\dot{x}^{(i)}\delta x^{(i)} + \dot{y}^{(i)}\delta yy^{(i)} + \dot{z}^{(i)}\delta z^{(i)}\right)\Big|_{t_1}^{t_2}.$$

However, the heat added to the system serves, on the one hand, to increase the internal energy and, on the other hand, to perform work against the external forces; therefore, we have

$$\delta E + \delta L = \delta Q.$$

We thereby obtain

$$2\delta\left(\tau\bar{\mathfrak{E}}\right) = \tau\delta Q + \sum_i m^{(i)}\left(\dot{x}^{(i)}\delta x^{(i)} + \dot{y}^{(i)}\delta yy^{(i)} + \dot{z}^{(i)}\delta z^{(i)}\right)\Big|_{t_1}^{t_2}.$$

To make the last term disappear, we require the above-mentioned overdetermined conditions. If the particles move over closed paths with a common period $\tau = t_2 - t_1$, then each individual term will be zero in the sum: the state of each particle is the same for times t_1 and t_2. A weaker condition would be to require the entire sum to be equal to zero. In this case, we obtain

$$2\delta\left(\tau\bar{\mathfrak{E}}\right) = \tau\delta Q.$$

However, the average kinetic energy is proportional to the temperature:

$$\bar{\mathfrak{E}} = KT,$$

that is,

$$\frac{\delta Q}{T} = \frac{2K\delta(\tau T)}{\tau T} = 2K\delta\ln\tau T.$$

This proves that—under the given condition—$\delta Q/T$ is a complete differential, that is, its integral depends only on the final state of the system and not on the manner in which this state was achieved. Here we are considering the initial state as fixed.

BOLTZMANN clearly understood in 1872 that "the problems of the mechanical theory of heat are also problems of probability theory." He presented his famous kinetic equation for the change in the general

distribution function $f(x, y, z; \dot{x}, \dot{y}, \dot{z}, t)$ in the following form:

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\dot{x} + \frac{\partial f}{\partial y}\dot{y} + \frac{\partial f}{\partial z}\dot{z} + \frac{X}{m}\frac{\partial f}{\partial \dot{x}} + \frac{Y}{m}\frac{\partial f}{\partial \dot{y}} + \frac{Z}{m}\frac{\partial f}{\partial \dot{z}} = \left[\frac{\partial f}{\partial t}\right]_{\text{collision}},$$

where X, Y, Z denote the components of the external force acting on the particles; the term on the right-hand side gives the change in the distribution function per unit time due to the collisions. This quantity depends in a rather complicated way on the form of the distribution function itself. It is at once clear that the collisions change nothing in the Maxwell–Boltzmann distribution; indeed, they ensure that a gas with such a complicated distribution function at a given point in time will enter into a state of equilibrium with the Maxwell–Boltzmann distribution if no external effects on the gas are permitted. However, that is but one of the consequences of the theorem presented by Boltzmann in 1872 and later called the H theorem, which states that if the distribution function $f(x, y, z; \dot{x}, \dot{y}, \dot{z}, t)$ satisfies the kinetic equation, then as a function of time, the quantity

$$E = \iiint\iiint f \ln f \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, \mathrm{d}\dot{x} \, \mathrm{d}\dot{y} \, \mathrm{d}\dot{z}$$

can only decrease (or remain constant at a minimal value), that is,

$$\frac{\partial E}{\partial t} \leq 0.$$

The integration extends over all possible values of the six variables. The thought immediately arises that E must be closely related to the entropy. A simple calculation shows that in the equilibrium state, E is proportional to the negative entropy.

As a curiosity, we should mention that, according to tradition and not completely verified, the name "H theorem" is based on a misunderstanding: The letter E had long been reserved to denote energy, and so the uppercase letter eta (H) was taken from the Greek alphabet.

Below, we sketch Boltzmann's ideas published in 1877 in a somewhat modernized form.

We consider a system consisting of N molecules, where the structure of the molecules can be arbitrarily complex. We therefore need more information in order to be able to characterize the state of the molecules.

The position of a molecule with f degrees of freedom can be uniquely specified by the f coordinates $q_1, q_2, \ldots, q_f$. To describe the state of such a molecule, the spatial coordinates x, y, z generally do not suffice; rather, one must introduce an f-dimensional space with coordinates $q_1, q_2, \ldots, q_f$ in which the molecule's position is represented by this point.

For a complete specification of the state of a molecule, we need in addition to the information about its location, information about its motion. The most useful approach is to use "canonically conjugate" momenta $p_1, p_2, \ldots, p_f$, associated with the spatial coordinates. The canonically conjugate quantities q and p also play a role, as we have seen (Section 4.2.7) in the Hamiltonian equations of motion in mechanics.

The position and motion of a molecule, or, more generally, its state of motion, are described by the set of $2f$ generalized coordinates $q_1, q_2, \ldots, q_f; p_1, p_2, \ldots, p_f$. We now use these as ordinary Cartesian coordinates in a $2f$-dimensional space. To this end, we place q_1 on the first axis, q_2 on the second, and finally p_f on the $2f$th. It should not cause any difficulty that we are only able to visualize up through coordinate q_3. We call the space thus obtained the *phase space* of the molecule. We can now say that the state of a molecule corresponds to a particular point in phase space because each point specifies both the spatial and momentum coordinates. We now divide the phase space into volume elements with volume $\mathrm{d}q_1 \cdots \mathrm{d}q_f \cdots \mathrm{d}p_f$ and number the resulting "cells" consecutively. Using this enumeration, we can identify individual locations in the phase space. We now know the state of a gas if at a given point in time we know, for each marked "individual" distinguishable molecule, the cell in phase space in which it is located. We call this state, which is determined by specifying all coordinates of the molecules that we consider to be

Quotation 4.49

O vast convocation, magnificent suns,
Gather and loosen your masses of gold;
Lead tenderly, sad, to solemn accords
The majestic funereal march of your sister, asleep.

The sands have run out! The Earth, forever dead,
After a final gasp (where a sob trembled!)
Among the nocturnal silence of echoless calm,
Float, an immense and solitary wreck.
That dream! swept away by the night, is it true?
You are only a shroud, inert and tragic mass;
And yet remember! Oh, epic saga, unique! ...
No, sleep; it's over now. Eternally, sleep.

...

And yet, remember, Earth, that primal age
When you could summon through the monotonous days
Only the wind's pantoums, the muffled clamor of waves,
And silvery whispers among the leaves.
But the rebel, frail and impure, appears!
He plunders sacred Maya's beautiful veils,
And the sob of Time springs upward toward the stars ...
But sleep; it's over now. Eternally, sleep.

...

Oh! can you forget the medieval night
When Terror would intone the "*Dies irae*"
While famine ground the old exhuméd bones
For Plague whose fury gluts the charnel house!
Remember that hour when man in his ultimate fear,
Under the barren sky, still clinging to Grace,
Cried "Glory to God the Just," and cursed his race!
But sleep; it's over now. Eternally, sleep.

...

And the stake! the irons! tortures! jails!
Bedlam and towers, brothels!
Alchemy! and music! and the arts!
Science! and the war to fertilize the land!
And luxuries! spleen, love, and charity!
Hunger and thirst, drink, ten thousand ills!
Oh, what a drama you lived, fast-cooling ashes!
But sleep; it's over now. Eternally, sleep.

...

And nothing remains! O marble Venus, rain etchings!
Insane Hegelian brain! sweet consoling songs!
Belfries of filigree woven, upspringing, consumed!
Books which held man's useless victories!
All that was born of your children's wrath,
All that was once your filth and your transient splendor,
O Earth, has become like a dream, a noble dream.
Go to sleep; it's over now. Eternally, sleep.

...

Eternally sleep; it's done. Believe if you will
The whole fantastic drama an evil dream.

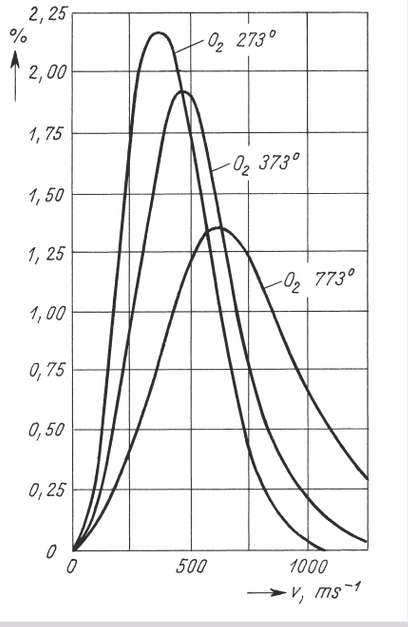
continued on next page

Quotation 4.49, ***continued***

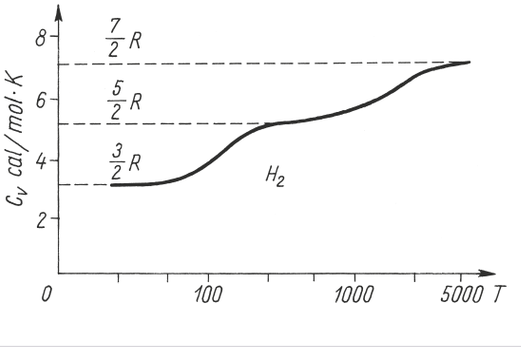
You have become a tomb which aimlessly wanders
.... Nameless within the unremembering darkness –
It *was* a dream! Oh no, you never lived!
All is alone! no witness! nothing to see or to think.
Nothing but darkness, time, and the silence...
Sleep; you've been dreaming. Eternally, sleep.

O vast convocation, magnificent suns,
Gather and loosen your masses of gold;
Lead tenderly, sadly, to solemn accords
The majestic funeral march of your sister, asleep.

—Jules Laforgue, "Funeral March for the Death of the Earth," 1881



▲ **Figure 4.137** The Maxwell velocity distribution.



▲ **Figure 4.138** Temperature dependence of the specific heat of hydrogen gas measured at constant volume. As the temperature increases, it appears that the number of degrees of freedom increases. This phenomenon will be explained later only by quantum theory.

distinguishable, the *microstate*. We now proceed from the assumption that *each microstate of a gas thus defined appears with equal probability*.

The measurable macroscopic state of a gas is given by the number of molecules in the various phase-space cells, or, in other words, by the density of the molecules in phase space.

If there are N_1, N_2, …, N_n, molecules in the N cells of phase space numbered 1, 2, …, N, then we are dealing with the macroscopic state characterized by the numbers N_1, N_2, …, N_n. The number of possible microstates associated with this macroscopic state is

$$P_{\text{td}} = \frac{N!}{N_1!N_2!\cdots N_n!}. \qquad (2)$$

This number is called the thermodynamic probability of the associated macroscopic state.

A gas left to itself progresses in a brief period of time from a given initial state to a state with maximal thermodynamic probability. For example, consider two parts of a container separated by a partition that contain gas at different pressures; for the volume of the entire container, the probability of this distribution is much less than the probability of a uniform distribution. Hence, if we remove the partition, then the pressures will equalize. A further example: If a jet of gas enters a gas-filled container at a uniform speed, then the velocities of the molecules in the jet will rapidly change: the molecules will scatter uniformly in every direction and the absolute values of their velocities will satisfy the Maxwell distribution because this state is much more probable than the initial state.

Expression (2) for the thermodynamic probability has its maximum at $N_1 = N_2 = \cdots = N_i = \cdots$, that is, at constant density, assuming that each element of phase space is assigned the same energy. However, if we associate energy E_i with the element that contains N_i particles and if the total energy E is a specified constant, then a different result is obtained. Expressed mathematically, we are now seeking a maximum for the thermodynamic probability P_{td} with the constraint that $\sum_i N_i E_i = E_0$; in addition, there is the natural (at least in classical problems) constraint $\sum_i N_i = N$ that the number of particles is constant. The solution with the greatest probability is the following "Boltzmann distribution":

$$N_i = A\mathrm{e}^{-E_i/kT}.$$

In the derivation, N and all the N_i are taken to be large; it is therefore possible to calculate with continuously variable values of P_{td}, and one can also use Stirling's asymptotic formula for $N!$:

$$N! \sim \sqrt{2\pi N}\left(\frac{N}{\mathrm{e}}\right)^N.$$

We have found in the thermodynamic probability a quantity that says something about the direction of processes. In nature, processes run in such a way that the total thermodynamic probability for all bodies involved in the process increases. This statement seems familiar to us because in thermodynamics, the same is said about the entropy: Entropy is the quantity that in macroscopic thermodynamics determines the direction of processes. We recall that according to the second law of thermodynamics, the entropy in a closed system can only increase. Therefore, there must be a close connection between the thermodynamic probability of the macroscopic state of a gas and the entropy. That is, the entropy should be a function of the thermodynamic probability.

This function can easily be determined. According to the laws of thermodynamics, the total entropy of two gases confined in containers isolated from each other

is equal to the sum of the individual entropies of the two gases: $S = S_1 + S_2$. The probability W that at a given time the entire system will be in a particular state is given by $W = W_1 \times W_2$, where W_1 and W_2 are the thermodynamic probabilities of the two gases. We know that the probability for the simultaneous occurrence of two independent processes is equal to the product of the probabilities of the individual processes, so from what we have said, it follows that $S = f(W)$ must be chosen such that $S = S_1 + S_2$ can be satisfied with $f(W_1 \times W_2) = f(W_1) + f(W_2)$. This requirement determines the function $f(W)$ as

$$S = k \ln W,$$

since the logarithm of a product is equal to the sum of the logarithms of the factors. Here k is the well-known universal Boltzmann constant. This relationship was first formulated by Boltzmann in 1877 and generalized in 1879 by Maxwell.

Having succeeded in establishing a relationship between entropy and the probability of the state of a gas, we may state that the second law of thermodynamics, with its statement about the growth in entropy, is no longer to be seen as absolute and valid under all circumstances. Namely, if every microstate of a gas is realized with equal probability, then over a sufficiently long period of time, the gas will take on every possible state. We know, however, that the maximal entropy is associated with the majority of these states; nevertheless, other states with a smaller entropy must occur, even if only very seldom.

One formulation of the second law—the one closest to our common sense—is that heat does not of its own accord pass from a colder body to a warmer body. By the above, this is not strictly true. It is not out of the question that when two bodies of different temperatures touch, the result will be that the colder one becomes colder and the warmer one becomes warmer. Within the kinetic theory of heat, this means that spheres moving at an average low velocity collide with the faster ones in such a way that the velocities of the faster spheres increase, while those of the slower ones correspondingly decrease. According to the laws of mechanics, this is possible: possible yes, but by the above, highly unlikely. To understand the order of magnitude involved, let us calculate the probability that the very small heat quantity of 1 erg flows by itself from one body at temperature 14° C to a body at 15° C: As the measure of change in entropy, $\Delta S/k$ is given by

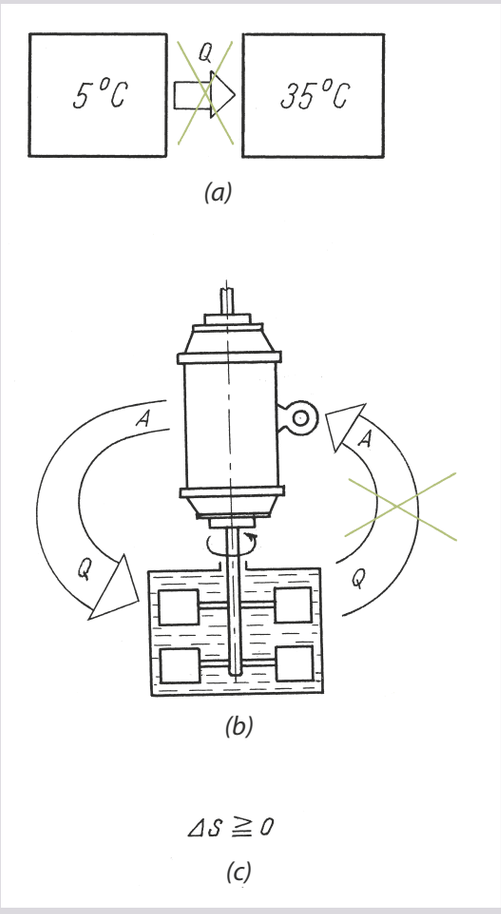
$$\frac{1}{k}\Delta S = \frac{1}{k}\left(\frac{\mathrm{d}Q}{T_1} - \frac{\mathrm{d}Q}{T_2}\right) = \left(\frac{1}{288} - \frac{1}{289}\right)\frac{1}{1.38\times 10^{-16}} \approx 10^{11},$$

from which we obtain the value

$$\frac{W_1}{W_2} = \mathrm{e}^{-\Delta S/k} = \mathrm{e}^{-10^{11}} = \frac{1}{\mathrm{e}^{10^{11}}}.$$

Now, $\mathrm{e}^{10^{11}}$ is an unimaginably large number, and we must execute correspondingly many experiments if we expect to observe the result indicated above. We may therefore be certain that we will never observe energy flowing from a colder body to a warmer body.

We can now state the second law of thermodynamics more precisely: If a gas is in a state whose entropy is significantly smaller than the maximum possible value, then the probability is very great that at a later time, a larger value for its entropy will be measured.



▲ **Figure 4.139** Various formulations of the second law of thermodynamics.

(a) Heat cannot *of itself* pass from one body to a hotter body without additional processes participating in the system. The possibility of other processes must be emphasized, because in a refrigeration system, a body can have heat extracted from it and passed to the warmer environment; however, there complex processes are involved.

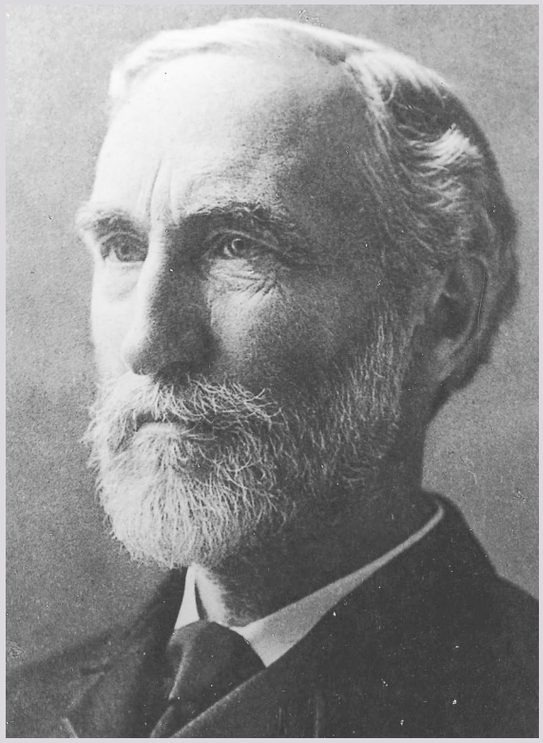
(b) A quantity of heat cannot be transformed into mechanical work with 100% efficiency. The work done by the motor here is transformed completely into heat, but this heat cannot be completely turned back into work. Hence, the arrangement shown will not work, even though it does not contravene the law of conservation of energy.

(c) The entropy of a closed system cannot decrease. Closed systems move on their own only into states of greater entropy, that is, states that are more probable or more disordered. Because entropy is a state quantity, this formulation is the best for deciding which end states in complex processes can be reached from given initial states. Thus, for example, recently the question was discussed whether the efficiency of fuel cells can be greater than 100%. Fuel cells are galvanic cells in which the chemical energy of fuels is transformed directly into electric energy using chemical reactions that may take place even at room temperature. At first glance we may think that an efficiency of more than 100% would be in violation of the first law of thermodynamics, the conservation of energy, because we would get more energy out of the system than we put in. However, we must consider that the

continued on next page

Figure 4.139 *continued*

environment belongs to the total system, and the apparatus could "pump" heat out of the environment. But what does the second law say about the possibility of such a process? By itself it cannot occur. However, if the entropy in the reaction products is greater than the entropy of the initial fuel, then this increase of entropy can compensate for the decrease of entropy due to the "pump" and the entropy of the entire system can increase. Consequently, the efficiency of certain fuel cells can—theoretically—be greater than 100%. In practice, however, such efficiencies are yet to be demonstrated.



▲ **Figure 4.140** Josiah Willard Gibbs (1839–1903): Studied at Yale University. First doctor of engineering in the United States (1862) with a dissertation titled *On the Form of the Teeth of Wheels in Spur Gearing*. After a three-year study of mathematics and physics at various European universities, from 1871 professor of theoretical physics at Yale. One of his most significant works is *A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces* (1873), in which the connections between volume, entropy, and energy are studied. Maxwell was so impressed by this method that he built a model and sent it to Gibbs. In 1876, Gibbs's best-known work, *On the Equilibrium of Heterogeneous Substances*, appeared, in which he introduces the thermodynamic potentials, presents the most general conditions for equilibrium, and discusses the thermodynamics of surface phenomena and of electrochemical processes. In 1902, his book on the fundamental principles of statistical mechanics was published. The *Great Soviet Encyclopedia* offers the following evaluation:

> In Gibbs's work, not a single error has thus far been discovered, and every one of his ideas is still valid in today's science.

continued on next page

Boltzmann's statistical interpretation of entropy, as presented above, made it possible to provide a convincing answer to all objections to the interpretation of thermodynamic phenomena within the framework of mechanics. These were the two most important objections:

1. Let the particles of the gas move as a mechanical system, which is to say that the gas assumes a series of states. According to the laws of mechanics, the opposite movement, where the gas traverses the same states in the opposite direction, is also possible. However, the laws of thermodynamics allow only changes in state in the direction in which entropy increases (Loschmidt, 1876).
2. Every closed mechanical system in its movement will repeat any of its states arbitrarily closely after sufficient time, so that every intermediate increase in entropy must be compensated for by a loss in entropy at some other time (Zermelo's objection, 1896, based on Poincaré's inversion theorem).

Yet another objection should be mentioned. We again start from the point of view that a gas can be seen as a mechanical system whose course of motion is determined by the laws of mechanics, and we also assume that every microstate of a gas is realized with equal probability. However, given the laws of mechanics, the specification of the initial state determines all future states of the gas over the course of time, so that the gas passes through *deterministic*, not *probabilistic*, states. But if we assume that the microstate is measured at stochastically chosen points in time, then it is reasonable to ask about the frequency with which we find that the gas obeys the deterministic laws of motion, in any given state. The question is whether we can then also show, using the equations of mechanics, that under such conditions every microstate of a gas will be found with the same frequency. Indeed, using Hamilton's equations of motion, John von Neumann was able to show that a system with a great many degrees of freedom runs through all the energetically allowed microstates (ergodic hypothesis) or at least approaches such states arbitrarily closely (quasiergodic hypothesis). With this it has been proven that, if we make our observations at random times, we will encounter a gas in every microstate with equal frequency, meaning that every microstate occurs with equal probability.

Yet before the fundamental problem of classical statistical mechanics could finally be explained in this way, it had lost its significance. It turned out that the phenomena of microphysics do not satisfy the equations of classical mechanics but must be described in terms of the probabilistic laws of quantum mechanics, which make possible a much simpler justification for the assertion that all microstates occur with the same frequency. The most important result for our further consideration is therefore not the complete reduction of the laws of thermodynamics to the laws of mechanics, but instead the derivation of strictly valid, apparently causal laws of macroscopic physics from probabilistic statements about elementary events, such as the occurrence of microstates.

The statistical treatment of thermodynamic processes in the framework of classical physics reached its zenith with the work of Gibbs (Figure 4.140) in the first years of the twentieth century. Gibbs envisioned a macroscopic system under investigation as a single object with a large number of degrees of freedom. Thus, a gas consisting of n pointlike particles has $N = 3n$ degrees of freedom.

The state of the entire system can therefore be characterized at any moment in time by specifying N generalized spatial coordinates $q_1, q_2, \ldots, q_N$ and the N generalized momentum coordinates $p_1, p_2, \ldots, p_N$. In the $2N$-dimensional phase space of the system, a single point therefore corresponds to the associated state. The movement of this point is prescribed by the Hamiltonian equations. The result is an extensive domain of validity for Gibbs's method applicable to every system whose laws can be given in Hamiltonian canonical form. This method can, for example, be extended even to electromagnetic fields.

Let us now picture our system as existing in a large number of instances with identical physical structure except for the initial state. Suppose that the corresponding point in $2N$-dimensional phase space is specified for each instance. We thus have before us a Gibbs ensemble that can be characterized by a distribution function $f(q_i, p_i, t)$ of these points. It then turns out that the various distribution functions correspond to various physical situations. The most important role is played by the Gibbs canonical ensemble (1901), whose distribution function has the form

$$f(q_i, p_i) = A e^{-\frac{H(p_i, q_i)}{kT}}.$$

This function describes the behavior of a system in thermal equilibrium with its environment; $H(p_i, q_i)$ is the well-known Hamiltonian function: the total energy expressed in terms of the conjugate spatial and momentum coordinates.

4.6 The Structure of Matter and Electricity: The Classical Atom

4.6.1 *Chemistry Hinting at the Atomic Structure of Matter*

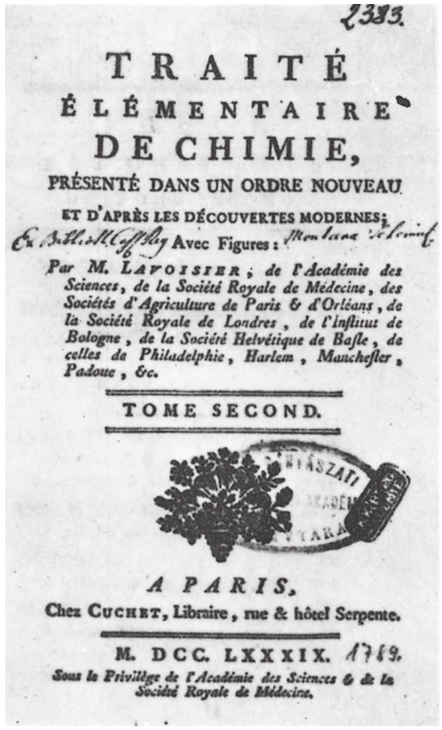
Earlier in this book, we became acquainted in some detail with the beginnings of atomic theory. We went back as far as Parmenides and his world model of a static homogeneous sphere that represented the denial of all change, but whose broken-up fragments we tried to link to the atomic ideas of Democritus. While for 2000 years, atomism was only able to explain qualitative, or philosophical, questions, we saw the first signs of a quantitative theory advanced by Daniel Bernoulli in the middle of the eighteenth century, which was subsequently expanded by Waterston and Joule, and then, above all, by Clausius, Maxwell, and Boltzmann in the completed form of the kinetic theory that is more or less valid today. To be

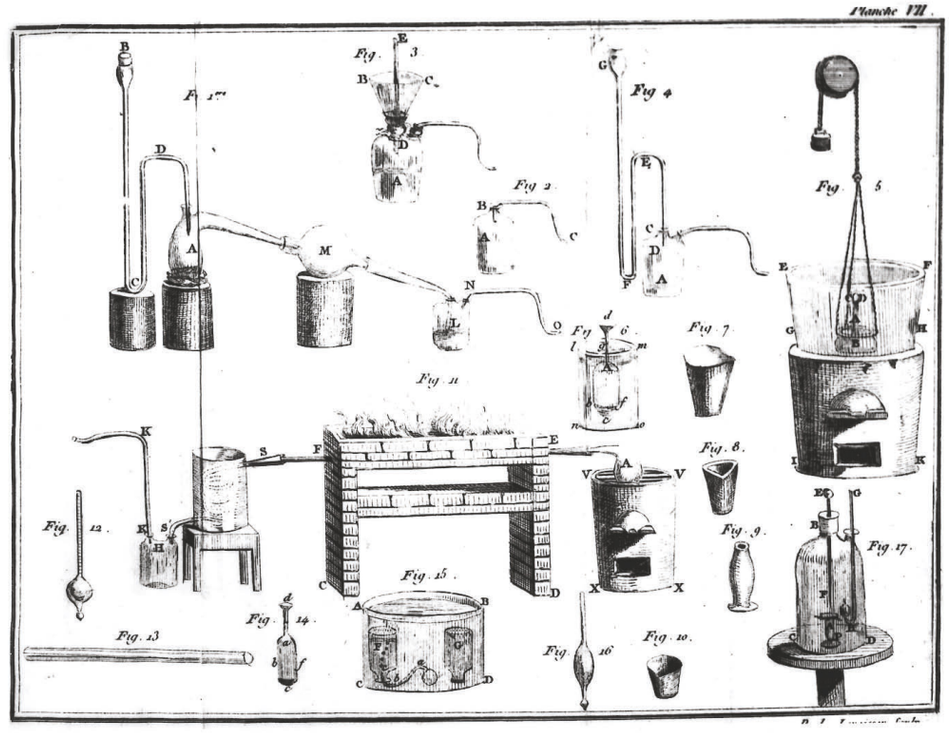
Figure 4.140 *continued*

The *Encyclopaedia Britannica* explains somewhat more concretely that Gibbs's principles have been formulated so generally and abstractly that the possibilities for their application were fully recognized only decades after the fact and that even quantum statistics can be built on this sturdy foundation. As an additional example of progress in this area, one might mention the thermodynamics of irreversible processes, or nonequilibrium thermodynamics, which developed only recently.

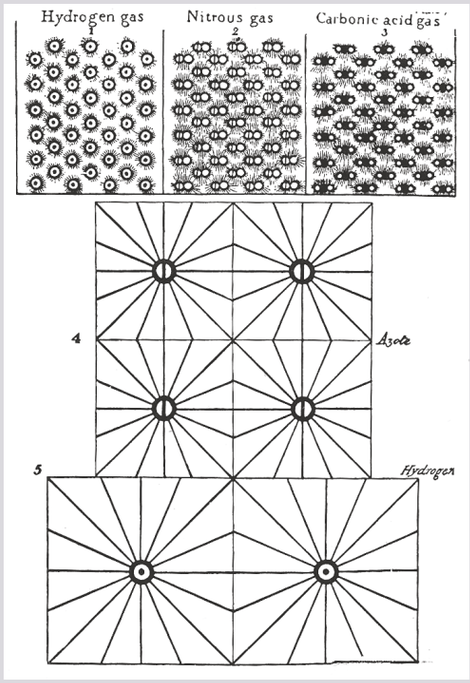
Thermodynamic potentials are, in general, functions of the state variables: volume (V), pressure (p), temperature (T), entropy (S), and internal energy (U). They make possible or simplify the calculation of characteristic quantities of a given system such as specific heat capacity or heat of reaction. Specifying the conditions for the equilibrium state is also simplified with their help:

1. The "free energy" introduced in 1882 by Helmholtz: $F = U - TS$, $dF = -SdT - pdV$. It yields, under constant temperature ($dT = 0$), the "freely transformable energy of the system." If $dT = 0$ and $dV = 0$, then $dF = 0$; that is, in constant-temperature, constant-volume systems, the free energy approaches an extreme (minimal) value.
2. The Gibbs heat function $H = U + pV$ (Heike Kamerlingh Onnes gave it the name enthalpy): For $dp = 0$, a change in enthalpy yields the heat of reaction of the chemical processes taking place.
3. The free enthalpy or Gibbs function: $G = H - TS = F + pV$ plays the same role for $dT = 0$, $dp = 0$—that is, for constant-temperature, constant-pressure processes—as free energy plays for constant-temperature, constant-volume systems; its value reaches a minimum at equilibrium.

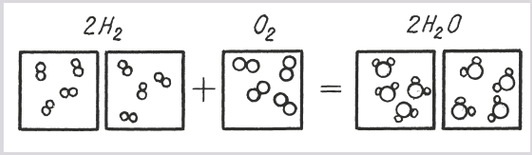




▲ ▶ **Figure 4.141** Title page of Lavoisier's epoch-making book on chemistry with a double page of figures. (Library of the University for Heavy Industry, Miskolc.)



▲ **Figure 4.142** The law of constant proportions from Dalton's book *A New System of Chemical Philosophy*, 1808.



▲ **Figure 4.143** Gay-Lussac's law on the volumetric relationships of chemically reacting gases.

sure, this theory does take into account the atomistic structure of matter, but it does not give any concrete details in regard to the atoms themselves.

The first specific contributions to the formation of the concept of atoms were made by chemists at the turn of the nineteenth century. Lavoisier succeeded in clarifying the notion of chemical element in experiments conducted around 1790 (Figure 4.141), and he determined that the elements of a chemical compound occur in fixed proportions by their weights. Joseph-Louis Proust (1754–1826) proposed the law of constant proportions, and Dalton (1766–1844) eventually arrived at the law of multiple proportions based on his experiments conducted between 1803 and 1808. Dalton at once explained the two laws on the basis of atomic theory: He stated that all atoms of an element are alike and that the smallest units of any material consist of only a small number of atoms. Dalton even published a table of atomic weights (Figure 4.142).

A further step was taken in 1808 by Gay-Lussac (1778–1850) with the formulation of the law of volumes for reacting gases (Figure 4.143). In connection with this, Avogadro (1776–1856) was able to state his famous law according to which at equal volume, temperature, and pressure, all gases contain an equal number of particles.

In the 1860s (Figure 4.144), the kinetic theory of gases was now based on more or less well-defined concrete ideas about the corpuscular structure of matter. However, it had no answer to the question of how the atoms of elements combine to form molecules, nor did it have anything to say about the internal structure of the atoms themselves. But, as we saw in the previous section, the kinetic theory of gases was able to claim enormous success by making concrete quantitative statements about the microphysical characteristics of molecules, for example, the mass and approximate size of a molecule.

Chemistry also offered the first hints that electrical phenomena must play an important role in the structure of materials. We spoke of this in Section 4.4, and here we are referring to Davy's conjecture and to Faraday's laws.

4.6.2 *The Electron: J. J. Thomson*

In Figure 4.144 we have sketched out some of the steps that led from the picture of the electrically neutral and structureless atom of kinetic gas theory (in which, in most treatments, the atoms were envisaged as smooth elastic spheres) to Rutherford's planetary model and beyond.

The primary breakthrough leading up to the explanation of atomic structure came from investigations of cathode rays that followed directly from experiments with gas discharge.

In the second half of the nineteenth century, a visitor entering a "modern" laboratory would have immediately noticed the vivid lights of the gas-discharge tubes, resembling the neon lights used in today's advertising signs.

The fact that a luminous phenomenon can be observed when electricity is passed through a rarefied gas was known as early as the mid-eighteenth century. The extraordinarily multi-talented Faraday left his mark in this area as well when, in 1830, he created an electric discharge in gas with the help of an electrification machine. His name is remembered in this area in the phenomenon called "Faraday dark space."

The investigations were accelerated by the discovery of a technological aid. In 1855, using the pump that bears his name, Heinrich Geissler (1814–1879) succeeded in obtaining a much more complete vacuum than had previously been achieved. In the following decade and a half, Julius Plücker (1801–1868) and

his student Johann Wilhelm Hittorf (1824–1914) conducted a thorough investigation of cathode rays (Figure 4.145), discovering the following properties:

- The cathode rays exit the cathode.
- They propagate in straight lines.
- They cause fluorescence when striking certain substances.
- Cathode rays can be deflected by a magnetic field.

The list of properties was further extended by Goldstein in 1871 as follows:

- Cathode rays exit perpendicular to the surface of the cathode, regardless of the shape of the cathode. Concave cathode shapes can focus the cathode rays.
- The properties of the cathode rays are independent of the material of which the cathode itself is made.
- Cathode rays can trigger chemical reactions, just like the Sun's ultraviolet rays.

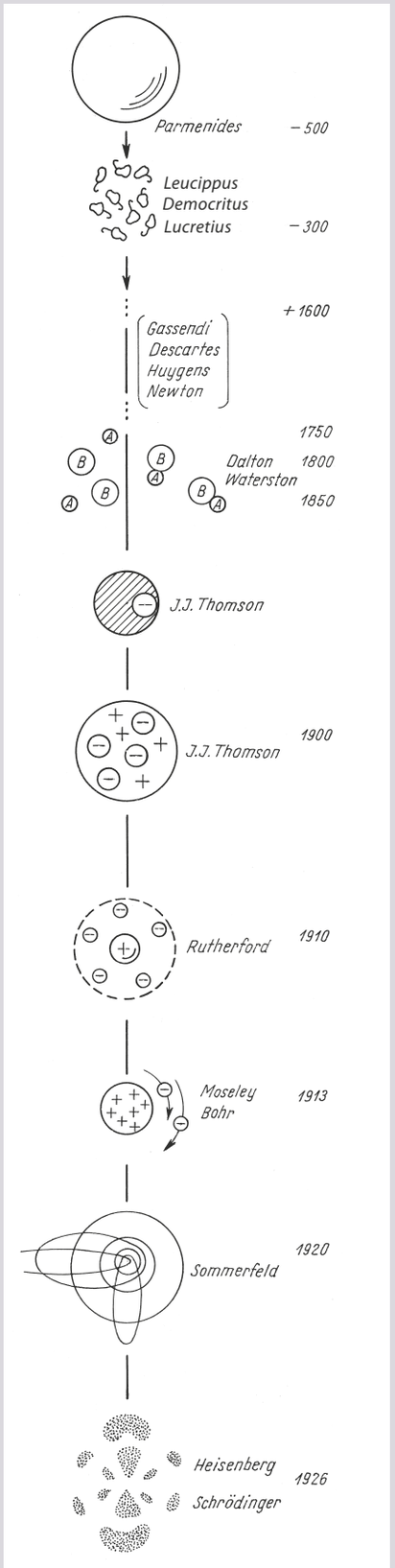
Based on the deflection of cathode rays by a magnetic field, Cromwell Varley in 1871 surmised that they might consist of negatively charged particles because negative charges in motion would be deflected in the same direction as given by experiments.

As a result of his 1879 experiments, Sir William Crookes (1832–1919), who is remembered by the name "Crookes tubes" for neon lights, found the following additional properties of cathode rays:

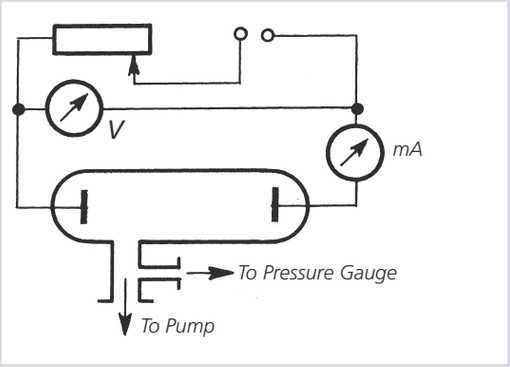
- When cathode rays strike a thin foil, the foil can become red hot; thus it is seen that cathode rays carry energy.
- Crookes also claimed that cathode rays possess momentum, since they can set a suitably placed paddle wheel in motion (later, however, J. J. Thomson determined that this rotation could also be caused by the heating effect from the impact of the rays).

Crookes was the first to attempt a theoretical explanation of the phenomena observed in cathode-ray experiments. He suggested that the cathode ray might be a stream of negatively charged molecules. These molecules would become negatively charged from a chance collision with the cathode and then be repelled from it with great velocity.

Around 1880, the German physicists Gustav Wiedemann (1826–1899) and Eugen Goldstein (1850–1930), as well as Heinrich Hertz, speculated that cathode rays might be a particular form of electromagnetic radiation. The deflection of cathode rays by a magnetic field, as observed in their experiments, was insufficient proof in their eyes of the assumption that the rays must consist of electrically charged particles. They referred in this connection to the rotation of the polarization plane of light by a magnetic field (the Faraday effect; Section 4.4)—that is, to the fact that light is affected by a magnetic field. Why, then, should it not be possible under certain conditions for a magnetic field to influence the direction of propagation of electromagnetic waves? We should not forget that Hertz conducted successful experiments with electromagnetic waves, which explains his bias. Hertz and his colleagues also performed what they called an *experimentum crucis* on whether cathode rays can be deflected by an electric field. If cathode rays consisted of electrically charged particles, then an electric field should also deflect them. True to Hertz's expectations, the experiment gave a negative result and reinforced his erroneous belief that cathode rays are just a kind of electromagnetic radiation (Quotation 4.50).



▲ **Figure 4.144** The principal stations in the development of our ideas about the structure of the atom.



▲ **Figure 4.145** Experimental setup for studying cathode rays.

Quotation 4.50

By experiments here described, I believe I have proved: ...

2. That cathode rays are only a phenomenon accompanying the discharge, and have nothing directly to do with the path of the current.
3. That the electrostatic and electromagnetic properties of the cathode rays are either *nil* or very feeble ... These cathode rays are electrically indifferent, and amongst known agents the phenomenon most nearly allied to them is light. The rotation of the plane of polarization of light is the nearest analog to the bending of cathode rays by a magnet.

—H. Hertz, *Miscellaneous Papers* [Anderson 1964, p. 37]

Today, of course, we know that the negative experimental result must have been due to poor experimental conditions. In all probability, the discharge tube had not been sufficiently evacuated, the residual gas was ionized by the electron beam, and space charges emerging in the ionized gas simply shielded it from the externally applied electric field that could have caused the deflection of the beam. Therefore, no deflection was observed. We should note for historical interest that the hypothesis of the wave nature of the electron—so important for the formulation of quantum mechanics—came under discussion immediately following the discovery of this particle. Figure 4.146 shows for the electron and, for comparison, the photon as well, how ideas about the nature and dual nature of these particles developed over time.

A further hint of the wave nature of cathode rays appeared to be the fact that they could penetrate thin foil. Using a discharge tube the far end of which (where the cathode rays impacted) was sealed by very thin metallic foil, Philipp Lenard (1862–1947), a pupil of Hertz, allowed cathode rays to exit into the open air. This foil was "vacuum-tight," that is, impermeable even to the smallest particles known at the time, namely, hydrogen molecules. If cathode rays consisted of particles, these must be significantly smaller than anything known before then.

A most serious objection to the hypothesis of the wave nature of cathode rays was that they carried a negative charge, which was demonstrated with ever greater reliability and which meant that they were a stream of some kind of negatively charged particles.

Already in 1880, Goldstein showed that the particles in this stream could not be negatively charged molecules, as envisioned by Crookes. In the experiment depicted in Figure 4.147, Goldstein alternately used one of two electrodes in a right-angled discharge tube as the cathode, so that in the upper, horizontal, arm of the discharge tube, the cathode rays could be made to propagate in either direction. As P. G. Tait had already pointed out, if Crookes' hypothesis were correct, a *Doppler shift* in the radiation spectrum of the moving molecules should be observable. However, in the experiment described above, Goldstein did not observe any such effect. In 1884, Arthur Schuster (1850–1934), working at the University of Manchester, had already come up with the correct interpretation: The cathode rays collided with molecules at rest and excited them, resulting in the emission of light.

The nearly half-century-old mystery of the nature of cathode rays was finally solved by J. J. Thomson in 1897 (Figure 4.148). Like Schuster before him, he investigated the deflection of cathode rays by magnetic fields, assuming that they consist of negatively charged ($-e$) particles of mass m, which means that the radius of the deflected circular path r and the magnetic field strength B are related as follows:

$$\frac{mv^2}{r} = evB \quad \Rightarrow \quad \frac{e}{m} = \frac{v}{Br}.$$

Because Schuster did not know the speed v of the particles, he could only estimate the particles' charge-to-mass ratio e/m. It was clearly necessary to find an additional relationship between this ratio and the speed of the particles. In his first experiments, Thomson determined the heat quantity and the electric charge that is conveyed by electrons striking an electrode. If n particles strike the electrode, then the resulting heat energy H will be the sum of the kinetic energies $\frac{1}{2}mv^2$ of the particles:

$$H = n\frac{1}{2}mv^2.$$

The total conveyed charge, in turn, is given by $Q = ne$. THOMSON measured the heat quantity with a sensitive bolometer and the charge with an electroscope. The above two equations then yield

$$\frac{e}{m} = \frac{2Qv^2}{H}.$$

Later THOMSON also developed a more precise method. The particles traveled through perpendicular electric and magnetic fields. Both fields deflected the charged particles, but in the opposite directions.

Figure 4.149(a) shows a sketch of the measuring device, and Figure 4.149(b) shows THOMSON's original apparatus.

Based on these experiments, THOMSON was able to conclude unambiguously that the nature of the particles in the cathode rays depended on neither the cathode's material nor the gas that filled the discharge tube. Therefore, these particles had to be components of the atoms of every element. We may view 1897 as the year in which the particle that plays the greatest role in the physics of everyday life, in the transformations of physical chemistry, and in light phenomena, the *electron*, was born.

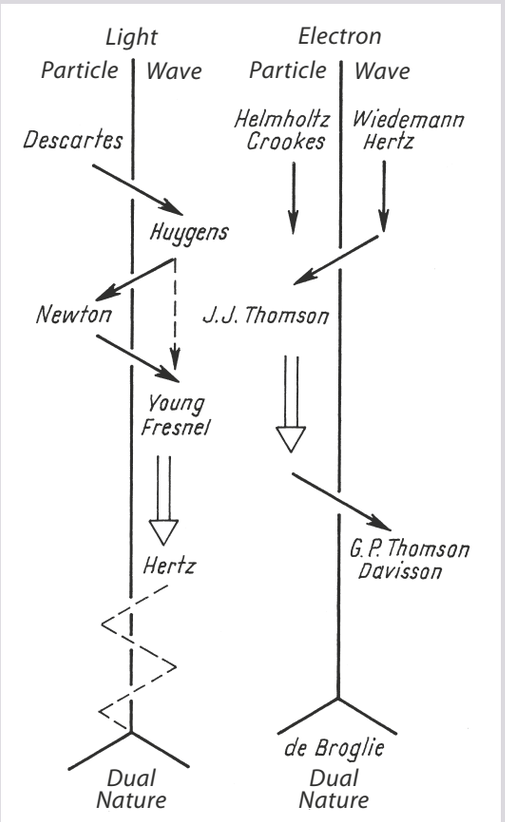
THOMSON could justifiably believe at first that these particles constituted the sought-for fundamental material of which all elements were composed. The name *electron* had already been coined in 1874 by GEORGE JOHNSTONE STONEY (1826–1911) together with HELMHOLTZ for the "atom" of electricity. They pointed out that if matter has atomic structure, then the laws of electrolysis imply with great probability that electricity also has atomic structure. The order of magnitude of this elementary charge could be easily determined from the FARADAY constant F as well as from the already well-accepted Avogadro number L:

$$F = Le, \quad e = F / L.$$

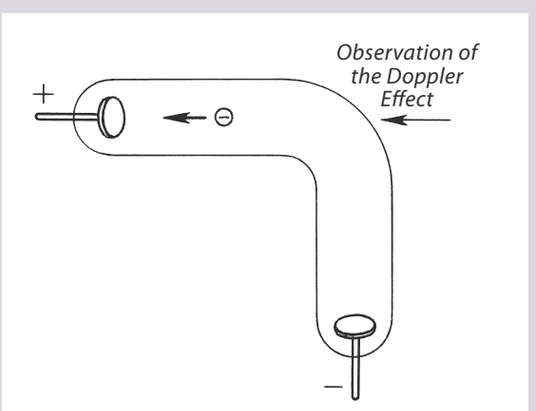
Combined with THOMSON's measurements for e/m, this number finally made it possible to estimate the mass of the electron. It was determined that this mass was three orders of magnitude smaller than that of the smallest atom, namely the hydrogen atom (Quotation 4.51).

To highlight the extent to which the cathode-ray experiments, which we can now rightfully refer to as electron-beam experiments, served as a stimulus to all of physics, we need only to mention here some of the discoveries from the four golden years of physics—1895, 1896, 1897, and 1898—that were direct or indirect consequences of such research. In 1895, ROENTGEN discovered x-rays (sometimes referred to as Roentgen rays); in 1896, BECQUEREL discovered radioactivity; 1897, as we have discussed, was the year in which the electron was discovered; and in 1898, the CURIES discovered radium and polonium.

ROENTGEN's experiments (Figure 4.150) were originally focused on cathode rays. He wrapped his brightly glowing discharge tube in opaque black paper, so he could observe the faint fluorescent light caused by the cathode rays. He noticed that fluorescent salts nearby were fluorescing whenever the discharge tube was turned on, despite the fact that neither visible light nor the cathode rays could have reached them. It is worth taking a closer look here at the apparent accidental nature of this discovery. After all, ROENTGEN had other goals in mind, and in addition the fluorescent salts had been placed there for a different experiment. Still, his comportment was remarkable in two respects. First, he noticed and judged worthy of further exploration a phenomenon that was peripheral to the main object of the investigation.



▲ **Figure 4.146** Particle–wave dualism. The changes in physicists' conception of the nature of light and the electron.



▲ **Figure 4.147** GOLDSTEIN's experiment. The intent of this experiment was to prove that cathode rays cannot be composed of negatively charged molecules.



◀ **Figure 4.148** Joseph John Thomson (1856–1940): Wished to become an engineer, but then studied mathematics and physics. 1884–1918: Cavendish professor (his predecessors in this chair were J. C. Maxwell and Lord Rayleigh; succeeding him were E. Rutherford and W. L. Bragg). 1906: Nobel Prize. His first significant contribution to physics: 1881—investigation into the electrodynamics of moving charges, in which he showed that charged particles show increased resistance to acceleration and therefore behave as if their masses had increased. This was the first hint of relativistic increase in mass. 1897: Discovery of the electron; 1907 development of what is now called the Thomson parabola method, with which for the first time the existence of isotopes could be verified. His son, Paget Thomson (1892–1975), was awarded a Nobel Prize in 1937 for his demonstration of the wave nature of light.

J. J. Thomson was the great teacher of the generation of scientists in the first quarter of the twentieth century. Among his students were E. Rutherford, P. Langevin, C. T. R. Wilson, W. L. Bragg, F. W. Aston, and E. Appleton.

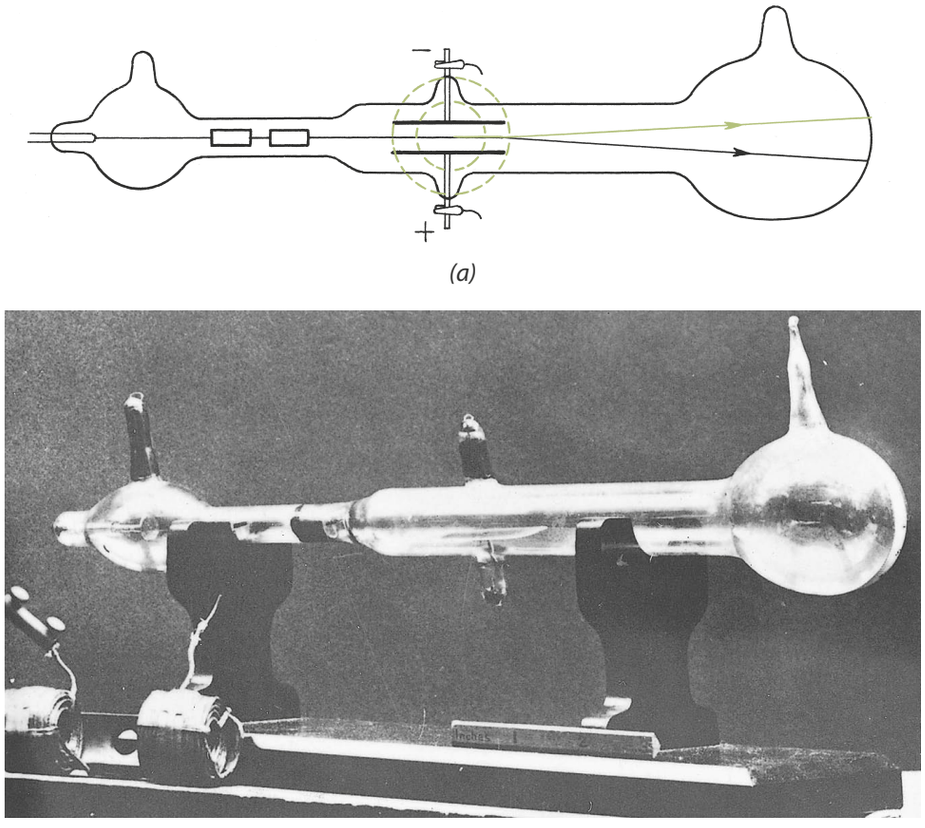
Quotation 4.51

From these determinations we see that the value of m/e is independent of the nature of the gas, and that its value 10^{-7} [using the units grams/emu here] is very small compared with the value 10^{-4}, which is the smallest value of this quantity previously known and which is the value for the hydrogen ion in electrolysis. ...

The smallness of m/e may be due to the smallness of m or the largeness of e, or to a combination of these two. ...

Thus on this view we have in the cathode rays matter in a new state, a state in which the subdivision of matter is carried very much further than in the ordinary gaseous state: a state in which matter—that is, matter derived from different sources such as hydrogen, oxygen, etc.—is of one and the same kind; this matter being the substance from which all the chemical elements are built up.

—J. J. Thomson, 1897 [Anderson 1964, pp. 46–47]



▲ **Figure 4.149** **(a)** Sketch and **(b)** apparatus for Thomson's experiment for measuring the specific charge of the electron [Harvard Project 1970].

Second, he very soon succeeded in discovering, through a comprehensive series of experiments, all the properties of the unknown new radiation required for immediate practical application (Quotation 4.52). It is remarkable how quickly in this case application followed discovery. Roentgen's first observation was recorded on November 8, 1895, and by December he had published his findings; on January 20 of the following year, in England, a broken arm was set with the help of an x-ray image (Quotation 4.53). In addition to their importance in medicine, x-rays are an indispensable aid in research into the structure of materials.

In the context of the fact that Roentgen did not simply dismiss his "accidental" observations of a peripheral phenomenon, we note that J. J. Thomson, among others, observed phenomena that later could be traced unambiguously to the effects of x-rays (Quotation 4.54). Thus, for example, several researchers noticed that photographic plates left in drawers in the vicinity of discharge tubes became fogged after a while even if they had been carefully wrapped in black paper. "Store the plates elsewhere," one researcher ordered his assistant.

In 1896, Becquerel was investigating the fluorescence of uranium salts in order to gain insight into the nature of x-rays; at that time, it was generally accepted that there must be a close connection between x-rays and fluorescence. We talk about this in detail later when we discuss the structure of the atomic nucleus.

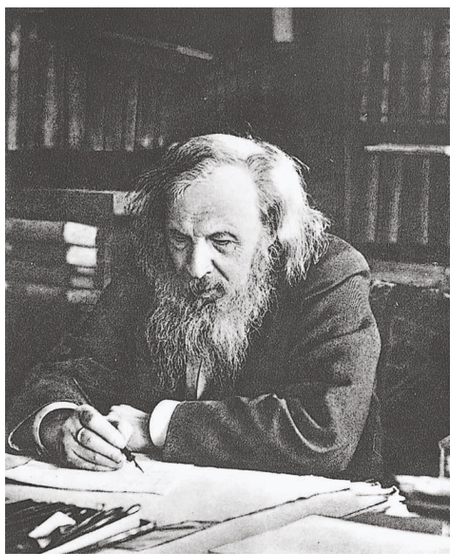
A further significant result of the cathode-ray—or, more precisely, gas-discharge—experiments was the discovery of positive anode rays. In 1886, Gold-

STEIN investigated particles passing through perforations (or canals, hence the original name "canal rays") in the cathode. The following year, WILHELM WIEN determined that these rays consist of positively charged particles of relatively large mass. This laid the foundation for mass spectrometry, a procedure for measuring atomic masses by making use of the deflection of particle streams in electric and magnetic fields. But this topic also belongs to the discussion of the nucleus.

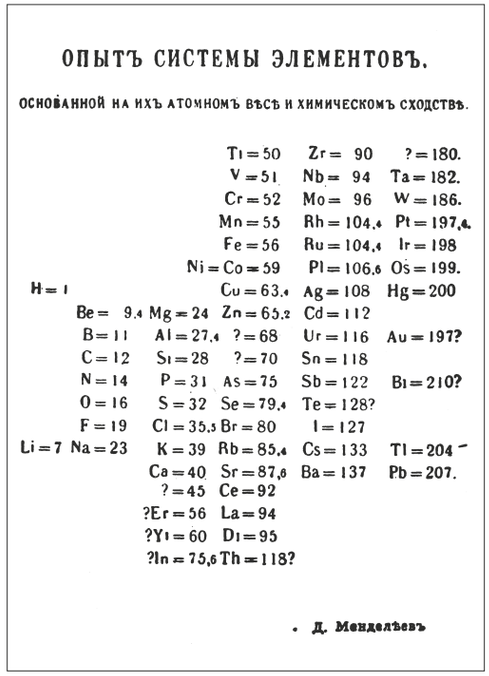
4.6.3 *Chemistry to the Rescue Again: The Periodic Table*

After the appearance of the electron, which was recognized and acknowledged as a constituent of all elements, the theoretical speculations regarding the structure of atoms could now begin on a more concrete experimental basis. It could be, of course, also possible in principle that each chemical element is qualitatively different from every other element, and that this difference is needed for creating the variety in our world. After all, the hypothesis of PROUT that every element was a composite of hydrogen atoms—based on the fact that, if we take the atomic weight of hydrogen as the unit, the atomic weights of the other elements are very nearly all integer numbers—ultimately foundered as the field of chemistry developed and ever more precise determinations of atomic weights' significant differences from such integer values were found. Today we know that PROUT's idea must be adjusted, but it was basically sound or at least it contained a germ of truth. The actual cause of the deviations from whole numbers is that chemically pure elements are generally mixtures of isotopes, but this became clear only later.

In the first half of the nineteenth century, it already had become clear that there are certain relationships among the chemical elements, pointing to a structure and requiring an explanation. Groups of elements such as the alkali metals and the halogens, whose members shared similar properties, were recognized. The deeper



▲ **Figure 4.151** DMITRI IVANOVICH MENDELEEV (1834–1907): Studies in chemistry in St. Petersburg; from 1857, instructor there. Worked in Heidelberg from 1859 to 1861; from 1864, professor of chemistry in St. Petersburg. In addition to his discovery of the periodic table (1869), which made him world famous, significant accomplishments include work in the area of physical chemistry and establishing the existence of a critical temperature.



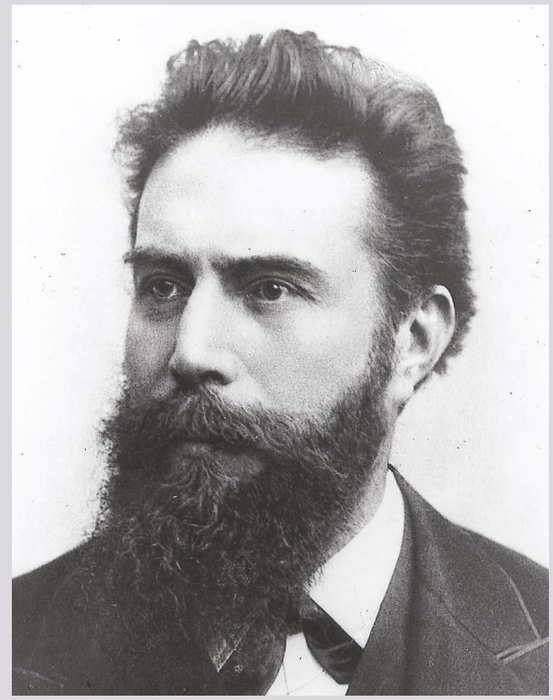
ОПЫТЪ СИСТЕМЫ ЭЛЕМЕНТОВЪ.

ОСНОВАННОЙ НА ИХЪ АТОМНОМЪ ВѢСѢ И ХИМИЧЕСКОМЪ СХОДСТВѢ.

				Ti = 50	Zr = 90	? = 180.
				V = 51	Nb = 94	Ta = 182.
				Cr = 52	Mo = 96	W = 186.
				Mn = 55	Rh = 104,4	Pt = 197,4.
				Fe = 56	Ru = 104,4	Ir = 198
				Ni = Co = 59	Pl = 106,6	Os = 199.
H = 1				Cu = 63,4	Ag = 108	Hg = 200
	Be = 9,4	Mg = 24	Zn = 65,2	Cd = 112		
	B = 11	Al = 27,4	? = 68	Ur = 116	Au = 197?	
	C = 12	Si = 28	? = 70	Sn = 118		
	N = 14	P = 31	As = 75	Sb = 122	Bi = 210?	
	O = 16	S = 32	Se = 79,4	Te = 128?		
	F = 19	Cl = 35,5	Br = 80	I = 127		
Li = 7	Na = 23	K = 39	Rb = 85,4	Cs = 133	Tl = 204	
		Ca = 40	Sr = 87,6	Ba = 137	Pb = 207.	
		? = 45	Ce = 92			
		?Er = 56	La = 94			
		?Yt = 60	Di = 95			
		?In = 75,6	Th = 118?			

Д. Менделѣевъ

▲ **Figure 4.152** The oldest form of the periodic table



▲ **Figure 4.150** WILHELM CONRAD ROENTGEN (1845–1923): After obtaining a degree as a mechanical engineer (1868, Swiss Federal Institute of Technology), he soon turned to physics. Professor at various German universities, including Giessen, Würzburg, then from 1900 until his retirement in 1920 at the University of Munich.

ROENTGEN became famous primarily through his discovery (1895) of x-rays (also called Roentgen rays). Another important accomplishment: investigation of the magnetic effect of a moving dielectric between charged capacitor plates (Roentgen current, 1888).

Quotation 4.52

1. If the discharge of a fairly large induction coil be made to pass through a ... [cathode ray tube] ... which has been sufficiently exhausted, the tube being covered with thin, black cardboard which fits it with tolerable closeness, and if the whole apparatus be placed in a completely darkened room, there is observed at each discharge a bright illumination of a paper screen covered with barium platinocyanide. ... This fluorescence is visible even when the paper screen is at a distance of two meters from the apparatus. It is easy to prove that the cause of the fluorescence proceeds from the discharge apparatus, and not from any other point in the conducting circuit. ...
2. [In this second section Röntgen described the wide variety of materials he found to be transparent, in differing degrees, to the rays. Paper, pack of cards, tin foil—single sheets were transparent, although several layers would show a shadow when placed between

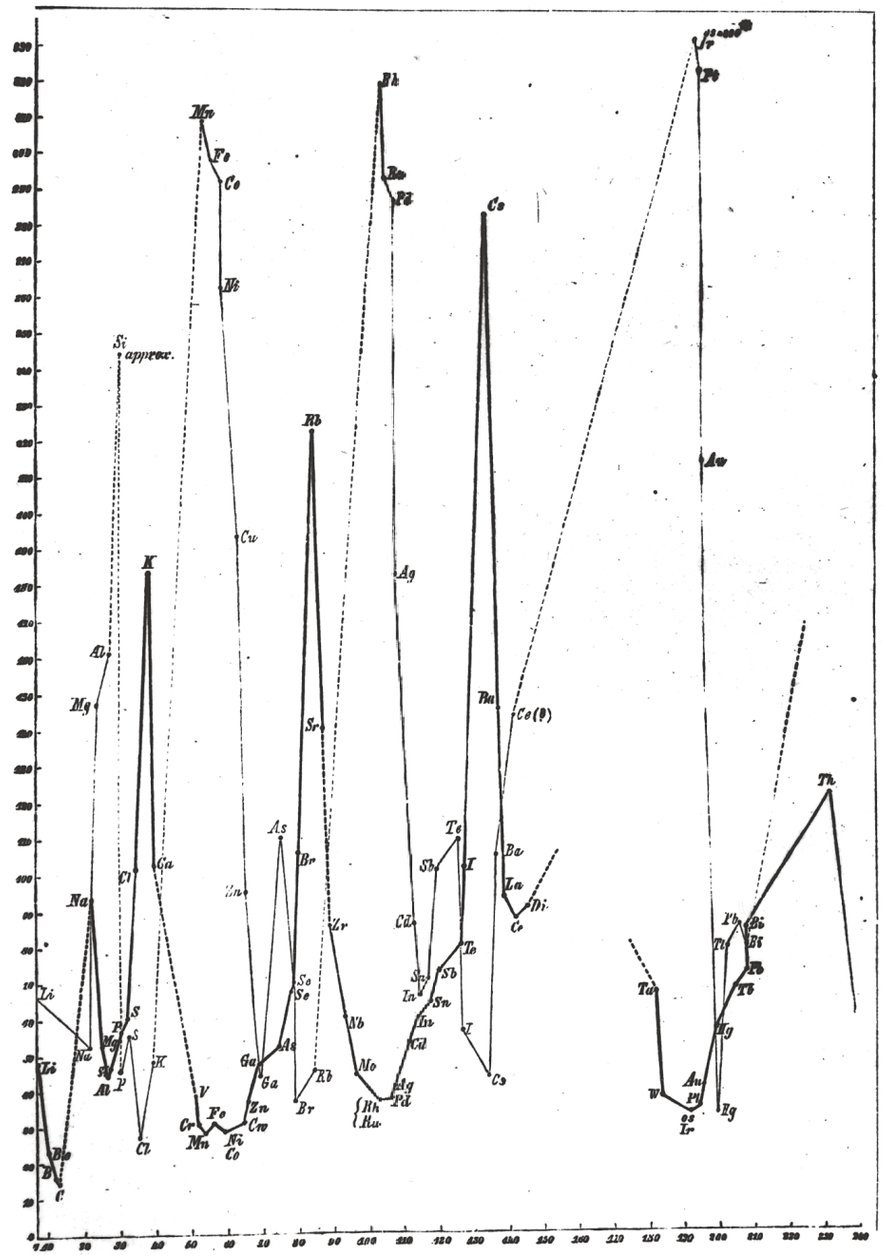
continued on next page

Quotation 4.52, ***continued***

the tube and the fluorescent screen. Thick blocks of wood, hard rubber, and glass that did not contain much lead were relatively transparent; lead glass was relatively opaque. Flesh he found to be relatively transparent compared with bones. Water, carbon disulphide, and various other liquids were transparent, but most metals were opaque in layers more than a few millimeters thick ...] ...

11. A further difference, and a most important once, between the behavior of cathode rays and of x-rays lies in the fact that I have not succeeded, in spite of many attempts, in obtaining a deflection of the x-rays by a magnet, even in very intense fields ...
12. According to experiments especially designed to test the question, it is certain that the spot on the wall of the discharge tube which fluoresces the strongest is to be considered as the main center from which the x-rays radiate in all directions ... If the cathode rays within the discharge apparatus are deflected by means of a magnet, it is observed that the x-rays proceed from another spot—namely, from that which is the new terminus of the cathode rays For this reason, therefore, the x-rays, which cannot be deflected [by magnetic fields], cannot be cathode rays simply transmitted or reflected without change by the glass wall. ... I therefore reach the conclusion that the x-rays are not identical with the cathode rays, but that they are produced by the cathode rays at the glass wall of the discharge apparatus.
17. In considering the question what are the x-rays, which, as we have seen, cannot be cathode rays—we may perhaps at first be led to think of them as ultraviolet light, owing to their active fluorescence and their chemical actions. But in so doing we find ourselves opposed by the most weighty considerations. If the x-rays are ultraviolet light, this light must have the following properties: [Röntgen then pointed out that the rays (a) cannot be refracted noticeably, (b) are not reflected, (c) are not polarizable, and (d) are absorbed by materials more or less in proportion to the density of the material.] That is to say, we must assume that *these* ultraviolet rays behave entirely differently from the infrared, visible, and ultraviolet rays which have been known up to this time.

—Wilhelm Conrad Röntgen, "On a New Kind of Rays, a Preliminary Communication" Part 1 [Anderson 1964, p. 56–58]



▲ **Figure 4.153** Atomic volumes as they depend on atomic weight. These curves from Lothar Meyer also suggest the periodicity of the structure of the elements. (From Ostwald's *Grundriß der allgemeinen Chemie*.)

cause for the existence of such groups was suggested in 1869 by Dmitri Mendeleev (Figure 4.151) and Lothar Meyer (1830–1895), independently of each other: If the chemical elements are ordered by increasing atomic weight, then their chemical properties reveal a characteristic periodicity (Figures 4.152, 4.153). This ordered system of elements is justifiably called Mendeleev's periodic table because Mendeleev understood the full implication of this discovery, even though Lothar Meyer established quantitative properties relating to the periodicity of the physical characteristics of atoms, such as atomic volume. In Figure 4.152, we see a facsimile of the list from 1869 that Mendeleev distributed to his physicists and chemists colleagues, and Quotation 4.55 reproduces selections from an article by Mendeleev published in German. It is clear that Mendeleev was aware of the

heuristic value of his system indicating directions for further research. In particular, he pointed out the "holes" in his table that needed to be filled. From the position of the unknown element in the periodic table, it was possible to predict nearly all of its physical and chemical properties, so that one could give precise indications as to where such a missing element might be sought and with what methods.

In relation to the technology of the twentieth century, it is an interesting fact that MENDELEEV predicted with great accuracy almost all the properties of the element germanium, unknown in his time, which a century later was used in the first transistor. MENDELEEV's contributions did not go unappreciated in Western European scientific circles. Together with LOTHAR MEYER, he was awarded the 1882 Davy medal of the Royal Society.

The periodic table gave chemists a roadmap with checkpoints; in other words, it soon became a practical tool. For the physicist, it served as a challenge to discover the underlying physical structures that would account for the recognized regularities.

4.6.4 *First Ideas about the Structure of the Atom*

The cathode-ray experiments made very plausible indeed the assumption that electrons play a key role in the structure of the atom. The idea suggested itself that the outwardly electrically neutral atoms must consist of equal numbers of electrons and positively charged particles of the same mass. Considering the small mass of the electron, it follows that every atom would have to consist of several thousand such particles. Additionally, the question of what would hold these particles together would still have to be answered.

THOMSON was the first to propose arguments that the number of electrons was not so large, about of the order of magnitude of the atomic weights. Correspondingly, he created an atomic model in 1904 that, until the introduction of the Rutherford atomic model in 1911 and the Bohr atom in 1913, was the most useful model, one that could explain the greatest number of experimental observations. According to THOMSON, the atom was composed of a homogeneous positively charged soup filling the entire atomic volume with very small, point-like, electrons embedded within it, not unlike plums mixed into pudding, hence the nickname "plum pudding model." There would have to be a force holding together the positively charged fluid similar to the cohesive force of ordinary fluids. The point-like electrons could be either at rest or in motion along definite circular paths (Figure 4.154). THOMSON thoroughly investigated the possibility of a stable configuration for both resting and moving electrons. He came up with a number of interesting ideas that were taken up in later models, although with differing justifications and in somewhat different forms. This had to do primarily with the arrangement of electrons in rings, which today we would call shells. THOMSON proved theoretically that if we embed a ring consisting of a number of electrons greater than a certain number in a positively charged fluid, it will remain stable only if there are electrons present in the interior of the ring as well. If the number of electrons exceeds another specific value, then a new ring must be constructed inside the original ring. The resulting periodicity in the arrangement of electrons can be seen as the first step toward an explanation of the periodic table. However, here the similarities are established by the electrons in the innermost ring, whereas today we know that the chemical properties of an element are determined by the electrons in the outermost shell.

THOMSON explained the emission of light from atoms in the following way: When a perturbation reaches an atomic system, the system will start to vibrate as a

Quotation 4.53

She is so tall, so slender; and her bones
Those frail phosphates, those carbonates of lime
Are well produced by cathode rays sublime;
By oscillations, amperes and by ohms,
Her dorsal vertebrae are not concealed
By epidermis, but are well revealed.

Around her ribs, those beauteous twenty-four
Her flesh a halo makes, misty in lime,
Her noseless, eyeless face looks into mine,
And I but whisper, "Sweetheart, je t'adore."
Her white and gleaming teeth at me do laugh,
Ah! lovely, cruel, sweet cathodagraph!

—LAURENCE K. RUSSEL, "Lines on an X-ray Portrait of a Lady," 1896 [Anderson 1964, p. 60]

Quotation 4.54

[T]he rays of the sun ... producing phosphorescence when incident upon such substances as German or uranium glass. It is perhaps worth while to observe, in passing, that the light produced in an ordinary discharge-tube by an intense discharge is very reach in phosphorogenic rays. I have been able to detect phosphorescence in pieces of ordinary German-glass tubing held at a distance of some feet from the discharge tube, though in this case the light had to pass through the glass walls of the vacuum tube and a considerable thickness of air before falling on the phosphorescent body.

—J. J. THOMSON, "On the Velocity of the Cathode-Rays"

Quotation 4.55

It was in March 1869 that I ventured to lay before the then youthful Russian Chemical Society the ideas upon the same subject which I had expressed in my just written *Principles of Chemistry*.

Without entering into details, I will give the conclusions I then arrived at in the very words I used: ...

1. The elements, if arranged according to their atomic weights, exhibit an evident *periodicity* of properties.
2. Elements which are similar as regards their chemical properties have atomic weights which are either of nearly the same value (e.g., platinum, iridium, osmium) or which increase regularly (e.g., potassium, rubidium, caesium).

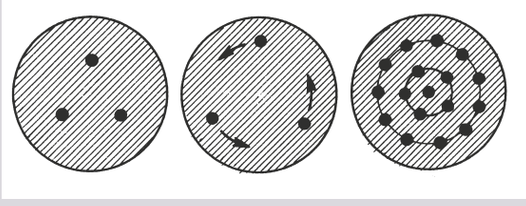
continued on next page

Quotation 4.55, ***continued***

3. The arrangement of the elements, or of groups of elements, in the order of their atomic weights, corresponds to their so-called *valencies* as well as, to some extent, to their distinctive chemical properties—as is apparent, among other series, in that of lithium, beryllium, barium, carbon, nitrogen, oxygen, and iron.
4. The elements which are the most widely diffused have *small* atomic weights.
5. The *magnitude* of the atomic weight determines the character of the element, just as the magnitude of the molecule determines the character of a compound.
6. We must expect the discovery of many yet *unknown* elements—for example, elements analogous to aluminium and silicon, whose atomic weight would be between 65 and 75.
7. The atomic weight of an element may sometimes be amended by a knowledge of those of the contiguous elements. Thus, the atomic weight of tellurium must lie between 123 and 126, and cannot be 128.
8. Certain characteristic properties of the elements can be foretold from their atomic weights. ...

The following year Professor Lothar Meyer, in his well-known paper, studied the same subject in more detail, and thus contributed to spread information about the periodic law.

—Dmitri Ivanovich Mendeleev, *The Principles of Chemistry*, part 4 [pp. 472, 486]



▲ **Figure 4.154** Thomson's atomic model.

whole. The vibrations will cause the emission of electromagnetic waves, which we perceive as light. He was, in fact, able in this way to derive vibrational frequencies for visible light that were the right order of magnitude. However, the Thomson model is completely helpless as far as explaining spectral lines, which at the time were already a well-known phenomenon with a detailed classification scheme.

We should recognize as a positive factor that in his atomic model, Thomson took into full account all the laws of classical electrodynamics, in particular that electrons moving in circular paths would emit electromagnetic radiation, so that atoms built up from such electrons could not be stable. Ten years later, this will be the most important objection to the Bohr atomic model. Thomson thoroughly investigated this problem and showed that the emissions from suitably arranged electrons can cancel each other through interference so that the effect of the radiation is so small that the atoms are essentially stable.

Thomson set even more ambitious goals. He attempted to explain the laws of radioactivity with his model of the atom. From his theory, it turns out that certain electron configurations are stable only above a critical rotational velocity. If over the course of time an atom loses some of its energy, and as a result its rotational velocity falls below the critical value, then the entire atom becomes unstable and completely rearranges itself, emitting a high-energy particle.

For his scientific work, above all on the particle nature of cathode rays, which led to the discovery of the electron as a particle, J. J. Thomson was awarded the 1906 Nobel Prize. Historians of science generally note at this juncture the interesting fact that Thomson's son, G. P. Thomson (together with C. J. Davisson), received the 1937 Nobel Prize for the experimental proof of the wave nature of the electron.

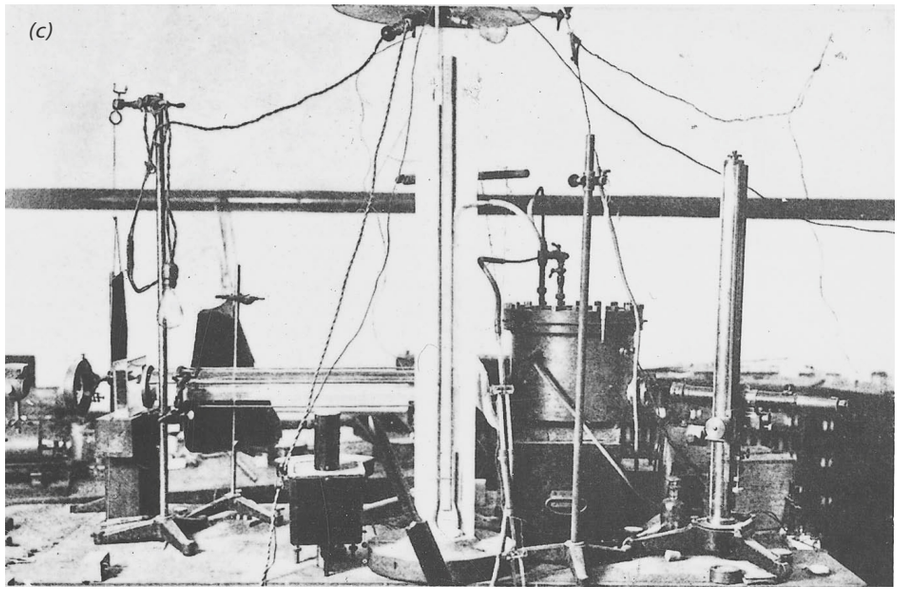
Let us now return to the electron. Thomson's measurements yielded a value for the charge-to-mass ratio e/m. In a series of experiments carried out between 1910 and 1916 (Figure 4.155), Millikan (1868–1953) determined the charge of the electron independently of any other microphysical data.

In the first years of the twentieth century, ideas about the structure of the atom were still very uncertain, but a set of key questions began to emerge (Quotation 4.56). It was clear that the electrons play a significant role in the atoms, but it remained unclear how the number of electrons and their spatial arrangement were to be reconciled with the chemical properties of the elements. Least comprehensible of all seemed to be the structure of the positive part that compensated for the negative charge of the electrons.

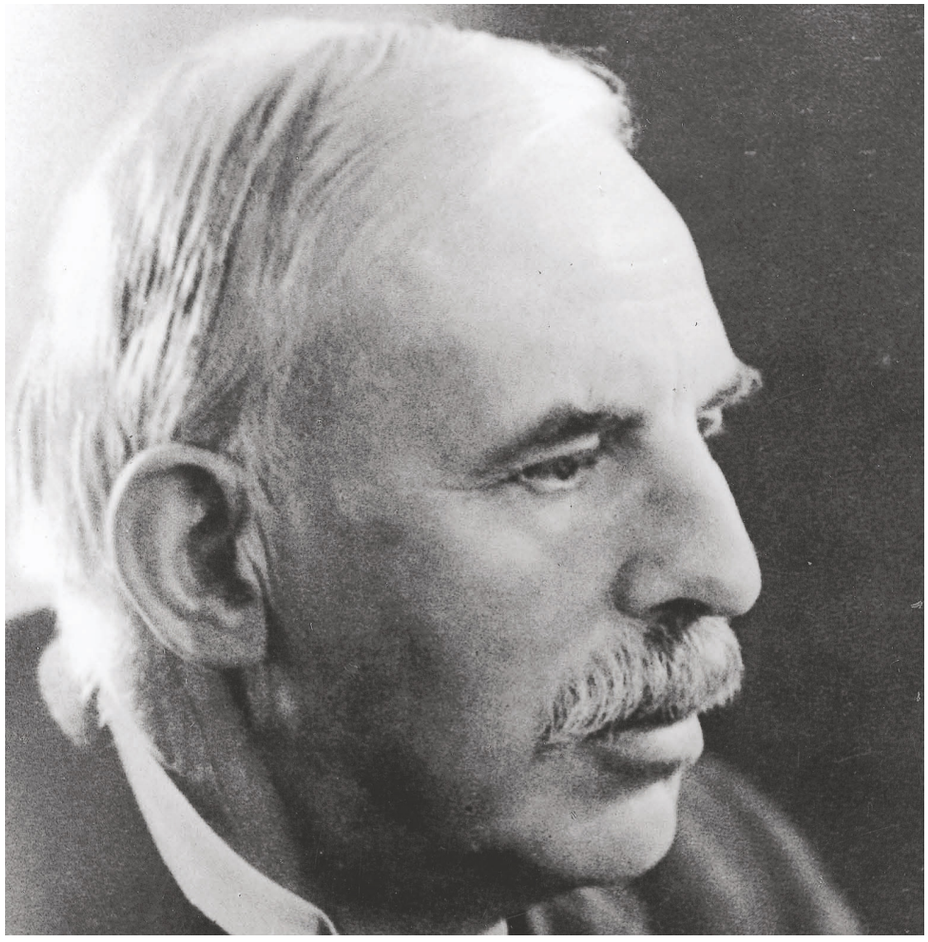
Hints began to accumulate that the number of electrons grows in proportion to the atomic weight. From this, it follows that the number of electrons is related to the location of the element in the periodic table.

Measurements taken by Rutherford in the years 1909 to 1911 (Figure 4.156) and colleagues led to a decisive turning point in our knowledge of atomic structure. Using alpha particles emitted during radioactive decay, Rutherford "plumbed the depths" of the atom. Here are two of the most important results of these renowned scattering experiments (Figure 4.157):

1. The greatest portion of the atomic mass is concentrated in a positively charged nucleus whose volume is very small in comparison to the total volume of the atom.
2. The number of electrons orbiting the nucleus identifies the element; it corresponds to the atomic number of the element and thus determines its place in the periodic table.



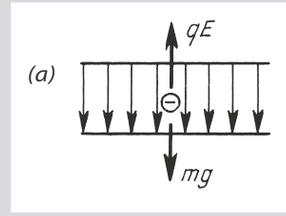
The Rutherford atomic model is like a miniature solar system. The role of gravitational force is assumed here by the attractive electrical force between the charge of the nucleus and that of the electrons, which keeps the electrons in their orbits. From this point in time, it was now possible to separate the physics of the atomic nucleus from the physics of the atom's periphery, or electron shell.

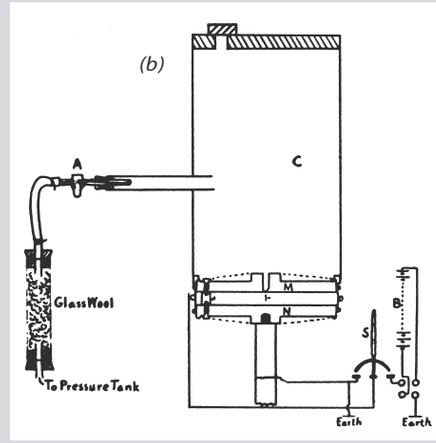


◀ ▼ **Figure 4.155** **(a)** The measuring technique used in Millikan's experiment.

(b) Schematic representation of Millikan's original experimental setup.

(c) Photograph of the experimental setup.





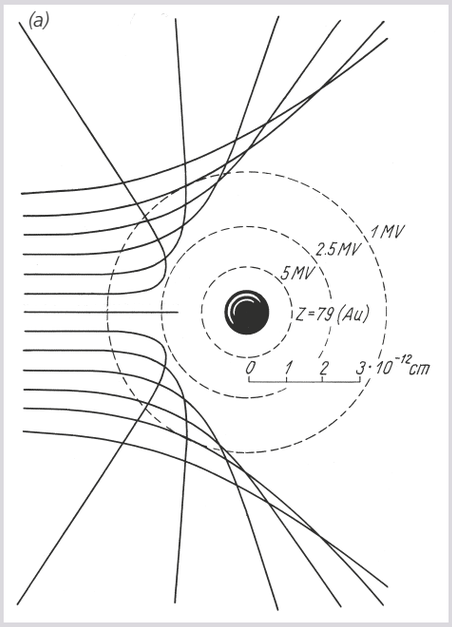
◀ **Figure 4.156** Ernest Rutherford (1871–1937): Born in New Zealand, the son of English and Scottish immigrants. Scholarship at the University of Christchurch. 1894: received doctorate degree, later colleague of J. J. Thomson in Cambridge. From 1898 to 1907: professor at McGill University in Montreal. Then worked in Manchester and Cambridge and in 1919 appointed director of the Cavendish Laboratory.

His dissertation was on the magnetic effect of high-frequency electromagnetic waves. With his magnetic detector, he received signals a half-mile away; this was a world record in 1894. 1895: properties of ionized gases; from 1896: study of radioactivity, discovery of alpha and beta radiation; 1900: exponential law; 1902: theory of radioactive decay; 1903–1908: explanation of the nature of alpha radiation; 1911: Rutherford atomic model; 1919: first artificially induced nuclear transformation; 1920: conjectured existence of neutrons.

Rutherford was the most important experimental physicist at the turn of the century, and the history of physics knows only one individual who surpassed him in this area—Faraday. Aside from his own accomplishments, he is significant as the intellectual leader of the Rutherford school, to which a large number of young scientists belonged, among others Bohr, Geiger, Chadwick, Hahn, Hevesy, Blackett, and Cockcroft.

An episode from his life: in 1909 Marsden and Geiger detected a backscatter of alpha-particles from a thin foil. This phenomenon eventually led to the discovery of the atomic nucleus. Rutherford remembers:

> It was about as credible as if you had fired a 15-inch shell at a piece of tissue paper and it came back and hit you. [Glasstone 1958, p. 93]



▲ **Figure 4.157** Rutherford's scattering experiment. **(a)** Motion of alpha particles in the field of the atomic nucleus. **(b)** Experimental setup.

Quotation 4.56

1. The main bulk of the atom may consist of ordinary matter (whatever unknown entity is hidden by that familiar phrase), associated with sufficient positive electricity (whatever that may be) to neutralize the charge belonging to the electron or electrons which undoubtedly exist in connection with each atom.
2. Or the bulk of the atom may consist of a multitude of positive and negative electrons, interleaved, as it were, and holding themselves together in a cluster by their mutual attractions, whether in a state of intricate orbital motion, or in some static geometrical configuration, kept permanent by appropriate connexions.
3. Or the bulk of the atom may be composed of an indivisible unit of positive electricity, constituting a presumably spherical mass or "jelly," in the midst of which an electrically equivalent number of point electrons are, as it were, "sown"; ...

continued on next page

The x-ray experiments of Henry Moseley (1887–1915) shored up support for the new model and added more detail to it (Quotation 4.57). From his measurements it was possible to clearly discern the patterns of the enigmatic period lengths in the periodic table of the elements. However, an interpretation of this periodicity came about only with the Bohr theory. Moseley had made contact with Bohr, and it was a great loss to science when he fell in 1915 at Gallipoli, on the Balkan front.

4.6.5 *The Line Spectrum and the Reappearance of the Integers*

From an atomic model we expect not only an explanation of the chemical properties, but also an interpretation of the line spectra emitted by atoms. From this perspective, it was not until Bohr's atomic model, based on the idea of quanta, that something novel was finally offered. One might even claim that the express purpose of Bohr's atomic model was to explain the observed regularities in the line spectra. However, this model goes beyond the limits of the classical atomic models; therefore, in what follows, we restrict our attention to the presentation of early experimental observations.

An important event in this regard was that in 1897, Pieter Zeeman (1865–1943) proved experimentally that magnetic fields influence spectral lines. The Zeeman effect, for which Lorentz provided a theoretical explanation, manifests itself in a broadening or splitting of spectral lines. With Lorentz's theory, the charge-to-mass ratio e/m of vibrating, and hence light-emitting, particles could be experimentally determined. The value of this ratio agrees closely with the value obtained by Thomson for the electron. Therefore, from this it follows that the electron must play the leading role in light emission as well.

As we have mentioned, the collective experimental findings on atomic spectra, and therefore also on atomic energy levels, supplied the actual impetus for the creation of the atomic theory of the twentieth century. It would be worthwhile, therefore, to list the most important discoveries in chronological order.

In Section 4.4, we spoke of the origins of the wave theory of light, which reached its classical apotheosis in the electromagnetic theory of light. In the execution of the experiments that we discussed, but also in parallel and intertwining experiments, knowledge about spectra was obtained. Newton's extensive investigation into the visible spectrum was extended at the beginning of the nineteenth century to the infrared and ultraviolet regions. The first measurements in the infrared were taken by William Herschel in 1801 using a sensitive thermometer; in the ultraviolet, the blackening of silver nitrate was investigated by J. Ritter (also in 1801). In 1802, William Wollaston (1766–1828) became aware of black lines in the Sun's spectrum, although he assigned no particular significance to the observation and never gave the phenomenon serious attention. Today, these lines are called the Fraunhofer lines. Joseph Fraunhofer (1787–1826) built the first modern spectroscope and used it, beginning in 1815, to conduct a thorough investigation of the black lines in the solar spectrum. He labeled the most striking of the 576 lines that he observed with the letters A, B, C, D, E, …, a, b, c, d, e, …. It is from this classification that we today call the yellow line of sodium the D line. Fraunhofer also recognized that the yellow D line of sodium emitted in an alcohol flame appears precisely at the location where a line is absent in the solar spectrum. By 1822, Herschel had already reached the conclusion that by analyzing the spectrum of various salts heated in a flame, the substances could be uniquely identified. This laid the foundation for spectral analysis. The increased use of Geissler tubes as

an experimental aid since 1856 served as a stimulus for the investigation of the spectra of a variety of substances. A significant contribution to our theoretical understanding of spectra was made by Gustav Robert Kirchhoff (Figure 4.158) and Robert Wilhelm Bunsen (1811–1899) with their joint article published in 1859. Summarizing these observations, they stated that the atoms of a gas or vapor emit characteristic line spectra when supplied with energy. The same atoms are also capable of absorbing light of the same wavelength as the emitted radiation. Thus, absorption lines appear at these frequencies in the continuous spectrum emitted by glowing solids and liquids.

This new knowledge accelerated the discovery of new elements. In 1861, cesium and rubidium were discovered, and in 1866, helium was discovered in the Sun's spectrum. This element had not yet been found on Earth, and so the element was named after *helios*, the Greek word for the Sun.

After the establishment of the periodic table, it was a matter of course for the question of the relationship between the spectra of the various elements and their place in the table to arise. The qualitative answer: the complexity of the spectrum increases with the element's atomic number.

A discovery by the Swiss high school teacher Johann Jakob Balmer (1825–1898) proved to be of fundamental importance to the systematization of spectral lines and hence to the atomic physics of the twentieth century. Balmer worked with the simplest of all spectra, that of hydrogen, and he attempted to represent the wavelengths of the lines of this spectrum in terms of a simple formula. (In this activity he would be followed in the next half-century by a host of outstanding physicists, such as Bohr, Heisenberg, Schrödinger, and Dirac, who all tested their own theories on hydrogen, leading to the maxim, popular among physicists, that "the hydrogen atom is invariant with respect to every theory.")

Balmer found that for the hydrogen spectrum, the wavelengths of the lines in the visible spectral range are given by the expression

$$\lambda = h\frac{m^2}{m^2-4}, \quad h = 3645.6 \text{ Å},$$

where m takes on the integer values $m = 3, 4, 5, 6$. The angstrom ($1 \text{ Å} = 10^{-8}$ cm) is common in spectroscopy as a unit of length. It is named for the Swedish physicist Anders Jonas Ångström (1814–1874), one of the founders of spectroscopy.

Most striking was that the values of the wavelengths calculated in this manner were not rough approximations to the experimental values, but agreed with them with excellent precision. Balmer offered no specific explanation for his formula. He was a man of broad knowledge with many interests, and he also investigated philosophical questions. He was also not immune to Pythagorean number mysticism. (While we are speaking of numerology, we should not forget to mention the fact that Niels Bohr was born in 1885, the same year in which Balmer's formula was worked out. Bohr's theory of the atom achieved decisive success with its justification of Balmer's formula; indeed, Bohr had set for himself the goal of explaining this formula.)

For our further discussion, we write Balmer's formula as

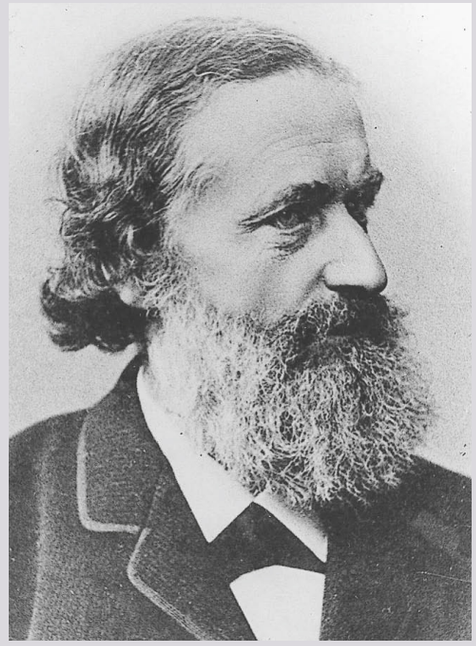
$$\frac{1}{\lambda} = A - \frac{R}{(n-\alpha)^2},$$

the form given in 1899 by Janne Robert Rydberg (1854–1919). For characterizing waves, instead of the wavelength Rydberg introduced its reciprocal, the wave

Quotation 4.56, ***continued***

4. It may consist of a kind of interlocked admixture of positive and negative electricity, indivisible and inseparable into units, and incapable of being appreciably sheared by applied forces, but incorporated together as a continuous mass; in the midst of which one or more isolated and individualised electrons may move about and carry on that display of external activity which confers upon the atom its observed properties.
5. A fifth view of the atom would regard it as a central "sun" of extremely concentrated positive electricity at the centre, with a multitude of electrons revolving in astronomical orbits, like asteroids, within its range of attraction. But this would give a law of inverse square for the force, and consequently periodic times dependent on distance, which appears not to correspond with anything satisfactorily observed.

—Oliver Lodge, *Electrons,* 1906 [Friedman and Satori 1965, p.66]



▲ **Figure 4.158** Gustav Robert Kirchhoff (1824–1887): Formulated Kirchhoff's law at age 21, and in 1850, together with Bunsen in Wrocław, laid the foundations for spectral analysis. In 1850, he showed that the quotient of emissivity and absorptivity for thermal radiation is constant. In 1860, he already recognized that the emissivity of a blackbody does not depend on the intrinsic properties of the material and is therefore of universal significance. From 1875, he worked alongside Helmholtz at the University of Berlin.

Quotation 4.57

1. Every element from aluminium to gold is characterized by an integer *N* which determines its X-ray spectrum. Every detail in the spectrum of an element can therefore be predicted from the spectra of its neighbours.
2. This integer *N*, the atomic number of the element, is identified with the number of positive units of electricity contained in the atomic nucleus.
3. The atomic numbers for all elements from Al to Au have been tabulated on the assumption that *N* for Al is 13.
4. The order of the atomic numbers is the same as that of the atomic weights, except where the latter disagrees with the order of the chemical properties.
5. Known elements correspond with all the numbers between 13 and 79 except three. There are here three possible elements still undiscovered.
6. The frequency of any line in the X-ray spectrum is approximately proportional to $A/(N-b)^2$, where *A* and *b* are constants.

—Henry G. J. Moseley, "The High-Frequency Spectra of the Elements," 1913

Quotation 4.58

If geometry were an experimental science, it would not be an exact science. It would be subjected to continual revision. Nay, it would from that day forth be proved to be erroneous, for we know that no rigorously invariable solid exists. *The geometrical axioms are therefore neither synthetic a priori intuitions nor experimental facts*. They are conventions. Our choice among all possible conventions is *guided* by experimental facts; but it remains *free*, and is only limited by the necessity of avoiding every contradiction, and thus it is that postulates may remain rigorously true even when the experimental laws which have determined their adoption are only approximate. In other words, *the axioms of geometry* (I do not speak of those of arithmetic) *are only definitions in disguise*. What, then, are we to think of the question: Is Euclidean geometry true? It has no meaning. We might as well ask if the metric system is true, and if the old weights and measures are false; if Cartesian co-ordinates are true and polar co-ordinates false. One geometry cannot be more true than another; it can only be more convenient. Now, Euclidean geometry is, and will remain, the most convenient: 1st, because it is the simplest, and it is not so only because of our mental habits or because of the kind of direct intuition that we have of Euclidean space;

continued on next page

number, and he found that the wave number can be written as the difference of quantities

$$\frac{R}{(n-c)^2}, \quad R = 109{,}677.7 \text{ cm}^{-1},$$

called "terms." It is understandable that following Rydberg's observation, attention was turned to the identification of the terms, and rightly so. And in fact, Bohr subsequently was able to identify them with atomic energy levels.

So that we will later be able to compare Balmer's formula with the formulas derived by Bohr, let us rewrite it somewhat:

$$\frac{1}{\lambda} = \frac{1}{h}\left(1 - \frac{4}{m^2}\right) = \frac{4}{h}\left(\frac{1}{2^2} - \frac{1}{m^2}\right) = \frac{1}{911.2}\left(\frac{1}{2^2} - \frac{1}{m^2}\right) \text{ Å}^{-1}.$$

The last step in the systematization of spectra prior to Bohr was the introduction of Ritz's combination principle (Walter Ritz, 1878–1909). Ritz's theory covered all the available knowledge about spectra, and it was even used by Heisenberg.

By the beginning of the second decade of the twentieth century, the classical atomic model had arrived at a form on which the quantum-physical models could directly build. The outlook for the quantum model was the most favorable in view of the fact that in 1912 Bohr was working alongside Rutherford in Cambridge.

4.6.6 *A Farewell to the Nineteenth Century*

The more closely we approach our own century after the great breakthroughs of the seventeenth century, the less frequently we encounter universal geniuses such as Descartes and Leibniz, to whom entire chapters would have to be devoted in the history of philosophy as well as in the history of mathematics and physics. Especially around the transition from the eighteenth to the nineteenth century, scientists who were great mathematicians and also great physicists carried the torch (Lagrange, Laplace, Cauchy, Fourier, Gauss). Perhaps the last of these was Poincaré, who worked almost a century later and whose contribution to the development of physics is not as highly appreciated as it should be. In the nineteenth century, "physicists" move into the foreground, those embodying in a single person a theoretician and an experimentalist (Maxwell, Boltzmann, Hertz, Kirchhoff), and at times also "engineers" (Kelvin, Siemens). But now specialization also begins, and we start to meet the great experimentalists (Faraday), so that in the twentieth century, the experimental physicists (such as Rutherford, who "used no mathematics beyond the rule of proportions") and the theoretical physicists (for example Pauli, whose very appearance in a laboratory was said to cause instruments to break) would work almost separately, in somewhat antagonistic camps. Only a few outstanding geniuses, like Fermi, could make major contributions in both the experimental and theoretical areas of physics. In contrast, the dual role of physicist-engineer was to take on added importance. Consequently, we find among the ranks of physics Nobel Prize winners names such as Richardson, Townes, Basov, Prochorov, Gabor, Bardeen, Shockley, and Brattain. (See also Plate XIX.)

In the foreground of this book, our interest is centered on the leading ideas of the great physicists, but we should not forget that these ideas were germinated and ripened by the collective efforts of many until they were seized by those geniuses and given their final form. If here and there an opportunity was overlooked by one of them, it was sure to be noticed by another; perhaps the general theory of relativity was the single great idea that might not have come to fruition—in the opinion of many—without one man, Einstein.

Not only the Aristotelian world view but also classical physics are closed systems insofar as every small detail has its place within them and every separate finding is essential. Let's take for example a trifling question in experimental technology: how many lines can a skilled experimental physicist or technician inscribe on a flake of quartz crystal of a given size? Yet this made the precise measurement of wavelengths possible. Another example: in the nineteenth century the vacuum pump was perfected, and the problem of creating extremely low temperatures was worked out, which made the study of the behavior of substances under extreme conditions possible. In fact, the increase in precision of measurements by the development of measurement techniques would deserve its own chapter in this book because it is closely connected with the theoretical questions. The following is one example.

Just as the telescope expanded human knowledge of the cosmos, one might have supposed that the microscope would have similarly furthered the penetration into the micro world, that is, into the world

of the atom. However, the microscope played no comparable role to that of the telescope; its influence was at best indirect. It is natural, then, to ask why it was not possible to fashion a microscope of such power that individual atoms would become visible. With respect to this question, the work of Ernst Karl Abbe (1840–1905) was of fundamental importance. Independently of the degree of perfection of the lenses and the removal of optical irregularities, the wave nature of light leads to a diffraction effect at the aperture of a lens. For this reason, for a pointlike object, we get a diffraction disk instead of a pointlike image, even if a perfect lens is used. Consequently, we cannot resolve arbitrarily fine structures, because of the overlapping of the diffraction disks. For the resolving power of a microscope (that is, the smallest distance between two points that the microscope can still reveal as two separate points), Abbe's theory yields the relation

$$d = 0.61\frac{\lambda}{n\sin\alpha},$$

where λ is the wavelength of the light used to illuminate the object, α is the angle between the optical axis of the microscope and the outermost light ray that reaches the microscope from the object, and n is the index of refraction of the medium in which the object is embedded. The value of the quantity $n \sin \alpha$, called the numerical aperture, is approximately 1, so we may then say, roughly speaking, that d is approximately equal to λ. The limit of the resolving power of a microscope is therefore given approximately by the wavelength of the light used, and because the wavelength of visible light is in the range from 3.8×10^{-7} m to 7.6×10^{-7} m, it is impossible to visualize atoms, whose diameters are on the order of 10^{-10} m, with an ordinary light microscope.

Abbe was a founder of the Zeiss optical company. The labor policies enacted under his leadership at Zeiss—such as contractually guaranteed eight-hour workday, profit-sharing, and right to a pension—would be praiseworthy even by today's standards.

In the development of physics, educational institutions at various levels became increasingly important. Today, we enjoy the blessings of the educational systems introduced in the nineteenth century, although we bear some of their encumbrances as well. When we think of their effect, we should think not only of the fact that universities stood at the center of scientific development and that physics teachers at the secondary level made significant contributions to science (Grassmann, Balmer), but we should remember as well that it is in the elementary and middle grades that the important foundations for the development of a scientific frame of mind and the formation of the necessary ideals and values are laid. So just as we remember not only the heroes in battle but also the unknown soldiers, so should we, acknowledge the contributions of the many unknown physicists and physics teachers.

We shall not go into detail about the mathematical tools developed in the support of physics because that would require a review of the mathematics of the nineteenth century in its entirety. But even such research as that of the Hungarian methematician János Bolyai into the independence of the axioms of geometry, which seems to be extremely abstract and deals with the foundations of mathematics, led to questions about the relationship of geometry to the real world and from there to the structure of space itself, ultimately contributing to the theory of general relativity.

The so-called Bolyai–Lobachevsky (non-Euclidean, hyperbolic) geometry arose as a solution to the *problem of parallels*. Euclid's fifth postulate, also known as the eleventh axiom (the parallel postulate) reads thus, according to Proclus: let a line and a point not on the line be given in the plane. Then there exists a unique line passing through this point that does not intersect the given line. In other words, there is a single parallel line through this point.

Even in antiquity it had been suspected that the parallel postulate might be redundant because it appeared that it should be derivable from the other axioms. However, attempts at such a derivation had too fail, as did attempts to prove that the negation of the axiom led to a contradiction; the postulate is, in fact, independent.

Around 1800, the parallel postulate problem again became current, this time for philosophical reasons affecting the worldview of the time: Kantian philosophy considers Euclidean geometry to be a priori valid as the totality of laws of (Euclidean) space, that is, one of the intuitive *Anschauungsformen* of our thought. It was suspected that the a priori validity of geometry was connected with the provability of the parallel postulate, that is, with its derivability from the other axioms. The independence of the axiom, in turn, would be somehow connected with reality, that is, with its a posteriori validity.

Despite the exhortations of his father, Farkas Bolyai—who was a close friend of Gauss and who had wrestled with the problem of the parallel postulate all his life—to leave the problem alone, the young Bolyai could write to his father in a letter of 1823, "out of nothing I created a new, different, world."

His findings first appeared in 1832, as an appendix to a textbook written by his father with the title, "Appendix. Scientiam spatii absolute veram exhibens: a veriate aut falsitate Axiomatis XI Euclidei (a priori haud unquam decidenda) independentem…" (The Science Absolute of Space: Independent of the Truth or Falsity of Euclid's Axiom XI (which can never be decided a priori)…).

Nikolai Lobachevsky, a Russian professor in Kazan, published his first work on this topic in 1829. Gauss had obtained results in this direction even earlier, around 1820, but did not publish them.

Quotation 4.58, ***continued***

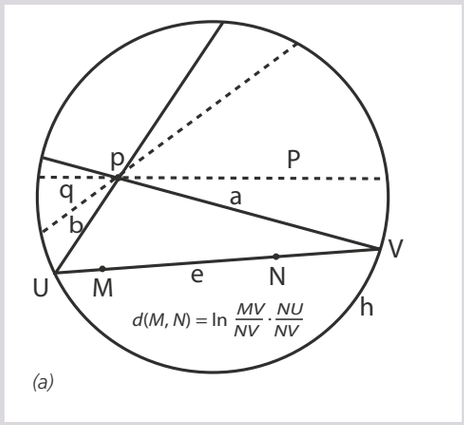
it is the simplest in itself, just as a polynomial of the first degree is simpler than a polynomial of the second degree; 2nd, because it sufficiently agrees with the properties of natural solids, those bodies which we can compare and measure by means of our senses. …

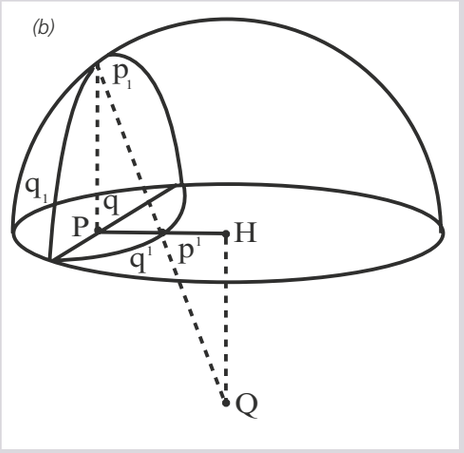
Suppose, for example, a world enclosed in a large sphere and subject to the following laws: The temperature is not uniform; it is greatest at the centre, and gradually decreases as we move towards the circumference of the sphere, where it is absolute zero. The law of this temperature is as follows: If R be the radius of the sphere, and r the distance of the point considered from the centre, the absolute temperature will be proportional to $R^2 - r^2$. Further, I shall suppose that in this world all bodies have the same co-efficient of dilatation, so that the linear dilatation of any body is proportional to its absolute temperature. Finally, I shall assume that a body transported from one point to another of different temperature is instantaneously in thermal equilibrium with its new environment. There is nothing in these hypotheses either contradictory or unimaginable. A moving object will become smaller and smaller as it approaches the circumference of a sphere. Let us observe, in the first place, that although from the point of view of our ordinary geometry the world is finite, to its inhabitants it will appear infinite. As they approach the surface of the sphere they become colder, and at the same time smaller and smaller. The steps they take are therefore also smaller and smaller, so that they can never reach the boundary of the sphere. If to us geometry is only the study of the laws according to which invariable solids move, to these imaginary beings it will be the study of the laws of motion of solids *deformed by the differences of temperature* alluded to.

No doubt, in our world, natural solids also experience variations of form and volume due to differences of temperature. But in laying the foundations of geometry we neglect these variations; for besides being but small they are irregular, and consequently appear to us to be accidental. In our hypothetical world this will no longer be the case, the variations will obey very simple and regular laws. On the other hand, the different solid parts of which the bodies of these inhabitants are composed will undergo the same variations of form and volume.

Let me make another hypothesis: suppose that light passes through media of different refractive indices, such that the index of refraction is inversely proportional to $R^2 - r^2$. Under these conditions it is clear that the rays of light will no longer be rectilinear but circular.

—Poincaré, *Science and Hypothesis* [pp. 49–50, 65–66]





▲ **Figure 4.159** **(a)** The Klein model of Bolyai–Lobachevsky geometry. The points of the circle do not belong to the non-Euclidean plane.

(b) The derivation of the Poincaré model from the Klein model (in part after F. Karteszi).

Figure 4.159 shows two representations in the Euclidean plane of the Bolyai–Lobachevsky geometry. They demonstrate the consistency of this geometry. However, we would like to briefly mention above all the Beltrami model because it is easy to comprehend. In 1868, Beltrami studied the question of whether there exists a (two-dimensional) surface embedded in three-dimensional space whose "internal geometry," that is, the geometry of the surface itself, is independent of its embedding. More precisely, the square ds^2 of the line element measured in this two-dimensional space, which determines the geometry, is identical to the expression for ds^2 associated with the plane of hyperbolic geometry. He did, in fact, find such a surface (Figure 5.33): This surface arises as a rotation of the tractrix curve (the curve traced by an object that is pulled by a thread from a point moving in a straight line, a "curve of pursuit"). We note as an aside that this surface is also characterized by another property: it possesses a constant negative curvature and is called a pseudosphere. The (straight) lines correspond to the geodesic (shortest) lines; the line elements correspond to the line elements along these curves, etc. Because only a part of the hyperbolic surface can be mapped onto this surface, the consistency of the hyperbolic geometry cannot be proven this way. In the Klein model (Figure 4.159(a), Felix Klein, 1871/72), the interior points of a circle in the Euclidean plane are taken as the points of the plane of the hyperbolic geometry. Thus, the points of the circle do not belong to the non-Euclidean plane. The points of the chord of the circle—minus its endpoints—form the lines of the hyperbolic plane. The distance between points M and N on this line are given by

$$d(M,N) = \ln \frac{MV}{NV} \times \frac{NU}{NV},$$

where MV, MU, NU, NV are the usual lengths of the given intervals. The significance of the cross ratio here is that it remains invariant under a projective transformation. This definition leads to a quantity that possesses all the properties required of a distance function:

$$d(M,M) = 0, \quad d(M,N) > 0, \quad d(M,N) + d(N,P) = d(M,P)$$

when M, N, and P lie on a straight line, and

$$\mathrm{d(M,N)} + \mathrm{d(N,P)} > \mathrm{d(M,\ P)}$$

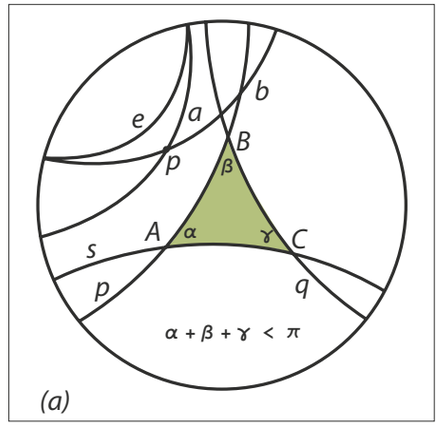
when they do not lie on a straight line (triangle inequality). It is also easy to see that if M is a fixed point on the chord UV and the chordal point N approaches the endpoint V, the distance between the points becomes arbitrarily large: $d(M,N) \to \infty$. Thus the points on the circle lie at infinity.

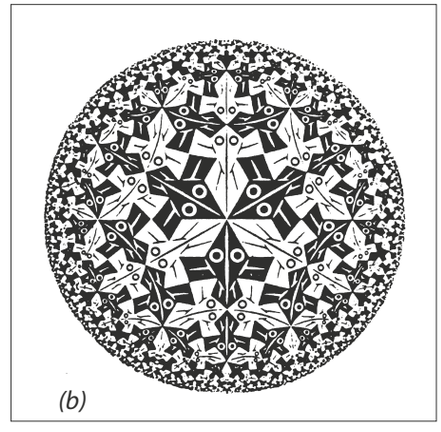
One may also see from the figure that the line a passing through the point P has no point in common with the line e (the points of the circle do *not* belong to the hyperbolic plane), and the lines b, p, and q also have no intersections, that is, in Bolyai–Lobachevsky geometry, one can have infinitely many parallels to a given line through an external point.

The Poincaré model (1882) is obtained from the Klein model in the following way (Figure 4.159(b)). We draw a hemisphere above the circle H and project points P onto the surface of the sphere perpendicularly to the base plane: The point P_1 thus obtained is then projected via a straight line passing through the pole of the lower hemisphere back onto the base plane (stereographic projection). The first projection takes the line q to a semicircle, and then the stereographic projection takes this semicircle to the arc of a circle that meets the base circle orthogonally.

These arcs represent the lines in the non-Euclidean geometry. Figure 4.160(a) shows the relationships in the Poincaré model: the parallels to the line e and the triangle ABC.

Figure 4.160(b) explores these relationships through an artistic vision, and Quotation 4.58 represents the no less thrilling vision of a scientist.





▶ **Figure 4.160** **(a)** In the Poincaré model, straight lines appear as circular arcs that meet the fundamental circle perpendicularly. The lines a and b through the point P are parallel to the line e. However, also parallel are all the lines in the shaded area. In the triangle ABC formed by the lines s, p, q, the sum of the interior angles is clearly less than 180°.

(b) Maurits Cornelis Escher (1898–1972), Dutch graphic artist, represents in this drawing a non-Euclidean (two-dimensional) space. The straight lines (as circular arcs) serve as an even grid for the fish, all of which therefore have the same measure; or, in everyday terms, they are all of the same size.

THE PHYSICS OF THE TWENTIETH CENTURY

"But this is dreadful," Grete Hermann said. "On the one hand, you claim that your knowledge of the Radium B atom is incomplete inasmuch as you cannot tell when and in what direction the electron will be emitted, while, on the other hand, you tell me that your knowledge is complete because, if there were further determinants, you would get into trouble with other experiments. But our knowledge cannot possibly be complete and incomplete at the same time. The whole thing is sheer nonsense... I was wondering why our inability to discover the causes of, say, the emission of an electron means that we must stop searching further. Admittedly, you don't forbid this search, but you claim that it is futile since no further determining factors can be found. Indeed, you contend that, if only it is formulated in precise mathematical language, the indeterminacy allows a definite prediction in another experiment. And this, too, you claim is borne out by the results. If you argue like that, you turn uncertainty into a physical reality, with an objective character, while, normally speaking, uncertainty is a synonym for ignorance, and as such something purely subjective."

Here I felt I must once again intervene in the discussion. "With your last remark," I said, "you have given a very precise description of the most characteristic feature of modern quantum theory. Whenever we try to deduce laws from our study of atomic phenomena, we discover that we no longer correlate objective processes in space and time, but only observational situations. Only for these can we derive empirical laws. The mathematical symbols with which we describe such observational situations represent possibilities rather than facts. One might say that they represent an intermediate stage between the possible and the factual, which can only be called objective in the sense that, say, temperature is called objective by statistical thermodynamics. Our knowledge about what is possible does admittedly enable us to make a few clear predictions, but, as a rule, it only allows us to speculate as to the probability of a future event. Kant could not possibly have foreseen that in an experimental realm so far beyond daily experience we could no longer treat observations as if they referred to 'Dinge an sich' or 'objects'; in other words, he could not foresee that atoms are neither things nor objects."

"In that case what are they?"

"We lack the right term, for our language is based on daily experience, and atoms are not. But if you dislike such evasions, we might say that atoms are parts of observational situations, parts that have a high explanatory value in the physical analysis of the phenomena involved."

—Heisenberg, *Physics and Beyond* [pp. 120, 122–123]

In our scientific expectation we have grown antipodes. You believe in God playing dice and I in perfect laws in the world of things existing as real objects, which I try to grasp in a wildly speculative way.

—Albert Einstein, Letter to Max Born, November 7, 1944 [p. 90]

DOI: 10.1201/9781032697697-6
This chapter has been made available under a CC-BY-NC-ND license.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

The Physics of the Twentieth Century

5.1 "Clouds on the Horizon of Nineteenth-Century Physics"

5.1.1 *A Conclusion or a New Start?*

The twentieth century will usher in the physics of the sixth decimal place, proclaimed the optimists at the turn of the century, elated at all of physics' accomplishments. By this they meant that the remaining task was simply to refine—based on the existing foundations—the measurements and calculations further in order to arrive at ever more precise numerical results (Quotations 5.1, 5.2).

However, the greatest physicists of the time saw clouds gathering on the horizon. The title of this section alludes to a lecture given by Lord Kelvin in 1900, in which he said:

> The beauty and clearness of the dynamical theory, which asserts heat and light to be modes of [mechanical] motion, is at present obscured by two clouds. The first came into existence with the undulatory theory of light... How could the Earth move through an elastic solid, such as essentially is the luminiferous ether? The second is the Maxwell–Boltzmann doctrine regarding the partition of energy.

We might conclude that both the optimism and the pessimism were prompted by a host of phenomena cropping up in the last quarter of the nineteenth century that did not fit into the classical framework. In fact, when we look at Figure 5.1, we can see a number of question marks, each signifying such a new phenomenon: the laws of spectra, the dependence on velocity of the mass of the electron, x-rays and radioactivity, to mention only a few.

A crisis in physics is never caused by a mere torrent of new phenomena, even when explanations for them have not yet been found. The real problem arises when we can find a definitive explanation based on the current theory—and "explanation" in physics naturally also means that the theory can connect the characteristic numerical values of the phenomenon in question to other already accepted values—but the predictions made by the theory do not agree with the experimentally measured values.

To put it plainly, we can speak of a crisis in a physical theory when in spite of sustained effort, a number of experimental observations cannot be reconciled with the theory.

In Figure 5.1 we have shown the division of the discipline of physics as it was generally accepted at the end of the nineteenth century. As we have already mentioned, this physics incorporates the physics of ordinary matter and the physics of the ether. To be sure, the idea of unification naturally arose as well: Many imagined the ether as a superfine material that should also have the properties of solid matter, in order for it to be able to propagate transverse electromagnetic waves. The converse idea was also proposed, that the ether is a continuous medium within which the particles of ordinary matter are formed as stable vortices in the manner of stable smoke rings in air.

Quotation 5.1

The more important fundamental laws and facts of physical science have all been discovered, and these are now so firmly established that the possibility of their ever being supplanted in consequence of new discoveries is exceedingly remote. ... Our future discoveries must be looked for in the 6th place of decimals.

—A. A. Michelson, *Light Waves and Their Uses,* 1903 [Holton 1988, p.88]

Quotation 5.2

Now to the field of physics as it presented itself at that time. In spite of all the fruitfulness in particulars, dogmatic rigidity prevailed in matters of principles: In the beginning (if there was such a thing) God created Newton's laws of motion together with the necessary masses and forces. This is all; everything beyond this follows from the development of appropriate mathematical methods by means of deduction. What the nineteenth century achieved on the strength of this basis, especially through the application of the partial differential equations, was bound to arouse the admiration of every responsive person.

—Albert Einstein, "Autobiographical Notes" [p. 19]

Quotation 5.3

I am never content until I have constructed a mechanical model of the object that I am studying. If I succeed in making one, I understand; otherwise I do not. Hence I cannot grasp the electromagnetic theory of light. I wish to understand light as fully as possible, without introducing things that I understand still less. Therefore I hold fast to simple dynamics for there, but not in the electromagnetic theory, can I find a model.

—Lord Kelvin, 1884 [Mason 1953, pp. 391–392]

Quotation 5.4

However, after the first brilliant results of the kinetic theory of gases, its recent progress seems not to have fulfilled expectations; at every attempt to place this theory on a firmer footing, the difficulties have increased at an alarming rate. Everyone who has studied the work of the two researchers who have delved most deeply into the analysis of molecular motion, Maxwell and Boltzmann, cannot avoid the impression that in overcoming these problems, the admirable effort of acumen in physics and mathematical facility expended to date does not stand in the desired relationship to the fruitfulness of the results obtained.

—Planck [1958, Vol. 1, pp. 372–373]

Quotation 5.5

For a long time, physicists have assumed that all properties of bodies ultimately can be reduced to combinations of geometric figures and local motions; the general principles to which all physical properties are subject should thus be none other than the principles that govern local motion, principles that underlie efficient mechanics. The general principles of physics would then be encoded in efficient mechanics.

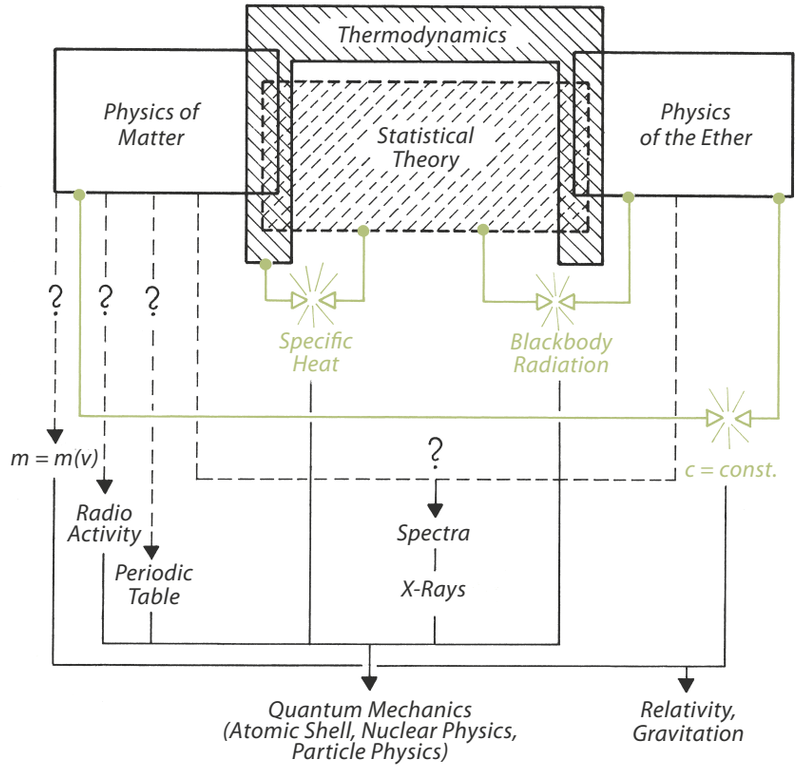
The reduction of all physical properties of combinations of geometric figures and local motions—what is usually called the *mechanical explanation* of the universe—seems today to have been refuted. And indeed, it is not to be refuted for a priori or metaphysical or mathematical reasons but because it was up to now only a project, a dream, and not a reality. Despite great efforts, physicists have not been able to devise such an arrangement of geometric figures and local motions that could provide, according to the rules of theoretical mechanics, a satisfactory representation of an only slightly expanded circle of the physical laws.

Will the attempt to reduce all of physics to theoretical mechanics, an attempt that always failed in the past, perhaps succeed in the future? Only a prophet could answer this question in a positive or negative sense.

Without giving preference to one or other of these answers, it seems much more reasonable to avoid efforts, at least provisionally, that have thus far been fruitless, such as the mechanical explanation of the universe.

We therefore attempt to formulate a system of general laws—laws that all physical properties must obey—without the a priori assumption that all these properties are reducible to geometric figures and local motions. The system of these general laws will in what follows not be reduced to the laws of efficient mechanics.

—P. Duhem, *Traité d´énergétique ou de thermodynamique générale*, 1911 [pp. 2, 3]



▲ **Figure 5.1** The structure of physics toward the end of the nineteenth century. Open problems are indicated with a question mark, and the phenomena that could not be explained within the confines of classical physics appear in color.

Thermodynamics existed as a separate discipline but it also provided a link between the two different groups of phenomena.

Finally, there was the statistical, or kinetic, theory of matter the goal of which was to reduce all physical phenomena to microphysical processes satisfying the laws of mechanics. The mechanistic explanation of the universe, proposed as a program by Descartes, remained an influence and a requirement even as late as the end of the nineteenth century (Quotation 5.2).

Earlier we described the attempts to understand the laws of electromagnetic fields and even entropy with the help of mechanical models. Even for Kelvin, the understanding of a physical phenomenon was complete only when a mechanical model for it could be provided (Quotation 5.3).

In Figure 5.1, the kinetic theory of matter is displayed with dashed borders to indicate that it was not fully accepted. Despite its rapid initial success, its weakness, which would later turn out to be of decisive import, was already evident in the pioneering works, namely, its failure to explain specific heat satisfactorily. Yet the theory of specific heat is based on a very general principle—the equipartition theorem. Maxwell formulated this theorem in 1878 in complete generality, using Lagrange's generalized coordinates with arbitrary force laws between the particles, assuming only the conservation of energy: In thermal equilibrium, energy is partitioned equally among the degrees of freedom of a particle. It is precisely a discrepancy between such a general theory and particular experimental observations that causes the most severe difficulties, for we cannot make the perhaps all too specialized assumptions of the theory responsible for the contradiction. Thus, we may understand why Planck himself (Quotation 5.4), and others as well (Quotation 5.5), looked upon the efforts in this area with a certain degree of skepticism.

5.1.2 *Mach and Ostwald*

Naturally, the question also arose—and with an intensity unusual for the end of the century—whether the problem is not in the program itself? Descartes's original program was put into concrete terms by Newton: The phenomena must be derived from motion of material particles and the forces between them. Laplace gave the most extreme and at the same time the most concrete formulation of this mechanistic worldview, whereby in this manner—at least in principle—one can describe not only present phenomena but the entire past and future as well.

The questioning of this program may be seen as a first attack on the authority of Newton.

At the same time as the initially successful kinetic theory began to reveal its shortcomings, efforts were made to "purify" physics—that is, to cleanse it of all traces of its metaphysical background. This positivistic tendency is not so much a consequence of a philosophical positivism as a direct reaction by natural scientists against philosophy itself. In other words, here for the first time, the scientists turned against the philosopher specialists and tried to solve the philosophical problems within their domains of study (Quotations 5.6, 5.7).

In his 1874 work *The Principles of Mechanics*, Kirchhoff broke with the idea of causality and wrote this about the task of mechanics:

> Mechanics is the science of motion; its subject is the complete description of motion in nature in the simplest possible manner.

This approach was taken up by Ernst Mach, whom we shall cite often in what follows. In his work *The Science of Mechanics: A Critical and Historical Account of Its Development* (1883), Mach explains his natural philosophy. He saw the task of science as discovering among experimental results those economical and easily used relationships that can be practically applied in the real world. For him, everything beyond this goal was metaphysics. One may speak of atoms as a mathematical model, but one should not view them as objects that actually exist. According to Mach, there are no causal relationships among phenomena, only functional ones, which merely describe the "course" of natural phenomena. Moreover, the "mania for proof," to derive everything from first principles, is absurd. If a law, or in Mach's words, a *rule*, is useful in describing experimental results, then no further proof is necessary. Mach's point of view can be summarized as follows: The task of physics is not to ask "why," but "how." Concepts and phenomena that are not immediately observable or measurable have no place in physics; in seeking functional relationships, one should be guided by the economy of the description (Figure 5.2).

There is probably no need to warn the reader of the danger in setting such a limitation on the province of physics, not to mention the fact that this limitation also finds itself in conflict with Mach's own demand for economy of description. We will nevertheless later see that according to the testimony of physicists of that era, Mach's philosophy was perhaps unique in having a serious influence on the development of twentieth-century physics. Among his critics was Lenin, whose polemics appeared in his 1909 work *Materialism and Empirio-Criticism: General Comments on a Reactionary Philosophy*. Here we will briefly mention a personal catharsis that befell Mach. During his whole life he fought against the existence of atoms: "Who has ever seen one?" he would constantly ask. On his deathbed, when he was shown a "spinthariscope" that presents the alpha particles emitted by radioactivity as sharp points of light, he finally had to admit, "they indeed exist."



▲ **Figure 5.2** Ernst Mach (1838–1916): Studies of mathematics, physics, and physiology in Vienna, then taught at Vienna and Graz. From 1867, professor of experimental physics at the German University in Prague, later its rector. From 1895 to 1901, again in Vienna as occupant of a chair created especially for him as the "Professor of the History and Theory of the Inductive Sciences."

Mach had a significant and lasting influence on the development of physics through his work in three different roles:

1. Mach's physics: He was an outstanding experimental physicist, above all through his work on gas dynamics, and here primarily through his research into phenomena associated with supersonic speeds. He achieved extraordinary results using new research techniques (stroboscope, photographs of density distributions): Mach front and Mach angle (the shock waves following objects moving at supersonic speeds) as well as the Mach number M, which is the ratio of an object's speed to the speed of sound: for $M < 1$, the speed is said to be subsonic; for $M > 1$, it is supersonic. The Mach number is an especially important characteristic of gas dynamics because it expresses the extent to which we can consider flowing gas incompressible ($M < 0.3$); it also characterizes the speed when completely new phenomena appear in shock waves ($M > 5$); in the hypersonic realm, chemical reactions within gases have to be considered as well.

   Mach's principle: The inertial mass of a body is a consequence of its interaction with the entire mass of the universe.
2. Mach's textbooks of "Critical and Historical Accounts" among which the most significant works are *The Science of Mechanics: A Critical and Historical Account of Its Development*, 1883; *The Principles of Thermodynamics: A Critical and Historical Account*, 1896; and *Knowledge and Error: Sketches on the Psychology of Enquiry*, 1905.
3. Mach's philosophy:

   > It is the object of science to replace, or *save*, experiences, by the reproduction and anticipation of facts in thought. Memory is handier than experience, and often answers the same purpose. This economical office of science, which fills its whole life, is apparent at first glance; and with its full recognition all mysticism in science disappears. …
   >
   > Sensations are not signs of things; but, on the contrary, a thing is a thought-symbol for a compound sensation of relative fixedness. Properly speaking, the world is not composed of "things" as its elements, but of colors, tones, pressures, spaces, times, in short what we ordinarily call individual sensations.
   >
   > —*The Science of Mechanics* [pp. 557, 579]

Quotation 5.6

Does one not know in general that scientists and philosophers are not currently good friends, at least in their scientific work? ...Who would not be inclined to favor the short, self-creative path of pure reason to the tedious, slowly progressing day labor of scientific research?

The scientific researchers were indicted by the philosophers for narrow-mindedness, and latter by the former for lack of sense. The scientists began now to place a certain weight on the fact that their work could be kept free of all philosophical influences, and it soon turned out that many of them, including men of exceptional importance, condemned all philosophy as useless, indeed as detrimental daydreaming. We cannot deny that the justified demands of philosophy, namely to exercise a critique of the sources of knowledge and to determine a standard of intellectual work, were thrown overboard with the unjustified demands raised by [Hegelian] philosophy to subordinate all other disciplines.

—Helmholtz, *Philosophische Vorträge und Aufsätze* (1855, 1862)

Quotation 5.7

That it could be a question whether it is even science is to be understood only from the development of the specifically modern sciences. These completed their development in the nineteenth century generally without philosophy, often in opposition to philosophy, and ultimately with indifference to it.

Several decades ago, there was a general opinion: Philosophy would have its day until all sciences were released from it, the initially universal science. Now that all that is amenable to investigation has been divided up, its time is over. After it became known how science obtains its compelling general validity, I have shown that under these criteria, philosophy breaks down. It carries out empty thoughts because it eschews experience, it tempts through illusions, it robs the forces for genuine research to use them for a trivial pursuit, for this general talk about the whole.

Such a picture of philosophy stood in the light of science as methodical, compelling, generally valid knowledge. Could any philosophy still consider itself a science at all?

Two reactions occurred:

First: The attack was recognized as justified. The representatives of philosophy therefore withdrew to pursue limited tasks. If philosophy is at an end because it has abandoned all its objects to the

continued on next page

However, the opponents of atomic theory did not stop at criticizing it; they proposed their own program. It is by no means coincidental that the advocates of this program, called energeticism, were primarily chemists or physical chemists. The leading light of this group was Wilhelm Ostwald (1853–1932). At the time, atomic theory had little to contribute to an explanation of chemical processes. Let us not forget that this problem, the physical explanation of chemical phenomena, is very difficult even for modern quantum mechanics, even if only in its technical aspects. Thus, Ostwald was able to assert with confidence that the mechanical interpretation of natural phenomena could not be the correct approach because even the interpretation of the simplest thermodynamic concept, namely, temperature, caused great difficulties in all but the most trivial cases. Ostwald and his followers saw in energy the physical quantity based on which a unified worldview could be constructed. According to them, energy is more basic and more real even than matter. Consequently, they proposed to replace in physics and chemistry the system of measurement based on units of mass, length, and time by a system based on units of energy, length, and time. Quotation 5.8 gives a good illustration of Ostwald's beliefs, according to which energy is the true embodiment of reality and the theories built on it can never fail, but can merely be made more precise.

Representatives of the two conceptions—atomic theory and energeticism—clashed with considerable passion at the Lübeck conference in 1895. The mood can be sensed from Sommerfeld's reminiscences:

> Among the energeticists, Helm, of Dresden, gave a talk. He was supported by Ostwald, and both of them were under the influence of Mach's philosophy, although Mach himself was not present. The debate between Boltzmann and Ostwald was reminiscent in content as well as form of the fight between a bull and an agile toreador. This time, however, despite all his swordsmanship, the toreador (Ostwald) was losing. Boltzmann's arguments were overwhelming. We mathematicians were all on Boltzmann's side.

In the interest of historical truth, we must mention that Ostwald lived long enough to see the complete triumph of the atomic theory and had the wisdom to admit his error publicly.

The nineteenth century achieved a wealth of partial results in all areas of physics, but it also made contributions to the general worldview of physics, which we now summarize.

Notwithstanding all its exaggerations, energeticism made a positive contribution by emphasizing the central role of the law of conservation of energy and thereby the central notion of energy itself.

With the second law of thermodynamics, a type of law was formulated for the first time in the history of physics that defines a direction for a process, and with this—contrary to the laws of classical mechanics—introduces the realm of irreversible processes and makes it possible to determine the direction of time based on physical laws.

Connected with this was the spread of statistical methods for analyzing phenomena, the notion of probability distribution becoming well established in physics, although for the time being in the context of ironclad determinism.

Finally, toward the end of the nineteenth century, physicists began to become accustomed to the idea that even though electromagnetic fields cannot be modeled with gears, rods, and joints and cannot be touched, they are as much part of reality as the pieces of matter that assault our senses in the crudest ways.

The key concept in the world of science of the nineteenth century outside of physics was that of progress, of evolution. At this time were born Hegel's conception of history, Darwin's theory of evolution, and the historical materialism of Marx and Engels.

Paradoxically, as noted above, the directionality of time also appeared in physics. However, physicists, or rather the philosophers popularizing the new theories, drew a conclusion opposite to that of evolution: a future of heat death of the cosmos (Section 4.6).

The attempts at resolving the contradictions inherent in the physics of the nineteenth century took place in the political and economic context of the times. Regional wars all over the world (Japan–China, Cuba–Spain–USA, Italy–Abyssinia, Greece–Turkey, Great Britain–Boers), the Boxer Rebellion in China, and the bourgeois revolution following the Russo-Japanese War were harbingers of the coming world war and revolutionary transformations. In France, the *Dreyfus Affair* caused a great uproar. A quotation from Thomas Mann (Quotation 5.9) characterizes the "fin-de-siècle" mood of the European intelligentsia.

5.2 The Theory of Relativity

5.2.1 *Antecedents: Failed Attempts at Measuring Absolute Velocity*

It would seem to be an easy matter to pinpoint the exact moment when the theory of relativity was born. In encyclopedias, works of popular science, and even in physics textbooks we often find statements like, "Relativity theory was created by Einstein in 1905." If we examine closely Einstein's 1905 work *Zur Elektrodynamik bewegter Körper* (On the Electrodynamics of Moving Bodies), then the assertion appears to be accurate, because this paper makes reference to no preceding work on the subject. Consequently, it is all the more striking that the physically most important result of the paper, which connects the characteristics of two coordinate systems moving in straight lines with uniform velocity relative to each other, is referred to as the *Lorentz transformation* in the scientific literature.

We may also refer to a proven principle in the history of science, namely that no revolutionary scientific idea ever springs fully formed from the mind of a single scientist in textbook-ready form, as it were. This was true even for Newton. Furthermore, as we have seen, this principle led the French historian of science Pierre Duhem to rediscover the results of medieval physics through the recognition that even such a universal genius as Leonardo da Vinci could not by himself, in a vacuum, create a whole new branch of science. When Einstein's role is weighed, we can even find the sort of exaggerated "undervaluing" that we saw earlier in the case of Galileo: There were some who doubted whether Einstein had made any significant contribution at all to the theory of relativity. We quote here E. Whittaker, who wrote in 1910:

> In the autumn of the same year [1905], in the same volume of the *Annalen der Physik* as his paper on the Brownian motion, Einstein published a paper which set forth the relativity theory of Poincaré and Lorentz with some amplifications, and which attracted much attention.

As the basis for an objective evaluation, one might even today invoke Wolfgang Pauli's justly famous 1921 monograph. Pauli wrote this work on the theory of relativity at the age of 20 at the request of Sommerfeld for the *Enzyklopädie der mathematischen Wissenschaften*. We shall essentially stick to that book in our appraisal.

Quotation 5.7, ***continued***

sciences, what remains is the knowledge of its history, first as a factor of the history of the sciences themselves, then as a phenomenon of intellectual history, as the history of errors, anticipation, the process of liberation in which philosophy has made itself superfluous.

Others followed the modern scientific cast of mind by rejecting previous forms of philosophy and desiring to formulate philosophy finally as a strict science. Thus they took on the task that remained to philosophy because it affected all sciences in common, namely questions of logic and epistemology, phenomenology. To win back its reputation, philosophy henceforth made itself by imitation and desire to serve into the handmaiden of the sciences. As such, it established epistemologically the right to determine what was scientifically valid, which was something that existed already—it in fact did something that was superfluous.

Against this zealous scientism stood a second reaction. The attack on the existence of philosophy was rejected by completely dismissing the demand that philosophy be a science. In fact, philosophy is not a science. Indeed, thus the argument, it relies on feeling and intuition, on imagination and genius. It is a conceptual invocation, not a discovery of the meaning of existence. It signifies an uplift of the spirit or it means longed-for death with open eyes. Indeed, some went even further: Philosophy is unsuited to concern itself with science, for it sees through the dubiousness of all scientific truth. Modern science is in any case fully in error, particularly through the ruinous consequences of rational life for the soul and for human existence. Philosophy itself is not a science, but because of this it approaches actual truth.

Both reactions—both the subjugation to and rejection of science, which was conceived as offering compelling, methodical, and generally valid knowledge—seem to be the end of philosophy. Whether it is sacrificed to science or whether it disowns all science, in no case is it still philosophy.

—Karl Jaspers, *Über Bedingungen und Möglichkeiten eines neuen Humanismus*, 1951

Quotation 5.8

Energy is thus contained within all real or concrete things as a significant component that is never lacking, and to that extent, we may say that *what is really real embodies itself as energy*.

continued on next page

Quotation 5.8, ***continued***

And indeed, energy is what is actual in two senses. It is the actual to the extent that it acts; wherever something happens, one can provide a reason for this occurrence through identifying the participating energies. And second, it is the actual to the extent that it makes it possible to specify the *content* of the event. Everything that occurs, occurs *through* energy and with respect to energy. It forms the stationary pole in the course of the phenomenon and simultaneously the driving force that allows the world of phenomena to circle this pole. Truly, if today, a poet were on the lookout for the greatest content of human thought and observation and if he wished to complain that no great thoughts any longer excite the soul to wide-ranging embrace, then I would say to him that the idea of energy is the greatest thing that mankind has seen on the horizon in the past century, and a poet who understood how to sing in worthy tones of the epic of energy would create a work that would be entitled to honor as the epic of humanity...

For this reason, a theory of energy in some area can never be refuted by the later development of science any more than progress in science could ever refute the theorems on the geometric similarity of triangles. The only thing that could happen in such a direction is an extension or even a sharpening of the content of the laws; in such cases, however, we are dealing with decorative work on the building, not, however, with tearing it down completely and building again from scratch, which, in contrast, has always unavoidably been the case with the usual mechanistic hypotheses. Think, for example, of the sequence of the hypotheses about light ...

—W. Ostwald, *Die Energie*g, 1908 [pp. 5, 6, 127, 128]

Quotation 5.9

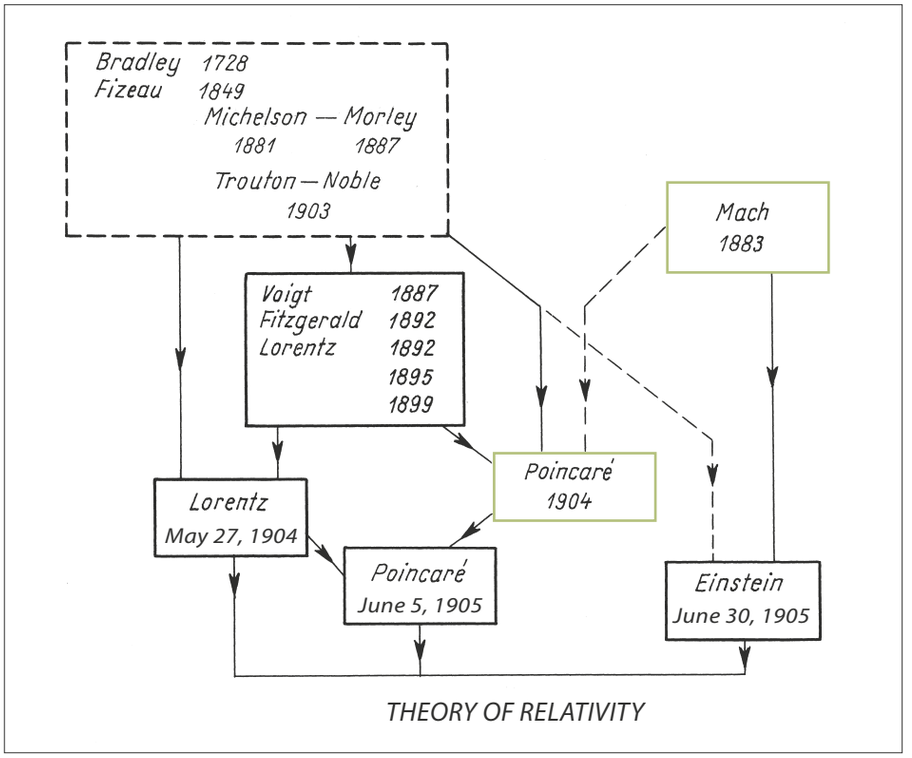
As on a previous occasion, Settembrini found this reprehensible; he at once proved to be very well informed about major current events—and approved of what was happening, since things were taking a course favorable to civilization. The general European atmosphere was imbued with ideas of peace and plans for disarmament. Democratic ideals were on the march. He claimed to have confidential information that the Young Turks had just completed preparations for their revolutionary uprising. Turkey as a constitutional nation-state—what a triumph for humanity!

"The liberalization of Islam," Naphta scoffed. "Excellent. Enlightened fanaticism—how fine." ...

"What you've said is, in any case, cynical. You are determined to interpret democracy's noble endeavors to assert itself internationally as mere political cunning." ...

continued on next page

Figure 5.3 shows the most significant experimental results, the relevant philosophical ideas, and the early attempts at physical theories that preceded the final formulation of the theory of relativity. It is difficult to determine whether experimental or philosophical considerations played a greater role. For Einstein, at least based on what he himself had to say in his later years, it seems that philosophical considerations were of paramount importance.

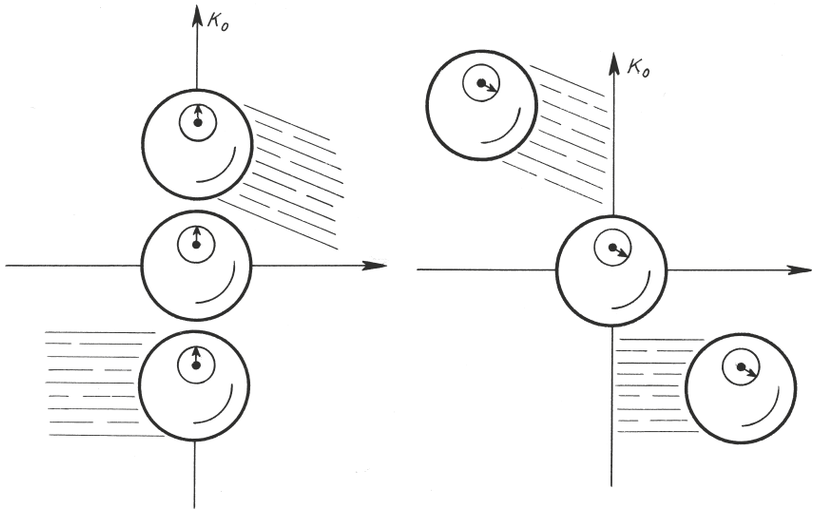


▲ **Figure 5.3** The development of the theory of relativity. Philosophical and qualitative ideas are indicated by the colored frames; the experimental results are shown in the dashed frame.

Let us begin by looking at the "commonsense" concepts at the end of the nineteenth century that served as a framework for accommodating all physical phenomena.

Physical processes take place in space and time. This is, of course, what we still believe even today. But physicists at the end of the nineteenth century also believed that there exist an absolute space and an absolute time, as formulated by Newton. Furthermore, following Paul Drude (1863–1906) and Max Abraham (1875–1922), they believed to have found the embodiment of absolute space in the ether that filled the universe, serving as the carrier of electrical phenomena and thus also of light. Consider a distinguished coordinate system K_0 fixed to absolute space. Phenomena described in this system deserve to be called "true" or "absolute." Thus, a velocity measured in this system can be called absolute velocity; and the rhythm of phenomena is determined by the absolute time that flows uniformly everywhere. This means that, at any point in space, under all circumstances, even on a rapidly moving body, this absolute time has a specific value.

As we have already mentioned, physicists saw their problem to be the investigation of the interaction of this ether with the material bodies moving within it. Bodies move in the ether, which is absolutely at rest, and the laws of motion and interaction are to be written down as functions of absolute time (Figure 5.4). As we see, this view corresponds to our natural perception of the world, our common



▲ **Figure 5.4** In Newtonian physics, one could speak of a distinguished "absolute" stationary coordinate system (K_0) and an absolute time valid in every coordinate system, as well as of simultaneity independent of the coordinate system. The luminiferous ether, filling up all of space, was assumed to give reality to this distinguished system.

sense, and the tried-and-true Newtonian principles. At the same time, it raises a number of questions that can be answered by experiment. Thus, as Maxwell suggested, we could determine the absolute velocity of Earth in space, not just its velocity relative to the Sun by a method analogous to measuring the speed of a ship relative to the surrounding air by determining the speed of sound in the direction of the ship's motion and also perpendicular to it; naturally, the measurement is to be carried out on board the ship. Such an experiment with light was first carried out in 1881 by Michelson (Figure 5.5), and then again in 1887 together with Morley. The experiment yielded the completely unexpected result that the speed of light, no matter what the angle between the directions of the light's propagation and the motion of Earth, is always the same! To get a feeling for how absurd this experimental result seems, let us imagine that we are sitting in a train that is moving at a particular velocity. As we all know, the speed of a train on a track next to ours will appear considerably greater if it is moving in the opposite direction than if it is overtaking us, moving in the same direction as our train. And now it turns out that light—and to avoid any misunderstanding let's emphasize that *only* light—behaves differently: Whether it is coming toward us or it is traveling in the same direction as us, its speed is always the same.

Understandably, people at first doubted the reliability of the measurements: Perhaps the expected differences in velocities were not seen only because the velocity of light ($c = 3 \times 10^8$ m/s) is too great in comparison to the other velocity in the experiment, namely, the speed of Earth in its orbit about the Sun (2.9×10^4 m/s). The Michelson–Morley experiment was replicated by many researchers using better and better measuring procedures. We mention here the contributions of Joos and his colleague Béla Pogány (1887–1943), a professor at the Budapest Technical University, whose 1932 experiment represented the ultimate precision attainable within the confines of "classical" optics. More recently, lasers have been employed to try to detect any dependence of the velocity of light on the direction of propagation. It is characteristic of the precision achieved by these experiments that even a few millimeters per second difference in the velocity of light, had it happened, would have been detected.

Quotation 5.9, ***continued***

"You want me to see such efforts as idealism or even some sort of religiosity, is that it? They are nothing but the last, feeble twitches of what little instinct for self-preservation a doomed international system still has. The catastrophe will, indeed must, follow—is coming toward us from all directions, in all guises." ...

"We know power is evil. But if the kingdom is to come, the dualism between good and evil, between this world and the next, between power and the Spirit, must be temporarily abrogated and transformed in a principle that unites asceticism and dominion. That is what I call the necessity of terror."

"But its agent! Its agent!"

"You even ask? Can it be your Manchester eyes have failed to notice the existence of a social theory that promises the victory of man over economics, a social theory whose principles and goals coincide exactly with those of the Christian City of God? ...

Well, then—after having been buried for centuries, all these economic principles and standards have been resurrected in the modern movement of communism. The correspondence is perfect, down to the meaning of international labor's claim to dominion over international marketeering and speculation. In the modern confrontation with bourgeois-capitalist rot, the world's proletariat embodies the humanity and criteria of the City of God. The point of the dictatorship of the proletariat, the political-economic demand for salvation in our time, is not dominion for its own sake and for all eternity, but only a temporary abrogation of the polarities of mind and Spirit under the sign of the cross. It is a way of overcoming the world by ruling the world, a transition; its point is transcendence, the kingdom itself. The proletariat has taken up Gregory the Great's task, his godly zeal burns within it, and its hands can no more refrain from shedding blood than could his. Its work is terror, that the world may be saved and the ultimate goal of redemption be achieved: the children of God living in a world without classes or laws."

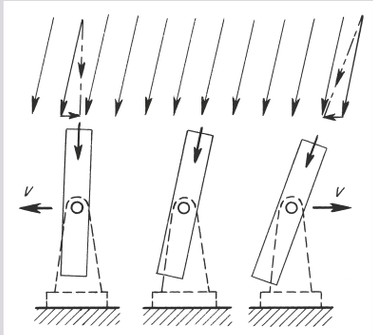
So ended Naphta's caustic oration. The little gathering was silent. The young men looked to Herr Settembrini—they would take their cue form him.

"Amazing," he said. "I certainly must admit my shock—I had not expected this. *Roma locuta*. And what a speech!"

—Thomas Mann, *The Magic Mountain* [pp. 374–396]



◀ Figure 5.5
ALBERT ABRAHAM MICHELSON (1852–1931): The first American physicist to receive the Nobel Prize. From 1880 to 1882, study trip through Europe; 1880: interferometer (Paris); 1881: first provisional measurements to determine the motion of Earth in relation to the ether (Berlin); from 1886: collaboration with EDWARD W. MORLEY; 1887: carried out experiment (together with MORLEY) that became world famous as a foundation for the theory of relativity. An additional important result: the unit of length can be defined in terms of measurements of wavelengths of spectral lines.



◀ Figure 5.6
The phenomenon of aberration: Due to the motion of Earth, the optical axis of the telescope must be aligned differently, depending on the direction of the motion and the direction of the light.

> At last I conjectured, that all the *Phenomena* hitherto mentioned, proceeded from the progressive Motion of Light and the Earth's annual Motion in its Orbit. For I perceived that, if Light was propagated in Time, the apparent Place of a fixt Object would not be the same when the Eye is at Rest, as when it is moving in any other Direction than that of the Line passing through the Eye and Object; and that, when the Eye is moving in different Directions, the apparent Place of the Object would be different. …
>
> The Velocity of Light [is] to the Velocity of the Eye (which in this Case may be supposed the same as the Velocity of the Earth's annual Motion in its Orbit) as 10,210 to One, from whence it would follow, that Light moves, or is propagated as far as from the Sun to the Earth in 8′12″.
>
> It is well known, that Mr. RÖMER, who first attempted to account for an apparent Inequality in the Times of the Eclipses of Jupiter's Satellites, by the Hypothesis of the progressive Motion of Light, supposed that it spent about 11 Minutes of Time in its Passage from the Sun to us: but it hath since been concluded by others from the like Eclipses, that it is propagated as far in about 7 Minutes. The Velocity of Light therefore deduced from the foregoing Hypothesis, is as it were a *Mean* betwixt what had at different times been determined from the Eclipses of Jupiter's Satellites.
>
> —JAMES BRADLEY, "An Account of a New Discovered Motion of the Fix'd Stars" [quoted in Elliott 1966, pp. 21, 23]

5.2.2 *Attempts at Adaptation*

For theoretical physicists, the explanation of such an anomaly by forcing it into a given framework is not a particularly difficult task. Let us suppose, for instance, that Earth simply carries the surrounding ether along with it on its course. If this were the case, then optical experiments carried out on Earth's surface should record results typical of those for the stationary ether and we would not have any reason to expect the speed of light to have different values in different directions. The trouble with this line of reasoning is that this was not the only phenomenon that could reveal a connection between the ether and ponderable matter. For instance, in Figure 5.3 we see the phenomenon of the aberration of light observed and correctly interpreted by BRADLEY as early as 1728. To observe the light emitted by a star, the orientation of the telescope's optical axis must be altered somewhat to compensate for the direction of Earth's motion relative to the direction of incidence of the light from the star (Figure 5.6). This phenomenon, in contrast, can be explained by assuming that Earth passes through the stationary ether *without* carrying it along.

Thus, to explain one phenomenon, we need to assume that the ether is carried along by Earth, but for the other, that it does not. To these two experimental facts, FIZEAU added an important third (Figure 5.7). He investigated how the speed of light changes in a moving medium, which in his experiments was water. He determined that the speed of light is indeed greater than in the stationary medium, but not by the amount that we would expect, namely, by the velocity of the moving medium.

The value c' that he measured for the speed of light is

$$c' = \frac{c}{n} + \left(1 - \frac{1}{n^2}\right)v,$$

where c/n is the speed of light in a medium with index of refraction n. The quantity v is the velocity of the medium—thus, in our case, the speed at which the water is flowing. What would be expected is for the speed of light in the moving medium to be $(c/n) + v$, whereas what was actually observed was that the fluid velocity has to be multiplied by a "drag coefficient" with a value less than 1 that depends on the index of refraction.

FIZEAU's experiment can be explained by assuming that the ether is carried along, to some degree, by the moving medium.

Although these three experimental observations may be individually explained within the classical framework, taken together they cannot possibly fit. This already implies that the classical framework is incapable of encompassing all of these facts.

The experiment of MICHELSON and MORLEY is so important for the formation of the theory of relativity that we now investigate more fully the principle of measurement as well as both the original experimental setup and a more modern version.

To present the experiment in a more understandable form and to make the role of the ether, either stationary or "carried along," more vivid, instead of light, we first consider sound waves. Imagine a ship with an open deck on which the speed of sound is to be measured. If the ship is at rest with respect to the air, which is the propagation medium for the sound waves, then, according to Figure 5.8(a), the speed of sound v_{sound} can be determined from the transit time $2t$ of a sound pulse that is required to

travel the distance l from the sound source to a reflecting surface and return:

$$2\tau = 2\frac{l}{v_{\text{sound}}}.$$

If now the ship is moving at velocity v_{ship} with respect to the stationary air, then the sound pulse requires the time

$$\tau_1 = \frac{l}{v_{\text{sound}} - v_{\text{ship}}}$$

to reach the reflecting surface. For the return path, the sound pulse requires the time

$$\tau_2 = \frac{l}{v_{\text{sound}} + v_{\text{ship}}},$$

so that the entire travel time of the sound pulse is given by

$$\tau_{\parallel} = \tau_1 + \tau_2 = \frac{l_{\parallel}}{v_{\text{sound}} - v_{\text{ship}}} + \frac{l_{\parallel}}{v_{\text{sound}} + v_{\text{ship}}} = \frac{2l_{\parallel}}{v_{\text{sound}}} \frac{1}{1 - v_{\text{ship}}^2 / v_{\text{sound}}^2}.$$

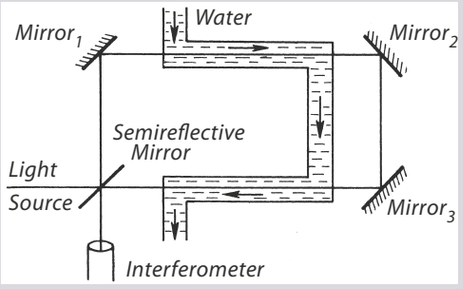
If we know the velocity of sound in stationary air, then from the measurement of this travel time, we can determine the speed of the ship with respect to the stationary air. The influence of the ship's motion is measurable even if the reflecting surface is perpendicular to the direction of travel, as shown in Figure 5.8(b). In such a case, the travel time is given by

$$\tau_{\perp} = \frac{2l_{\perp}}{\sqrt{v_{\text{sound}}^2 - v_{\text{ship}}^2}} = 2\frac{l_{\perp}}{v_{\text{sound}}} \frac{1}{\sqrt{1 - v_{\text{ship}}^2 / v_{\text{sound}}^2}}.$$

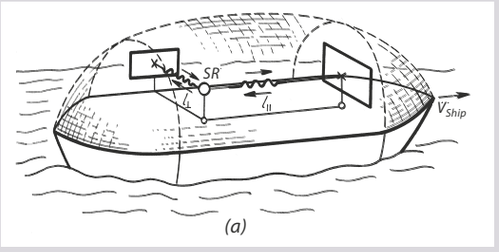
From this measurement, too, the speed of the ship can be determined.

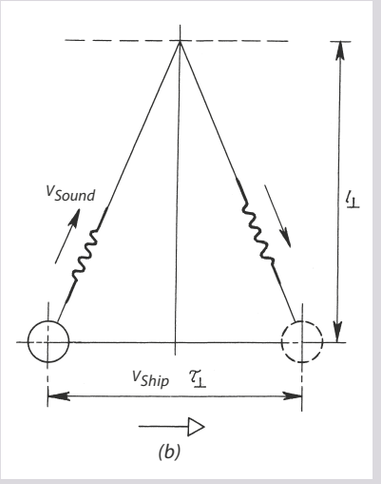
Transferring this model to measuring the velocity of light, we first have to replace the ship by Earth, and the air by the "ether." Moreover, a more precise measurement technique must be used because the direct measurement of the travel times involved is very problematic due to the high value for the speed of light. Therefore, we use the interference between two light waves, a well-known precision method in optics. A correspondingly modified measuring procedure, shown in Figure 5.9, is suited to the investigation of the influence of the motion of the coordinate system on the velocity of light. A light ray emitted from a source L is decomposed into two parts by a parallel-plane glass plate. One of the partial rays falls on a mirror with its reflective surface perpendicular to the direction of motion, and the other partial ray falls on a mirror that is oriented parallel to the direction of motion. After reflection in the mirrors, the two partial rays meet on the parallel-plane plate, and from there, a certain portion of both rays arrives at an interferometer, where they create an interference image corresponding to the path difference. If we rotate the experimental setup by 90°, so that the mirror that was originally standing perpendicular to the direction of motion moves into a position that is parallel to the direction of motion and vice versa, then the path difference will change sign. Consequently, the interference lines should also shift. However, despite repeated experiments carried out with the greatest possible care and with a great variety of measurement techniques, a shift in the interference lines was not observed. From this, we can conclude that the velocity of light is independent of the coordinate system.

Let us now look at a modern version of the Michelson experiment in which laser beams are used. The experimental setup is depicted in Figure 5.10. At this point, it is



▲ **Figure 5.7** Schematic representation of FIZEAU's experiment to measure the speed of light in flowing water.

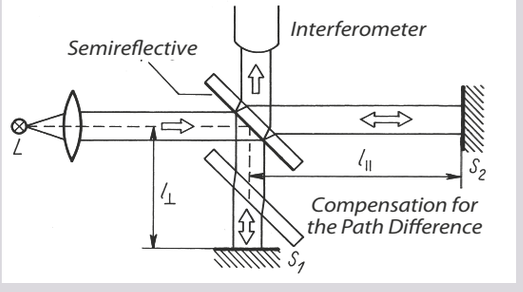


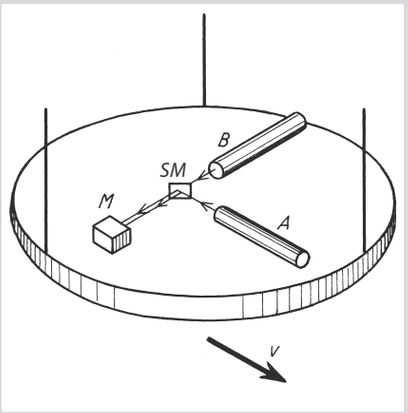


▲ ◀ **Figure 5.8** **(a)** For a stationary ship, we can measure the speed of sound using reflected sound pulses from a send-and-receive unit (SR). If the ship is in motion, we can measure its speed with respect to the stationary air. Of course, we have to make sure that the state of motion of the air is not disturbed, which can be achieved, for example, by using nets as the sound-reflecting surfaces. If a canopy were to cover the entire ship, as shown with dashed lines, then the measurements on the moving ship would be the same as on the ship at rest. This arrangement corresponds to the argument that "the moving Earth carries the ether along."

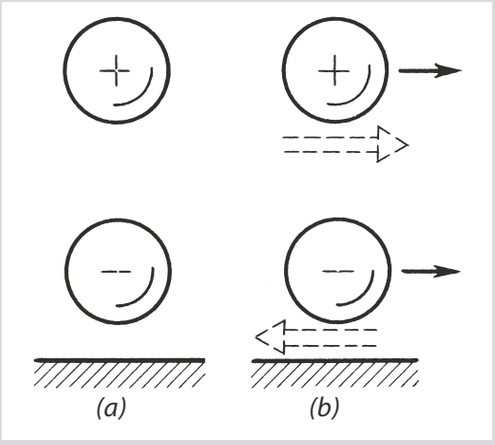
(b) Reflection from a surface whose normal is perpendicular to the direction of motion.

▼ **Figure 5.9** Principle of the Michelson–Morley experiment: The entire experimental setup can be turned about a vertical axis in such a way that $l_\parallel$ and $l_\perp$ are interchanged.



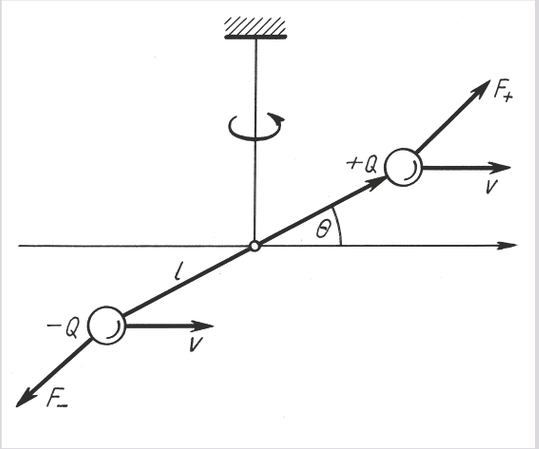


◀ **Figure 5.10** Experiment by Jaseja, Javan, Murray, and Townes. With the help of the semireflective mirror *SM*, light from the gas lasers *A* and *B* reach the mixing unit *M*, which measures the difference in frequencies. As in Figure 5.9, the experimental apparatus can be rotated.



▲ **Figure 5.11** **(a)** An electrostatic force acts between two stationary charges.

(b) Moving charges represent electric currents, and the magnetic force between the two currents opposes the electrostatic force.



▲ **Figure 5.12** A torque acts on two charges of opposite sign moving parallel to each other bound by a rigid rod; the torque acts to maintain the rod perpendicular to the direction of motion.

sufficient for us to know that in contrast to ordinary light rays, laser beams are coherent and can be described by a single sinusoidal wave. In a gas laser, for example, the laser beam develops between two opposing mirrors that act as an optical resonator. The wavelength and frequency therefore can be determined from the relationship

$$n\lambda = 2l, \quad v = \frac{c}{\lambda} = \frac{c}{2l/n} = \frac{nc}{2l}.$$

Within the classical framework, according to which light propagates in the stationary ether with velocity c, it can be shown that the frequency emitted by a laser oriented parallel to the direction of Earth's motion is given by

$$v_{\parallel} = \frac{nc}{2}\frac{1}{l_{\parallel}\kappa^2}, \quad \text{where } \kappa^2 = \frac{1}{1 - v^2/c^2},$$

whereas a laser oriented perpendicularly to the direction of Earth's motion should emit rays of frequency

$$v_{\perp} = \frac{nc}{2l_{\perp}}\frac{1}{\kappa}.$$

If now the rays of both lasers are superimposed in a mixer, as shown in Figure 5.10, where the beat frequency of the difference in frequencies can be measured, no difference in frequencies is observed, and the constancy of the velocity of light has been proven to such a degree of precision that even a deviation of a few millimeters per second relative to its magnitude of 3×10^8 m/s would be detectable.

The ether was considered a carrier not only of electromagnetic waves changing rapidly over time, but also of static electric and magnetic fields. The question naturally arose whether it would be possible to show experimentally "absolute motion," that is, motion with respect to the ether of electrically charged bodies. Consider, for example, as in Figure 5.11(a,b), one metal sphere charged positively and one negatively, so they will attract each other according to Coulomb's law. If we now move the charges in parallel at the same speed, since moving charges represent an electric current, a magnetic force will also arise between the two charges. But the charges at rest in a laboratory are moving along with Earth at a large velocity, so these magnetic forces should already be there. Even though the magnetic force is very small in comparison to the electrostatic force, it is nonetheless, considering Earth's great velocity, within the limits of detectability. The effect of the magnetic force is such that if we suspend the two oppositely charged spheres separated by a rod on a torsion wire—in the manner shown in Figure 5.12—it will tend to turn the rod perpendicular to the direction of Earth's motion. In 1903, Trouton and Noble carried out the experiment depicted in Figure 5.13, and, despite a measuring instrument sensitive enough to detect the expected force, no such effect was observed.

This was yet another failed attempt to measure an absolute velocity with respect to the stationary ether. As a consequence of many *negative* experimental results—and not for the first time in the history of science—the outlines of a *positive* experimental law gradually began to emerge: The natural phenomena conspired against all the physicists' efforts to detect the stationary ether, the distinguished coordinate system K_0.

However, before finally concluding that such a distinguished coordinate system and consequently also the ether did not exist and that the laws of nature must be precisely of a form to take this fact into account, the theoretical physicists tried to save the day using a variety of ad hoc assumptions. Interestingly, the first

attempt already came fairly close to the correct solution. In 1887, WOLDEMAR VOIGT (1850–1919) theoretically investigated the question of the relationship that would have to exist between the coordinates of a stationary coordinate system and those of a moving one if we wanted the velocity of light to be the same in both coordinate systems. It is of interest that VOIGT had already given up the concept of absolute time and *had also transformed the time coordinate* in the transition from one coordinate system into the other.

We now go into greater detail in discussing VOIGT's train of thought, which is impressive in its originality and audacity.

The propagation of light in a stationary ether is described by the differential equation

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}=\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2},\quad \psi=\psi(x,y,z;t). \tag{1}$$

Experiments seem to prove that light satisfies the same differential equation in a moving ether as well. Let us assume that the coordinate system K' moves parallel to the x-axis with respect to the coordinate system K (Figure 5.14). Let us now invoke the Galilean transformation

$$x'=x-vt,\quad x=x'+vt,$$
$$y'=y,\qquad y=y',$$
$$z'=z,\qquad z=z',$$

and substitute the new coordinates x', y', z' as well as the seemingly obvious relationship $t'=t$ into the wave equation (1). We then obtain a differential equation of a completely different form. From this approach, it necessarily follows that the velocity of light should depend on the motion of the coordinate system with respect to the ether. VOIGT, however, had the idea that if instead of the Galilean transformation, we use the transformation formulas

$$x'=x-vt,$$
$$y'=y\sqrt{1-\beta^2},$$
$$z'=z\sqrt{1-\beta^2},$$
$$t'=t-\frac{v}{c^2}x,\quad \beta=v/c,$$

then we obtain for the differential equation, expressed in the new coordinates,

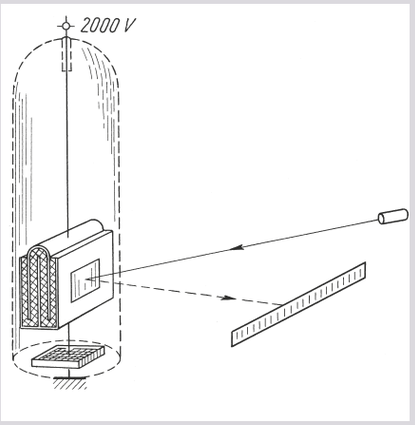
$$\frac{\partial^2\psi}{\partial x'^2}+\frac{\partial^2\psi}{\partial y'^2}+\frac{\partial^2\psi}{\partial z'^2}=\frac{1}{c^2}\frac{\partial^2\psi}{\partial t'^2},\quad \psi=\psi(x',y',z';t').$$

That is, the speed of propagation of light is written in both coordinate systems as equations that are identical in form.

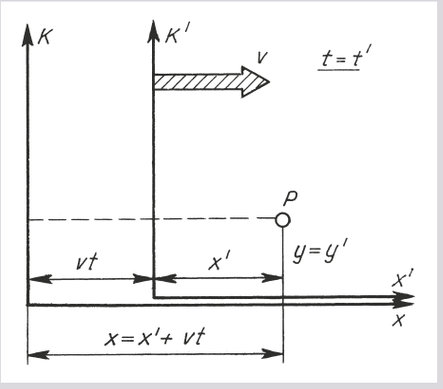
The transformation found by VOIGT differs from the correct Lorentz transformation only by a constant factor. To obtain the correct Lorentz transformation, we need to multiply the right sides of the transformation equations by $1/\sqrt{1-\beta^2}$.

So the advantage of the solution found by VOIGT is that he found a mathematical transformation formula that leaves the form of the wave equation for the propagation of light invariant, and in the process he discarded the relationship $t=t'$, which until then had seemed obvious and necessary, opening the possibility for the first time in the history of physics to call into question the concept of absolute time on the basis of both theory and experiment

But the theory was in need of correction. First of all, it did not satisfy an important requirement, namely, the fundamental requirement of relativity: If we wanted to express the quantities x, y, z, t in terms of the quantities x', y', z', t', transforming in the opposite direction, we would get a different set of transformation formulas. In other words, the distinguished absolute coordinate system, the ether representing



◀ **Figure 5.13** Schematic representation of TROUTON and NOBLE'S experiment. The charge pair is provided by the oppositely charged plates of a mica capacitor hanging on a torsion wire.



▲ **Figure 5.14** The Galilean transformation based on "common sense." Time is "obviously" the same in both coordinate systems.

Quotation 5.10

Let us imagine two observers who wish to synchronize their clocks using optical signals... When station B receives a signal from station A, its clock should not register the same time as that of station A at the moment the signal was sent, but a time later by the time during which the signal traveled. Let us suppose, for example, that A sends its signal at the moment the indicator of its clock points to o and that station B receives this signal at time t. The clocks will then be synchronized if B sets is clock to t, the travel time of the signal. In order to check the synchronization, B sends out a signal when its clock stands at o. When A receives the signal, its clock should read t. Now the clocks are synchronized... on the assumption that the two stations are stationary...

If we synchronize moving clocks by this method, they will not show the correct time; they show what one might call local time, so that one is slower than the other. However, this means nothing... since as required by the principle of relativity, there is no method to determine whether something is at rest or in absolute motion...

continued on next page

Quotation 5.10, ***continued***

And now I move on to the Lavoisier principle of conservation of mass; in fact, one cannot touch this principle without upsetting all of classical mechanics. Namely, today, many are of the opinion that classical mechanics appears to be true only because we calculate only with small velocities but that it loses its validity as soon as we allow speeds near that of light.

From all these results—assuming that they will be further confirmed—an entirely new mechanics arises characterized first of all by the fact that no velocity can exceed that of light... since bodies resist forces attempting to accelerate them with ever greater inertia; and this inertia becomes infinite as the velocity of light is approached.

Mass has two aspects: it simultaneously appears as a characteristic of inertia and as the important factor in Newton's law of gravitation for the attractive force. If it does not represent a constant as a characteristic of inertia, how could the gravitational mass be considered constant?

—Poincaré, Lecture in St. Louis, 1904

The laws of nature must be identical for stationary bodies and those moving at constant velocity, since there is no method, nor can there be any, to determine whether we are in such motion or stationary.

—Poincaré, "L'état actuel et l'avenir de la physique mathématique," 1904

Quotation 5.11

It would appear as though Newton in the remarks here cited still stood under the influence of the medieval philosophy, as though he had grown unfaithful to his resolves to investigate only actual facts. When we say a thing *A* changes with the time, we mean simply that the conditions that determine a thing *A* depend on the conditions that determine another being *B*. The vibrations of a pendulum take place *in time* when its excursion *depends* on the position of the earth. Since, however, in the observation of the pendulum, we are not under the necessity of taking into account its dependence on the position of the earth, but may compare it with any other thing (the conditions of which of course also depend on the position of the earth), the illusory notion easily arises that *all* the things with which we compare it are unessential. Nay, we may, in attending to the motion of a pendulum, neglect entirely other external things, and find that for every position of it our thoughts and sensations are different. Time, accordingly, appears to be some particular and independent thing, on the progress of which the position of the pendulum depends ... .

continued on next page

absolute space, is still implicitly present in the whole approach. Connected with this is the fact that the physicists continued to suppose for almost two decades longer—and Lorentz believed it even longer than that—that the time lapse of physical processes is controlled by the absolute time of K_0, and the transformed time t' should be seen merely as a "calculational quantity," or, in a somewhat more realistic interpretation, as *local time*, which differs from the true absolute time t due to whatever causes.

Voigt's work went unnoticed and had no subsequent influence on the development of the theory of relativity.

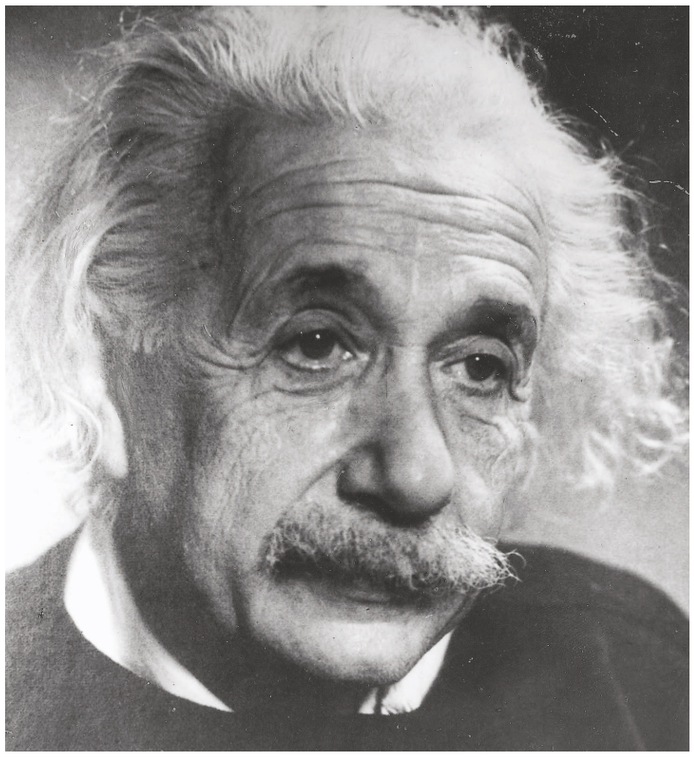
Starting already in 1892, Lorentz tried to interpret, by using a variety of ideas, the negative result of the experiment by Michelson and Morley. Simultaneously with George Francis Fitzgerald (1851–1901), he proposed the so-called Lorentz–Fitzgerald contraction hypothesis: The lengths of objects will somehow contract in the direction of motion, as a result of interaction with the ether, by the factor $\sqrt{1-\beta^2}$. Continually refining his theory, in 1899—in the last year of the century—Lorentz already arrived at the correct transformation formulas for the spatial coordinates; although he gave the incorrect transformation for time, he at least realized, like Voigt, that time has to be transformed as well.

We can see from a talk given by Henri Poincaré (1854–1912) in 1904 in St. Louis before an audience of interested lay people (Quotation 5.10) the extent to which the leading physicists at the beginning of the twentieth century were eagerly awaiting the birth of the theory of relativity. We may imagine that behind all of these qualitative considerations there were many concrete supporting quantitative investigations. As we see, here the term "relativity principle" appears for the first time and here we also read for the first time about the synchronization of clocks, which will reappear later with Einstein. Despite his many brilliant conjectures, we also see that Poincaré was still unable to free himself from the notion of absolute time and absolute motion. Although he professed that there is no mechanism for observing an absolute state of motion, he was not yet ready to conclude from this that no distinguished coordinate system K_0 could exist.

Mach approached the problem from a philosophical perspective. Continuing in the tradition of Leibniz, he characterized not only the Newtonian views on absolute space as metaphysical speculation, as did Leibniz, but the views on absolute time as well (Quotation 5.11). As Einstein himself confessed, Mach had a great influence on him (Quotation 5.12). Although Einstein (Figure 5.15) later completely rejected Mach's epistemological views, there is considerable evidence (oral communications, the article introducing the general theory of relativity, etc.) that in his formulation of the both the special and general theories of relativity, Einstein was guided less by the constraining force of physical facts than by philosophical, or in particular epistemological, considerations.

5.2.3 *The Protagonists: Lorentz, Einstein, and Poincaré*

Relativity arrived at its current formulation in 1905. In Figure 5.3, under the names Lorentz, Poincaré, and Einstein, we show the dates when each of them made their work public. At the times indicated, Lorentz presented his fundamental work to the Amsterdam Academy, Poincaré (Figure 5.16) submitted his work to the French Academy of Sciences, and Einstein's article was received by the editors of the *Annalen der Physik*. If instead we consider the dates of appearance of the printed articles, then the order of precedence changes somewhat: Lorentz, 1904; Einstein, 1905; Poincaré, 1906 (Figures 5.17, 5.18, 5.19).



◀ **Figure 5.15** Albert Einstein (1879–1955): A leading representative of physics and the intellectual life of the twentieth century. In 1900, diploma for teaching physics; in 1905, while working at the Swiss patent office as a "technical expert, third class," he submitted four famous papers to the journal *Annalen der Physik*. From 1909, taught at the University of Zurich, then in Prague, and finally again in Zurich at the Swiss Federal Institute of Technology (ETH). From 1914, professor at Berlin, director of the Kaiser Wilhelm Institute of Physics and member of the Prussian Academy of Sciences. In 1933, emigration to the United States, where until his death he was a professor at the Institute for Advanced Study, in Princeton, New Jersey.

His four foundational papers of 1905 deal with the following topics: the theory of Brownian motion, foundations of the special theory of relativity, connections between the theory of relativity and mass–energy equivalence, and the theory of the photoelectric effect. 1916: general relativity; 1916–1917: derivation of Planck's radiation formula from statistical assumptions. From 1920 until his death, intensive work on a "unified theory of matter." His unsuccessful attempts in this area as well as his skepticism regarding the generally accepted interpretation of quantum mechanics led him in his later years into a certain degree of scientific isolation.

continued on next page

Lorentz began his work with the assumption that the fundamental equations of electrodynamics must have the same form in all coordinate systems that move uniformly and linearly with respect to one another. In this way, he arrived at the system of equations, generally accepted and considered valid even today, that can be used in transforming the characteristic quantities. Because of their relatively simple form, we can present here in full detail the connections between the space and time coordinates of two frames of reference moving with respect to each other at the constant velocity v:

$$
\begin{aligned}
x' &= \frac{1}{\sqrt{1-\beta^2}}(x - vt), & x &= \frac{1}{\sqrt{1-\beta^2}}(x' + vt'),\\
y' &= y, & y &= y',\\
z' &= z, & z &= z',\\
t' &= \frac{1}{\sqrt{1-\beta^2}}\left(t - x\frac{v}{c^2}\right), & t &= \frac{1}{\sqrt{1-\beta^2}}\left(t' + x'\frac{v}{c^2}\right).
\end{aligned}
\qquad (2)
$$

As we mentioned above, in this article and also later in his life, Lorentz maintained that the concepts of an absolute space and an absolute time were fundamental. He saw local time as a computational quantity without deeper content and a necessary evil useful only in accounting for the observed phenomena.

Quotation 5.11, ***continued***

It is utterly beyond our power to *measure* the changes in abstraction by *time*. Quite the contrary, time is an abstraction, at which we arrive by means of the changes of things; made because we are not restricted to any one *definite* measure, all being interconnected. A motion is termed uniform in which equal increments of space described correspond to equal increments of space described by some motion with which we form a comparison, as the rotation of the earth. A motion may, with respect to another motion, be uniform. But the question whether a motion is *in itself* uniform, is senseless.

With just as little justice, also, may we speak of an "absolute time"—*of a time independent of* change. This absolute time can be measured by comparison with no motion; it has therefore neither a practical nor a scientific value; and no one is justified in saying that he knows aught about it. It is an idle metaphysical conception.

—Ernst Mach, *The Science of Mechanics* [pp. 272–273]

Quotation 5.12

[A]ll physicists of the last century saw in classical mechanics a firm and final foundation for all physics, yes, indeed, for all natural science, and that they never grew tired in their attempts to base Maxwell's theory of electro-magnetism, which, in the meantime, was slowly beginning to win out, upon mechanics as well. Even Maxwell and H. Hertz, who in retrospect appear as those who demolished the faith in mechanics as the final basis of all physical thinking, in their conscious thinking adhered throughout to mechanics as the secured basis of physics. It was Ernst Mach who, in his *History of Mechanics*, shook this dogmatic faith; this book exercised a profound influence upon me in this regard while I was a student. I see Mach's greatness in his incorruptible skepticism and independence; in my younger years, however, Mach's epistemological position also influenced me very greatly, a position which today appears to me to be essentially untenable ... For he did not place in the correct light the essentially constructive and speculative nature of thought and more especially of scientific thought; in consequence of which he condemned theory on precisely those points where its constructive-speculative character unconcealably comes to light, as for example in the kinetic atomic theory.

—Einstein, "Autobiographical Note" [p. 21]

Figure 5.15 *continued*

His philosophical views are best expressed in the following quotation:

> I see on the one side the totality of sense-experiences, and, on the other, the totality of the concepts and propositions which are laid down in books. The relations between the concepts and propositions among themselves and each other are of a logical nature, and the business of logical thinking is strictly limited to the achievement of the connection between concepts and propositions among each other according to firmly laid down rules, which are the concern of logic. The concepts and propositions get "meaning," viz., "content," only through their connection with sense-experiences. The connection of the latter with the former is purely intuitive, not itself of a logical nature. The degree of certainty with which this relation, viz., intuitive connection, can be undertaken, and nothing else, differentiates empty phantasy from scientific "truth." The system of concepts is a creation of man together with the rules of syntax, which constitute the structure of the conceptual systems. Although the conceptual systems are logically entirely arbitrary, they are bound by the aim to permit the most nearly possible certain (intuitive) and complete co-ordination with the totality of sense-experiences; secondly they aim at greatest possible sparsity of their logically independent elements (basic concepts and axioms), i.e., undefined concepts and underived propositions.
>
> A proposition is correct if, within a logical system, it is deduced according to the accepted logical rules. A system has truth-content according to the certainty and completeness of its co-ordination-possibility to the totality of experience. A correct proposition borrows its "truth" from the truth-content of the system to which it belongs.
>
> —Albert Einstein, "Autobiographical Notes" [pp. 11–13]

Einstein characterizes himself as follows:

> My passionate interest in social justice and social responsibility has always stood in curious contrast to a marked lack of desire for direct association with men and women. I am a horse for single harness, not cut out for tandem or teamwork. I have never belonged wholeheartedly to country or state, to my circle of friends or even to my own family. These ties have always been accompanied by a vague aloofness, and the wish to withdraw into myself increases with the years.
>
> Such isolation is sometimes bitter, but I do not regret being cut off from the understanding and sympathy of other men. I lose something by it, to be sure, but I am compensated for it in being rendered independent of the customs, opinions and prejudices of others and am not tempted to rest my peace of mind upon such shifting foundations.
>
> [Infeld 1941, pp. 287–288]

A characterization by C. P. Snow:

> There were other paradoxes. He was the voice of liberal science, the prophet of reason and peace, for a generation. At the end, he believed, without bitterness, in the depth of his gentle and tranquil spirit, that it had

continued on next page

Poincaré extended Lorentz's work and, being an excellent mathematician, he noticed the special properties of the Lorentz transformation: among other things, the transformations form a "group" in the mathematical sense, now called the *Lorentz group*, which has the property of leaving the quantity $x^2 + y^2 + z^2 - c^2t^2$ unchanged, or invariant, and therefore it can be interpreted as a rotation in the *x, y, z, ict* "four-dimensional" space. Poincaré correctly identified the transformation rules for a host of physical quantities. These include, in addition to the coordinates *x, y, z*, the charge density r, the current density **J**, the scalar potential *j*, and the vector potential **A**. With Poincaré we find many results that would be systematized by Minkowski only later, in 1908.

In Poincaré's work, the mathematical formalism of the theory of relativity now stands before us essentially in its full battle raiment.

But this leaves us with the question of what was left for Einstein to do? It is simplest if we quote Einstein himself:

> There is no doubt, that the special theory of relativity, if we regard its development in retrospect, was ripe for discovery in 1905. Lorentz had already observed that for the analysis of Maxwell's equations the transformations which later were known by his name were essential, and Poincaré had even penetrated deeper into these connections. Concerning myself, I knew only Lorentz's important work of 1895 (the two papers quoted above in the German text) but not Lorentz's later work, nor the consecutive investigations by Poincaré. In this sense my work of 1905 was independent. The new feature of it was the realization of the fact that the bearing of the Lorentz transformation transcended its connection with Maxwell's equations and was concerned with the nature of space and time in general. [Born 1956, p. 194]

What, then, are these general connections and to what extent can Einstein's interpretation of the acknowledged correct relationships be seen as a revolutionary contribution? We quote from Pauli's monograph, *Theory of Relativity*:

> The formal gaps left by Lorentz's work were filled by Poincaré. He stated the relativity principle to be generally and rigorously valid. Since he, in common with the previously discussed authors, assumed Maxwell's equations to hold for the vacuum, this amounted to the requirement that all laws of nature must be covariant with respect to the "Lorentz transformation."
>
> It was now Einstein, finally, who in a way completed the basic formulation of this new discipline. His paper of 1905 was submitted at almost the same time as Poincaré's article and had been written without previous knowledge of Lorentz's paper of 1904. It includes not only all the essential results contained in the other two papers, but shows an entirely novel, and much more profound, understanding of the whole problem. This will now be demonstrated in detail. ...
>
> Lorentz and Poincaré had taken Maxwell's equations as the basis of their considerations. On the other hand, it is absolutely essential to insist that such a fundamental theorem as the covariance law should be derivable from the simplest possible basic assumptions. The credit for having succeeded in doing just this goes to Einstein. He showed that only the following simple axiom in electrodynamics need be assumed: *The velocity of light is independent of the motion of the light source*. [pp. 3, 5]

Einstein was the first to break consciously and decisively with the notions of absolute space and absolute time and thus to reject the existence of a distinguished reference system K_0 and hence also the necessity of the ether, and therefore its existence. He viewed every inertial system as completely equivalent. In other words, an observer in any given coordinate system can accept the time shown by stationary clocks as the correct, true time; nonetheless, the observer must also realize that the correct and true time for other systems will be different.



Lorentz himself was ill disposed toward Einstein's interpretation for a long time. Nonetheless, in the following decade, he contributed greatly to the promotion of the theory of relativity. Poincaré, however, unexpectedly ceased active work in the area and in effect ignored the progress being made in his popular writings.

Beyond his concrete achievements, Einstein was also the catalyst of progress in the first decades of the twentieth century. There is scarcely a significant advance in relativity or quantum theory in which he did not play a significant role. The citation from the committee awarding Einstein the Nobel Prize emphasized this wide-ranging activity in theoretical physics; as a particular achievement, it highlighted the explanation of the photoelectric effect, not the theory of relativity.

In a 1911 letter, Poincaré characterized Einstein as follows:

> Monsieur Einstein is one of the most original minds I have known; in spite of his youth he already occupies a very honorable position among the leading scholars of his time. We must especially admire in him the ease with which he adapts himself to new concepts and his ability to infer all the consequences from them. [Pais 2005, p. 170]

With a genius of Einstein's stature, there is no need to cast him in the role of a "fairy-tale Hercules to whom every heroic deed is ascribed regardless of who actually

Figure 5.15 *continued*

> all been in vain. He was the most complete of internationalists: he broke away from the Jewish community, he hated all separatisms and nationalisms: yet he was compelled to take his place as the most eminent Jew alive, the committed Zionist. He wanted to lose his personality in the world of nature; but that personality became one of the most publicized of the century, and his face—at first glance the face of an inspired and saintly golliwog—as well known as a film star's. [1980, p. 4]

Finally, Sommerfeld writes about Einstein thusly:

> The battle over the theory of relativity had also slightly reached over into America, where youth was warned by a Boston Cardinal to beware of Einstein the atheist. Thereupon Rabbi Herbert S. Goldstein of New York cabled Einstein: "Do you believe in God?" Einstein cabled back: "I believe in Spinoza's God, who reveals himself in the harmony of all being, not in a God who concerns himself with the fate and actions of men." It would have been impossible for Einstein to give the rabbi a more pointed reply or one which came closer to his own innermost convictions. Many a time, when a new theory appeared to him arbitrary or forced, he remarked: "God doesn't do anything like that." I have often felt and occasionally also stated that Einstein stands in a particularly intimate relation to the God of Spinoza.
>
> —"To Albert Einstein's Seventieth Birthday" [p. 103]

◀ **Figure 5.16** Jules Henri Poincaré (1854–1912): Studies at the École Polytechnique, dissertation at a mining academy on the theory of differential equations. Brief teaching position in Caen, then from 1881, professor at the Université de Paris. In 1906, elected president of the Académie des Sciences.

Poincaré is considered, first of all, a mathematician and, indeed, its most outstanding representative at the turn of the century: *Analysis situs* (1895)—a unified theory of topology; elimination method to solve the equation $\Delta u = 0$ with given boundary conditions (Dirichlet problem); *On Curves Determined by Differential Equations* (from 1878, classification of singular points, notion of limit cycle, which is familiar to today's engineers from the theory of nonlinear oscillations); theory of automorphic functions and their relation to non-Euclidean geometry.

In his comprehensive work *Les méthodes nouvelles de la mécanique céleste* (three volumes, 1892, 1893, 1899), the *n*-body problem is considered. The methods given here are also used in classical quantum mechanics.

His contribution to the theory of relativity is described in the main text. Although Poincaré at first opposed quantum theory, in his later writings one can see an appreciation of the fundamentally statistical character of natural laws. Even in 1919, C. G. Darwin, in a letter to Bohr, bemoaned the fact that Poincaré was no longer alive, for he certainly would have found a way around the difficulties of quantum theory.

In his popular philosophical books, which are considered masterpieces of French prose (*La science et l'hypothèse*, 1903; *La valeur de la science*, 1905; *Science et méthode*, 1908), he expounds his philosophy known as conventionalism.

Some historians of science hold the opinion that Poincaré understood the theory of relativity as deeply as Einstein and that, with regard to the mathematical aspects, he went even further, but his "hypersensitivity" hindered him from fighting for recognition of his ideas.

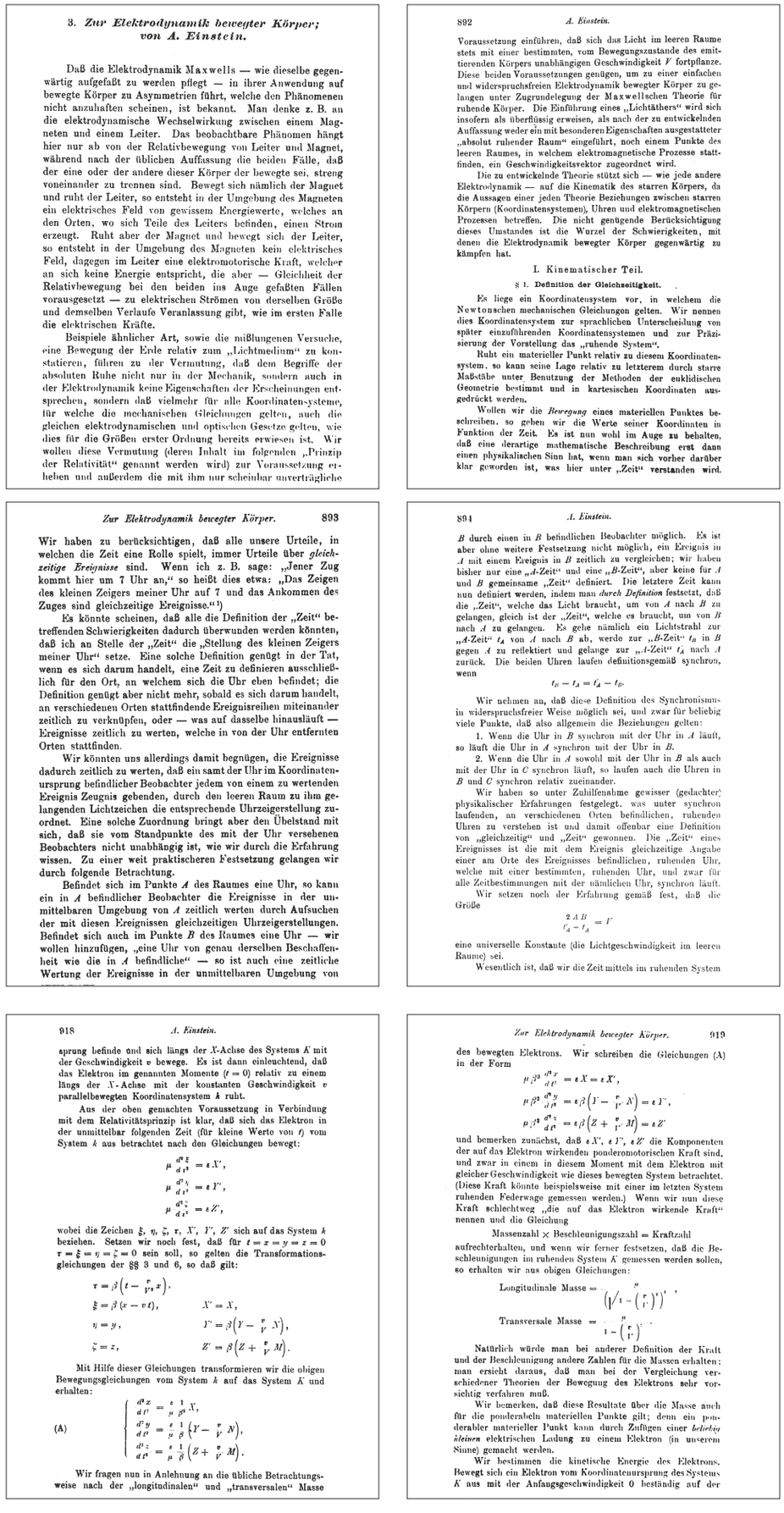
continued on next page

Figure 5.16 *continued*

The concepts of conventionalism are illustrated by the following quotes:

> Can we maintain that certain phenomena which are possible in Euclidean space would be impossible in non-Euclidean space, so that experiment in establishing these phenomena would directly contradict the non-Euclidean hypothesis? I think that such a question cannot be seriously asked. To me it is exactly equivalent to the following, the absurdity of which is obvious: There are lengths which can be expressed in metres and centimetres, but cannot be measured in toises, feet, and inches …
>
> —Poincaré, *Science and Hypothesis* [p. 73]

> Suppose, for example, a world enclosed in a large sphere and subject to the following laws: The temperature is not uniform; it is greatest at the centre, and gradually decreases as we move towards the circumference of the sphere, where it is absolute zero. The law of this temperature is as follows: If R be the radius of the sphere, and r the distance of the point considered from the centre, the absolute temperature will be proportional to $R^2 - r^2$. Further, I shall suppose that in this world all bodies have the same co-efficient of dilatation, so that the linear dilatation of any body is proportional to its absolute temperature. Finally, I shall assume that a body transported from one point to another of different temperature is instantaneously in thermal equilibrium with its new environment. There is nothing in these hypotheses either contradictory or unimaginable. A moving object will become smaller and smaller as it approaches the circumference of the sphere. Let us observe, in the first place, that although from the point of view of our ordinary geometry this world is finite, to its inhabitants it will appear infinite. As they approach the surface of the sphere they become colder, and at the same time smaller and smaller. The steps they take are therefore also smaller and smaller, so that they can never reach the boundary of the sphere. If to us geometry is only the study of the laws according to which invariable solids move, to these imaginary beings it will be the study of the laws of motion of solids *deformed by the differences of temperature* alluded to. …
>
> Let me make another hypothesis: suppose that light passes through media of different refractive indices, such that the index of refraction is inversely proportional to $R^2 - r^2$. Under these conditions it is clear that the rays of light will no longer be rectilinear but circular.
>
> —Poincaré, *Science and Hypothesis* [pp. 65–66]



3. *Zur Elektrodynamik bewegter Körper; von A. Einstein.*

Daß die Elektrodynamik Maxwells — wie dieselbe gegenwärtig aufgefaßt zu werden pflegt — in ihrer Anwendung auf bewegte Körper zu Asymmetrien führt, welche den Phänomenen nicht anzuhaften scheinen, ist bekannt. Man denke z. B. an die elektrodynamische Wechselwirkung zwischen einem Magneten und einem Leiter. Das beobachtbare Phänomen hängt hier nur ab von der Relativbewegung von Leiter und Magnet, während nach der üblichen Auffassung die beiden Fälle, daß der eine oder der andere dieser Körper der bewegte sei, streng voneinander zu trennen sind. Bewegt sich nämlich der Magnet und ruht der Leiter, so entsteht in der Umgebung des Magneten ein elektrisches Feld von gewissem Energiewerte, welches an den Orten, wo sich Teile des Leiters befinden, einen Strom erzeugt. Ruht aber der Magnet und bewegt sich der Leiter, so entsteht in der Umgebung des Magneten kein elektrisches Feld, dagegen im Leiter eine elektromotorische Kraft, welcher an sich keine Energie entspricht, die aber — Gleichheit der Relativbewegung bei den beiden ins Auge gefaßten Fällen vorausgesetzt — zu elektrischen Strömen von derselben Größe und demselben Verlaufe Veranlassung gibt, wie im ersten Falle die elektrischen Kräfte.

Beispiele ähnlicher Art, sowie die mißlungenen Versuche, eine Bewegung der Erde relativ zum „Lichtmedium" zu konstatieren, führen zu der Vermutung, daß dem Begriffe der absoluten Ruhe nicht nur in der Mechanik, sondern auch in der Elektrodynamik keine Eigenschaften der Erscheinungen entsprechen, sondern daß vielmehr für alle Koordinatensysteme, für welche die mechanischen Gleichungen gelten, auch die gleichen elektrodynamischen und optischen Gesetze gelten, wie dies für die Größen erster Ordnung bereits erwiesen ist. Wir wollen diese Vermutung (deren Inhalt im folgenden „Prinzip der Relativität" genannt werden wird) zur Voraussetzung erheben und außerdem die mit ihm nur scheinbar unverträgliche

892 *A. Einstein.*

Voraussetzung einführen, daß sich das Licht im leeren Raume stets mit einer bestimmten, vom Bewegungszustande des emittierenden Körpers unabhängigen Geschwindigkeit V fortpflanze. Diese beiden Voraussetzungen genügen, um zu einer einfachen und widerspruchsfreien Elektrodynamik bewegter Körper zu gelangen unter Zugrundelegung der Maxwellschen Theorie für ruhende Körper. Die Einführung eines „Lichtäthers" wird sich insofern als überflüssig erweisen, als nach der zu entwickelnden Auffassung weder ein mit besonderen Eigenschaften ausgestatteter „absolut ruhender Raum" eingeführt, noch einem Punkte des leeren Raumes, in welchem elektromagnetische Prozesse stattfinden, ein Geschwindigkeitsvektor zugeordnet wird.

Die zu entwickelnde Theorie stützt sich — wie jede andere Elektrodynamik — auf die Kinematik des starren Körpers, da die Aussagen einer jeden Theorie Beziehungen zwischen starren Körpern (Koordinatensystemen), Uhren und elektromagnetischen Prozessen betreffen. Die nicht genügende Berücksichtigung dieses Umstandes ist die Wurzel der Schwierigkeiten, mit denen die Elektrodynamik bewegter Körper gegenwärtig zu kämpfen hat.

I. Kinematischer Teil.

§ 1. Definition der Gleichseitigkeit.

Es liege ein Koordinatensystem vor, in welchem die Newtonschen mechanischen Gleichungen gelten. Wir nennen dies Koordinatensystem zur sprachlichen Unterscheidung von später einzuführenden Koordinatensystemen und zur Präzisierung der Vorstellung das „ruhende System".

Ruht ein materieller Punkt relativ zu diesem Koordinatensystem, so kann seine Lage relativ zu letzterem durch starre Maßstäbe unter Benutzung der Methoden der euklidischen Geometrie bestimmt und in kartesischen Koordinaten ausgedrückt werden.

Wollen wir die *Bewegung* eines materiellen Punktes beschreiben, so geben wir die Werte seiner Koordinaten in Funktion der Zeit. Es ist nun wohl im Auge zu behalten, daß eine derartige mathematische Beschreibung erst dann einen physikalischen Sinn hat, wenn man sich vorher darüber klar geworden ist, was hier unter „Zeit" verstanden wird.

Zur Elektrodynamik bewegter Körper. 893

Wir haben zu berücksichtigen, daß alle unsere Urteile, in welchen die Zeit eine Rolle spielt, immer Urteile über *gleichzeitige Ereignisse* sind. Wenn ich z. B. sage: „Jener Zug kommt hier um 7 Uhr an," so heißt dies etwa: „Das Zeigen des kleinen Zeigers meiner Uhr auf 7 und das Ankommen des Zuges sind gleichzeitige Ereignisse.")

Es könnte scheinen, daß alle die Definition der „Zeit" betreffenden Schwierigkeiten dadurch überwunden werden könnten, daß ich an Stelle der „Zeit" die „Stellung des kleinen Zeigers meiner Uhr" setze. Eine solche Definition genügt in der Tat, wenn es sich darum handelt, eine Zeit zu definieren ausschließlich für den Ort, an welchem sich die Uhr eben befindet; die Definition genügt aber nicht mehr, sobald es sich darum handelt, an verschiedenen Orten stattfindende Ereignisreihen miteinander zeitlich zu verknüpfen, oder — was auf dasselbe hinausläuft — Ereignisse zeitlich zu werten, welche in von der Uhr entfernten Orten stattfinden.

Wir könnten uns allerdings damit begnügen, die Ereignisse dadurch zeitlich zu werten, daß ein samt der Uhr im Koordinatenursprung befindlicher Beobachter jedem von einem zu wertenden Ereignis Zeugnis gebenden, durch den leeren Raum zu ihm gelangenden Lichtzeichen die entsprechende Uhrzeigerstellung zuordnet. Eine solche Zuordnung bringt aber den Übelstand mit sich, daß sie vom Standpunkte des mit der Uhr versehenen Beobachters nicht unabhängig ist, wie wir durch die Erfahrung wissen. Zu einer weit praktischeren Festsetzung gelangen wir durch folgende Betrachtung.

Befindet sich im Punkte A des Raumes eine Uhr, so kann ein in A befindlicher Beobachter die Ereignisse in der unmittelbaren Umgebung von A zeitlich werten durch Aufsuchen der mit diesen Ereignissen gleichzeitigen Uhrzeigerstellungen. Befindet sich auch im Punkte B des Raumes eine Uhr — wir wollen hinzufügen, „eine Uhr von genau derselben Beschaffenheit wie die in A befindliche" — so ist auch eine zeitliche Wertung der Ereignisse in der unmittelbaren Umgebung von

894 *A. Einstein.*

B durch einen in B befindlichen Beobachter möglich. Es ist aber ohne weitere Festsetzung nicht möglich, ein Ereignis in A mit einem Ereignis in B zeitlich zu vergleichen; wir haben bisher nur eine „A-Zeit" und eine „B-Zeit", aber keine für A und B gemeinsame „Zeit" definiert. Die letztere Zeit kann nun definiert werden, indem man *durch Definition* festsetzt, daß die „Zeit", welche das Licht braucht, um von A nach B zu gelangen, gleich ist der „Zeit", welche es braucht, um von B nach A zu gelangen. Es gehe nämlich ein Lichtstrahl zur „A-Zeit" t_A von A nach B ab, werde zur „B-Zeit" t_B in B gegen A zu reflektiert und gelange zur „A-Zeit" t'_A nach A zurück. Die beiden Uhren laufen definitionsgemäß synchron, wenn

$$t_B - t_A = t'_A - t_B.$$

Wir nehmen an, daß diese Definition des Synchronismus in widerspruchsfreier Weise möglich sei, und zwar für beliebig viele Punkte, daß also allgemein die Beziehungen gelten:

1. Wenn die Uhr in B synchron mit der Uhr in A läuft, so läuft die Uhr in A synchron mit der Uhr in B.

2. Wenn die Uhr in A sowohl mit der Uhr in B als auch mit der Uhr in C synchron läuft, so laufen auch die Uhren in B und C synchron relativ zueinander.

Wir haben so unter Zuhilfenahme gewisser (gedachter) physikalischer Erfahrungen festgelegt, was unter synchron laufenden, an verschiedenen Orten befindlichen, ruhenden Uhren zu verstehen ist und damit offenbar eine Definition von „gleichzeitig" und „Zeit" gewonnen. Die „Zeit" eines Ereignisses ist die mit dem Ereignis gleichzeitige Angabe einer am Orte des Ereignisses befindlichen, ruhenden Uhr, welche mit einer bestimmten, ruhenden Uhr, und zwar für alle Zeitbestimmungen mit der nämlichen Uhr, synchron läuft.

Wir setzen noch der Erfahrung gemäß fest, daß die Größe

$$\frac{2\,AB}{t'_A - t_A} = V$$

eine universelle Konstante (die Lichtgeschwindigkeit im leeren Raume) sei.

Wesentlich ist, daß wir die Zeit mittels im ruhenden System

918 *A. Einstein.*

sprung befinde und sich längs der X-Achse des Systems K' mit der Geschwindigkeit v bewege. Es ist dann einleuchtend, daß das Elektron im genannten Momente ($t = 0$) relativ zu einem längs der X-Achse mit der konstanten Geschwindigkeit v parallelbewegten Koordinatensystem k ruht.

Aus der oben gemachten Voraussetzung in Verbindung mit dem Relativitätsprinzip ist klar, daß sich das Elektron in der unmittelbar folgenden Zeit (für kleine Werte von t) vom System k aus betrachtet nach den Gleichungen bewegt:

$$\mu \frac{d^2\xi}{dt^2} = \varepsilon X',$$
$$\mu \frac{d^2\eta}{dt^2} = \varepsilon Y',$$
$$\mu \frac{d^2\zeta}{dt^2} = \varepsilon Z',$$

wobei die Zeichen $\xi, \eta, \zeta, \tau, X', Y', Z'$ sich auf das System k beziehen. Setzen wir noch fest, daß für $t = x = y = z = 0$ $\tau = \xi = \eta = \zeta = 0$ sein soll, so gelten die Transformationsgleichungen der §§ 3 und 6, so daß gilt:

$$\tau = \beta\left(t - \frac{v}{V^2}x\right), \qquad X' = X,$$
$$\xi = \beta(x - vt), \qquad Y' = \beta\left(Y - \frac{v}{V}N\right),$$
$$\eta = y, \qquad Z' = \beta\left(Z + \frac{v}{V}M\right),$$
$$\zeta = z,$$

Mit Hilfe dieser Gleichungen transformieren wir die obigen Bewegungsgleichungen vom System k auf das System K und erhalten:

$$\text{(A)} \qquad \begin{cases} \frac{d^2x}{dt^2} = \frac{\varepsilon}{\mu}\frac{1}{\beta^3}X, \\ \frac{d^2y}{dt^2} = \frac{\varepsilon}{\mu}\frac{1}{\beta}\left(Y - \frac{v}{V}N\right), \\ \frac{d^2z}{dt^2} = \frac{\varepsilon}{\mu}\frac{1}{\beta}\left(Z + \frac{v}{V}M\right). \end{cases}$$

Wir fragen nun in Anlehnung an die übliche Betrachtungsweise nach der „longitudinalen" und „transversalen" Masse

Zur Elektrodynamik bewegter Körper. 919

des bewegten Elektrons. Wir schreiben die Gleichungen (A) in der Form

$$\mu\beta^3 \frac{d^2x}{dt^2} = \varepsilon X = \varepsilon X',$$
$$\mu\beta^2 \frac{d^2y}{dt^2} = \varepsilon\beta\left(Y - \frac{v}{V}N\right) = \varepsilon Y',$$
$$\mu\beta^2 \frac{d^2z}{dt^2} = \varepsilon\beta\left(Z + \frac{v}{V}M\right) = \varepsilon Z'$$

und bemerken zunächst, daß $\varepsilon X'$, $\varepsilon Y'$, $\varepsilon Z'$ die Komponenten der auf das Elektron wirkenden ponderomotorischen Kraft sind, und zwar in einem in diesem Moment mit dem Elektron mit gleicher Geschwindigkeit wie dieses bewegten System betrachtet. (Diese Kraft könnte beispielsweise mit einer im letzten System ruhenden Federwage gemessen werden.) Wenn wir nun diese Kraft schlechtweg „die auf das Elektron wirkende Kraft" nennen und die Gleichung

$$\text{Massenzahl} \times \text{Beschleunigungszahl} = \text{Kraftzahl}$$

aufrechterhalten, und wenn wir ferner festsetzen, daß die Beschleunigungen im ruhenden System K gemessen werden sollen, so erhalten wir aus obigen Gleichungen:

$$\text{Longitudinale Masse} = \frac{\mu}{\left(\sqrt{1 - \left(\frac{v}{V}\right)^2}\right)^3},$$

$$\text{Transversale Masse} = \frac{\mu}{1 - \left(\frac{v}{V}\right)^2}.$$

Natürlich würde man bei anderer Definition der Kraft und der Beschleunigung andere Zahlen für die Massen erhalten; man ersieht daraus, daß man bei der Vergleichung verschiedener Theorien der Bewegung des Elektrons sehr vorsichtig verfahren muß.

Wir bemerken, daß diese Resultate über die Masse auch für die ponderabeln materiellen Punkte gilt; denn ein ponderabler materieller Punkt kann durch Zufügen einer *beliebig kleinen* elektrischen Ladung zu einem Elektron (in unserem Sinne) gemacht werden.

Wir bestimmen die kinetische Energie des Elektrons. Bewegt sich ein Elektron vom Koordinatenursprung des Systems K aus mit der Anfangsgeschwindigkeit 0 beständig auf der

continued on next page

▶ **Figure 5.17** Some facsimile pages from Einstein's first article on the theory of relativity. The titles of all three protagonists' articles indicate that the theory of relativity arose from problems in the field of electrodynamics of moving bodies. Lorentz wrote his article in English (Figure 4.115): "Electromagnetic Phenomena in a System Moving with Any Velocity Smaller than That of Light"; Poincaré's was, of course, written in French: *Sur la dynamique de l'électron*; and finally, Einstein's work bore the title *Zur Elektrodynamik bewegter Körper*.

accomplished it." Indeed, Einstein must share the renown for establishing the special theory of relativity among notable predecessors and contemporaries, particularly with Lorentz and Poincaré. In working out the details and in the further development of the theory, Planck and Minkowski also made important contributions.

5.2.4 *The Measurement of Distance and Time*

Next we shall examine briefly the consequences of the relativity principle as clearly enunciated by both Poincaré and Einstein.

The universal principle of the theory of special relativity is contained in the following postulate: Physical laws are invariant with respect to the Lorentz transformations if we switch from one inertial system to another arbitrary inertial system. From this, it follows directly that the speed of light (in vacuum) has a constant value independent of the coordinate system.

How this fact inevitably leads to the overthrow of the classical notion of simultaneity is shown clearly in the following thought experiment, advanced by Einstein:

Imagine a train, as shown in Figure 5.20, moving with constant speed past an observer standing by the track. A second observer is located at the precise midpoint of the passing train. At exactly the time at which the observer in the train passes the observer standing by the track, both observers independently note the simultaneous flashes of two light pulses sent from the front and back of the train. What conclusions do the two observers draw with regard to the time at which the light pulses from the front and back of the train were sent? The observer on the train—taking into account that by the measurements taken with his ruler, he is standing in the middle of the train, and also that the velocity of light is independent of the state of motion—can conclude from the fact that the two pulses arrived simultaneously that they must have started at the same time, in other words, that the flashes must have been sent at the same time from the front and back of the train.

The observer standing alongside the track also knows that the speed of light is constant and that light signals require a finite amount of time to traverse a finite distance. From the simultaneity of the arrival of the two signals he concludes that the signal from the back end of the train must have been sent first because when the signals were sent, the back end of the train was farther from him than the front. The light pulse leaving the front of the train and propagating with constant speed can arrive at the same time as the pulse leaving the back of the train only if the former was emitted at a later time. The stationary observer thus concludes that the two flashes must have been sent at different times.

In the theory of relativity, we imagine that at every point in both coordinate systems we place a corresponding clock, which we can compare with the clock of the momentarily coinciding point of the other coordinate system. But we also synchronize the clocks in any given coordinate system so that they will run together. How can we do this? The simplest way is to exploit the constancy of the speed of light and send a light signal from the origin of the coordinate system at time $t = 0$. On receiving the signal, an observer standing at distance r from the coordinate origin sets his clock to time r/c, and assuming identical physical construction, the clocks will now run synchronously. The formulas of the Lorentz transformation are derived from the case in which the clocks in the coordinate systems K and K' are all synchronized with a light pulse that was emitted when the origins of the two coordinate systems coincided.

As we have seen, the most surprising consequence of the Lorentz transformation is that simultaneity is a relative notion. Directly related to this are the contraction

Figure 5.17 *continued*

As his starting point, Lorentz chooses the negative result of the Michelson experiment and emphasizes the ad hoc character of the solutions given thus far by himself and Fitzgerald. He cites the experiment by Trouton and Noble from the previous year (1903) and refers to an opinion expressed by Poincaré already in 1900:

> It would be more satisfactory if it were possible to show by means of certain fundamental assumptions and without neglecting terms of one order of magnitude or another, that many electromagnetic actions are entirely independent of the motion of the system. Some years ago, I already sought to frame a theory of this kind. I believe it is now possible to treat the subject with a better result. The only restriction as regards the velocity will be that it be less than that of light.
>
> I shall start from the fundamental equations of the theory of electrons. [pp. 12–13]

As is typical for him, Einstein begins by emphasizing the asymmetry of the Maxwell equations for the case of moving bodies. It is clear that only the relative motion of a magnet and a conductor determines the observable phenomena, but we still interpret differently what happens when the magnet moves and the conductor is stationary (here there will be an electric field and the current is induced) versus when the magnet is stationary and the conductor moves (there is no electric field but a force will act on the electrons in the conductor and the current will be the same as in the first case). Then Einstein mentions generally—without reference to concrete cases—the unsuccessful experimental efforts to determine the motion of Earth relative to the absolutely stationary "light medium." Then comes the critical insight; the conjecture that there is no distinguished inertial system is raised to the level of an axiom together with a second underlying assumption: The velocity of light is independent of the state of motion of the light source and has the same value for all inertial systems. The article continues in a radically new path: the definition of simultaneity followed by a derivation of the Lorentz transformation by interpreting what it means to measure time and length.

This is perhaps what excited physicists more than anything else. Even among the leading physicists, there were only a few who were truly comfortable with the electrodynamics of Maxwell and Lorentz, and even fewer were capable of recognizing the epoch-making significance of a transformation that leaves the Maxwell equations invariant. And now, with Einstein, all one had to do was pay careful attention to how clocks are synchronized, and one could not escape the conclusion that all of physics had to be reconsidered. After all, even a high school physics book starts with the discussion of how to measure time, distance, and mass.

The article further discusses the addition theorem of velocities, the transformation rules for electromagnetic field quantities, the Doppler effect, the theory of aberration, and the transformation rules for the energy of light rays; this is especially important because the later article on the equivalence of energy and mass continues from this point.

continued on next page

Figure 5.17 *continued*

What is *not* to be found in the article: Even though the dynamics of the electron are discussed, the establishment of the equations of motion had to wait for PLANCK. EINSTEIN still speaks of longitudinal and transversal mass, like LORENTZ. However, the formula

$$m_0 c^2 \left[\left(1 - v^2 / c^2\right)^{-1/2} - 1 \right]$$

for the kinetic energy is derived here.

The derivation of the formula for the change of mass,

$$m = m_0 \left(1 - v^2 / c^2\right)^{-1/2},$$

on the basis of mechanical considerations, that is, from the example of collision processes—observed in two inertial systems and on the assumption of conservation of energy and momentum, just as it was done by HUYGENS in the classical case—is the contribution of LEWIS and TOLMAN (1908).

We have mentioned in the main text a characteristic feature of EINSTEIN's article: Not a single name, not a single earlier publication, is mentioned, other than the name M. BESSO, a friend and colleague who is thanked for his assistance.

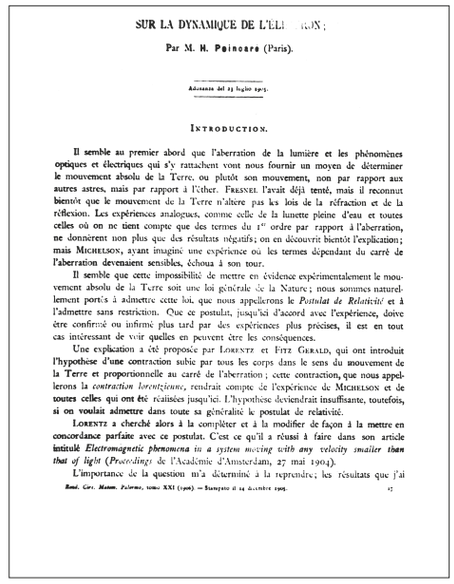
▶ **Figure 5.18** Some facsimile pages from POINCARÉ's article. It was submitted on July 28, 1905 and appeared in 1906 (*Rendiconti del Circolo Matematico di Palermo* 23, p. 129). In his monograph, PAULI writes:

> The formal gaps left by Lorentz's work were filled by Poincaré. He stated the relativity principle to be generally and rigorously valid. [1981, p. 3]

In addition to results already mentioned in the main text, we would emphasize the following conclusions of the article: The Lorentz transformation can be interpreted as a rotation in four-dimensional space *x*, *y*, *z*, *ict*; other quadruples of numbers could be formed that obey the same transformation rules as *x*, *y*, *z*, *t*. This yields a simple method for finding Lorentz-invariant quantities. IVANENKO refers emphatically to a fact that had essentially been forgotten: POINCARÉ was the first to apply the theory of relativity to the problem of gravitation; it was the first realistic and correct step in the direction of the modification of Newton's law of gravitation. POINCARÉ recognized the finite speed of the gravitational effect and attempted to find a relativistic correction to Newton's law.

The question is still raised from time to time as to why the discovery of the theory of relativity is ascribed *solely* to EINSTEIN. We have offered some explanations in the main text but here is another: EINSTEIN published in one of the leading journals of physics, while POINCARÉ, in contrast, wrote for mathematicians, so to speak, in a relatively obscure Italian journal. POINCARÉ's abstract style and especially the fact that he tended to undervalue his own work may also have played a role:

> The importance of this question induced me to investigate it recently; the results that I obtained agree on all major points with those of Lorentz; only in some isolated questions I feel compelled to modify or supplement them; in the future, their differences will come to light; they are of only secondary significance.



SUR LA DYNAMIQUE DE L'ÉLECTRON ;

Par M. H. Poincaré (Paris).

Adoptée dans la Séance du 23 Juillet 1905.

INTRODUCTION.

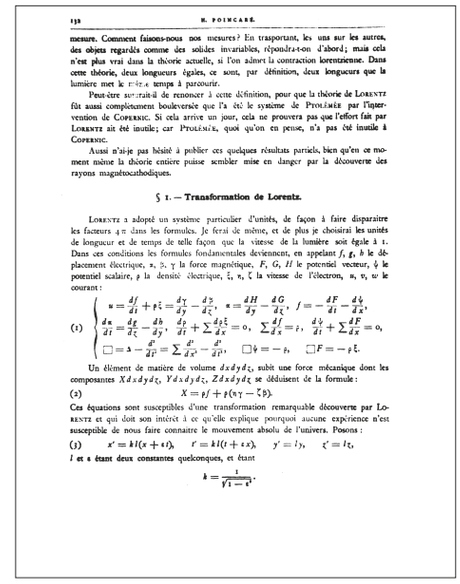
Il semble au premier abord que l'aberration de la lumière et les phénomènes optiques et électriques qui s'y rattachent vont nous fournir un moyen de déterminer le mouvement absolu de la Terre, ou plutôt son mouvement, non par rapport aux autres astres, mais par rapport à l'éther. FRESNEL l'avait déjà tenté, mais il reconnut bientôt que le mouvement de la Terre n'altère pas les lois de la réfraction et de la réflexion. Les expériences analogues, comme celle de la lunette pleine d'eau et toutes celles où on ne tient compte que des termes du 1er ordre par rapport à l'aberration, ne donnèrent non plus que des résultats négatifs; on en découvrit bientôt l'explication; mais MICHELSON, ayant imaginé une expérience où les termes dépendant du carré de l'aberration devenaient sensibles, échoua à son tour.

Il semble que cette impossibilité de mettre en évidence expérimentalement le mouvement absolu de la Terre soit une loi générale de la Nature; nous sommes naturellement portés à admettre cette loi, que nous appellerons le *Postulat de Relativité* et à l'admettre sans restriction. Quoi qu'en soit, jusqu'ici d'accord avec l'expérience, doive être confirmé ou infirmé plus tard par des expériences plus précises, il est en tout cas intéressant de voir quelles en peuvent être les conséquences.

Une explication a été proposée par LORENTZ et FITZ GERALD, qui ont introduit l'hypothèse d'une contraction subie par tous les corps dans le sens du mouvement de la Terre et proportionnelle au carré de l'aberration; cette contraction, que nous appellerons la *contraction lorentzienne*, rendrait compte de l'expérience de MICHELSON et de toutes celles qui ont été réalisées jusqu'ici. L'hypothèse deviendrait insuffisante, toutefois, si on voulait admettre dans toute sa généralité le postulat de relativité.

LORENTZ a cherché alors à la compléter et à la modifier de façon à la mettre en concordance parfaite avec ce postulat. C'est ce qu'il a réussi à faire dans son article intitulé *Electromagnetic phenomena in a system moving with any velocity smaller than that of light* (*Proceedings* de l'Académie d'Amsterdam, 27 mai 1904).

L'importance de la question m'a déterminé à la reprendre; les résultats que j'ai



mesure. Comment faisons-nous nos mesures? En transportant, les uns sur les autres, des objets regardés comme des solides invariables, répondra-t-on d'abord; mais cela n'est plus vrai dans la théorie actuelle, si l'on admet la contraction lorentzienne. Dans cette théorie, deux longueurs égales, ce sont, par définition, deux longueurs que la lumière met le même temps à parcourir.

Peut-être suffirait-il de renoncer à cette définition, pour que la théorie de LORENTZ fût aussi complètement bouleversée que l'a été le système de PTOLÉMÉE par l'intervention de COPERNIC. Si cela arrive un jour, cela ne prouvera pas que l'effort fait par LORENTZ ait été inutile; car PTOLÉMÉE, tout qu'on en pense, n'a pas été inutile à COPERNIC.

Aussi n'ai-je pas hésité à publier ces quelques résultats partiels, bien qu'en ce moment même la théorie entière puisse sembler mise en danger par la découverte des rayons magnétocathodiques.

§ 1. — Transformation de Lorentz.

LORENTZ a adopté un système particulier d'unités, de façon à faire disparaître les facteurs 4π dans les formules. Je ferai de même, et de plus je choisirai les unités de longueur et de temps de telle façon que la vitesse de la lumière soit égale à 1. Dans ces conditions les formules fondamentales deviennent, en appelant f, g, h le déplacement électrique, α, β, γ la force magnétique, F, G, H le potentiel vecteur, ψ le potentiel scalaire, ρ la densité électrique, ξ, η, ζ la vitesse de l'électron, u, v, w le courant :

$$(1)\quad \begin{cases} u = \dfrac{df}{dt} + \rho\xi = \dfrac{d\gamma}{dy} - \dfrac{d\beta}{dz}, & \alpha = \dfrac{dH}{dy} - \dfrac{dG}{dz}, \quad f = -\dfrac{dF}{dt} - \dfrac{d\psi}{dx}, \\ \dfrac{dg}{dt} = \dfrac{d\alpha}{dz} - \dfrac{dh}{dx}, & \dfrac{dp}{dt} + \sum \dfrac{df}{dx} = 0, \quad \sum \dfrac{df}{dx} = \rho, \quad \dfrac{d\psi}{dt} + \sum \dfrac{dF}{dx} = 0, \\ \square = \Delta - \dfrac{d^2}{dt^2} = \sum \dfrac{d^2}{dx^2} - \dfrac{d^2}{dt^2}, & \square\psi = -\rho, \quad \square F = -\rho\xi. \end{cases}$$

Un élément de matière de volume $dx\,dy\,dz$, subit une force mécanique dont les composantes $X\,dx\,dy\,dz$, $Y\,dx\,dy\,dz$, $Z\,dx\,dy\,dz$ se déduisent de la formule :

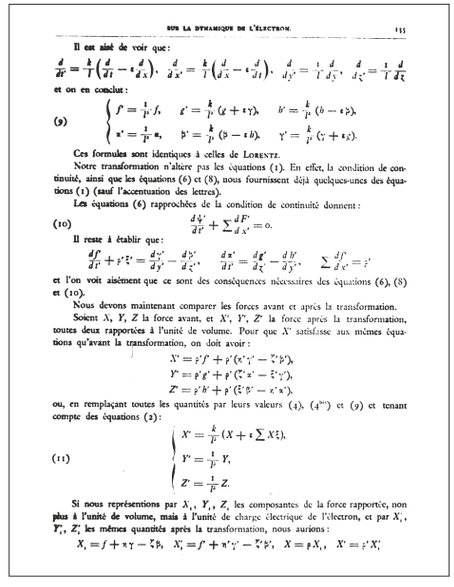
$$(2)\qquad X = \rho f + \rho(u\gamma - v\beta).$$

Ces équations sont susceptibles d'une transformation remarquable découverte par LORENTZ et qui doit son intérêt à ce qu'elle explique pourquoi aucune expérience n'est susceptible de nous faire connaître le mouvement absolu de l'univers. Posons :

$$(3)\qquad x' = kl(x + \varepsilon t), \qquad t' = kl(t + \varepsilon x), \qquad y' = ly, \qquad z' = lz,$$

l et ε étant deux constantes quelconques, et étant

$$k = \frac{1}{\sqrt{1 - \varepsilon^2}}.$$



Il est aisé de voir que :

$$\frac{d}{dt'} = \frac{1}{l}\left(\frac{d}{dt} - \varepsilon\frac{d}{dx}\right), \quad \frac{d}{dx'} = \frac{k}{l}\left(\frac{d}{dx} - \varepsilon\frac{d}{dt}\right), \quad \frac{d}{dy'} = \frac{1}{l}\frac{d}{dy}, \quad \frac{d}{dz'} = \frac{1}{l}\frac{d}{dz}$$

et on en conclut :

$$(5)\qquad \begin{cases} f' = \dfrac{1}{k^2}f, & g' = \dfrac{k}{l}(g + \varepsilon\gamma), & h' = \dfrac{k}{l}(h - \varepsilon\beta), \\ \alpha' = \dfrac{1}{k^2}\alpha, & \beta' = \dfrac{k}{l}(\beta - \varepsilon h), & \gamma' = \dfrac{k}{l}(\gamma + \varepsilon g). \end{cases}$$

Ces formules sont identiques à celles de LORENTZ.

Notre transformation n'altère pas les équations (1). En effet, la condition de continuité, ainsi que les équations (6) et (8), nous fournissent déjà quelques-unes des équations (1) (sauf l'accentuation des lettres).

Les équations (6) rapprochées de la condition de continuité donnent :

$$(10)\qquad \frac{d\psi'}{dt'} + \sum \frac{dF'}{dx'} = 0.$$

Il reste à établir que :

$$\frac{df'}{dt'} + \rho'\xi' = \frac{d\gamma'}{dy'} - \frac{d\beta'}{dz'}, \qquad \frac{dx'}{dt'} = \frac{dg'}{dz'} - \frac{dh'}{dy'}, \qquad \sum \frac{df'}{dx'} = \rho'$$

et l'on voit aisément que ce sont des conséquences nécessaires des équations (6), (8) et (10).

Nous devons maintenant comparer les forces avant et après la transformation.

Soient X, Y, Z la force avant, et X', Y', Z' la force après la transformation, toutes deux rapportées à l'unité de volume. Pour que X' satisfasse aux mêmes équations qu'avant la transformation, on doit avoir :

$$X' = f'f + \rho'(\eta'\gamma' - \zeta'\beta'),$$
$$Y' = g'g + \rho'(\zeta'\alpha' - \xi'\gamma'),$$
$$Z' = f'h' + \rho'(\xi'\beta' - \eta'\alpha').$$

ou, en remplaçant toutes les quantités par leurs valeurs (4), (4bis) et (9) et tenant compte des équations (2) :

$$(11)\qquad \begin{cases} X' = \dfrac{k}{l^2}(X + \varepsilon\sum X\xi), \\ Y' = \dfrac{1}{l^2}Y, \\ Z' = \dfrac{1}{l^2}Z. \end{cases}$$

Si nous représentons par X_1, Y_1, Z_1 les composantes de la force rapportée, non plus à l'unité de volume, mais à l'unité de charge électrique de l'électron, et par X_1', Y_1', Z_1' les mêmes quantités après la transformation, nous aurions :

$$X_1 = f + \eta\gamma - \zeta\beta, \quad X_1' = f' + \eta'\gamma' - \zeta'\beta', \quad X = \rho X_1, \quad X' = \rho' X_1'$$



Si alors nous posons :

$$\square' = \sum \frac{d^2}{dx'^2} - \frac{d^2}{dt'^2},$$

il viendra :

$$\square' = \square l^{-2}.$$

Considérons une sphère entraînée avec l'électron dans un mouvement de translation uniforme et soit :

$$(x - \xi t)^2 + (y - \eta t)^2 + (z - \zeta t)^2 = r^2$$

l'équation de cette sphère mobile dont le volume sera $\frac{4}{3}\pi r^3$.

La transformation la changera en un ellipsoïde, dont il est aisé de trouver l'équation. On déduit aisément en effet des équations (3) :

$$(3^{bis})\quad x = \frac{k}{l}(x' - \varepsilon t'), \qquad t = \frac{k}{l}(t' - \varepsilon x'), \qquad y = \frac{y'}{l}, \qquad z = \frac{z'}{l}.$$

L'équation de l'ellipsoïde devient ainsi :

$$k^2(x' - \varepsilon t' - \xi t' + \varepsilon\xi x')^2 + (y' - \eta k t' + \eta\varepsilon k x')^2 + (z' - \zeta k t' + \zeta\varepsilon k x')^2 = l^2 r^2.$$

Cet ellipsoïde se déplace avec un mouvement uniforme; pour $t' = 0$, il se réduit à

$$k^2 x'^2(1 + \xi\varepsilon)^2 + (y' + \eta k\varepsilon x')^2 + (z' + \zeta k\varepsilon x')^2 = l^2 r^2$$

et a pour volume :

$$\frac{4}{3}\pi r^3 \frac{l^3}{k(1 + \xi\varepsilon)}.$$

Si l'on veut que la charge d'un électron ne soit pas altérée par la transformation et si l'on appelle ρ' la nouvelle densité électrique, il viendra :

$$(4)\qquad \rho' = \frac{k}{l^3}(\rho + \varepsilon\rho\xi).$$

Que seront maintenant les nouvelles vitesses ξ', η', ζ'; on devra avoir :

$$\xi' = \frac{dx'}{dt'} = \frac{d(x + \varepsilon t)}{d(t + \varepsilon x)} = \frac{\xi + \varepsilon}{1 + \varepsilon\xi},$$

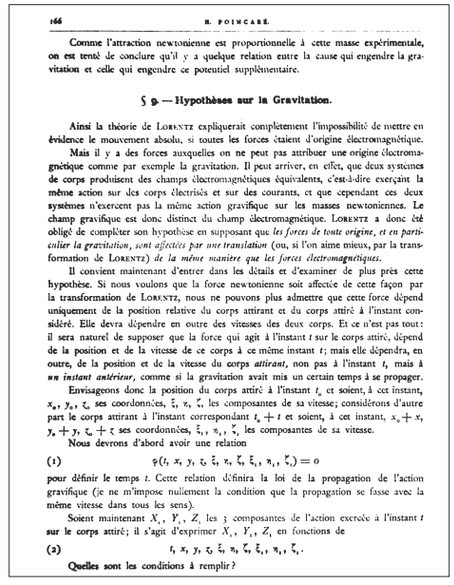
$$\eta' = \frac{dy'}{dt'} = \frac{dy}{k\,d(t + \varepsilon x)} = \frac{\eta}{k(1 + \varepsilon\xi)}, \qquad \zeta' = \frac{\zeta}{k(1 + \varepsilon\xi)},$$

d'où :

$$(4^{bis})\qquad \rho'\xi' = \frac{k}{l^3}(\rho\xi + \varepsilon\rho), \qquad \rho'\eta' = \frac{1}{l^3}\rho\eta, \qquad \rho'\zeta' = \frac{1}{l^3}\rho\zeta.$$

C'est ici que je dois signaler pour la première fois une divergence avec LORENTZ. LORENTZ pose (à la différence des notations près) (loco citato, page 813, formules 7 et 8) :

$$\rho' = \frac{1}{k^2 l^3}\rho, \qquad \xi' = k^2(\xi + \varepsilon), \qquad \eta' = k\eta, \qquad \zeta' = k\zeta.$$



Comme l'attraction newtonienne est proportionnelle à cette masse expérimentale, on est tenté de conclure qu'il y a quelque relation entre la cause qui engendre la gravitation et celle qui engendre ce potentiel supplémentaire.

§ 9. — Hypothèses sur la Gravitation.

Ainsi la théorie de LORENTZ expliquerait complètement l'impossibilité de mettre en évidence le mouvement absolu, si toutes les forces étaient d'origine électromagnétique.

Mais il y a des forces auxquelles on ne peut pas attribuer une origine électromagnétique comme par exemple la gravitation. Il peut arriver, en effet, que deux systèmes de corps produisent des champs électromagnétiques équivalents, c'est-à-dire exerçant la même action sur des corps électrisés et sur des courants, et que cependant ces deux systèmes n'exercent pas la même action gravifique sur les masses newtoniennes. Le champ gravifique est donc distinct du champ électromagnétique. LORENTZ a donc été obligé de compléter son hypothèse en supposant que *les forces de toute origine, et en particulier la gravitation, sont affectées par une translation* (ou, si l'on aime mieux, par la transformation de LORENTZ) *de la même manière que les forces électromagnétiques*.

Il convient maintenant d'entrer dans les détails et d'examiner de plus près cette hypothèse. Si nous voulons que la force newtonienne soit affectée de cette façon par la transformation de LORENTZ, nous ne pouvons plus admettre que cette force dépend uniquement de la position relative du corps attirant et du corps attiré à l'instant considéré. Elle devra dépendre en outre des vitesses des deux corps. Et ce n'est pas tout : il sera naturel de supposer que la force qui agit à l'instant t sur le corps attiré, dépend de la position et de la vitesse de ce corps à ce même instant t; mais elle dépendra, en outre, de la position et de la vitesse du corps *attirant*, non pas à l'instant t, mais à *un instant antérieur*, comme si la gravitation avait mis un certain temps à se propager.

Envisageons donc la position du corps attiré à l'instant t_0 et soient, à cet instant, x_0, y_0, z_0 ses coordonnées, ξ, η, ζ, les composantes de sa vitesse; considérons d'autre part le corps attirant à l'instant correspondant $t_0 + t$ et soient, à cet instant, $x_0 + x$, $y_0 + y$, $z_0 + z$ ses coordonnées, ξ_1, η_1, ζ_1 les composantes de sa vitesse.

Nous devrons d'abord avoir une relation

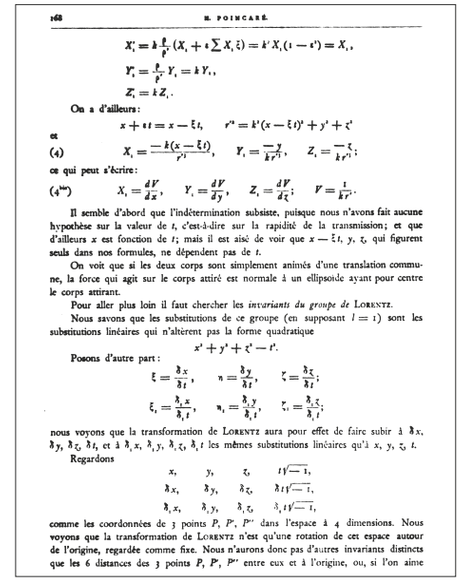
$$(1)\qquad \varphi(t, x, y, z, \xi, \eta, \zeta, \xi_1, \eta_1, \zeta_1) = 0$$

pour définir le temps t. Cette relation définira la loi de la propagation de l'action gravifique (je ne m'impose nullement la condition que la propagation se fasse avec la même vitesse dans tous les sens).

Soient maintenant X_1, Y_1, Z_1 les 3 composantes de l'action exercée à l'instant t sur le corps attiré; il s'agit d'exprimer X_1, Y_1, Z_1 en fonctions de

$$(2)\qquad t, x, y, z, \xi, \eta, \zeta, \xi_1, \eta_1, \zeta_1.$$

Quelles sont les conditions à remplir?



$$X_1' = k\frac{\rho}{\rho'}(X_1 + \varepsilon\sum X_1\xi) = k'X_1(1 - \varepsilon^2) = X_1,$$

$$Y_1' = \frac{\rho}{\rho'}Y_1 = kY_1,$$

$$Z_1' = kZ_1.$$

On a d'ailleurs :

$$x + \varepsilon t = x - \xi t, \qquad r'^2 = k^2(x - \xi t)^2 + y^2 + z^2$$

et

$$(4)\qquad X_1 = \frac{-k(x - \xi t)}{r^3}, \qquad Y_1 = \frac{-y}{kr^3}, \qquad Z_1 = \frac{-z}{kr^3};$$

ce qui peut s'écrire :

$$(4^{bis})\qquad X_1 = \frac{dV}{dx}, \qquad Y_1 = \frac{dV}{dy}, \qquad Z_1 = \frac{dV}{dz}; \qquad V = \frac{1}{kr}.$$

Il semble d'abord que l'indétermination subsiste, puisque nous n'avons fait aucune hypothèse sur la valeur de t, c'est-à-dire sur la rapidité de la transmission; et que d'ailleurs x est fonction de t; mais il est aisé de voir que $x - \xi t$, y, z, qui figurent seuls dans nos formules, ne dépendent pas de t.

On voit que si les deux corps sont simplement animés d'une translation commune, la force qui agit sur le corps attiré est normale à un ellipsoïde ayant pour centre le corps attirant.

Pour aller plus loin il faut chercher les *invariants du groupe de* LORENTZ.

Nous savons que les substitutions de ce groupe (en supposant $l = 1$) sont les substitutions linéaires qui n'altèrent pas la forme quadratique

$$x^2 + y^2 + z^2 - t^2.$$

Posons d'autre part :

$$\xi = \frac{\delta x}{\delta t}, \qquad \eta = \frac{\delta y}{\delta t}, \qquad \zeta = \frac{\delta z}{\delta t};$$

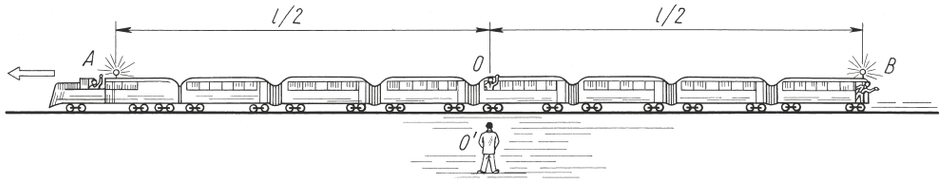
$$\xi_1 = \frac{\delta_1 x}{\delta_1 t}, \qquad \eta_1 = \frac{\delta_1 y}{\delta_1 t}, \qquad \zeta_1 = \frac{\delta_1 z}{\delta_1 t};$$

nous voyons que la transformation de LORENTZ aura pour effet de faire subir à δx, δy, δz, δt, et à $\delta_1 x$, $\delta_1 y$, $\delta_1 z$, $\delta_1 t$ les mêmes substitutions linéaires qu'à x, y, z, t.

Regardons

$$x, \quad y, \quad z, \quad t\sqrt{-1},$$
$$\delta x, \quad \delta y, \quad \delta z, \quad \delta t\sqrt{-1},$$
$$\delta_1 x, \quad \delta_1 y, \quad \delta_1 z, \quad \delta_1 t\sqrt{-1},$$

comme les coordonnées de 3 points P, P', P'' dans l'espace à 4 dimensions. Nous voyons que la transformation de LORENTZ n'est qu'une rotation de cet espace autour de l'origine, regardée comme fixe. Nous n'aurons donc pas d'autres invariants distincts que les 6 distances des 3 points P, P', P'' entre eux et à l'origine, ou, si l'on aime



▲ **Figure 5.20** Flashes of light from lamps at the front and rear of the train arrive at the observer in the middle of the moving train and the observer on the platform simultaneously. The observer on the train concludes from this that both flashes of light were emitted at the same time, whereas according to the observer on the platform, the lamp at the end of the train must have flashed first.

of lengths and the dilation of time. The former means that a distance measured in a moving frame of reference by an observer moving along with the frame, if measured from a stationary frame, will be shorter. At first glance, this appears paradoxical because we can use the terms "moving" and "stationary" only in the relative sense, so this statement also means that every observer will see shorter distances in other systems moving relative to his own.

Consider a rod on the x'-axis of the coordinate system K' with one end at the origin and the other end at $x_2' = l_0$. Therefore, its length measured in the coordinate system K' is l_0. If a stationary observer in the coordinate system K wishes to measure the length of the rod, he must, of course, be careful to determine the location of the beginning and end points at the same time. This self-evident requirement and the fact that the notion of simultaneity is dependent on the coordinate system leads us immediately to the above-mentioned length contraction. It suffices at time $t = 0$ to determine the value of the x-coordinate associated with $x_2' = l_0$ in the system K because at this time, the origins of the two coordinate systems coincide. The time t' in system K' associated according to equation (2) with time $t = 0$ in system K yields $t' = (-v/c^2)x'$. From this, we obtain

$$x = \kappa\left(x' - \frac{v^2}{c^2}x'\right) = \sqrt{1-\frac{v^2}{c^2}}x', \quad \kappa = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}.$$

The length of the rod measured in coordinate system K thus turns out to be

$$l = \sqrt{1-\frac{v^2}{c^2}}l_0.$$

We see from this formula that from the point of view of a stationary coordinate system, the length of a moving rod is reduced by a factor of $\sqrt{1-v^2/c^2} < 1$.

Even if we interchange the roles of the two coordinate systems, the same argument leads us to the identical conclusion.

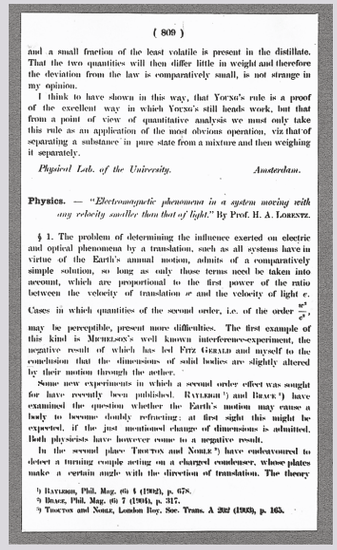
If we now measure in coordinate system K', at a given location, let's say at $x' = 0$, the time T_0 in which a particular process occurs or the time interval between two events, then from the coordinates $x' = 0$ and $t' = T_0$ in coordinate system K at location x, we obtain a time $t = T$ given by

$$T = \kappa\left(T_0 + \frac{v}{c^2}x'\right) = \kappa T_0 > T_0.$$

If we combine the above results, we conclude that a rod in a moving system measured from a stationary system shrinks and that moving clocks go more slowly (Figures 5.21, 5.22).

The Lorentz transformation can be interpreted quite clearly in geometrical terms. Let us call one axis of a planar coordinate system the x-axis, and the other axis the ct-axis. Then a point in this coordinate system fixes the location and time in the frame of reference K. The quantities x', ct', which are associated with coordinates x, ct through the Lorentz transformation, should correspond in system K' to the coordinates x, ct. It is easy to see that the image of the x'-axis ($ct' = 0$) in system K is given by a line that forms an angle $\varphi = \arctan\frac{v}{c}$ with the x-axis. Likewise, the ct' axis ($x' = 0$) forms an angle j with the ct-axis (Figure 5.23). In fact, the points $t' = 0$ (x' arbitrary) satisfy, because

$$x = (x' + vt')\kappa \quad \text{and} \quad t = \left(\frac{v}{c^2}x' + t\right)\kappa,$$



(809)

and a small fraction of the least volatile is present in the distillate. That the two quantities will then differ little in weight and therefore the deviation from the law is comparatively small, is not strange in my opinion.

I think to have shown in this way, that VOCKO's rule is a proof of the excellent way in which VOCKO's still heads work, but that from a point of view of quantitative analysis we must only take this rule as an application of the most obvious operation, viz that of separating a substance in pure state from a mixture and then weighing it separately.

Physical Lab. of the University. Amsterdam.

Physics. — *"Electromagnetic phenomena in a system moving with any velocity smaller than that of light."* By Prof. H. A. LORENTZ.

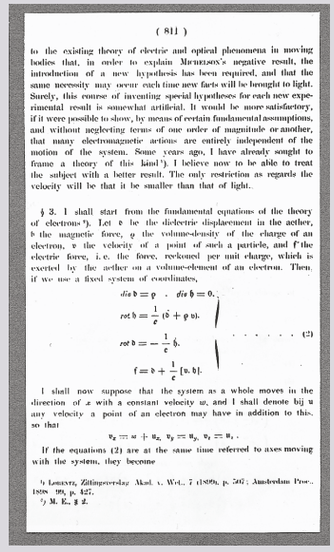
§ 1. The problem of determining the influence exerted on electric and optical phenomena by a translation, such as all systems have in virtue of the Earth's annual motion, admits of a comparatively simple solution, so long as only those terms need be taken into account, which are proportional to the first power of the ratio between the velocity of translation w and the velocity of light c.

Cases in which quantities of the second order, i.e. of the order $\frac{w^2}{c^2}$, may be perceptible, present more difficulties. The first example of this kind is MICHELSON's well known interference-experiment, the negative result of which has led FITZ GERALD and myself to the conclusion that the dimensions of solid bodies are slightly altered by their motion through the aether.

Some new experiments in which a second order effect was sought for have recently been published. RAYLEIGH [1]) and BRACE [2]) have examined the question whether the Earth's motion may cause a body to become doubly refracting; at first sight this might be expected, if the just mentioned change of dimensions is admitted. Both physicists have however come to a negative result.

In the second place TROUTON and NOBLE [3]) have endeavoured to detect a turning couple acting on a charged condenser, whose plates make a certain angle with the direction of translation. The theory

[1]) RAYLEIGH, Phil. Mag. (6) 4 (1902), p. 678.
[2]) BRACE, Phil. Mag. (6) 7 (1904), p. 317.
[3]) TROUTON and NOBLE, London Roy. Soc. Trans. A 202 (1903), p. 165.



(811)

to the existing theory of electric and optical phenomena in moving bodies that, in order to explain MICHELSON's negative result, the introduction of a new hypothesis has been required, and that the same necessity may occur each time new facts will be brought to light. Surely, this course of inventing special hypotheses for each new experimental result is somewhat artificial. It would be more satisfactory, if it were possible to show, by means of certain fundamental assumptions, and without neglecting terms of one order of magnitude or another, that many electromagnetic actions are entirely independent of the motion of the system. Some years ago, I have already sought to frame a theory of this kind [1]). I believe now to be able to treat the subject with a better result. The only restriction as regards the velocity will be that it be smaller than that of light.

§ 3. I shall start from the fundamental equations of the theory of electrons [2]). Let $\mathfrak{d}$ be the dielectric displacement in the aether, $\mathfrak{h}$ the magnetic force, ϱ the volume-density of the charge of an electron, $\mathfrak{v}$ the velocity of a point of such a particle, and $\mathfrak{f}$ the electric force, i.e. the force, reckoned per unit charge, which is exerted by the aether on a volume-element of an electron. Then, if we use a fixed system of coordinates,

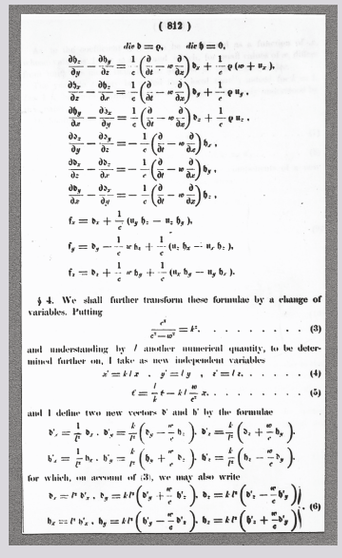
$$\left.\begin{aligned} &div\,\mathfrak{d} = \varrho \quad . \quad div\,\mathfrak{h} = 0, \\ &rot\,\mathfrak{h} = \frac{1}{c}(\mathfrak{d}' + \varrho\,\mathfrak{v}), \\ &rot\,\mathfrak{d} = -\frac{1}{c}\mathfrak{h}', \\ &\mathfrak{f} = \mathfrak{d} + \frac{1}{c}[\mathfrak{v}.\mathfrak{h}]. \end{aligned}\right\} \quad \ldots\ldots\ldots (2)$$

I shall now suppose that the system as a whole moves in the direction of x with a constant velocity w, and I shall denote by u any velocity a point of an electron may have in addition to this, so that

$$v_x = w + u_x, \; v_y = u_y, \; v_z = u_z.$$

If the equations (2) are at the same time referred to axes moving with the system, they become

[1]) LORENTZ, Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegten Körpern, Leiden 1895; Zittingsverslag Akad. v. Wet., 7 (1899), p. 507; Amsterdam Proc., 1899, p. 427.
[2]) M. E., § 2.



(812)

$$div\,\mathfrak{b} = \varrho, \qquad div\,\mathfrak{h} = 0,$$

$$\frac{\partial \mathfrak{d}_z}{\partial y} - \frac{\partial \mathfrak{d}_y}{\partial z} = \frac{1}{c}\left(\frac{\partial}{\partial t} - w\frac{\partial}{\partial x}\right)\mathfrak{d}_x + \frac{1}{c}\varrho\,(w + u_x),$$

$$\frac{\partial \mathfrak{d}_x}{\partial z} - \frac{\partial \mathfrak{d}_z}{\partial x} = \frac{1}{c}\left(\frac{\partial}{\partial t} - w\frac{\partial}{\partial x}\right)\mathfrak{d}_y + \frac{1}{c}\varrho\,u_y,$$

$$\frac{\partial \mathfrak{d}_y}{\partial x} - \frac{\partial \mathfrak{d}_x}{\partial y} = \frac{1}{c}\left(\frac{\partial}{\partial t} - w\frac{\partial}{\partial x}\right)\mathfrak{d}_z + \frac{1}{c}\varrho\,u_z,$$

$$\frac{\partial \mathfrak{h}_x}{\partial y} - \frac{\partial \mathfrak{h}_y}{\partial z} = -\frac{1}{c}\left(\frac{\partial}{\partial t} - w\frac{\partial}{\partial x}\right)\mathfrak{d}_x,$$

$$\frac{\partial \mathfrak{h}_y}{\partial z} - \frac{\partial \mathfrak{h}_z}{\partial x} = -\frac{1}{c}\left(\frac{\partial}{\partial t} - w\frac{\partial}{\partial x}\right)\mathfrak{d}_y,$$

$$\frac{\partial \mathfrak{h}_z}{\partial x} - \frac{\partial \mathfrak{h}_x}{\partial y} = -\frac{1}{c}\left(\frac{\partial}{\partial t} - w\frac{\partial}{\partial x}\right)\mathfrak{d}_z,$$

$$\mathfrak{f}_x = \mathfrak{d}_x + \frac{1}{c}(u_y\,\mathfrak{b}_z - u_z\,\mathfrak{b}_y),$$

$$\mathfrak{f}_y = \mathfrak{d}_y - \frac{1}{c}w\,\mathfrak{b}_z + \frac{1}{c}(u_z\,\mathfrak{b}_x - u_x\,\mathfrak{b}_z),$$

$$\mathfrak{f}_z = \mathfrak{d}_z + \frac{1}{c}w\,\mathfrak{b}_y + \frac{1}{c}(u_x\,\mathfrak{b}_y - u_y\,\mathfrak{b}_x).$$

§ 4. We shall further transform these formulae by a change of variables. Putting

$$\frac{c^2}{c^2 - w^2} = l^2. \quad \ldots\ldots\ldots (3)$$

and understanding by l another numerical quantity, to be determined further on, I take as new independent variables

$$x' = l\,x \;, \quad y' = l\,y \;, \quad t' = l\,\epsilon. \quad \ldots\ldots\ldots (4)$$

$$\epsilon = \frac{l}{c}t - l\frac{w}{c^2}x. \quad \ldots\ldots\ldots (5)$$

and I define two new vectors $\mathfrak{d}'$ and $\mathfrak{h}'$ by the formulae

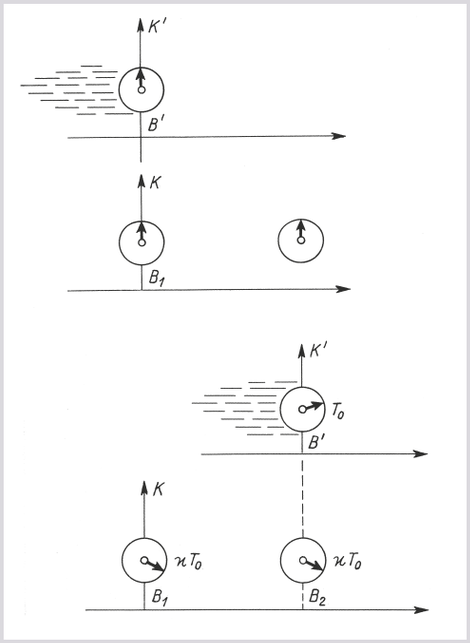
$$\mathfrak{d}_x' = \frac{1}{l^2}\mathfrak{d}_x \;, \quad \mathfrak{d}_y' = \frac{k}{l^2}\left(\mathfrak{d}_y - \frac{w}{c}\mathfrak{h}_z\right), \quad \mathfrak{d}_z' = \frac{k}{l^2}\left(\mathfrak{d}_z + \frac{w}{c}\mathfrak{h}_y\right),$$

$$\mathfrak{h}_x' = \frac{1}{l^2}\mathfrak{h}_x \;, \quad \mathfrak{h}_y' = \frac{k}{l^2}\left(\mathfrak{h}_y + \frac{w}{c}\mathfrak{d}_z\right), \quad \mathfrak{h}_z' = \frac{k}{l^2}\left(\mathfrak{h}_z - \frac{w}{c}\mathfrak{d}_y\right).$$

for which, on account of (3), we may also write

$$\left.\begin{aligned} &\mathfrak{d}_x = l^2\,\mathfrak{d}_x' \;, \quad \mathfrak{d}_y = k l^2\left(\mathfrak{d}_y' + \frac{w}{c}\mathfrak{h}_z'\right), \quad \mathfrak{d}_z = k l^2\left(\mathfrak{d}_z' - \frac{w}{c}\mathfrak{h}_y'\right), \\ &\mathfrak{h}_x = l^2\,\mathfrak{h}_x' \;, \quad \mathfrak{h}_y = k l^2\left(\mathfrak{h}_y' - \frac{w}{c}\mathfrak{d}_z'\right), \quad \mathfrak{h}_z = k l^2\left(\mathfrak{h}_z' + \frac{w}{c}\mathfrak{d}_y'\right). \end{aligned}\right\} \quad (6)$$

◀ ▲ ▼
Figure 5.19 Three pages from LORENTZ's article on the theory of relativity. The fundamental equations of electron theory serve as the starting point for his treatment. They can be seen on the second of the three pages displayed here.



▲ **Figure 5.21** Seen from a system at rest, clocks in the moving system run slower.

It is easy to draw the wrong conclusion from an explanatory picture like this one. It might seem that observers B' and B_1 mutually read off the respective times $t' = 0$ and $t = 0$; then observers B' and B_2 read off the times $t' = T_0$ and $t = \kappa T_0$, and since $\kappa T_0 > T_0$, *both* observers would conclude that the clocks in K' are slow. The error in this train of thought is the following: If we put ourselves in the place of observer B', we read the clocks of observers B_1 and B_2 as these clocks race past us, as opposed to reading one moving clock in the K coordinate system and comparing this single clock with the synchronized clocks in the K' systems that have been placed at different locations. We must keep in mind that simultaneity is a notion that is relative to the coordinate system. Let us consider the clock of observer B_2, which is moving relative to us (observer B'): When our clock displays the time T_0, the hands of this clock indicate κT_0; but at the time when the hands of all the synchronized clocks on the x'-axis showed the time $t' = 0$, this clock stood *not* at $t = 0$, but at $\tau = \kappa(v^2/c^2)T_0$; namely, the transformation formulas give the clock at point $x = v\kappa T_0$ at time $t' = 0$ the value $\tau = \kappa(v^2/c^2)T_0$. Thus, according to our observation (that is, on the clock of observer B'), the time interval is $0 \rightarrow T$; on the clock of observer B_2 that is moving relative to us, this interval will be $\kappa(v^2/c^2)T_0 \rightarrow \kappa T_0$. Therefore, to the time interval T_0 measured in K' is associated an interval in K of length

$$\kappa T_0 - \kappa\frac{v^2}{c^2}T_0 = \kappa\left(1-\frac{v^2}{c^2}\right)T_0 = \kappa\frac{1}{\kappa^2}T_0 = \frac{T_0}{\kappa}.$$

Since $T_0/\kappa < T_0$, observer B' can also maintain that the clocks moving relative to him are slow. These relationships are pictured geometrically in Figure 5.25.

Instead of the rate of advance of a moving clock, we could well be looking at the rate of the observer's heartbeat, or the average rate of decay of a radioactive atom. All of these would be measured as slower by a stationary observer.

the equation

$$ct = \frac{v}{c}x.$$

The points satisfying the condition $x' = 0$ lie in the coordinate system K on the line $x = \frac{v}{c}(ct)$, where again the validity of the LORENTZ transformation is assumed. We obtain the coordinates of an arbitrary point in the coordinate system K' from the axis intercepts of the lines parallel to the coordinate axes x' and ct' and intersecting at the desired point. Nevertheless, we must make sure that the unit lengths are correctly entered on the coordinate axes. First, it can be shown by simple substitution that the difference

$$x^2 - (ct)^2 = x'^2 - (ct')^2$$

is invariant with respect to the Lorentz transformation. However, this means that this difference always has the same value for related coordinates x, t and x', t'.

We see at once that the hyperbola

$$x^2 - (ct)^2 = 1$$

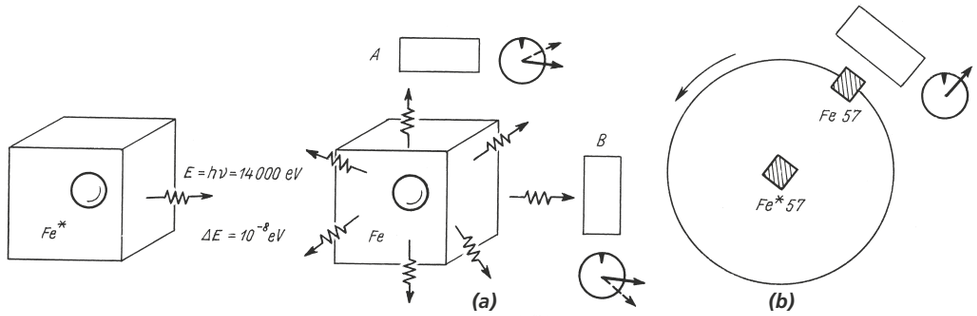
intersects the x-axis at the unit mark. The associated coordinates (x', ct') also satisfy this equation of a hyperbola, so that in particular for $t' = 0$, we obtain the value $x' = 1$; that is, in system K' as well, this hyperbola fixes the unit length. Analogously, for $x = 0$, the hyperbola

$$(ct)^2 - x^2 = (ct')^2 - x'^2 = 1$$

gives the value $ct = 1$; and for $x' = 0$, the value $ct' = 1$. Therefore, with this hyperbola the unit length can be fixed on the time axis (Figure 5.24).

Based on these ideas, we can provide each coordinate system with the appropriate coordinate grid and read off the corresponding space and time points (Figure 5.25).

If we plot the position of a point at different times, we get the *world line* for that point. The lines of the coordinate grid parallel to the t-axis are the world lines of the points that are at rest on the x-axis at unit distance from one another. The line leaving the origin of the coordinate system at an angle of 45° describes the world line of



▲ **Figure 5.22** Demonstration of time dilation using the Mössbauer effect: The isotope Co-57 changes via beta decay with a half-life of 280 days into the excited state of the isotope Fe*-57, which then returns to the ground state with a half-life of 10^{-7} seconds through the emission of a gamma quantum with energy 14,000 eV with an extremely narrow spectrum (line width of only 10^{-8} eV). If the nucleus were in a free atom, then on emission of the photon, the atom would rebound and would receive a part of the energy, resulting in a broadening of the spectral line (10^{-3} eV). If, on the other hand, the nucleus is part of a crystalline lattice, then the lattice effectively serves as a support for the nucleus, and the sharpness of the emission line is preserved. A Fe-57 nucleus in the ground state can absorb the gamma quanta emitted from the Fe*-57 nuclei due to a resonance effect, and the absorber will then re-radiate the absorbed energy equally in all spatial directions. In this case, shown in diagram **(a)**, the counter *A* located perpendicular to the primary gamma rays entering from the left will show a strong signal. If for some reason the conditions for resonance absorption were not satisfied, then the signal from counter *A* would be smaller because of the reduced absorption and reemission, while counter *B* would register more gamma quanta for the same reason as the scattering due to the absorption would be less (shown by the dashed meter positions). With the arrangement shown in diagram **(b)**, the Mössbauer effect can be used to demonstrate the time dilation between two systems moving relatively to each other. The nuclei in the ground state here move relative to the excited nuclei, and due to the different time scales, the frequency of the gamma quantum changes, which will prevent the resonance absorption and create the conditions that show the effect.

a flash of light. Because the speed of light is the greatest speed achievable, possible motions are between being at rest and this line; that is, the slope of their world lines must be greater than 45°. In the coordinate system K, we have the relationship

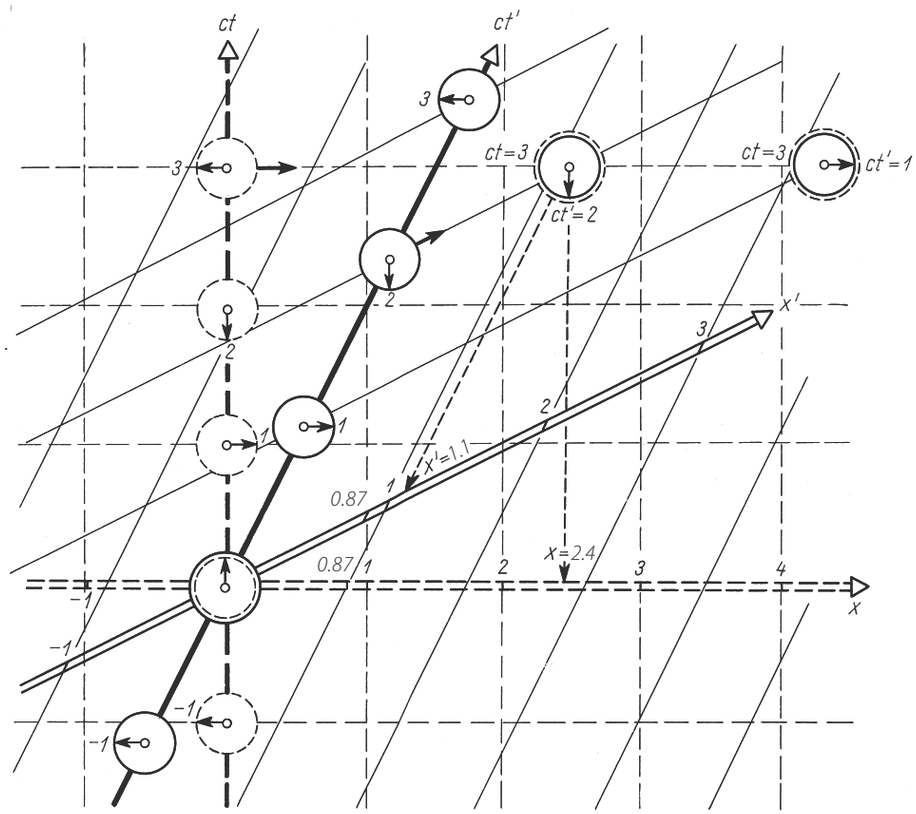
$$\tan\varphi = \frac{v}{c}$$

between the slope of the world line and the speed of the point, where φ is the angle enclosed by the world line and the ct-axis.

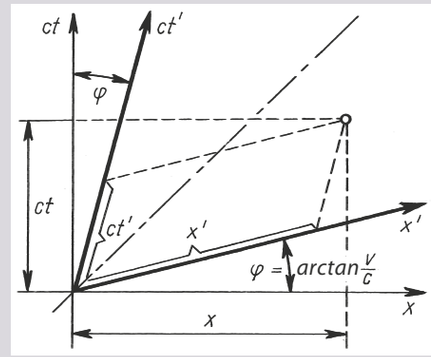
With the help of Figure 5.26, we can clearly describe all the connections discussed above. Thus, for example, we consider events in the coordinate system K to be simultaneous if the associated points lie on a line parallel to the x-axis. We see at once that these events are not simultaneous in the coordinate system K' because they are associated with different values of t'. Similarly, length contraction and time dilation are vividly illustrated as well as their relative nature, specifically in that an observer in either coordinate system would see a unit-length rod in the other system equally shortened.

5.2.5 *The Equivalence of Energy and Mass*

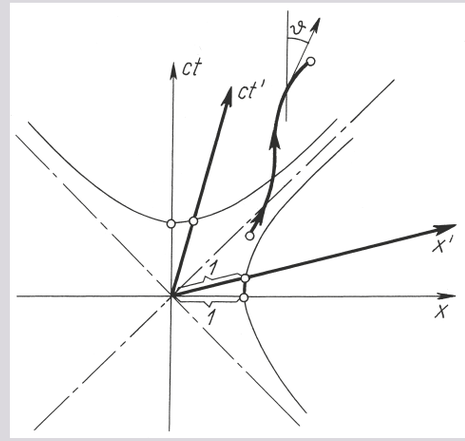
The derivation of the transformation formulas of the theory of relativity was based on the premise that the fundamental laws of electrodynamics, including the laws relating to the propagation of light, have the same form in all frames of reference moving uniformly in a straight line with respect to each other. From this it fol-



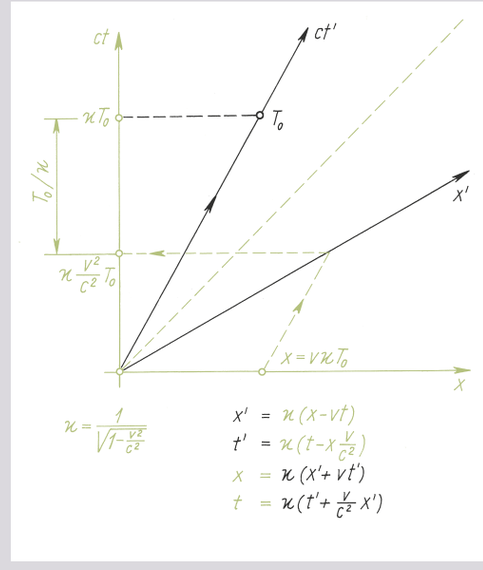
▲ **Figure 5.26** The dashed coordinate grid and the dashed clocks belong to the coordinate system K, and the clocks and coordinate system drawn with solid lines belong to K'. The figure was drawn for $\beta = v/c = ½$. It is easily seen that viewed from either system, the unit of length in the other system contracts to 0.87.



▲ **Figure 5.23** A geometric depiction of the LORENTZ transformation.



▲ **Figure 5.24** Establishing the unit lengths on the time-coordinate and spatial-coordinate axes using the hyperbolas $c^2t^2 - x^2 = 1$ and $x^2 - c^2t^2 = 1$.

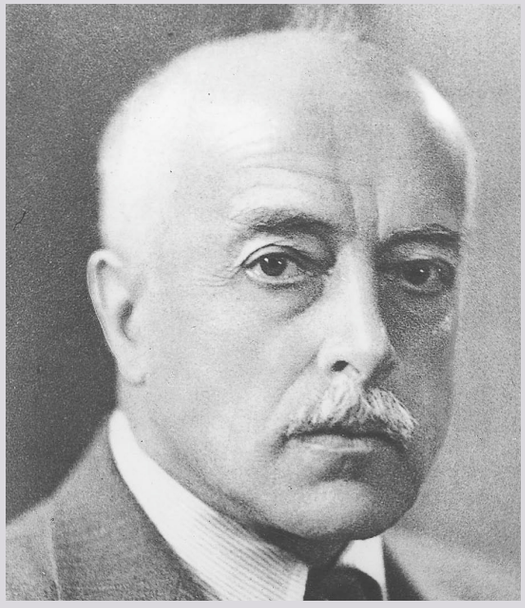


▲ **Figure 5.25** Geometric depiction of the solution of the thought experiment shown in Figure 5.21.

Quotation 5.13

The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.

—Minkowski, "Space and Time," 1908 [p. 75]



▲ **Figure 5.27** Max von Laue (1879–1960): Studies under Woldemar Voigt in Göttingen and under Max Planck in Berlin. 1905, assistant to Planck; 1909 to 1912, Private lecturer under Sommerfeld. From 1919, professor in Berlin. Max von Laue is known primarily for his research into the diffraction and interference of x-rays in crystals (1912: Laue diagrams). With his work, it was confirmed that x-rays are a form of electromagnetic waves, and it became possible to directly study the structure of crystals. In addition to von Laue, W. H. Bragg and W. L. Bragg (father and son) did fruitful work in this area.

Von Laue was from the beginning an adherent, a contributor, and an advocate of the theory of relativity. He showed that the value of the speed of light measured in flowing water directly follows from the theory of relativity. In 1911, he wrote the first survey monograph on the theory of relativity, which together with the second volume that appeared in 1919 forms a fundamental work on this topic.

As an admirer of Einstein and one of his best friends, he continued to support Einstein and his theory even after Hitler's rise to power.

lows as a matter of course that the fundamental laws of electrodynamics will also remain valid without change in the theory of relativity. But the Maxwell equations could now be written in a form that makes their invariance with respect to a Lorentz transformation particularly clear. This problem was solved by Minkowski in 1908 with the introduction of the four-dimensional *Minkowski "world,"* in which space and time coordinates are closely linked (Quotation 5.13). In 1911, Max von Laue (Figures 5.27, 5.28), expanding on Minkowski's work, found a unified representation for electric and magnetic fields by using a higher-order mathematical notion, known as tensors. These tensors contain the components of both the electric and magnetic fields, making it clear that though it is possible to decompose a unified electromagnetic field into an electric field and a magnetic field, this decomposition depends on the coordinate system used. It is possible in one coordinate system to observe an electrostatic field only, while in another the same field is seen both as an electrostatic field and *also* as a magnetic field.

This is not the case with the laws of Newtonian mechanics. The form of Newton's equation of motion remains invariant in passing from one coordinate system to another that is moving uniformly and linearly relative to the first *only if the coordinates are transformed according to the Galilean transformation*. But this means that for the theory of relativity, an entirely new mechanics must be created whose fundamental equations have the same form in every inertial system *if the transition between the inertial systems is carried out with the help of the Lorentz transformation*. The need for a new mechanics had already been noted by Poincaré (Quotation 5.10). Of course, any new theory must take into account the fact that all during the two centuries of its existence, Newtonian mechanics had enjoyed dazzling success. Its validity for problems of daily life that show up in engineering practice was indisputable. Therefore, relativistic mechanics would have to be formulated in such a way—another brilliant prognosis made by Poincaré (Quotation 5.10)—that for speeds small in relation to the speed of light it closely approximated classical Newtonian mechanics. The crucial difference between relativistic mechanics and classical mechanics becomes visible when the dependence of mass on velocity is taken into account (Figure 5.29). In classical mechanics, mass is a constant quantity independent of velocity, whereas in relativistic mechanics, it grows larger with increasing speed and becomes infinitely large as the limit of the speed of light is approached. The development of relativistic mechanics and the increase of mass with velocity is closely coupled with the assertion of the theory of relativity that has been of the greatest importance for everyday life: the equivalence of mass and energy. It is worthwhile to trace the path from the first approaches to the eventual general formulation of the equivalence theorem.

The question of what additional force is needed to accelerate a body when it is electrically charged was raised almost a generation prior to the precise formulation of the theory of relativity. That we need a greater force to accelerate a charged body than a neutral one is clear, for example, from simple considerations of the energy balance: Moving charges imply a current flow, which means an associated magnetic field, which in turn has a field energy. This additional field energy could have been provided only by an additional accelerating force. In 1881, J. J. Thomson calculated the force necessary to accelerate an electron and found that it corresponds to a mass m that is directly related to the energy of the electrical field surrounding the electron:

$$m = \frac{4E}{3c^2}.$$

The name Poincaré comes up once again in this context. As early as 1900, Poincaré raised the question of the inertia of energy propagating in the form of electromagnetic radiation. Assuming the validity of the law of conservation of momentum, he found that in the interaction between radiation of energy E that propagates at velocity c and a body of mass m, the momentum balance can be written as

$$p = c\frac{E}{c^2}.$$

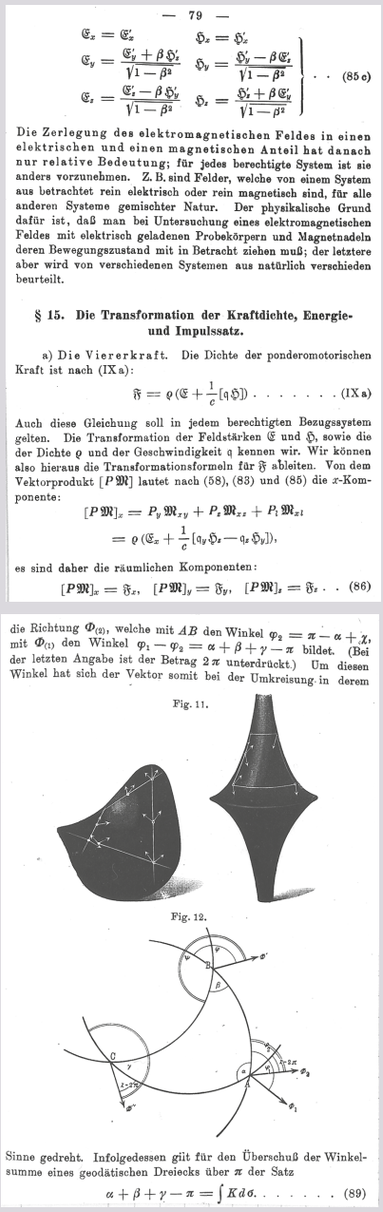
Poincaré noted that the quantity E/c^2 has the dimension of mass, and in this manner, one can talk about the inertia of energy. This discovery by Poincaré was quoted—uncharacteristically—by Einstein in an article published in 1906:

> In an article published in a previous year (A. Einstein, *Annalen der Physik* 18 (1905), p. 639), I showed that the Maxwell electromagnetic equations together with the relativity principle and energy principle lead to the conclusion that the mass of a body changes with a change in its energy content, regardless of the type of energy change. It turned out that to an energy change of magnitude ΔE there must correspond a change in mass in the same direction of magnitude $\Delta E/V^2$, where V denotes the velocity of light.
>
> In this article I would like to show that that theorem is a necessary and sufficient condition for the law of conservation of motion of the center of gravity (at least to a first approximation) to hold as well for systems in which in addition to mechanical processes, electromagnetic processes also occur. Although the simple formal treatment that must be carried out is by and large already contained in an article by Poincaré (H. Poincaré, *Lorentz-Festschrift*, 1900, p. 252), for the sake of clarity I shall not rely on this work.

In 1903, Poincaré went so far as to speculate that perhaps mass might be of electromagnetic origin. Hidden in this conjecture is the intimate connection between mass and energy. Friedrich Hasenöhrl (1874–1915) showed in 1904 that if we enclose radiation of energy E in a box and then accelerate the box, the presence of the electromagnetic energy will have the same effect as if the mass of the box had increased by an amount proportional to E.

The connection between mass and energy had thus already been suspected by many physicists at the time of the birth of the theory of relativity, but in each specific case they arrived at the connection by using different considerations. In contrast, the equivalence of matter and energy follows naturally and in a general form from the basic conditions of relativity. Nevertheless, the path to the solution was not smooth or without detours. In his 1905 article "Does the Inertia of a Body Depend on Its Energy Content?" Einstein proves, using the transformation formulas of relativity theory, that if a body emits energy E in the form of radiation, its mass will decrease by the amount E/c^2. At the close of this article, Einstein concludes that "the mass of a body is a measure of its energy content." It is customary to refer to this article when claiming that in 1905, Einstein found the equivalence principle of mass and energy in a completely general form.

However, in that article, the connection between change in mass and energy emitted is derived in a special case only, and with a questionable argument. The statement of the general principle of equivalence of energy and mass appears here merely as a brilliant conjecture that was still awaiting a precise formulation and rigorous proof (Plate XXII).



— 79 —

$$\left.\begin{aligned} \mathfrak{E}_x &= \mathfrak{E}'_x & \mathfrak{H}_x &= \mathfrak{H}'_x \\ \mathfrak{E}_y &= \frac{\mathfrak{E}'_y + \beta\mathfrak{H}'_z}{\sqrt{1-\beta^2}} & \mathfrak{H}_y &= \frac{\mathfrak{H}'_y - \beta\mathfrak{E}'_z}{\sqrt{1-\beta^2}} \\ \mathfrak{E}_z &= \frac{\mathfrak{E}'_z - \beta\mathfrak{H}'_y}{\sqrt{1-\beta^2}} & \mathfrak{H}_z &= \frac{\mathfrak{H}'_z + \beta\mathfrak{E}'_y}{\sqrt{1-\beta^2}} \end{aligned}\right\} \quad (85\,c)$$

Die Zerlegung des elektromagnetischen Feldes in einen elektrischen und einen magnetischen Anteil hat danach nur relative Bedeutung; für jedes berechtigte System ist sie anders vorzunehmen. Z. B. sind Felder, welche von einem System aus betrachtet rein elektrisch oder rein magnetisch sind, für alle anderen Systeme gemischter Natur. Der physikalische Grund dafür ist, daß man bei Untersuchung eines elektromagnetischen Feldes mit elektrisch geladenen Probekörpern und Magnetnadeln deren Bewegungszustand mit in Betracht ziehen muß; der letztere aber wird von verschiedenen Systemen aus natürlich verschieden beurteilt.

§ 15. Die Transformation der Kraftdichte, Energie- und Impulssatz.

a) Die Viererkraft. Die Dichte der ponderomotorischen Kraft ist nach (IX a):

$$\mathfrak{F} = \varrho\,(\mathfrak{E} + \tfrac{1}{c}[\mathfrak{q}\mathfrak{H}]) \quad (IX\,a)$$

Auch diese Gleichung soll in jedem berechtigten Bezugssystem gelten. Die Transformation der Feldstärken $\mathfrak{E}$ und $\mathfrak{H}$ sowie die der Dichte ϱ und der Geschwindigkeit $\mathfrak{q}$ kennen wir. Wir können also hieraus die Transformationsformeln für $\mathfrak{F}$ ableiten. Von dem Vektorprodukt $[P\mathfrak{M}]$ lautet nach (58), (83) und (85) die x-Komponente:

$$[P\mathfrak{M}]_x = P_y\mathfrak{M}_{xy} + P_z\mathfrak{M}_{xz} + P_t\mathfrak{M}_{xt}$$
$$= \varrho\,(\mathfrak{E}_x + \tfrac{1}{c}[\mathfrak{q}_y\mathfrak{H}_z - \mathfrak{q}_z\mathfrak{H}_y]),$$

es sind daher die räumlichen Komponenten:

$$[P\mathfrak{M}]_x = \mathfrak{F}_x, \quad [P\mathfrak{M}]_y = \mathfrak{F}_y, \quad [P\mathfrak{M}]_z = \mathfrak{F}_z \quad (86)$$

die Richtung $\Phi_{(2)}$, welche mit AB den Winkel $\varphi_2 = \pi - \alpha + \chi$, mit $\Phi_{(1)}$ den Winkel $\varphi_1 - \varphi_2 = \alpha + \beta + \gamma - \pi$ bildet. (Bei der letzten Angabe ist der Betrag 2π unterdrückt.) Um diesen Winkel hat sich der Vektor somit bei der Umkreisung in derem

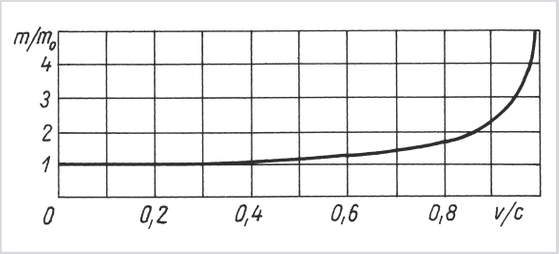
Fig. 11.

Fig. 12.

Sinne gedreht. Infolgedessen gilt für den Überschuß der Winkelsumme eines geodätischen Dreiecks über π der Satz

$$\alpha + \beta + \gamma - \pi = \int K\,d\sigma. \quad (89)$$

Figure 5.28
Pages from Laue's books on the special and general theories of relativity.



▲ **Figure 5.29** Measurements by Kaufmann of the increase in the mass of electrons as a function of their velocity.

Walter Kaufmann (1871–1947): Studies in Berlin and Munich; 1908 to 1935: professor at Königsberg (today Kaliningrad); 1897: measurement of the specific charge of the electron, with the greatest precision at the time. First reported on measurements of very fast beta particles in the 1902 article "Die magnetische und elektrische Ablenkbarkeit der Becquerelstrahlen und die scheinbare Masse der Elektronen" (The magnetic and electric deflectability of Becquerel rays and the apparent mass of electrons). His measurements were later repeated and refined by A. H. Bucherer.

▶ **Figure 5.30** Illustration of the equivalence of mass and energy. After combustion, the measured mass of the end products in a cooled state is reduced by a portion of order of magnitude 10^{-10} times smaller than the original mass. This lost mass will appear as heat energy of the combustion products: In heating, the energy of motion increases, and therefore the masses of the individual particles increase as well. In the third reaction—one that is not practically realizable in such a mass scale—all of the mass will appear as energy.

Quotation 5.14

What is Geometry?

A conversation between an experimental physicist, a pure mathematician and a relativist, who advocates the newer conceptions of time and space in physics:

Relativist: There is a well know proposition of Euclid which states that "Any two sides of a triangle are together greater than the third side." Can either of you tell me whether nowadays there is good reason to believe that this proposition is true?

Mathematician: For my part, I am quite unable to say whether the proposition is true or not. I can deduce it by trustworthy reasoning from certain other propositions or axioms, which are supposed to be still more elementary. If these axioms are true, the proposition is true; if the axioms are not true, the proposition is not true universally. Whether the axioms are true or not I cannot say, and it is outside my province to consider.

Physicist: But is it not claimed that the truth of these axioms is self-evident?

Mathematician: They are by no means self-evident to me; and I think the claim has been generally abandoned.

Physicist: Yet since on these axioms you have been able to found a logical and self-consistent system of geometry, is not this indirect evidence that they are true?

continued on next page

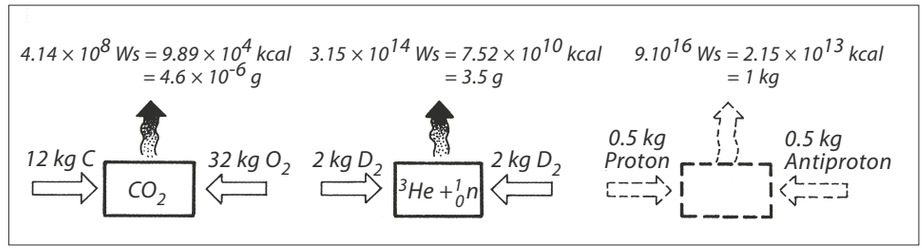
Nevertheless, the effect of this article in terms of stimulating further research was very great indeed. Characteristically, the leading minds—such as Planck and Lorentz—recognized the possibilities for extension, and they continually enriched the theory of relativity along the way to its completion with new discoveries. We may mention here as an interesting fact that in the following years, until the 1908 introduction of the four-dimensional formulation by Minkowski, it was Planck who played the leading role in the development of the theory. In a like manner, it was Einstein who continued the advancement of Planck's ideas relating to quantum theory. Thus, in 1906, Planck, building on Einstein's ideas, found the equation of motion of mechanics

$$\mathbf{F} = \frac{\mathrm{d}\,\mathbf{p}}{\mathrm{d}\,t}, \quad m = \frac{m_0}{\sqrt{1 - v^2 / c^2}}, \quad \mathbf{p} = m\mathbf{v},$$

which is valid in the relativistic domain, and determined the formula

$$E = mc^2$$

for the total energy of a moving body.



We now comment on these formulas in more detail. The equation of motion is, on its face, identical to the Newtonian equation of motion, namely,

force is equal to the change in momentum per unit time,

but with the crucial difference that in the formula for momentum, $\mathbf{p} = m\mathbf{v}$, the mass is now dependent on velocity. If we look at the expression for the total energy, then we can observe, after the mathematical transformation

$$m_0 c^2 \left(1 - \frac{v^2}{c^2}\right)^{-1/2} \approx m_0 c^2 \left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) = m_0 c^2 + \frac{1}{2} m_0 v^2,$$

that we obtain the Newtonian kinetic energy $\frac{1}{2} m_0 v^2$ as an additional energy above the rest energy $m_0 c^2$ if the velocity is small compared to the speed of light.

Finally, in 1907, Planck—reaching back to the conservation of momentum already investigated by Poincaré—derived the law of the equivalence of mass and energy in a more general case. Nonetheless, Pauli (in his 1921 monograph that we have already cited several times) still felt it necessary to rely on Lorentz's later paper from the year 1911 for his conclusion:

> It may therefore be considered as proven that the relativity principle in conjunction with the laws of conservation of momentum and energy leads to the fundamental principle of the inertia of all energy. We may agree with Einstein that this is the most important principle of special relativity. A quan-

titative experimental proof has not yet been possible. Already in his first publication on this subject, Einstein suggested the possibility of a verification of the theory using radioactive processes. But the expected defects in the atomic weights of radioactive elements are too small to be determined empirically. However, the possibility of explaining the deviations of the atomic weights (with respect to $H = 1$) of elements from integer values to the extent that they are not determined by isotopes through the interaction energy of nuclear components and their inertia, which Langevin was the first to suggest, has recently been much discussed. Perhaps in the future, the law of inertia of energy will be able to be established through observations related to the stability of nuclei. Indications of a quantitative agreement are present.

Thus, the equivalence law

$$E = mc^2$$

asserts that to every mass belongs an energy, and to every energy a mass. This equation—one of the fundamental relationships accessible to the general public even in quantitative form—can be seen as the symbol of twentieth-century physics (Plate XXII).

Einstein's crucial contribution to the discovery of this law can be seen as his having been the first to attempt to apply the general methods of relativity theory to this problem, thereby pointing the way to the final result.

Another half a century would have to pass before a part of the nuclear mass, in accordance with the above formula, could be transformed into electrical energy in atomic power plants, but even for the establishment of the energy balance in particular nuclear reactions and for the spectacular evidence of particle pair production and annihilation, we had to wait three decades (Figure 5.30). In 1919, Einstein, in his popular book *Special and General Relativity*, was only able to write as follows:

> A direct comparison of this relation with experiment is not possible at the present time, owing to the fact that the changes in energy E_o to which we can subject a system are not large enough to make themselves perceptible as a change in the inertial mass of the system. E_o/c^2 is too small in comparison with the mass m, which was present before the alteration of the energy. It is owing to this circumstance that classical mechanics was able to establish successfully the conservation of mass as a law of independent validity.

5.2.6 *Matter and the Geometry of Space*

The general theory of relativity is woven from three strands of thought. The first goes back to the doubts expressed at the beginning of the nineteenth century regarding the logical necessity of the axioms of Euclidean geometry. As we have discussed earlier, it was here that János Bolyai, Nikolai Lobachevsky, and Carl Friedrich Gauss played crucial roles. Both Gauss and Lobachevsky, as well as F. Bolyai (the father of János), raised the question of whether Euclidean geometry is in fact realized in nature. The problem of the relationship between geometry and physical reality was thus for the first time seen as suggesting that the validity of the axioms of geometry should be decided *through measurement* (Figure 5.31).

As an introduction to this topic, it is worth reading through the prologue that A. S. Eddington (1882–1944) wrote in his 1920 book *Space, Time, and Gravitation: An Outline of the General Relativity Theory*. His writing makes the connection

Quotation 5.14, ***continued***

Mathematician: No. Euclid's geometry is not the only self-consistent system of geometry. By choosing a different set of axioms I can, for example, arrive at Lobachevsky's geometry, in which many of the propositions of Euclid are not in general true. From my point of view there is nothing to choose between these different geometries.

Relativist: How is it then that Euclid's geometry is so much the most important system?

Mathematician: I am scarcely prepared to admit that it is the most important. But for reasons which I do not profess to understand, my friend the Physicist is more interested in Euclidean geometry than in any other, and is continually setting us problems in it. Consequently we have tended to give an undue share of attention to the Euclidean system. There have, however, been great geometers like Riemann who have done something to restore a proper perspective.

Relativist (to *Physicist*): Why are you specially interested in Euclidean geometry? Do you believe it to be the true geometry?

Physicist: Yes. Our experimental work proves it true.

Relativist: How, for example, do you prove that any two sides of a triangle are together greater than the third side?

Physicist: I can, of course, only prove it by taking a very large number of typical cases, and I am limited by the inevitable inaccuracies of experiment. My proofs are not so general or so perfect as those of the pure mathematician. But it is a recognized principle in physical science that it is permissible to generalize from a reasonably wide range of experiment; and this kind of proof satisfies me.

Relativist: It will satisfy me also. I need only trouble you with a special case. Here is a triangle *ABC*; how will you prove that *AB* + *BC* is greater than *AC*?

Physicist: I shall take a scale and measure the three sides.

Relativist: But we seem to be talking about different things. I was speaking of a proposition of geometry—properties of space, not of matter. Your experimental proof only shows how a material scale behaves when you turn it into different positions.

Physicist: I might arrange to make the measures with an optical device.

Relativist: That is worse and worse. Now you are speaking of properties of light.

Physicist: I really cannot tell you anything about it, if you will not let me make measurements of any kind; measurement is my only means of finding out about nature. I am not a metaphysicist.

Relativist: Let us agree then that by *length* and *distance* you always mean a quantity arrived at by measurements with material or optical appliances.

continued on next page

Quotation 5.14, ***continued***

You have studied experimentally the laws obeyed by these *measured lengths*, and have found the geometry to which they conform ... Let us proceed to examine the laws of natural geometry. I have a tape-measure, and here is the triangle. $AB = 39\frac{1}{2}$ in., $BC = \frac{1}{8}$ in., $CA = 39\frac{7}{8}$ in. Why, your proposition does not hold!

Physicist: You know very well what is wrong. You gave the tape-measure a big stretch when you measured *AB*.

Relativist: Why shouldn't I?

Physicist: Of course, a length must be measured with a rigid scale.

Relativist: That is an important addition to our definition of length. But what is a rigid scale?

Physicist: A scale which always keeps the same length.

Relativist: But we have just defined length as the quantity arrived at by measures with a rigid scale; so you will want another rigid scale to test whether the first one changes length; and a third to test the second; and so *ad infinitum*. ...

Mathematician: The view has been widely held that space is neither physical nor metaphysical, but conventional. Here is a passage from Poincaré's *Science and Hypothesis*, which describes this alternative view of space:

"If Lobachevsky's geometry is true, the parallax of a very distant star will be finite. If Riemann's is true, it will be negative. These are the results which seem within the reach of experiment, and it is hoped that astronomical observations may enable us to decide between the two geometries. But what we call a straight line in astronomy is simply the path of a ray of light. If, therefore, we were to discover negative parallaxes, or to prove that all parallaxes are higher than a certain limit, we should have a choice between two conclusions: we could give up Euclidean geometry, or modify the laws of optics, and suppose that light is not rigorously propagated in a straight line. It is needless to add that everyone should look upon this solution as the more advantageous. Euclidean geometry, therefore, has nothing to fear from fresh experiments."

Relativist: Poincaré's brilliant exposition is a great help in understanding the problem now confronting us. He brings out the interdependence between geometrical laws and physical laws, which we have to bear in mind continually. We can add on to one set of laws that which we subtract from the other set. I admit that space is conventional---for that matter, the meaning of every word in the language is conventional. Moreover, we have actually arrived at the parting of the ways imagined by Poincaré, though the crucial experiment is not precisely the one he mentions.

—Sir Arthur Eddington, *Space, Time and Gravitation*, 1921 [pp. 1–4, 9]

between geometry and physical reality seem almost self-evident (Quotation 5.14, Figure 5.32).

Of particular significance on the path to the general theory of relativity is Gauss's 1827 *Disquisitiones circa superficies curvas*, which he was inspired to write through his practical work as a surveyor. In this work, Gauss investigated the inherent geometry of curved surfaces, that is, the properties of curved surfaces that are independent of the surface's embedding in three-dimensional space and that can therefore be established by using measurements carried out on the surface itself—in other words, those properties that intelligent two-dimensional beings living on such a surface would themselves be able to determine—as this problem is typically explained in the popular scientific literature. For the purposes of future development, the most important result was that of Gauss that such two-dimensional beings could determine the surface's *curvature*. In a more precise formulation, we may say that the Gaussian curvature of a curved surface embedded in three-dimensional Euclidean space can be determined with the help of measurements that are taken exclusively within the surface. The Gaussian curvature is the reciprocal of the product of the so-called *principal curvature radii* of the surface (Figures 5.33, 5.34).

In his fundamental 1854 work *On the Hypotheses that Underlie Geometry*, Riemann extends Gauss's work and generalizes the Gaussian idea of a space with an arbitrary number of dimensions. He begins with the definition that the distance between two neighboring points in an arbitrary n-dimensional "curved" space depends on coordinate differences of the form

$$\mathrm{d}s^2 = \sum g_{\mu\nu}\,\mathrm{d}x_\mu\,\mathrm{d}x_\nu .$$

This relationship can be viewed as a generalization of the formula

$$\mathrm{d}s^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2 ,$$

valid in Euclidean space, which is just the Pythagorean theorem.

In a simple "two-dimensional curved space" familiar to everyone, namely, the surface of a sphere of radius r (Figure 5.34), the formula looks like this:

$$\mathrm{d}s^2 = r\,\mathrm{d}\vartheta^2 + r^2 \sin^2\vartheta\,\mathrm{d}\varphi^2 .$$

Here we would like to emphasize the last paragraph from Quotation 5.15, taken from Riemann's book:

> ...that in a discrete manifold, the ground of its metric relations is given in the notion of it, while in a continuous manifold, this ground must come from outside. Either therefore the reality which underlies space must form a discrete manifold, or we must seek the ground of its metric relations outside it, in binding forces which act upon it.... This leads us into the domain of another science, of physics...

W. K. Clifford, who translated Riemann's book into English, speaks more decisively:

> That this variation of the curvature of space is what really happens in that phenomenon which we call the *motion of matter*, whether ponderable or ethereal. (Quotation 5.16)

The second line of thought leading to the general theory of relativity is again of an epistemological nature. EINSTEIN was himself probably guided primarily by epistemological considerations, to which he calls attention in his 1916 work on the general theory of relativity. In this work, EINSTEIN devotes a separate chapter to the question of why the general theory of relativity is necessary above and beyond the special theory. As we have already mentioned, EINSTEIN's mature works are also considered exemplary from a pedagogical point of view. This holds for both the 1905 foundational work on the special theory of relativity and the 1916 survey article. Therefore, we can do no better in explaining the necessity for a generalization of the special theory of relativity than to refer to EINSTEIN's own words (Quotation 5.17).

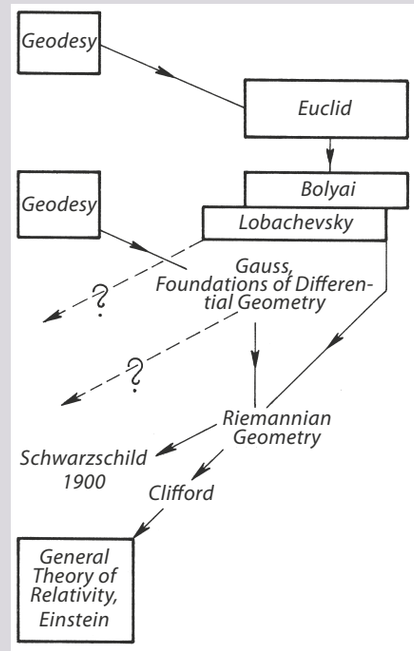
The basis for the equivalence of coordinate systems that are accelerating with respect to each other was found in an experimental result that had been known since NEWTON but had gone largely unnoticed until the end of the nineteenth century, namely, the equivalence of *gravitational* mass and *inertial* mass. This was shown with great precision by LORÁND EÖTVÖS (Figure 5.35). The precision of his measurements was not surpassed until recently.

The geometry of curved spaces, the epistemological requirement that the distinction of coordinate spaces have an observable and measurable basis, and the experimental fact of the equivalence of the gravitational and inertial mass—these were the three strands of thought from which EINSTEIN created the general theory of relativity.

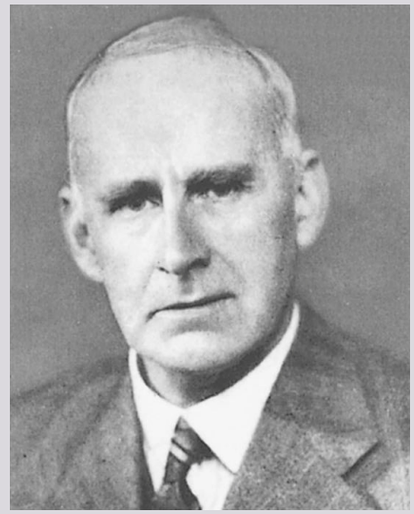
We now expend a few words here as to why Euclidean geometry had to be abandoned. Let us imagine a coordinate system K' that rotates relative to a coordinate system K with constant angular velocity (Figure 5.36). A rotating coordinate system, of course, means that it is an accelerated coordinate system, in which the special theory of relativity is applicable only "locally," that is, in the investigation of a given small region. If we draw a circle in the rotating coordinate system whose circumference we cover with measuring rods, then each of the rods undergoes a *Lorentz contraction* if we view it from the coordinate system K. The length of the measuring rods covering the circle's diameter, however, will not change because they are oriented perpendicular to the direction of motion. A stationary observer in system K determines that Euclidean geometry cannot be valid in system K' because there the ratio between the circumference and the diameter of a circle is not π but a somewhat smaller number. It is therefore necessary to use the more general *Riemannian metric*, which differs from the Euclidean one. It is evident that the characteristics of this curved space should be somehow proportional to the acceleration in the coordinate system K', and since we can substitute the gravitational force field for the acceleration, they should be related somehow with the "cause" of the force field, that is, with mass. Of course, this connection is not simple, and from the first approaches (1907) to the final formulation of the theory (1916), countless solutions miscarried. In the end, the fundamental equations of general relativity took the following form:

$$R_{ik} - \frac{1}{2} g_{ik} R = -\kappa T_{ik}, \quad i, k = 1, 2, 3, 4.$$

We have already met the quantities g_{ik} that appear in this equation in the formula for the square of the infinitesimal distance; they are the metric coefficients in the four-dimensional space-time continuum. The number R_{ik} is the curvature tensor, a complicated mathematical quantity formed from the metric coefficients and their derivatives. Finally, R is the curvature scalar, which can be calculated from the



◀ **Figure 5.31** Connections between physics and geometry.



◀ **Figure 5.32** Sir (since 1930) ARTHUR STANLEY EDDINGTON (1882–1944), British astronomer and physicist. 1913, professor of physics; 1914, director of the observatory at Cambridge, supported the development of physics between the two world wars in three areas. He adopted the newest results in physics, so to speak, *in statu nascendi* and made them known in outstanding books of popular science (see, for example, Quotation 5.14), wrote influential textbooks (*The Mathematical Theory of Relativity*, 1923), and in his own area, astrophysics, created a new picture of the internal structures of the stars (*The Internal Constitution of the Stars*, 1926). He suggested that alongside gas pressure and gravitation, radiation pressure contributes in significant measure to equilibrium. Thus he was able to provide a theoretical basis to the empirically formulated mass–luminosity relation. According to his theory, all stars end their lives, regardless of their mass, as white dwarfs.

EDDINGTON was named Astronomer Royal of England (Quotation 5.18). He organized the expedition to the island Principe on May 28–29, 1919, where the historic experiment on the bending of light rays in the Sun's gravitational field was carried out, which was the first proof of the correctness of the theory of general relativity.

components of the curvature tensor and the metric coefficients. In two-dimensional space, this quantity is directly connected with the Gaussian curvature. The quantity T_{ik} is the energy momentum tensor, which accounts for the distribution of mass in the space, and k is a constant. ($k = 8\pi G/c^4$, where $G = 6.67 \times 10^{-11}$ Nm^2kg^{-2} is universal gravitational constant, c is the velocity of light, hence $k = 2.07 \times 10^{-45}$ $Ns^4kg^{-2}m^{-2}$.)

In solving the difficult mathematical problems that arose in the process of working out of the general theory of relativity, EINSTEIN was afforded considerable support during his brief tenure as a professor in Zurich from his friend and colleague MARCEL GROSSMANN. Both men investigated the covariant properties of the field equations in the years 1913 and 1914. EINSTEIN published the fundamental equation of gravitation at the November 18, 1915 session of the Prussian Academy in the form $R_{ik} = -\gamma T_{ik}$. At the next meeting—one week later, on November 25, 1915—he gave a corrected, essentially final, form:

$$R_{ik} = -\gamma\left(T_{ik} - \frac{1}{2}g_{ik}T\right).$$

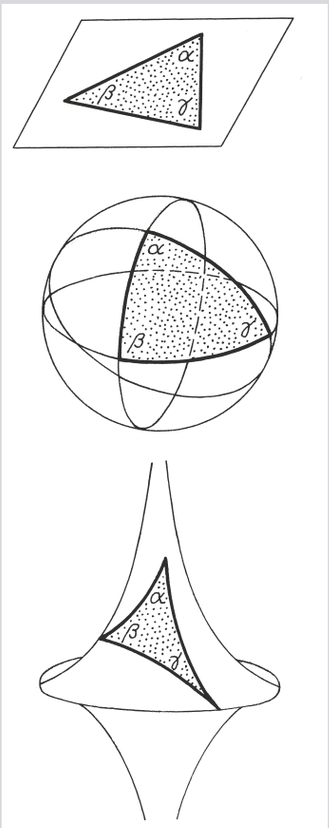
And now, in a role similar to that played by POINCARÉ, another genius steps into the limelight—DAVID HILBERT, the new *princeps mathematicorum*: On November 20, 1915, that is, five days earlier than EINSTEIN, he presented his gravitation equation to the Royal Academy in Göttingen,

$$\sqrt{g}\left(R_{ik} - \frac{1}{2}Rg_{ik}\right) = \frac{\partial\sqrt{g}L}{\partial g_{ik}},$$

and, as was noted later, it agrees with EINSTEIN's version (Figure 5.37). "I was under the impression that, simultaneously with EINSTEIN, HILBERT also found the now accepted equations of general relativity. Is this correct? If so, is there a reason no one seems to mention this now?" asks WIGNER in a letter to JAGDISH MEHRA. "The only reason, probably, is that people do not read the literature;" another reason is "the sociology of science," replied MEHRA [1973, p. 178].

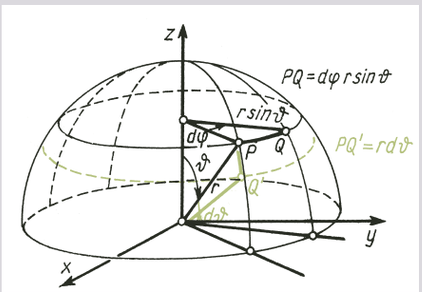
If we recall the fundamental postulate of the special theory of relativity, then we see that, other than the requirement for shape invariance in physical laws in passing from one inertial system to another, no significant concrete statements about particular physical relationships can be made. The general theory of relativity, in contrast, not only establishes the equivalence of coordinate systems that move arbitrarily with respect to each other but moreover, *leads to a new theory of gravitation*, wherein the gravitational effects do not appear as force effects whose sources are the masses, but rather the masses change the geometry of space according to the fundamental equations. How will this fact now appear in the general theory of relativity once the gravitational force is eliminated? Clearly, here the curvature of space is to assume the role of the gravitational force. In the space "curved" by a mass, a free body will move in a "straight line." But the role of the "straight line" in the curved space is, in turn, assumed by the "geodesic" line, the shortest path along the curvature. It can be shown that, to a first approximation, we obtain in this way the same results as those from the Newtonian theory of gravitation. The discrepancy between the results, once we compute to a better approximation, is very small but not insignificant. Thus, with the help of general relativity, it was possible to correctly interpret the perihelion shift of Mercury's orbit, about which we have already spoken in Section 4.4.7.

The general theory of relativity has two additional observable consequences: the red shift of the spectral lines in very strong gravitational fields and the curving of the path of light rays in similarly strong gravitational fields. The latter effect is of measurable magnitude when light from distant stars passes close to the Sun on its way to Earth. Of course, this can be observed only during a total solar eclipse (Figure 5.38). Quotation 5.18 gives an impression of the enthusiastic response to



◀ **Figure 5.33** Three distinct spaces for two-dimensional life forms.

The three visualizable surfaces, or two-dimensional "spaces," may serve as models to illustrate processes in the curved four-dimensional space-time continuum. The *plane* surface represents a space with zero curvature. Its three-dimensional analogue is the space of "common sense." This space was viewed as the only possible one for describing all physical processes until the beginning of the nineteenth century. The *spherical* surface is an example of a space with positive curvature. A spherical surface steadily expanding like a soap bubble is a model for an expanding universe. Finally, the third surface has constant negative curvature; it can be generated by the rotation of a special curve (tractrix, a pursuit curve) about the x-axis. The "triangles" drawn on the three surfaces are meant to suggest that one can determine the curvature of a surface through measurements that can be carried out exclusively within the surface itself. The sums of the interior angles of a triangle in the three cases are respectively equal to, greater than, and less than 180°. To be able to draw a triangle, one first needs a definition of a straight line on a curved surface: By a line, we understand a curve that connects two points by the shortest possible path. The curvature of space in this way leads to particular physical consequences.



▲ **Figure 5.34** On the surface of a sphere—for example on the surface of Earth—the location of a point can be specified by two angular coordinates; these are essentially identical to the longitude and latitude. Here the values of g_{ik} in the formula for the line element ds^2 are

$$g_{11} = r^2, \quad g_{12} = g_{21} = 0, \quad g_{22} = r^2 \sin^2 \vartheta.$$

the news that measurements taken by an expedition sent to observe a solar eclipse had confirmed Einstein's theory.

Einstein hoped to expand the general theory of relativity to include not only phenomena of gravitation, but also those of electrodynamics. This grandiose undertaking of condensing all physical phenomena into one equation did not succeed. Let us hear the account of this by Gamow, an active participant in twentieth-century physics and an acquaintance of Einstein (Quotation 5.19).

The basic postulate of the general theory of relativity says—as we have seen—that all frames of reference are equivalent for the formulation of the laws of nature, regardless of their states of motion. In accelerated systems, appropriate gravitational fields can be introduced; this possibility is ensured by the equivalence of gravitational and inertial mass. Therefore, every observer can feel justified in considering his description correct as well as that of every other observer. Thus, for example, a terrestrial observer can correctly represent the motion of the celestial bodies, such as the Sun and planets, even if he chooses the stationary Earth as a frame of reference. In their book *The Evolution of Physics*, Einstein and Infeld express this fact concisely thus: "The struggle, so violent in the early days of science, between the views of Ptolemy and Copernicus would then be quite meaningless" because both are correct.

In addition to what we have just said, we would like to address the hopes of many lay people, and even some physicists that—in view of the complexity of the abstractions and laws of modern physics—we might one day be able to return to the old "common sense" way of looking at things. The rehabilitation of an old theory, as in the example above, is typical of the arguments that might be raised to foster such hopes. But let us for a moment imagine that Cardinal Barberini, the head of the commission of the Holy Inquisition investigating Galileo, had somehow obtained knowledge of general relativity. Couldn't he then with an ironic smile chide Galileo, saying that all his anguish and clever arguments were in vain because posterity will in any case judge both sides as being correct? But then again, if Barberini had also taken the time to leaf through a book on general relativity and had realized what the price might be for bringing the Ptolemaic and Copernican worldviews into concert—in other words, if he had learned about the four-dimensional curved space in which time and space are merged almost inseparably and in which acceleration and the gravitational field are interchangeable—then his smirk would surely disappear, and in dismay he would probably admit that, in such a case, he would prefer Galileo's arguments.

5.2.7 *Einstein on Space, Ether, and the Field Problem of Physics*

The mechanical properties of the ether were at first a mystery. Then came H. A. Lorentz's great discovery. All the phenomena of electromagnetism then known could be explained on the basis of two assumptions: that the ether is firmly fixed in space—that is to say, unable to move at all—and that electricity is firmly lodged in the mobile elementary particles. Today his discovery may be expressed as follows: physical space and the ether are only different terms for the same thing; fields are physical states of space. But the physicists were still far removed from such a way of thinking; space was still, for them, a rigid, homogeneous something, incapable of changing or assuming various states. Only the genius of Riemann, solitary and uncomprehended, had already won its way by the middle of the last century to a new conception of space, in which space was deprived of its rigidity, and the possibility

Quotation 5.15

It is known that geometry assumes, as things given, both the notion of space and the first principles of constructions in space. It gives definitions of them which are merely nominal, while the true determinations appear in the form of axioms. The relation of these assumptions remains consequently in darkness; we neither perceive whether and how far their connection is necessary, nor *a priori*, whether it is possible. ...

... the properties which distinguish space from other conceivable triply extended magnitudes are only to be deduced from experience. Thus arises the problem, to discover the simplest matters of fact from which the metric relations of space may be determined; ... These matters of fact are—like all matters of fact—not necessary, but only of empirical certainty; they are hypotheses ...

... in a discrete manifold, the ground of its metric relations is given in the notion of it, while in a continuous manifold, this ground must come from outside. Either therefore the reality which underlies space must form a discrete manifold, or we must seek the ground of its metric relations outside it, in binding forces which act upon it ... This leads us into the domain of another science, of physics ...

—Riemann, "On the Hypotheses Which Lie at the Foundations of Geometry" [pp. 652–653,660, 661]

Quotation 5.16

Riemann has shewn that as there are different kinds of lines and surfaces, so there are different kinds of space of three dimensions; and that we can only find out by experience to which of these kinds the space in which we live belongs. In particular, the axioms of plane geometry are true within the limits of experiment on the surface of a sheet of paper, and yet we know that the sheet is really covered with a number of small ridges and furrows, upon which (the total curvature not being zero) these axioms are not true. ...

I wish here to indicate a manner in which these speculations may be applied to the investigation of physical phenomena. I hold in fact

1. That small portions of space *are* in fact of a nature analogous to little hills on a surface which is on the average flat; namely, that the ordinary laws of geometry are not valid in them.
2. That this property of being curved or distorted is continually being passed on from one portion of space to another after the manner of a wave.

continued on next page

Quotation 5.16, ***continued***

3. That this variation of the curvature of space is what really happens in that phenomenon which we call the *motion of matter*, whether ponderable or ethereal ...

—W. K. Clifford, "On The Space Theory of Matter" [pp. 523–524]

Quotation 5.17

The Need for an Extension of the Postulate of Relativity

In classical mechanics, and no less in the special theory of relativity, there is an inherent epistemological defect which was, perhaps for the first time, clearly pointed out by Ernst Mach. We will elucidate it by the following example: Two fluid bodies of the same size and nature hover freely in space at so great a distance from each other and from all other masses that only those gravitational forces need be taken into account which arise from the interaction of different parts of the same body. Let the distance between the two bodies be invariable, and in neither of the bodies let there be any relative movements of the parts with respect to one another. But let either mass, as judged by an observer at rest relatively to the other mass, rotate with constant angular velocity about the line joining the masses. This is a verifiable relative motion of the two bodies. Now let us imagine that each of the bodies has been surveyed by means of measuring instruments at rest relatively to itself, and let the surface of S_1 prove to be a sphere, and that of S_2 an ellipsoid of revolution. Thereupon we put the question: What is the reason for this difference in the two bodies? No answer can be admitted as epistemologically satisfactory, unless the reason given is an *observable fact of experience*. ...

Newtonian mechanics does not give a satisfactory answer to this question. It pronounces as follows: The laws of mechanics apply to the space R_1, in respect to which the body S_1 is at rest, but not the space R_2, in respect to which the body S_2 is at rest. But the privileged space R_1 of Galileo, thus introduced, is a merely *factitious* cause, and not a thing that can be observed. It is therefore clear that Newton's mechanics ... makes the factitious cause R_1 responsible for the observable difference in the bodies S_1 and S_2.

The only satisfactory answer must be that the physical system consisting of S_1 and S_2 reveals within itself no imaginable cause to which the differing

continued on next page

of its partaking in physical events was recognized. This intellectual achievement commands our admiration all the more for having preceded Faraday's and Maxwell's field theory of electricity. Then came the special theory of relativity with its recognition of the physical equivalence of all inertial systems. The inseparability of time and space emerged in connection with electrodynamics, or the law of the propagation of light. Hitherto it had been silently assumed that the four-dimensional continuum of events could be split up into time and space in an objective manner, i.e., that an absolute significance attached to the "now" in the world of events. With the discovery of the relativity of simultaneity, space and time were merged in a single continuum in a way similar to that in which the three dimensions of space had previously been merged into a single continuum. Physical space was thus extended to a four-dimensional space which also included the dimension of time. The four-dimensional space of the special theory of relativity is just as rigid and absolute as Newton's space.

The theory of relativity is a fine example of the fundamental character of the modern development of theoretical science. The initial hypotheses became steadily more abstract and remote from experience. On the other hand, it gets nearer to the grand aim of all science, which is to cover the greatest possible number of empirical facts by logical deduction from the smallest possible number of hypotheses or axioms. Meanwhile, the train of thought leading from the axioms to the empirical facts or verifiable consequences gets steadily longer and more subtle. The theoretical scientist is compelled in an increasing degree to be guided by purely mathematical, formal considerations in his search for a theory, because the physical experience of the experimenter cannot lead him up to the regions of highest abstraction. The predominantly inductive methods appropriate to the youth of science are giving place to tentative deduction. Such a theoretical structure needs to be very thoroughly elaborated before it can lead to conclusions which can be compared with experience. Here, too, the observed fact is undoubtedly the supreme arbiter; but it cannot pronounce sentence until the wide chasm separating the axioms from their verifiable consequences has been bridged by much intense, hard thinking. The theorist has to set about this Herculean task fully aware that his efforts may only be destined to prepare the death blow to his theory. The theorist who undertakes such a labor should not be carped at as "fanciful"; on the contrary, he should be granted the right to give free rein to his fancy, for there is no other way to the goal. His is no idle daydreaming, but a search for the logically simplest possibilities and their consequences. This plea was needed in order to make the listener or reader more inclined to follow the ensuing train of ideas with attention; it is the line of thought which has led from the special to the general theory of relativity and thence to its latest offshoot, the unified field theory. In this exposition the use of mathematical methods cannot be completely avoided.

We start with the special theory of relativity. This theory is still based directly on an empirical law, that of the constancy of the velocity of light. Let P be a point in empty space, P' an infinitely close point at a distance $d\sigma$. Let a flash of light be emitted from P at a time t and reach P' at a time $t + dt$. Then

$$d\sigma^2 = c^2\, dt^2.$$

If dx_1, dx_2, dx_3 are the orthogonal projections of $d\sigma$ and the imaginary time coordinate

$$\sqrt{-1}\cdot ct = x_4$$

is introduced, then the above-mentioned law of the constancy of the velocity of light propagation takes the form

$$\mathrm{d}s^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2 + \mathrm{d}x_4^2 = 0.$$

Since this formula expresses a real situation, we may attribute a real meaning to the quantity d*s*, even if the neighboring points of the four-dimensional continuum are so chosen that the corresponding d*s* does not vanish. This may be expressed by saying that the four-dimensional space (with an imaginary time-coordinate) of the special theory of relativity possesses a Euclidean metric.

The fact that such a metric is called Euclidean is connected with the following. The postulation of such a metric in a three-dimensional continuum is fully equivalent to the postulation of the axioms of Euclidean geometry. The defining equation of the metric is nothing but the Pythagorean theorem applied to the differentials of the coordinates.

In the special theory of relativity those coordinate changes (by transformation) are permitted for which also in the new coordinate system the quality ds^2 (fundamental invariant) equals the sum of the squares of the coordinate differentials. Such transformations are called Lorentz transformations.

The heuristic method of the special theory of relativity is characterized by the following principle: only those equations are admissible as an expression of natural law which do not change their form when the coordinates are changed by means of a Lorentz transformation (covariance of equations with respect to Lorentz transformations).

This method led to the discovery of the necessary connection between momentum and energy, between electric and magnetic field strength, electrostatic and electrodynamic forces, inert mass and energy; and the number of independent concepts and fundamental equations in physics was thereby reduced.

This method pointed beyond itself. Is it true that the equations which express natural laws are covariant with respect to Lorentz transformations only and not with respect to other transformations? Well, formulated in that way the question really has no meaning, since every system of equations can be expressed in general coordinates. We must ask: Are not the laws of nature so constituted that they are not materially simplified through the choice of any one *particular* set of coordinates?

We will only mention in passing that our empirical law of the equality of inert and gravitational masses prompts us to answer this question in the affirmative. If we elevate the equivalence of all coordinate systems for the formulation of natural laws into a principle, we arrive at the general theory of relativity, provided we retain the law of the constancy of the velocity of light or, in other words, the hypothesis of the objective significance of the Euclidean metric at least for infinitely small portions of four-dimensional space.

This means that for finite regions of space the (physically meaningful) existence of a general Riemannian metric is postulated according to the formula

$$\mathrm{d}s^2 = \sum_{\mu\nu} g_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu,$$

where the summation is to be extended to all index combinations from 1,1 to 4,4.

The structure of such a space differs quite basically in *one* respect from that of a Euclidean space. The coefficients $g_{\mu\nu}$ are really known. One can also say: the structure of such a space is as such completely undetermined. It is only

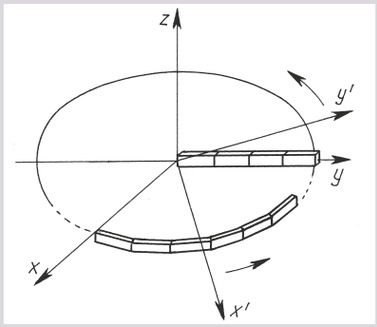
Quotation 5.17, ***continued***

behaviour of S_1 and S_2 can be referred. The cause must therefore lie *outside* this system. We have to take it that the general laws of motion, which in particular determine the shapes of S_1 and S_2, must be such that the mechanical behaviour of S_1 and S_2 is partly conditioned, in quite essential respects, by distant masses which we have not included in the system under consideration. These distant masses and their motions relative to S_1 and S_2 must then be regarded as the seat of the causes (which must be susceptible to observation) of the different behaviour of our two bodies S_1 and S_2. ... Of all imaginable spaces R_1, R_2, etc., in any kind of motion relatively to one another, there is none which we may look upon as privileged *a priori* without reviving the above-mentioned epistemological objection. *The laws of physics must be of such a nature that they apply to systems of reference in any kind of motion.* Along this road we arrive at an extension of the postulate of relativity.

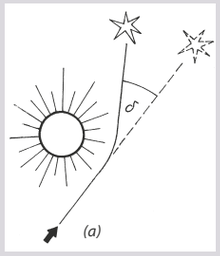
—Albert Einstein, "The Foundation of the General Theory of Relativity" [pp. 148–149]

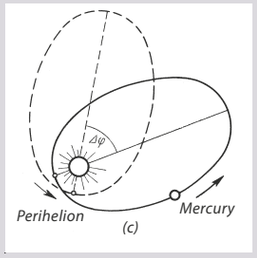


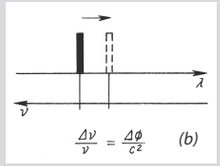
▲ **Figure 5.35** Loránd Eötvös (1848–1919): famous for his work on gravitation. With the Eötvös torsion balance, the smallest changes in the gravitational field could be determined, for example, for locating petroleum deposits. The equivalence of inertial mass and gravitational mass was established by Eötvös to a precision of 1 part in 100 million, an accuracy that was exceeded only recently. Einstein referred to this result in his work on the general theory of relativity. Eötvös also had new solutions for the problem of surface tension.



◀ **Figure 5.36** The special theory of relativity already implies that Euclidean geometry is not valid in coordinate systems accelerating with respect to one another.







◀ ▲ **Figure 5.38** The power of the general theory of relativity stems from its inner coherence and aesthetic beauty. For the experimental physicists here are three concrete empirical facts that prove its validity: **(a)** the bending of light rays in strong gravitational fields; **(b)** the red shift of spectral lines in strong gravitational fields; **(c)** the perihelion precession of the planet Mercury.

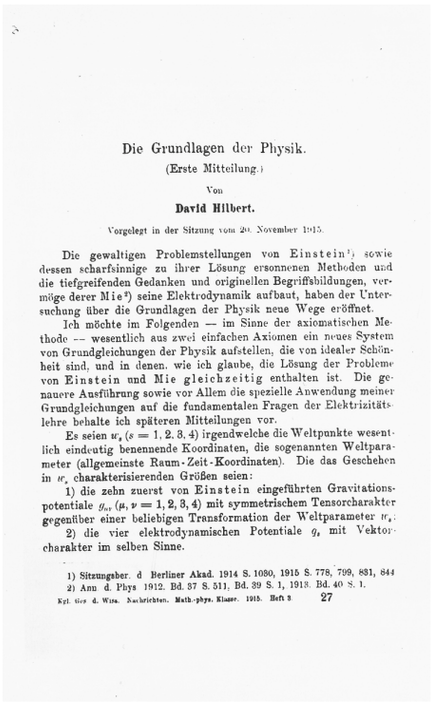
The equivalence of inertial and gravitational mass—the fundamental fact for the general theory of relativity—has been verified over the course of history with ever greater precision. Galileo's free-fall experiments provided a precision of 10^{-2}, and Newton's pendulum experiments permitted a conclusion of 10^{-3}. Bessel refined Newton's method in 1832 to obtain a value of 1.5×10^{-5}. In 1909, Eötvös achieved a record precision of 10^{-8}, which was surpassed only in recent years: Bartlett and van Buren give a precision of 4×10^{-12} (1986).

Light coming from a distant object (galaxy, quasar, etc.) in the direction of Earth can be bent by gravitational fields acting as a lens, resulting in multiple images of the same source. The first gravitational lens was discovered in 1979. A striking example of this gravitational lensing is the "Einstein Cross," where the gravitational field of a galaxy produces four separate images of a single distant quasar.

Quotation 5.18

The pilgrim fathers of the scientific imagination as it exists to-day, are the great tragedians of ancient Athens, Aeschylus, Sophocles, Euripides. Their vision of fate, remorseless and indifferent, urging a tragic incident to its inevitable issue, is the vision possessed by science. Fate in Greek Tragedy becomes the order of nature in modern thought. The absorbing interest in the particular heroic incidents, as an example and a verification of the workings of fate, reappears in our epoch as concentration of interest on the crucial experiments. It was my good

continued on next page



Die Grundlagen der Physik.

(Erste Mitteilung.)

Von

David Hilbert.

Vorgelegt in der Sitzung vom 20. November 1915.

Die gewaltigen Problemstellungen von Einstein[1], sowie dessen scharfsinnige zu ihrer Lösung ersonnenen Methoden und die tiefgreifenden Gedanken und originellen Begriffsbildungen, vermöge derer Mie[2] seine Elektrodynamik aufbaut, haben der Untersuchung über die Grundlagen der Physik neue Wege eröffnet.

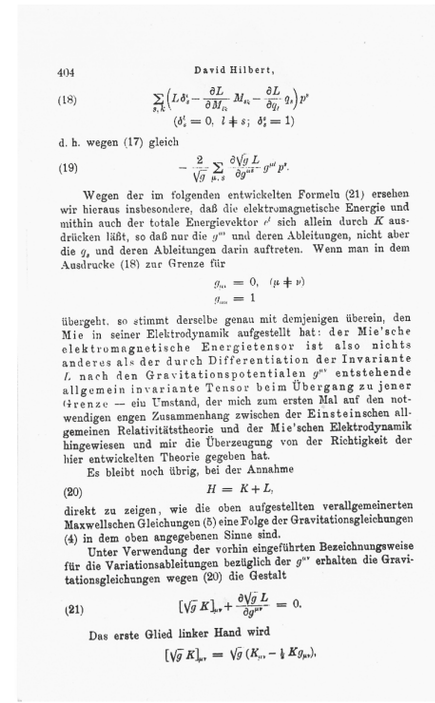
Ich möchte im Folgenden — im Sinne der axiomatischen Methode — wesentlich aus zwei einfachen Axiomen ein neues System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der Probleme von Einstein und Mie gleichzeitig enthalten ist. Die genauere Ausführung sowie vor Allem die spezielle Anwendung meiner Grundgleichungen auf die fundamentalen Fragen der Elektrizitätslehre behalte ich späteren Mitteilungen vor.

Es seien w_s ($s = 1, 2, 3, 4$) irgendwelche die Weltpunkte wesentlich eindeutig benennende Koordinaten, die sogenannten Weltparameter (allgemeinste Raum-Zeit-Koordinaten). Die das Geschehen in w_s charakterisierenden Größen seien:

1) die zehn zuerst von Einstein eingeführten Gravitationspotentiale $g_{\mu\nu}$ ($\mu, \nu = 1, 2, 3, 4$) mit symmetrischem Tensorcharakter gegenüber einer beliebigen Transformation der Weltparameter w_s;

2) die vier elektrodynamischen Potentiale q_s mit Vektorcharakter im selben Sinne.

1) Sitzungsber. d. Berliner Akad. 1914 S. 1080, 1915 S. 778, 799, 831, 844
2) Ann. d. Phys. 1912. Bd. 37 S. 511, Bd. 39 S. 1, 1913. Bd. 40 S. 1.



$$\sum_{s,k}\left(L\delta_i^s - \frac{\partial L}{\partial M_{is}}M_{ik} - \frac{\partial L}{\partial q_i}q_s\right)p^s \quad (\delta_s^s = 0,\ l \neq s;\ \delta_s^s = 1) \tag{18}$$

d. h. wegen (17) gleich

$$-\frac{2}{\sqrt{g}}\sum_{\mu,s}\frac{\partial\sqrt{g}\,L}{\partial g^{\mu s}}g^{\mu\nu}p^s. \tag{19}$$

Wegen der im folgenden entwickelten Formeln (21) ersehen wir hieraus insbesondere, daß die elektromagnetische Energie und mithin auch der totale Energievektor e^i sich allein durch K ausdrücken läßt, so daß nur die $g^{\mu\nu}$ und deren Ableitungen, nicht aber die q_s und deren Ableitungen darin auftreten. Wenn man in dem Ausdrucke (18) zur Grenze für

$$g_{\mu\nu} = 0, \quad (\mu \neq \nu)$$
$$g_{\mu\mu} = 1$$

übergeht, so stimmt derselbe genau mit demjenigen überein, den Mie in seiner Elektrodynamik aufgestellt hat: der Mie'sche elektromagnetische Energietensor ist also nichts anderes als der durch Differentiation der Invariante L nach den Gravitationspotentialen $g^{\mu\nu}$ entstehende allgemein invariante Tensor beim Übergang zu jener Grenze — ein Umstand, der mich zum ersten Mal auf den notwendigen engen Zusammenhang zwischen der Einsteinschen allgemeinen Relativitätstheorie und der Mie'schen Elektrodynamik hingewiesen und mir die Überzeugung von der Richtigkeit der hier entwickelten Theorie gegeben hat.

Es bleibt noch übrig, bei der Annahme

$$H = K + L, \tag{20}$$

direkt zu zeigen, wie die oben aufgestellten verallgemeinerten Maxwellschen Gleichungen (5) eine Folge der Gravitationsgleichungen (4) in dem oben angegebenen Sinne sind.

Unter Verwendung der vorhin eingeführten Bezeichnungsweise für die Variationsableitungen bezüglich der $g^{\mu\nu}$ erhalten die Gravitationsgleichungen wegen (20) die Gestalt

$$[\sqrt{g}\,K]_{\mu\nu} + \frac{\partial\sqrt{g}\,L}{\partial g^{\mu\nu}} = 0. \tag{21}$$

Das erste Glied linker Hand wird

$$[\sqrt{g}\,K]_{\mu\nu} = \sqrt{g}\,(K_{\mu\nu} - \tfrac{1}{2}Kg_{\mu\nu}),$$



Die Feldgleichungen der Gravitation.

Von A. Einstein.

In zwei vor kurzem erschienenen Mitteilungen[1] habe ich gezeigt, wie man zu Feldgleichungen der Gravitation gelangen kann, die dem Postulat allgemeiner Relativität entsprechen, d. h. die in ihrer allgemeinen Fassung beliebigen Substitutionen der Raumzeitvariabeln gegenüber kovariant sind.

Der Entwicklungsgang war dabei folgender. Zunächst fand ich Gleichungen, welche die Newtonsche Theorie als Näherung enthalten und beliebigen Substitutionen von der Determinante 1 gegenüber kovariant waren. Hiernach fand ich, daß diesen Gleichungen allgemein kovariante entsprechen, falls der Skalar des Energietensors der »Materie« verschwindet. Das Koordinatensystem war dann nach der einfachen Regel zu spezialisieren, daß $\sqrt{-g}$ zu 1 gemacht wird, wodurch die Gleichungen der Theorie eine eminente Vereinfachung erfahren. Dabei mußte aber, wie erwähnt, die Hypothese eingeführt werden, daß der Skalar des Energietensors der Materie verschwinde.

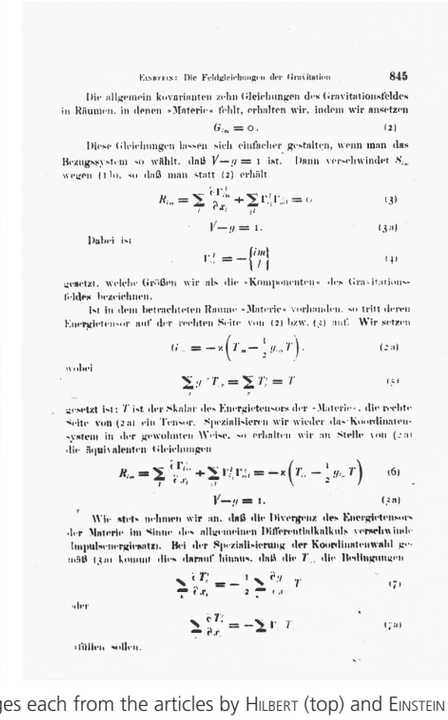
Neuerdings finde ich nun, daß man ohne Hypothese über den Energietensor der Materie auskommen kann, wenn man den Energietensor der Materie in etwas anderer Weise in die Feldgleichungen einsetzt, als dies in meinen beiden früheren Mitteilungen geschehen ist. Die Feldgleichungen für das Vakuum, auf welche ich die Erklärung der Perihelbewegung des Merkur gegründet habe, bleiben von dieser Modifikation unberührt. Ich gebe hier nochmals die ganze Betrachtung, damit der Leser nicht genötigt ist, die früheren Mitteilungen zusammenzusuchen.

Aus der bekannten Riemannschen Kovariante vierten Ranges leitet man folgende Kovariante zweiten Ranges ab:

$$G_{im} = R_{im} + S_{im} \tag{1}$$

$$R_{im} = -\sum_l \frac{\partial\left\{\begin{matrix} im \\ l \end{matrix}\right\}}{\partial x_l} + \sum_{ls}\left\{\begin{matrix} il \\ s \end{matrix}\right\}\left\{\begin{matrix} ms \\ l \end{matrix}\right\} \tag{1a}$$

$$S_{im} = \sum_l \frac{\partial\left\{\begin{matrix} il \\ l \end{matrix}\right\}}{\partial x_m} - \sum_{ls}\left\{\begin{matrix} im \\ s \end{matrix}\right\}\left\{\begin{matrix} sl \\ l \end{matrix}\right\} \tag{1b}$$



Die allgemein kovarianten zehn Gleichungen des Gravitationsfeldes in Räumen, in denen »Materie« fehlt, erhalten wir, indem wir ansetzen

$$G_{im} = 0. \tag{2}$$

Diese Gleichungen lassen sich einfacher gestalten, wenn man das Bezugssystem so wählt, daß $\sqrt{-g} = 1$ ist. Dann verschwindet S_{im} wegen (1b), so daß man statt (2) erhält

$$R_{im} = \sum_l \frac{\partial\Gamma^l_{im}}{\partial x_l} + \sum_{ls}\Gamma^l_{is}\Gamma^s_{ml} = 0 \tag{3}$$

$$\sqrt{-g} = 1. \tag{3a}$$

Dabei ist

$$\Gamma^l_{ik} = -\left\{\begin{matrix} im \\ l \end{matrix}\right\} \tag{4}$$

gesetzt, welche Größen wir als die »Komponenten« des Gravitationsfeldes bezeichnen.

Ist in dem betrachteten Raume »Materie« vorhanden, so tritt deren Energietensor auf der rechten Seite von (2) bzw. (3) auf. Wir setzen

$$G_{im} = -\varkappa\left(T_{im} - \frac{1}{2}g_{im}T\right), \tag{2a}$$

wobei

$$\sum_s g^{is}T_{is} = \sum_s T^i_i = T \tag{5}$$

gesetzt ist; T ist der Skalar des Energietensors der »Materie«, die rechte Seite von (2a) ein Tensor. Spezialisieren wir wieder das Koordinatensystem in der gewohnten Weise, so erhalten wir an Stelle von (2a) die äquivalenten Gleichungen

$$R_{im} = \sum_l \frac{\partial\Gamma^l_{im}}{\partial x_l} + \sum_{ls}\Gamma^l_{is}\Gamma^s_{ml} = -\varkappa\left(T_{im} - \frac{1}{2}g_{im}T\right) \tag{6}$$

$$\sqrt{-g} = 1. \tag{2a}$$

Wie stets nehmen wir an, daß die Divergenz des Energietensors der Materie im Sinne des allgemeinen Differentialkalkuls verschwinde (Impulsenergiesatz). Bei der Spezialisierung der Koordinatenwahl gemäß (2a) kommt dies darauf hinaus, daß die T_{im} die Bedingungen

$$\sum_i \frac{\partial T^i_i}{\partial x_i} = -\frac{1}{2}\sum_i \frac{\partial g}{\partial x_i}T \tag{7}$$

oder

$$\sum_i \frac{\partial T^i_i}{\partial x_i} = -\sum_{ik}\Gamma^i_{ik}T^k_i \tag{7a}$$

erfüllen sollen.

▲ **Figure 5.37** Two characteristic facsimile pages each from the articles by Hilbert (top) and Einstein (bottom) in which the fundamental equation of general relativity is established. One can see that Hilbert fully acknowledges Einstein's contribution in proposing the basic ideas.

determined more closely by specifying laws which the metrical field of the $g_{\mu\nu}$ satisfy. On physical grounds it was assumed that the metrical field was at the same time the gravitational field.

Since the gravitational field is determined by the configuration of masses and changes with it, the geometric structure of this space is also dependent on physical factors. Thus, according to this theory space is—exactly as Riemann guessed—no longer absolute; its structure depends on physical influences.

(Physical) geometry is no longer an isolated self-contained science like the geometry of Euclid.

The problem of gravitation was thus reduced to a mathematical problem: it was required to find the simplest fundamental equations which are covariant with respect to arbitrary coordinate transformation. This was a well-defined problem that could at least be solved.

I will not speak here of the experimental confirmation of this theory, but explain at once why the theory could not rest permanently satisfied with this success. Gravitation had indeed been deduced from the structure of space, but besides the gravitational field there is also the electromagnetic field. This had, to begin with, to be introduced into the theory as an entity independent of gravitation. Terms which took account of the existence of the electromagnetic field had to be added to the fundamental field equations. But the idea that there exist two structures of space independent of each other, the metric-gravitational and the electromagnetic, was intolerable to the theoretical spirit. We are prompted to the belief that both sorts of field must correspond to a unified structure of space.

—Einstein, *Ideas and Opinions* [pp. 281–285]

5.3 Quantum Theory

5.3.1 *Blackbody Radiation in Classical Physics*

In general, the theory of relativity should be seen not so much as the first achievement of the modern physics of the twentieth century but rather as the culmination of classical physics. Indeed, the theory of relativity did not bring about any fundamental changes in the character of the physical laws, and in the case of the Maxwell equations, it did not even alter their form; it merely provided a framework for the laws with the condition that they must be invariant with respect to the Lorentz transformation. Within the framework of relativity theory, just as before, differential equations yield unambiguous space-time causation rules for the description of physical phenomena. The classical character of the theory of relativity is also indicated by the fact that it is—at least in the derivation given by Einstein—the logical consequence of a typical nineteenth-century philosophical principle that can be traced back at least as far as Leibniz. With some exaggeration, it might even be said that the theory of relativity was born not so much from physical facts as from critical epistemological considerations. In quantum theory and quantum mechanics, the situation is just the reverse. The experimental facts forced new mathematical relationships that correctly described the phenomena, and these relationships in turn led inevitably to a completely new philosophical interpretation of the processes within the atomic realm.

In contrast to the theory of relativity, in the case of quantum theory, one can pinpoint almost to the day when the fundamental idea of the theory first appeared. In hindsight, it seems that the history of science could have taken a simpler path to the discovery of quantum phenomena: With line spectra, nature practically flaunts the existence of well-determined, discrete energy states in the atomic realm. But this is clear to us only today. Physicists in the second half of the nineteenth century could have very well imagined instead that these discrete lines resulted from radiation of oscillators with well-defined frequencies and that there was no connection between the energy of the oscillator and the frequency of the radiation.

Figure 5.39 shows the leading physicists of the era—the founders and promoters of quantum theory and the theory of relativity.

Quotation 5.18, ***continued***

fortune to be present at the meeting of the Royal Society in London when the Astronomer Royal for England announced that the photographic plates of the famous eclipse, as measured by his colleagues in Greenwich Observatory, had verified the prediction of Einstein that rays of light are bent as they pass in the neighborhood of the sun.

The whole atmosphere of tense interest was exactly that of the Greek drama: we were the chorus commenting on the decree of destiny as disclosed in the development of a supreme incident. There was dramatic quality in the very staging: the traditional ceremonial, and in the background the picture of Newton to remind us that the greatest of scientific generalizations was now, after more than two centuries, to receive its first modification.

—A. N. Whitehead, *Science and the Modern World* [p. 13]

Quotation 5.19

For almost four decades, until his death in 1955, Einstein worked on the so-called "unified field theory," i.e. the theory which would unify the gravitational and electromagnetic fields on a single geometrical basis. But, as the years dragged by, the task became more and more hopeless. Every so often Einstein came out with new sets of formulas which, he claimed, would be destined to solve the riddle of the unified field theory, and these complicated tensor expressions were printed on the front pages of the *New York Times* and other papers all over the world. But the formulas always turned out to be unfit for the job, and silence fell again until the next revelation. Theoretical physicists, old and young, gradually lost confidence in the possibility of giving to the electromagnetic field a purely geometrical interpretation. It would have been fine if this could have been done, but Nature cannot be forced to do what is not in her nature. On the other hand, physics was making rapid progress in the newly discovered fields, new fields introduced by wave mechanics took a firm position in science. If a purely geometrical interpretation is given to the electromagnetic field, we would have to subdue meson fields, hyperon fields, and many other new fields, in order to be able to say: physics is nothing but geometry. Einstein himself became more and more touchy about his thesis, and more and more reluctant to discuss these problems with other physicists. During one of his visits to Great Britain in the early thirties, he gave a lecture on the unified field theory at a girls' school in northern England (the blackboard with the complicated tensor formulas has been preserved by the school authorities), but refused to talk in Cambridge University. His attention became more and more occupied by the problems of Zionism and world peace, but his scientific agility remained as sharp as ever.

—George Gamow, *Biography of Physics* [pp. 207–208]



▶ **Figure 5.39** The first Solvay Congress in Brussels in 1911 (the name of the congress is in honor of the Belgian industrialist Ernest Solvay, whose foundation underwrote the costs of the conference; he is in the photograph): The topic of the conference was the Planck radiation law for blackbodies that is at the "origin" of quantum theory. Most of the leading physicists at the turn of the century were working on both the theory of relativity and quantum theory, so in this picture we see the principals of both fields.

Such a group portrait spontaneously excites the viewer's imagination. In old paintings or newer formal photographs of physicists, the artist—the painter or the photographer—typically tries to create an image of the scientist in one of the stereotypes: we see wise, all-knowing and benevolent faces, or the iron will, or the scholar lost in thought searching for new ideas. The emphasis is on the relationship between scientist and science; the connections between the scientists themselves or between the scientist and society are relegated to the background. But look at this group picture! One cannot but wonder what Einstein may be thinking about Poincaré, who is sitting in front of him and with a sly smile is explaining to Mme. Curie what might be a particularly amusing paradox. The otherwise strong-willed Mme. Curie, however, is clearly tired and tries to listen with all her strength, while her thoughts are perhaps with her two daughters whom she left home. Mme. Curie has been a widow for six years now, and she is not only a scientist but also a mother. Hasenöhrl (1874–1915) is gazing out into the world with justified pride; he has achieved significant results on the equivalence of mass and energy. Just this year (1911), he conjectured that there must be a relationship between the phase integral and the quantum state. He can be most proud, however, of his outstanding lectures at the University of Vienna, which he held as Boltzmann's successor, as well as of his exceptional students (among them Schrödinger who graduated in 1910). Four years later, Hasenöhrl would perish in the First World War on the Italian front.

We also see in the photograph a generational contrast. The younger ones have moustaches but no beards. Jeans, whom we shall meet again later, looks with his smooth face and necktie as though he had somehow wandered in from the 1950s.

Two names are familiar to today's physicists, Brillouin and de Broglie, but they refer to others than whom we might expect: Marcel Brillouin is the father of Léon Brillouin (1889–1969), the originator of the theory of energy zones in solids. Maurice de Broglie is the elder brother of Louis de Broglie, who introduced the notion of matter waves. Yet the conference was also crucially important for Louis, at the time 19 years old (see the caption to Figure 5.59).

continued on next page

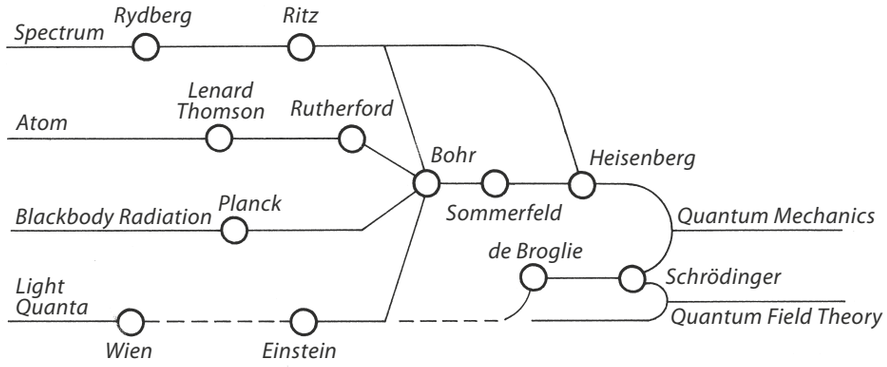
Figure 5.40 shows the path of the creation of quantum theory as well as the relationships between experimental results and the theoretical ideas derived from them. We can see that the line spectra were only of crucial importance for the formulation of the Bohr theory (1913) or rather for the derivation of the Heisenberg equations (1925). We also see that the concept of a discrete energy quantum, as well as the statement regarding the proportionality between energy and frequency, that is, the fundamental relationship $E = h\nu$, arose for the first time in Planck's investigation of blackbody radiation (Figure 5.41).

The problem of blackbody radiation is important both theoretically and for its role in the history of science, so it is worth discussing in detail.

Applying the second law of thermodynamics, in 1859–1860 Kirchhoff succeeded in deriving the following simple but powerful relationship between the emissivity and absorptivity of bodies. If a body at a given temperature and at a given frequency emits energy e per unit time, and the same body at the same temperature and frequency absorbs a fraction a ($0 \le a \le 1$) of the incident energy, then the quotient e/a is independent of the material composition of the body; in other words, it is the same for all bodies. If the emissivities $e_1, e_2, \ldots, e_n$, and absorptivities $a_1, a_2, \ldots, a_n$, are determined for several bodies, then it follows that

$$\frac{e_1}{a_1} = \frac{e_2}{a_2} = \cdots.$$

Simply stated, the law says that if a body radiates very strongly at a given temperature and frequency in comparison to other bodies, then this body will also strongly absorb that radiation in comparison to the others. Kirchhoff proved this law by showing that if the law was not satisfied, then it would be possible to set up an



▲ **Figure 5.40** The main steps on the path to understanding the quantum world (after[Hund 1974]).

experiment that would let heat flow from a cooler body to a warmer body, in violation of the second law of thermodynamics.

It is here that the unique significance of blackbodies becomes clear. By a blackbody we mean a body that absorbs all the radiation striking it, and therefore its absorptivity is $a = 1$. Thus, the relationship given above yields

$$\frac{e_1}{a_1} = \frac{e_2}{a_2} = \cdots = \frac{e_{\text{black}}}{1} = e_{\text{black}},$$

which means that the ratio of emissivity to absorptivity of an arbitrary body is equal to the emissivity of a blackbody. From what we have just said, the quotients e/a are *independent of any material constant* of the body under investigation, so this law is a *universal law of nature.* This universal character gives the radiation law for blackbodies its fundamental significance. Planck himself was intrigued by this universal character (Quotation 5.20). It was soon recognized that blackbody radiation can be realized in a cavity with metallic walls (of platinum–iridium). If the walls of the enclosure are kept at a fixed temperature, then within the cavity, the radiation reaches a state of equilibrium, which is the blackbody radiation. If now an opening is made so small that the thermal equilibrium is not disturbed, then the radiation emitted through this opening can be taken as equal to the blackbody radiation to a very good approximation (Figure 5.42). A very simple and geometrically clearly understandable relationship exists between the energy density of the radiation in the cavity and the intensity of the radiation emitted from the opening in the wall of the enclosure. They are proportional to each other, and it is easy to see that the constant of proportionality contains the velocity c with which the energy propagates. We therefore have

$$B_\nu = \frac{c}{4\pi} u_\nu,$$

where the factor $1/4p$ comes from geometric considerations. In this formula, $u_\nu = u_\nu(\nu, T)$ is the energy within a unit volume of the cavity in a unit frequency range, and $B_\nu(\nu, T)$ is the radiant power emitted perpendicularly per unit area of the cavity opening per unit time into a unit solid angle.

The next significant step forward in the understanding of the laws of blackbody radiation came with the Stefan–Boltzmann law. In 1879, while analyzing measurements taken by Tyndall, Stefan had the following insight: According to those measurements, a particular test body at a temperature of 1473 K radiated 11.7

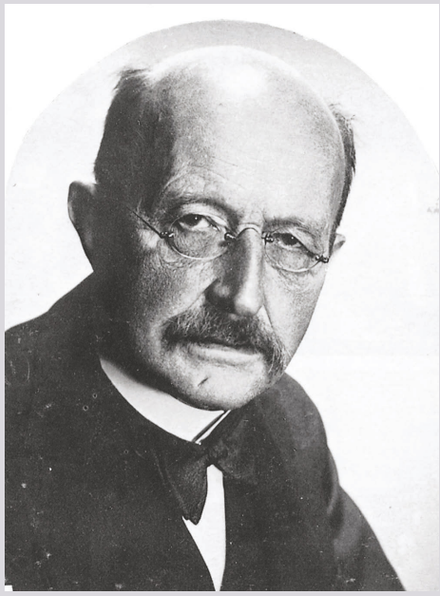
Figure 5.39 *continued*

The year 1911 was also a signal year for Kamerlingh-Onnes, for in that year, he discovered the phenomenon of superconductivity (Plate XXV) in his already world-renowned low-temperature laboratory. Today, we have unjustifiably forgotten such experimental physicists as Knudsen (1871–1948) and Emil Warburg (1846–1931). Knudsen invented the manometer now named for him and investigated gas flow at very low pressures. Warburg discovered the phenomena of hysteresis in ferromagnetic materials and established the relationship between hysteresis loss and the area of the hysteresis loop. Finally, we have Paul Langevin (1872–1946), whose contributions to the development of French and even European physics went far beyond his concrete results (properties of ionized gases; relationship between mobility and recombination, the theory of paramagnets). For example, the twin paradox is due to him. His role is comparable to that of Paul Ehrenfest, who also failed to achieve the kinds of outstanding individual results that might have brought him the greatest international recognition: the Nobel Prize.

Quotation 5.20

My original decision to devote myself to science was a direct result of the discovery which has never ceased to fill me with enthusiasm since my early youth—the comprehension of the far from obvious fact that the laws of human reasoning coincide with the laws governing the sequences of the impressions we receive from the world about us; that, therefore, pure reasoning can enable man to gain an insight into the mechanism of the latter. In this connection, it is of paramount importance that the outside world is something independent from man, something absolute, and the quest for the laws which apply to this absolute appeared to me as the most sublime scientific pursuit of life.

—Max Planck, *Scientific Autobiography...* [p. 13]



▲ **Figure 5.41** Max Planck (1858–1947): Studies in physics at Munich and Berlin. 1879, dissertation on the second law of thermodynamics. After a period at Munich and then at Kiel, called to Berlin in 1889 as Kirchhoff's successor. Planck's most important achievement is the derivation of the law for blackbody radiation, which led to the discovery of energy quantization (1900). He viewed the continuing development of quantum theory with some reservations, but took part in the work on the theory of relativity with great intensity (1907, relativistic kinematics, precise formulation of the energy–mass equivalence). With his five-volume work *Einführung in die theoretische Physik* (Introduction to theoretical physics; 1916–1932), he contributed substantially to the worldwide education of young physicists.

Planck's personal life was overshadowed by a number of tragic events. One of his sons was killed in the First World War, his twin daughters died shortly thereafter, and another son was executed for participating in the assassination plot against Hitler. Toward the end of the war, his house, with irreplaceable items of scientific historical value, was destroyed by fire. His conscientiousness and love of order as well as his instilled loyalty to Prussian tradition and the power of the state made it difficult but did not prevent him from bringing forth revolutionarily new ideas and maintaining his upright and humane behavior under trying circumstances. We quote the following thoughtful sentence from his autobiography:

> A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.
>
> —*Scientific Biography and Other Papers* [pp. 33–34]

times the energy that it radiated at 798 K. Stefan observed that

$$11.7 \approx \left(\frac{1473}{798}\right)^4,$$

and from this he surmised that the total energy radiated by a blackbody is proportional to the fourth power of the temperature, that is,

$$E = \sigma T^4,$$

where the quantity s is the Stefan–Boltzmann constant. It is worthy of note that the test body used by Tyndall in his measurements was not even close to being a blackbody, and in later measurements the factor 11.7 was also found to be completely wrong, the correct value being 18.6. Nevertheless, the faulty measurement had already done excellent service: It led Stefan to the correct law. Following this, Boltzmann was able to give a rigorous thermodynamic explanation. We have already drawn attention to the fact that the laws of macroscopic thermodynamics are valid independently of the assumptions regarding the microscopic structure of matter. Thus, they remain valid in quantum physics as well; indeed, they made it possible to test the validity of new relationships derived from quantum theory.

On the basis of classical physics, another law that could be used as a control function of this kind was established by Wien in 1894. Wien analyzed the following thought experiment: Imagine a cylinder "filled with blackbody radiation" that is sealed by a reflective movable piston. Using the piston, we slowly reduce the volume of the cavity so that the compression is adiabatic, that is, the equilibrium is maintained. The frequency of the waves reflected by the moving piston will change according to the Doppler effect while the energy contents and thus the temperature of the enclosed space will increase due to the work performed by the movement of the piston. A quantitative evaluation of this process allowed Wien to find a relationship describing the dependence of the spectral distribution of the radiation on the temperature. The most important result of his investigations can be understood with the help of Figure 5.43, in which the intensity of the emitted radiation is plotted as a function of wavelength. The intensity distribution has a maximum at a particular wavelength, where the location of the maximum changes with the temperature in the cavity. Wien's displacement law says that the product of the wavelength l_{max} at which the intensity achieves its maximum and the temperature T is constant:

$$\lambda_{max} T = \text{constant}.$$

In other words, with increasing temperature, the maximum in the curve is displaced toward the shorter wavelengths, hence the name of this law.

From these considerations, Wien arrived at the important conclusion that the spectral distribution function must be of the form

$$u_\nu = \nu^3 F\left(\frac{\nu}{T}\right). \qquad (1)$$

From thermodynamic considerations alone, Wien was not able to obtain an explicit expression for the function $F(\nu/T)$. Nevertheless, the formula is still extremely significant. First of all, by simple differentiation, the displacement law can be derived from it. Moreover, it set a goal for the physicists—including Planck—to find a physical law that has this form.

5.3.2 *Planck: Entropy Points the Way to the Solution*

Wien's law was the last valid theoretical statement that classical physics could make about the problem of blackbody radiation. Thereafter, classical physics was just fumbling in the dark. So it was that in 1896, Wien himself attempted to determine the function $F(\nu/T)$ that appears in formula (1). As he was now relying on microphysical conjectures, it is understandable that his results only had a limited range of validity. In particular, Wien assumed that there must be some sort of connection between the radiation law and Maxwell's velocity distribution, where the velocity should be replaced by a quantity depending only on the frequency. Considering that the law would have to have the form of equation (1), he finally obtained the explicit form for the Wien radiation law:

$$u_\nu = \alpha \nu^3 \, \mathrm{e}^{-\beta\nu/T} . \qquad (2)$$

Planck himself believed that the Wien radiation law was correct. Measurements also seemed to confirm its validity. Therefore, Planck's efforts during the years 1897 to 1899 were focused on providing this law with a more serious theoretical foundation than the one that had been given by Wien himself. To this end, Planck made use of the fact that blackbody radiation develops in a cavity independently of the makeup of the cavity walls, so that he was free to choose as the wall material the simplest imaginable model: Let the wall consist of oscillators with frequency ν, or, in other words, of positive and negative charges that are coupled by elastic forces. The radiation of the oscillators and their interaction with the radiation field would be described by the laws of electrodynamics.

Here we have to note that even if Planck's argument in itself was not revolutionary, he employed the most recent results of Hertz's theory of electromagnetic waves for a completely new situation. With his model, Planck derived the following relationship between the energy density of the radiation and the average energy of the oscillators in the cavity wall:

$$u_\nu = \frac{8\pi\nu^2}{c^3} U(\nu, T). \qquad (3)$$

This is to be understood as saying that in its interaction with the oscillators, the electric field of the radiation excites them in dissimilar ways: in equilibrium state some are given greater amplitudes, some smaller, but in a definite distribution that is independent of time. If we add the energies of all the oscillators and divide by the number of oscillators, we obtain the average energy per oscillator $U(\nu, T)$.

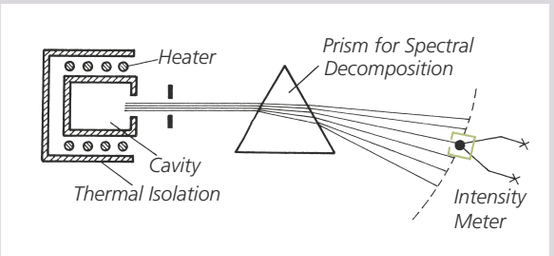
At this point, let us pause a moment. A very general law of classical statistics, the equipartition law, practically cries out to be applied here. According to this law, for each degree of freedom, the average kinetic energy is $\frac{1}{2}kT$, and further, taking into account that the oscillators also have potential energy whose average is the same, we obtain

$$U(\nu, T) = kT.$$

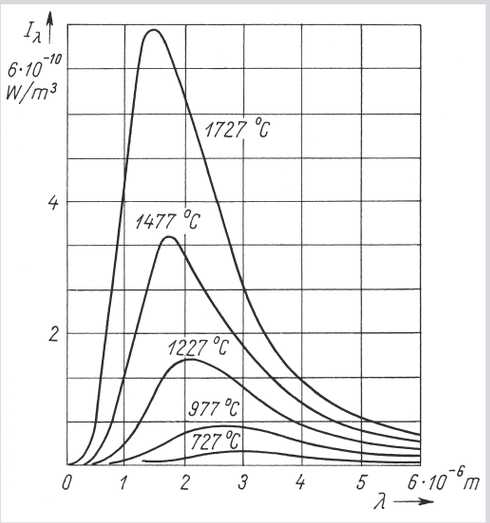
For the energy density of the radiation, we could then obtain

$$u_\nu = \frac{8\pi\nu^2}{c^3} kT.$$

However, Planck did not take this route. In June 1900, Rayleigh and Jeans derived this very law, now known as the Rayleigh–Jeans law. In contrast to our



▲ **Figure 5.42** Design of an experimental setup for measuring the intensity distribution of blackbody radiation.



▲ **Figure 5.43** Graphs for Wien's displacement law.

The discoverer of this law, Wilhelm Wien (1864–1928)—see also Figure 5.39—began his career as Helmholtz's assistant. After 20 years teaching at Würzburg, he worked from 1920 as an experimental physicist in Munich. Wien belongs to the group of physicists who made significant contributions to both theoretical and experimental physics. With his research into the motion of charged particles under the simultaneous influence of electric and magnetic fields, he laid the foundations of mass spectrometry; and with the post-deflection acceleration of canal rays (ion beams) in vacuum, he paved the way for nuclear reactions using accelerators.

If we compare this figure with Figure 5.45, we see that the shape of the graphs near their maximums can be understood only with the help of quantum theory. The question then to be asked is if it is possible that classical physics can make assertions about maximal values that also hold true for quantum mechanics. The question was formulated in this form in 1906 by Paul Ehrenfest and then answered with the help of the adiabatic hypothesis, which in its general form says that we have to seek quantities that remain invariant under adiabatic changes to the system, that is to say, changes are slow with respect to internal activity. These *adiabatic invariants* are precisely the quantities that need to be quantized. The law $U = \nu\Phi(\nu/T)$ introduced in Section 5.3.2 can be written in the form $U/\nu = \Phi(\nu/T)$, where both U/ν and ν/T are adiabatic invariants of the system. They then remain constant during the thought experiment for the derivation of the displacement law; if the energy at the initial state has the quantized values $U = nh\nu$, then after a displacement of the piston, the relation $U' = nh\nu'$ still remains satisfied.

continued on next page

Figure 5.43 *continued*

The adiabatic hypothesis can be visualized with the help of a simple oscillatory system: a pendulum swinging on a thread. We can adiabatically change the pendulum length l, which determines the frequency according to the formula $\nu = \frac{1}{2\pi}\sqrt{g/l}$, by slowly winding up the thread on a spool with a very small radius at the suspension point. Here "slowly" means that the pendulum can make many oscillations during the time that the length changes appreciably. Due to the change in the pendulum length, the frequency will change, but the energy of the pendulum will also change: by raising the pendulum bob we are putting work into the system. If the initial energy was E_0 and the frequency ν_0, and at some arbitrary later time they are E and ν, then it can be shown relatively easily that $E_0/\nu_0 = E/\nu$. Thus, we have found the adiabatic invariant that must be quantized: $E/\nu = nh$, or $E = nh\nu$.



▲ **Figure 5.44** Lord Rayleigh, John William Strutt (1842–1919): Studies in applied mathematics and physics at Cambridge under Stokes. Worked primarily in his private laboratory at his country estate, but was also, from 1879 to 1884, director of the Cavendish Laboratory as successor to Maxwell. The principal object of his research was acoustics (*The Theory of Sound*, 1878, 1879). In 1904, he received the Nobel Prize for his discovery of argon. One of his better-known contributions is an explanation of the blue color of the sky using the Rayleigh scattering law. The intensity of scattering is inversely proportional to the fourth power of the wavelength, and therefore the atmosphere scatters the blue component of white sunlight with greater intensity, which is what we see.

Bohr mentions Rayleigh's comment on Bohr's theory:

> In my young days I took many views very strongly and among them that a man who has passed his sixtieth year ought not to express himself about modern ideas. Although I must confess that today I do not take this view quite so strongly, I keep it strongly enough not to take part in this discussion!
>
> —Lord Rayleigh [Bohr 1963, p. 44]

presentation above, Rayleigh (Figure 5.44) did not obtain this relationship via the laws of classical electrodynamics; instead, he simply counted the possible number of standing waves in a cavity with the wave frequencies in a tight range around the frequency ν. We must add at once that the Rayleigh–Jeans law was still faulty.

Let's return to the work that Planck undertook a year earlier. Planck wanted to keep within the framework of macroscopic thermodynamics and therefore based his further investigations on entropy as the crucial quantity (Quotation 21). At this time, he still believed in the correctness of Wien's law as given in equation (2) and was trying to find a thermodynamic derivation. Without any specific justification, he gave the formula

$$S = \frac{U}{av}\ln\frac{U}{\mathrm{e}bv} \tag{4}$$

for the entropy of the oscillators (a and b are constants; e is the basis of the natural logarithm). If we apply the thermodynamic relationship

$$\frac{\partial S}{\partial U} = \frac{1}{T}$$

as well as the relationship between the energy density in the cavity and the average energy per oscillator,

$$u_\nu = \frac{8\pi\nu^2}{c^3}U,$$

we immediately obtain the Wien radiation law. In all probability, Planck reached the formula for the entropy by traversing the path we have just described backward, and, knowing the result that he wanted, found the appropriate starting function that worked, doubtless only after several tries. Naturally, Planck was not satisfied with the argument and saw the entire process for the time being as completely formal. For Planck, computing the second derivative of entropy with respect to energy was particularly important because in the equilibrium state, the entropy must be at a maximum, and the second derivative plays a crucial role in the determination of the maximum. From equation (4), we obtain at once

$$\frac{\partial^2 S}{\partial U^2} = \frac{\text{constant}}{U}.$$

In the meantime, however, the experimental physicists were not idle, either. Otto Lummer (1860–1925), Pringsheim, Rubens, Kurlbaum, and others verified the radiation law experimentally with ever increasing precision and in an ever broader frequency range. These measurements demonstrated unambiguously that the Wien law, or rather the Wien–Planck law, is not a good one because it conforms to the experimental results only at the higher frequencies. At this time, the Rayleigh–Jeans law also appeared on the scene, but in contrast, this one described the phenomena correctly only for the smaller frequencies (Figure 5.45, Quotation 5.22).

Thus, in the autumn of 1900, the situation regarding blackbody radiation was as follows:

There exist two laws derived theoretically: the Wien–Planck law and the Rayleigh–Jeans law. The theoretical derivation of the Wien–Planck law is problematic, and agreement with experiment is achieved only for high frequencies. For small frequencies, the law gives a result that deviates sharply from experiment, which was described at the time as an *infrared catastrophe*.

The Rayleigh–Jeans law stands on firmer theoretical ground, as it is based on a very general principle of classical statistics, the equipartition law. However, this law also describes only a limited frequency range correctly. This law suffers an *ultraviolet catastrophe*.

Planck took stock of the new situation and accordingly constructed an entropy function for the Raleigh–Jeans law as well. For our purposes, the important thing

is only that the second derivative of this function is inversely proportional to the square of the energy. So we can assemble the following schema:

Wein – Planck	Rayleigh – Jeans
$u_\nu = \alpha \nu^3 e^{-\beta \nu / T}$	$u_\nu = \frac{8\pi}{c^3} \nu^2 kT$
$\frac{\partial^2 S}{\partial U^2} = \frac{\text{constant}}{U}$	$\frac{\partial^2 S}{\partial U^2} = \frac{\text{constant}}{U^2}$

$$\frac{\partial^2 S}{\partial U^2} = \frac{a}{U(U+b)}$$

How can the two solutions be combined so that a single expression agrees with experiment in the entire frequency range? Following a lucky hunch, Planck combined the second derivatives of the entropy functions in the simplest possible way. This combination is given at the end of the above schema. It can be seen that at small values of U, it is equal to one of the expressions, and at large values, to the other one.

Some call Planck's insight the most productive interpolation ever in the history of physics. Still others consider this the birth of quantum theory.

From the relationship

$$\frac{\partial S}{\partial U} = a' \ln \frac{U+b}{U}$$

we obtain, with the help of the well-known thermodynamic equation

$$\frac{1}{T} = \frac{\partial S}{\partial U}$$

for the average kinetic energy of the oscillators, the expression

$$U = \frac{b}{e^{1/(a'T)} - 1}.$$

If we consider in addition that equations (1) and (3) compel U to take the form

$$U = \nu \Phi\left(\frac{\nu}{T}\right),$$

then it follows that $b = c\nu$ and $1/a' = b\nu$, and we finally obtain

$$u_\nu = A \frac{\nu^3}{e^{\beta \nu / T} - 1}$$

for the energy density of the blackbody radiation.

Planck also consulted with the experimental physicists, in particular with Rubens. They determined that the law advanced by Planck correctly described experimental results *in the entire frequency range*.

Let us compare the laws once again. On October 19, 1900, the situation was the following (after [Hund 1967]):

Quotation 5.21

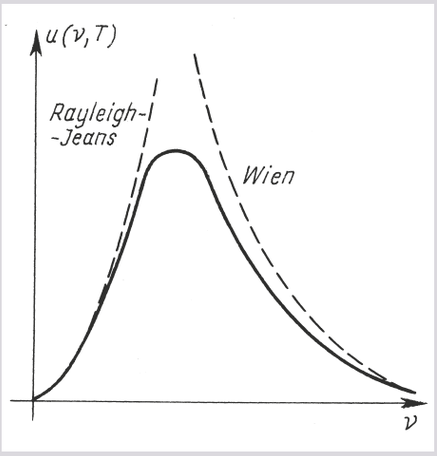
It was an odd jest of fate that a circumstance which on former occasions I had found unpleasant, namely, the lack of interest of my colleagues in the direction taken by my investigations, now turned out to be an outright boon. While a host of outstanding physicists worked on the problem of spectral energy distribution, both from the experimental and theoretical aspect, every one of them directed his efforts solely toward exhibiting the dependence of the intensity of radiation on the temperature. On the other hand, I suspected that the fundamental connection lies in the dependence of entropy upon energy. As the significance of the concept of entropy had not yet come to be fully appreciated, nobody paid any attention to the method adopted by me, and I could work out my calculations completely at my leisure, with absolute thoroughness, without fear of interference or competition.

—Planck, *Scientific Autobiography...* [p. 38]

Quotation 5.22

[W]e may assert without exaggeration that there was no period in the history of physics when theory and experience were linked so closely together as there are now. Conversely, it was the facts learned from experiments that shook and finally overthrew the classical theory. Each new idea and each new step were suggested to investigators, where it was not actually thrust upon them, as the result of measurements. The theory of relativity was led up to by Michelson's experiments on optical interference, and the quantum theory by Lummer's, Pringsheim's, Ruben's and Kurlbaum's measurements of the spectral distribution of energy, by Lenard's experiments on the photoelectric effect, and by Franck and Hertz's experiments on the impact of electrons. It would lead me too far if I were to enter on the numerous and surprising results which have compelled physical theory to abandon the classical standpoint and to enter on a definite new course.

—Planck, *The Universe in the Light of Modern Physics* [pp. 56–57]



▲ **Figure 5.45** Regions of validity for the laws of Wien and of Rayleigh–Jeans.

The solid line shows the results of the measurements and also corresponds to Planck's formula.

James H. Jeans (1877–1946; see also Figure 5.39): Taught at Cambridge and Princeton (1904–1012). From 1923 to 1944: Mount Wilson Observatory. His main interest was astrophysics. He is the originator of the "catastrophe" theory of the origin of the solar system, according to which the planets were created as the result of a collision between the Sun and another star. Jeans studied the energy of stars, double stars, and spiral nebulae, and he is responsible for the theory of continuous creation. He wrote a number of popular-scientific works that were translated into a number of languages. His textbooks have also found a wide audience.

Even in 1911, Jeans had not yet accepted Planck's theory, but the success of Bohr's theory convinced him, and he became an advocate of the new ideas.

	Wein-Planck	Planck	Rayleigh—Jeans
	$u_\nu = \alpha\nu^3 \, e^{-\beta\nu/T}$	$u_\nu = A\frac{\nu^3}{e^{\beta\nu/T}-1}$	$u_\nu = \frac{8\pi\nu^2}{c^3}kT$
Theoretical basis:	unsatisfactory	lacking	good
Region of validity:	high frequencies	all frequencies	low frequencies

5.3.3 *The Appearance of the Energy Quantum*

Experimental and theoretical physicists, to the extent that they were informed of this provisional situation, as well as Planck himself, were entirely dissatisfied by the utter lack of theoretical foundation for the new law. At this time, Planck, who according to his recollection was going through the most intensive and tiring period of his life (from October 10 to December 14, 1900), now chose a new direction. Although he stayed with the entropy concept, he abandoned phenomenological thermodynamics. Following Boltzmann, he proceeded from the relationship between thermodynamic probability and entropy that we have already seen (Section 4.5.11):

$$S = k \ln W.$$

As we know, here W is the thermodynamic probability of a macrostate; and we know further that this probability is proportional to the number of realization possibilities of that state.

However, in order to be able to apply the combinatorial procedure that serves to determine the number of realization possibilities of a macrostate, *the total energy must be divided into a finite number of elements. Consequently, a given oscillator may possess some integer multiple of these energy elements of some finite size.*

At this time and in this form, for the first time in the history of physics, the notions of energy element or energy quantum appeared. We quote from Planck's work that he presented on December 14, 1900 [Planck 1958, p. 700]:

> In a diathermal medium enclosed by reflective walls with speed of propagation of light c, a large number of linear monochromatically vibrating resonators are located an appropriate distance from one another, to be precise, N with vibrational frequency ν (per second), N' with vibrational frequency ν', N'' with vibrational frequency ν'', etc., where all the N are large numbers. Suppose the system contains a given quantity of energy: the total energy E_t, in ergs, which occurs in part in the medium as progressive waves, in part in the resonators as vibration of the same. The question is how in the stationary state this energy is distributed among the vibrations of the resonators and among the individual colors of the radiation present in the medium and what temperature the entire system then possesses.
>
> To answer this question, we first consider only the vibrations of the resonators and provisionally assign them particular arbitrary energies, namely to the N resonators ν say the energy E, the N' resonators ν' the energy E', etc.

Of course, the sum

$$E + E' + E'' + \cdots = E_0$$

must be less than E_t. The difference $E_t - E_o$ is associated with the radiation located in the medium. Now the distribution of the energy is to be carried out for the individual resonators within each type, first the distribution of the energy E to the N resonators with vibration number ν. If E is seen as an infinitely divisible quantity, the distribution to infinitely many types is possible. However, we consider—and this is the crucial point of the entire calculation—E as being composed of a specific number of finite equal parts and make use in this regard of the constant of nature $h = 6.55 \times 10^{-27}$ [erg $\times$ sec]. This constant multiplied by the common vibration number ν of the resonators yields the energy element ε in ergs, and through division of E by ε we obtain the number P of energy elements that are to be allocated among the N resonators. If the quotient thus calculated is not an integer, then one takes for P a nearby integer.

It is now plausible that the distribution of P energy elements among the N resonators can be carried out in only a finite, strictly determined, number of ways. We shall call each such distribution, adopting an expression used by Boltzmann for a similar concept, a "complexion." If we label the resonators 1, 2, 3, through N, write the numbers in sequence, and write under each one the number of associated energy elements, then we obtain for each complexion a symbol of the following form:

1	2	3	4	5	6	7	8	9	10
7	38	11	0	9	2	20	4	4	5

Here we have assumed $N = 10$, $P = 100$. The number of all possible complexions is clearly equal to the number of all possible arrangements of numbers that one can obtain for the bottom row for particular values of N and P. To avoid any possible misunderstanding, we note that two complexions will be considered different if the same numbers appear but in a different order. From combinatorics, we can write the number of all possible complexions as

$$\frac{N(N+1)(N+2)\cdots(N+P-1)}{1\cdot 2\cdot 3\cdots P} = \frac{(N+P-1)!}{(N-1)!P!},$$

which to good approximation is equal to

$$\frac{(N+P)^{N+P}}{N^N P^P}.$$

Having determined the number of possible complexions for every type of resonator and multiplying these together, we get the total number of complexions $\mathfrak{R}$ that belong to a given E, E', E'' … energy distribution. For a different energy distribution there will be a different R. We now search for the distribution to which the greatest number of complexions belong. Let us call this number $\mathfrak{R}_0$. This distribution then characterizes the stationary state of the resonators.

If we now form the quotients E/N, E'/N', E''/N'' …, we obtain the energies $U_\nu, U_\nu', U_\nu'', \ldots$ of each of the corresponding resonators. We can also determine the energy density of the radiation field with the help of the relationship

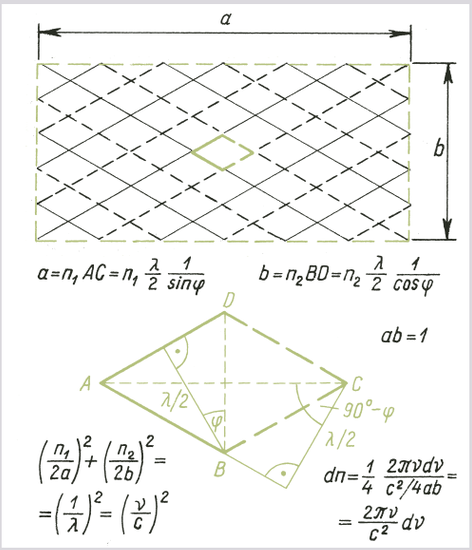
$$u_\nu \, \mathrm{d}\nu = \frac{8\pi\nu^2}{c^3} U_\nu \, \mathrm{d}\nu.$$

Quotation 5.23

I tried immediately to weld the elementary quantum of action h somehow into the framework of the classical theory. But in the face of all such attempts, this constant showed itself to be obdurate. So long as it could be regarded as infinitesimally small, i.e. when dealing with higher energies and longer periods of time, everything was in perfect order. But in the general case difficulties would arise at one point or another, difficulties which became more noticeable as higher frequencies were taken into consideration. The failure of every attempt to bridge this obstacle soon made it evident that the elementary quantum of action plays a fundamental part in atomic physics, and that its introduction opened up a new era in natural science. For it heralded the advent of something entirely unprecedented, and was destined to remodel basically the physical outlook and thinking of man, which, ever since Leibniz and Newton laid the groundwork for infinitesimal calculus, were founded on the assumption that all causal interactions are continuous.

My futile attempts to fit the elementary quantum of action somehow into the classical theory continued for a number of years, and they cost me a great deal of effort. Many of my colleagues saw in this something bordering on a tragedy. But I feel differently about it. For the thorough enlightenment I thus received was all the more valuable. I now knew for a fact that the elementary quantum of action played a far more significant part in physics than I had originally been inclined to suspect, and this recognition made me see clearly the need for the introduction of totally new methods of analysis and reasoning in the treatment of atomic problems. The development of such methods—in which, however, I could no longer take an active part—was advanced mainly by the efforts of Niels Bohr and Erwin Schrödinger.

—Planck, *Scientific Autobiography...* [p. 44–45]



▲ **Figure 5.46** Determination of the number of possible vibrational states per frequency interval in some space. We may imagine the space as being filled either with a solid or with air, in which case we are discussing acoustic waves, or we could also consider electric vibrational states in an evacuated cavity with metallic walls, or even an electron enclosed within the cavity, where we are interested in the electron's quantum states. In the one-dimensional and two-dimensional cases, the vibrational states and the counting of them is easily illustrated, whereas in three-dimensional problems we have to rely on the mathematical derivation. If we keep in mind that according to the counting procedure given by RAYLEIGH, in three dimensions there are $(8\pi\nu^2/c^3)d\nu$ possible vibrational states in the frequency interval between ν and $d\nu$, then the following three derivations of Planck's formula can be easily followed:

1. DEBYE in 1910 gave the derivation that is usually found in modern textbooks. The average energy of a vibrational state of frequency ν at temperature T is given by

$$U(\nu,T) = \frac{\sum_n nh\nu\, e^{-nh\nu/kT}}{\sum_n e^{-nh\nu/kT}} = \frac{h\nu}{e^{h\nu/kT}-1}.$$

   For $h\nu \ll kT$ the exponential function can be replaced by its series expansion and we obtain the classical expression $h\nu/(h\nu/kT) = kT$. If we now multiply the average energy by the number of possible states $(8\pi\nu^2/c^3)d\nu$, then we obtain Planck's formula.

2. The derivation given by EINSTEIN (1916/1917) is given its own discussion in the main text of this book (Section 5.3.6) due to its special significance.

3. In 1924, BOSE considered photons to be independent particles of energy $h\nu$ and applied to them the statistical methods of kinetic gas theory, with the significant difference that in the counting of the possible distributions, he considered the photons to be indistinguishable from one another, in other words, the

continued on next page

Finally, we can introduce the temperature (J) by using the relation

$$\frac{1}{\vartheta} = k\frac{d\log\Re_0}{dE_0}.$$

However, in his foundational article "Über das Gesetz der Energieverteilung im Normalspectrum" (On the Law of Energy Distribution in the Normal Spectrum), *Annalen der Physik* 4:9 (1901), 564–566, PLANCK does not follow this program. Rather, he writes down the entropy of N resonators with frequency ν in the form

$$S_N = k\log\Re.$$

Using Stirling's approximation formula

$$\Re = \frac{(N+P)^{N+P}}{N^N P^P},$$

the above expression can be reformulated as

$$S_N = k[(N+P)\ln(N+P) - N\ln N - P\ln P].$$

After a simple rearrangement, one obtains

$$S_N = kN\left[\left(1+\frac{U}{\varepsilon}\right)\ln\left(1+\frac{U}{\varepsilon}\right) - \frac{U}{\varepsilon}\ln\frac{U}{\varepsilon}\right],$$

which yields the entropy associated with a single oscillator,

$$S = \frac{S_N}{N} = k\left[\left(1+\frac{U}{\varepsilon}\right)\ln\left(1+\frac{U}{\varepsilon}\right) - \frac{U}{\varepsilon}\ln\frac{U}{\varepsilon}\right].$$

If we now also consider the relationship known from thermodynamics between the derivative with respect to the energy and the absolute temperature, without any difficulty we obtain

$$U = \frac{\varepsilon}{e^{\varepsilon/kt}-1}.$$

From this relationship, using

$$\varepsilon = h\nu \quad \text{and} \quad u_\nu = \frac{8\pi\nu^2}{c^3}U,$$

we obtain Planck's formula

$$u_\nu = \frac{8\pi\nu^2}{c^3}\frac{h\nu}{e^{h\nu/kT}-1}$$

in its final form.

Planck's radiation formula in itself has great persuasive power. Consider the following:

1. Directly, that is, using a series development, it yields the Rayleigh–Jeans formula for low frequencies, whereas for high frequencies, it yields Wien's law.
2. If we use Planck's formula to calculate the radiated power over the entire frequency range, we obtain the Stefan–Boltzmann law.
3. Wien's displacement law can be derived by simple differentiation.
4. With the help of this formula, PLANCK succeeded in bringing the fundamental natural constants together into a coherent system. Thus, given the known constants of the Stefan–Bolztmann law and the Wien displacement law, we

can determine the values for Planck's constant and Boltzmann's constant:

$$\left.\begin{matrix}\sigma \\ \lambda_m T\end{matrix}\right] \rightarrow \begin{matrix} h = 6.55\times10^{-27}\ \text{erg s}, \\ k = 1.346\times10^{-16}\ \text{erg/K}. \end{matrix}$$

From the Boltzmann constant and the gas constant, we can then obtain the Loschmidt number; and finally, the Loschmidt number and the Faraday constant yield the charge of the electron:

$$\left.\begin{matrix}k \\ R\end{matrix}\right] \rightarrow L = 6.17\times10^{23}\ /\ \text{mol}, \qquad \left.\begin{matrix}L \\ F\end{matrix}\right] \rightarrow e = 4.69\times10^{-10}\ \text{esu}.$$

5. Most importantly, Planck's radiation law satisfies the requirement that it was designed to satisfy: It describes the spectral energy distribution of the blackbody radiation correctly in the entire frequency range, in complete agreement with experiment.

After all this, we might suppose that the theory was immediately accepted as a complete success. Yet that was not the case. In a summary of the discoveries of 1900 published in the journal *Fortschritten der Physik*, Planck's work was described only in broad outline. Planck's great rival, Jeans, did not even mention Planck's formula in his 1904 book *Dynamical Theory of Gases*. Lorentz, as late as 1909, was still reproaching Planck for not having proposed any justification for the assumptions underlying his theory.

Yet, what is most surprising is that Planck himself did not view the case as closed, and for many years attempted to find a suitable explanation of the radiation formula within the confines of classical physics (Quotation 5.23). Nevertheless, according to his son, Planck clearly understood the importance of his breakthrough, which was "perhaps comparable only to that of Newton." The introduction of the novel concept of the discrete energy quantum was not the only reason for aversion to the acceptance of the theory. As pointed out by Einstein in 1906, the theory had some heterogeneous elements. For instance, to establish the connection between the average energy and the field energy, Planck used the equations of classical electrodynamics with the assumption of continuous energy changes. At the same time, in the expression for the average energy, he assumed that the energy is quantized. And there is yet another inconsistency in Planck's train of thought. Boltzmann derived the distribution corresponding to the equilibrium state from the requirement that this distribution make the thermodynamic probability a maximum. Planck used a different counting method and did not arrive at his final formula by searching for a maximum.

The first two decades of the twentieth century resolved the contradictions and clarified what was vague. As a result of this work, the theory's derivation that we encounter in today's textbooks is much simpler and is based on different considerations (Figure 5.46).

A derivation of the radiation formula by Einstein in 1916 is worth special mention because this derivation—with certain changes—constituting as it does the theoretical basis for all laser technology, can be found in the introduction of almost all modern textbooks on quantum electronics. However, this derivation took full account of the progress made until that time, so we will return to its discussion later.

Figure 5.46 *continued*

exchange of two photons does not count as a different state. He also derived the factor $8\pi\nu^2/c^3$ from the new assumption that the size of an elementary cell of the phase space should be equal to h^3. To distinguish it from the classical Maxwell–Boltzmann statistics, these new statistics are called the Bose–Einstein statistics and particles obeying these statistics are called *bosons*.

Satyendra Nath Bose (1894–1974): Diploma in 1915 from the University of Calcutta; 1924–1925: Work with Mme. Curie in Paris, then returned to India and a teaching position at Dacca and at Calcutta. Sent his work to Einstein, who then extended the ideas therein to monatomic gases.

The fundamental character of the quantum of action was emphasized from another point of view by Planck himself. In 1906, he wrote this in his book *Vorlesungen über die Theorie der Wärmestrahlung* (Lectures on the Theory of Thermal Radiation):

> … that using both … constants *h* and *k*, one has the possibility to set units for length, mass, time, and temperature that independently of special objects or substances, maintain their meaning for all times and for all cultures, even those beyond the Earth and beyond mankind, and which therefore may be called "natural units of measurement."

These words sound even more ambitious than the declaration of the French National Assembly, which collected a decades-long study into a law for the creation of a unified system of measurement, the metric system. This system was to hold "for all peoples and for all times," in that the unit of length could be determined by the same object as the unit of time, namely, Earth.

The great circle between Dunkirk and Montjuich (near Barcelona) was measured with great precision (J. B. S. Delambre and P. F. A. Méchain), and from this arose the definition of the unit of length (legally adopted in 1799, for compulsory use starting in 1801): The meter is one ten-millionth of the meridian quadrant through Paris. This length (international prototype meter) was realized as the distance between two etched lines in a platinum–iridium bar. Through a refinement of measuring techniques, a discrepancy was discovered between the unit as set by law and the international prototype. From 1889, the latter of these was used as the definition. The meter is the distance between the etched lines on the international prototype. The suggestion was made by J. Babinet and Maxwell to define the unit length in terms of the wavelength of certain spectral lines. In 1923, Michelson, Fabry, and Perot proposed a measurement system that was based on the red line of cadmium. It served as a norm for comparison from 1927 to 1960. Then the unit of length was redefined by law in terms of the orange–red line of the krypton isotope Kr-86.

Since 1983, primarily through the research and lobbying of Zoltán Bay, the definition of the unit of length has been based on the velocity of light in vacuum because it is spatially and temporally constant, independent of the state of motion:

The meter is the distance that light travels in 1/299792458 seconds.

This implies, of course, that the velocity of light in vacuum has, according to this definition, the exact value c = 299792458 m/s. One cannot speak here of an error in measurement. And now back to Planck:

continued on next page

Figure 5.46 *continued*

The means of establishing the four units for length, mass, time, and temperature are given by the two mentioned constants h and k, and further by the size of the light propagation speed c in vacuum and by the gravitation constant f:

$$h = 6.548 \times 10^{-27} \text{ g cm}^2 / \text{s},$$

$$k = 1.346 \times 10^{-16} \text{ g cm}^2 / \text{s}^2 \text{ K}^{-1},$$

$$c = 3 \times 10^{10} \text{ cm} / \text{s},$$

$$f = 6.685 \times 10^{-8} \text{ cm}^3 / \text{g s}^2.$$

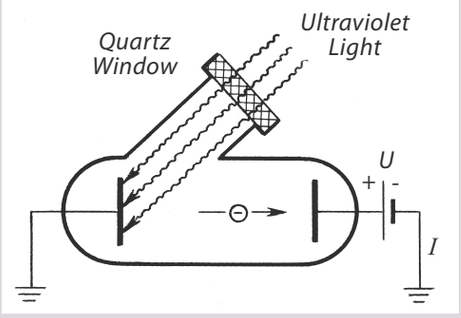
If one now chooses the "natural units" in such a way that in the new system of measurement, each of the above four constants assumes the value 1, then one obtains the units

$$\sqrt{\frac{fh}{c^3}} = 4.03 \times 10^{-33} \text{ cm}, \qquad \sqrt{\frac{ch}{f}} = 5.42 \times 10^{-5} \text{ g},$$

$$\sqrt{\frac{fh}{c^5}} = 1.34 \times 10^{-43} \text{ s}, \qquad \frac{1}{k}\sqrt{\frac{c^5 h}{f}} = 3.63 \times 10^{32} \text{ °K}.$$

These quantities retain their natural meaning as long as the laws of gravitation, propagation of light in vacuum, and the two principal laws of thermodynamics remain valid; they must therefore, according to the most varied intelligence and measured by the most varied methods, always be those given here.

The quantities introduced here play an important role in today's cosmology and particle physics under the names of Planck mass and Planck time. The time from 0 to 10^{-43} seconds after the big bang is known as the Planck era, when (probably) all four fundamental interactions were combined into a single interaction (see also Figures 5.155, 5.187).



▲ **Figure 5.47** The photoelectric effect.

The number of electrons that exit the photocathode due to the light can be determined from the strength of the current in the anode circuit. To determine the energy of the electrons, the voltage on the anode is made increasingly negative until the current stops flowing. This voltage is characteristic of the energy of the electrons exiting the cathode.

Philipp Lenard (1862–1947): Studies in Heidelberg; as assistant to Heinrich Hertz from 1886 investigated cathode rays; 1894 in Breslau, and from 1907, professor of physics in Heidelberg.

continued on next page

It must be emphasized that the classical equipartition principle remains applicable in many cases, such as, for example, in Brownian motion. Following Einstein (1905) and Smoluchowski, for the stochastic motion of the suspended particle, one has

$$\frac{1}{2} m \overline{x^2} = \frac{1}{2} kT.$$

From this quantity, usable for empirical confirmation, the (average) distance a particle moves from its initial position in time t can be calculated as (where r is the radius of the particle, h the viscosity of the fluid):

$$\overline{x^2} = (kT \cdot 3\pi\eta r)\tau.$$

5.3.4 *Einstein: Light Is Also Quantized*

Einstein was the first to recognize the new possibilities in the idea of the quantum. In 1905, in the same issue of *Annalen der Physik* in which his groundbreaking work on relativity also appeared, he gave an explanation of the photoelectric effect. To wit, Lenard made the completely unexpected experimental discovery in 1902 that the energy of an electron emitted from the surface of a metal through the action of light *is independent of the intensity* of the incident light, as was to be expected from the classical theory, but instead, that it depends *on the color,* or more precisely on the frequency of the light. The intensity of the light determines only the *number* of electrons emitted (Figure 5.47). Einstein theorized that light consists of energy quanta $h\nu$, that is, of *photons*, and viewed the emission of an electron as the result of an interaction between the photon and an electron in the metal during which the electron takes on the energy of the photon. In this way it is easy to understand why the energy of the emitted electrons increases with the frequency of the photons. Today, this explanation seems obvious, and we might even be inclined to think that it is a natural consequence of Planck's quantum hypothesis. But that is not at all the case, for in his original article, Planck *had not yet spoken of the quantum nature of radiation, that is, of light*: Only the energy of the oscillators forming the walls of the cavity containing the blackbody radiation assumed discrete values. Such phenomena could be envisioned even if the radiation energy were continuous: Perhaps some accumulation process controls the stepwise changes in the oscillator energies. In this connection, we can also cite the document signed by Planck and his colleagues in which they recommend Einstein's election to the Prussian Academy of Sciences. After listing Einstein's scientific contributions, they apologetically add that everyone is liable at times to overshoot the mark, as, for example, in the case of Einstein's hypothesis of light quanta. Almost a decade later, Einstein was awarded the Nobel Prize primarily for this work.

Another great contradiction of the classical statistical theory shown in Figure 5.1 was the discrepancy between the theoretically computed and experimentally measured specific heats of solids. In 1906 Einstein was the first to apply the quantum hypothesis to this problem. His results were later refined by Debye.

5.3.5 *Bohr: The "Classical" Quantum Theory of the Atom*

As we mentioned in the introduction to this section, the quantum hypothesis was created to explain a very complicated phenomenon in which nature's predilection for whole numbers was only very indirectly apparent. In contrast, this favoring of discrete values is much more directly evident in atomic line spectra. It was on the strength of its explanation of the atomic spectra that quantum theory took over

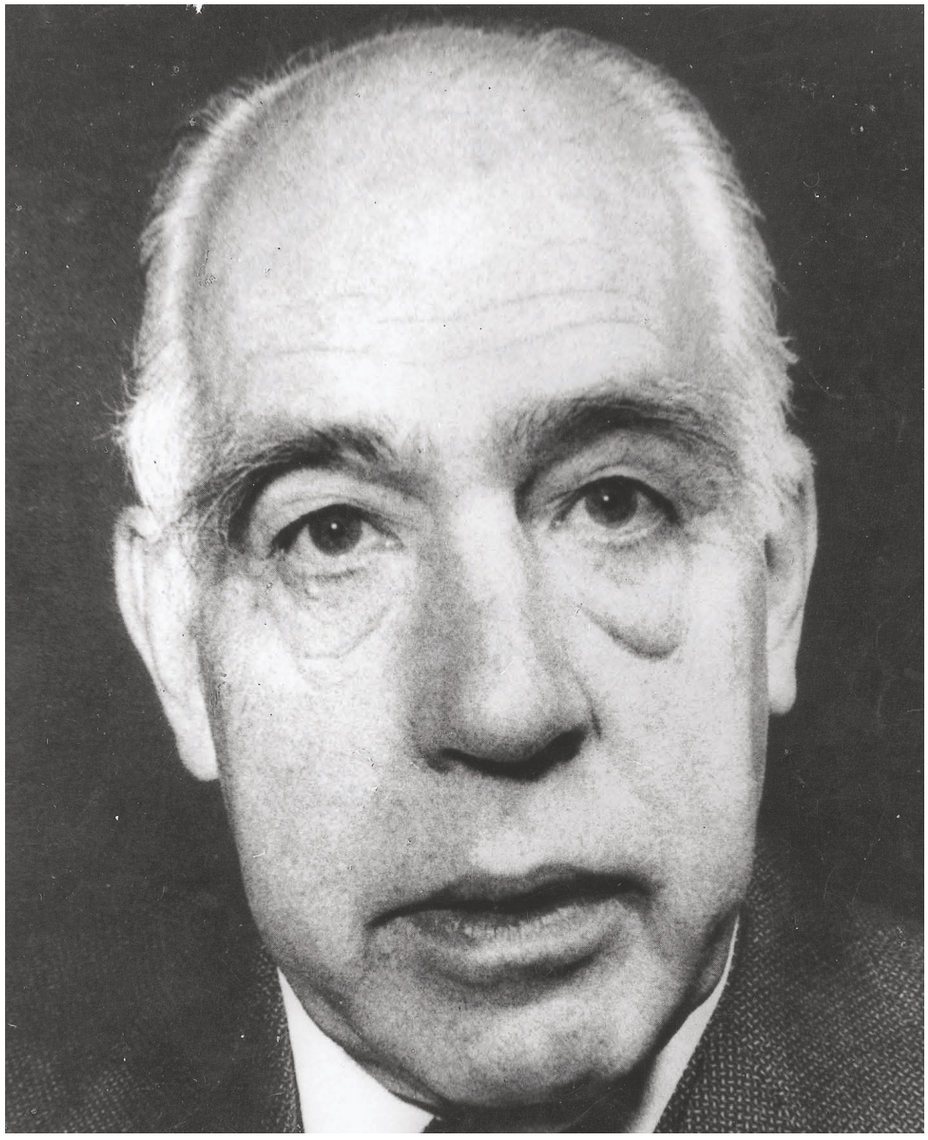


Figure 5.47 *continued*

Lenard's other significant results aside from work on the photoelectric effect are the Lenard window, with which cathode rays can be brought into the open air; absorption of cathode rays (1905 Nobel Prize); and the conjecture that the "force center" of the atom—what today is called the nucleus—is concentrated in a very small volume.

Lenard vigorously attacked Einstein and what he called "dogmatic Jewish" physics, in contrast to a "pragmatic German" variety. All the leading physicists, with the sole exception of Stark, condemned Lenard's action.

◀ **Figure 5.48** Niels Bohr (1885–1962): Son of a professor of physiology and brother of the well-known mathematician Harald Bohr. 1911, doctorate with a thesis on the theory of electrons of metal, then a brief period with J. J. Thomson, and finally, in 1912, in Manchester with Rutherford. In the same year, return to Copenhagen, there first assistant and then in 1913 private lecturer. 1914 to 1916 again in Manchester, then again professor in Copenhagen, where an institute of theoretical physics was established for him in 1920. In 1943 flight from occupied Denmark to Sweden by sailboat. From there, the British secret service brought him to England. Later, he took part in the atomic bomb program in the United States, and in 1945, returned to his homeland. After his death, his successor as director of his institute was his son Aage Bohr.

The basic ideas underlying Bohr's quantum theory are contained in his article *On the Constitution of Atoms and Molecules*, which appeared in 1913 in *Philosophical Magazine*. The Bohr–Sommerfeld model was created between 1915 and 1922. 1916: Bohr's correspondence principle. Bohr's principal contribution was the explanation of the physical background of quantum mechanics and its philosophical interpretation (Bohr's complementarity principle, Copenhagen interpretation). 1936–1943, work on nuclear physics (drop model, theory of nuclear reactions, theory of nuclear fission).

Bohr played an outstanding role in the physics of the twentieth century, and in the 1920s, Copenhagen was a Mecca for theoretical physicists. The wish of every theoretical physicist was to be able to belong to Bohr's circle at least once in their lifetime, even if only for a short while.

The correspondence principle:

> [A]lthough the process of radiation can not be described on the basis of the ordinary theory of electrodynamics, according to which the nature of the radiation emitted by an atom is directly related to the harmonic components occurring in the motion of the system, there is found, nevertheless, to exist a far-reaching correspondence between the various types of possible transitions between the stationary states on the one hand and the various harmonic components of the motion on the other hand. This correspondence is of such a nature, that the present theory of spectra is in a certain sense to be regarded as a rational generalization of the ordinary theory of radiation. [Bohr 1924, pp. 23–24]

continued on next page

the limelight and became the central problem of interest for the second and third decades of the twentieth century. Here entered upon the stage a great personality of the first half of the century, who not only with his own scientific work but also by forming a tight community with young researchers, made decisive contributions to our knowledge and the formation of our worldview: Niels Bohr (Figure 5.48). We may view it as a fortunate circumstance that Bohr began his career under Rutherford. Rutherford liked Planck's quantum theory because with its help one could determine the charge of the electron, and as an experimental physicist, that is what impressed him the most. So he looked favorably at the theoretical physicist Bohr's efforts to combine his—that is, Rutherford's—atomic model with the quantum hypothesis.

The Bohr atomic theory in the form that is found today in introductory physics textbooks for high-school and college students first saw the light of day in 1913. The following are the fundamental postulates of the theory:

Electrons can orbit the atomic nucleus only along certain paths. In such orbits—contrary to the laws of classical electrodynamics—an electron does not radiate. In

Figure 5.48 *continued*

Complementarity:

> I proposed a point of view that could be called "complementarity" and is suited to subsume the typical characteristics of the individuality of quantum phenomena and simultaneously to clarify the particular aspects of the problem of observation within this realm of experience. Here it is crucial that *however far the phenomena might go beyond the domain of classical physical explanation, the representation of all experience must be possible using classical concepts*. The basis of this is simply that with the word "experiment" we refer to a situation in which we can communicate to others what we have done and what we have learned, and that therefore, the experimental setup and results of observation must be described in clearly understandable language using a suitable application of the terminology of classical physics.
>
> From this decisive point follows the *impossibility of a sharp separation between the behavior of atomic objects and the interaction with the measuring equipment that serves to define the conditions under which the phenomena appear*.
>
> —Bohr, *Atomphysik und menschliche Erkenntnis* (Atomic Physics and Human Knowledge) [p. 39]

Quotation 5.24

That is nonsense; the Maxwell equations hold under all circumstances; an electron on a circular trajectory must radiate.

—Von Laue

Very remarkable! There must be something behind it. I do not believe that the derivation of the absolute value of the Rydberg constant is purely fortuitous.

—Einstein [Jammer 1966, p. 86]

the case of circular paths, the radius of the orbit is determined so that the angular momentum of an orbiting electron can assume only integer multiples of $h/2\pi$:

$$mv_n r_n = n\frac{h}{2\pi}, \quad n = 1, 2, \ldots.$$

An atom radiates only when an electron jumps from one orbit to another. The frequency of radiated light is given by the *Bohr frequency condition*

$$h\nu = E_{n2} - E_{n1},$$

where E_{n1} is the energy of an electron in an orbit with quantum number n_1, and E_{n2} is the energy associated with quantum number n_2; $h = 6.626 \times 10^{-34}$ J·s is Planck's constant. Thus, in passing from one path to another, the electron radiates the energy difference between the two path energies in the form of a single photon. Of course, in the absorption of radiation, the reverse process takes place. An atom can absorb only the photons $h\nu$ that have just the right energy to raise an electron from a lower-energy orbit to a higher-energy orbit.

Bohr's groundbreaking work appeared in three parts for a total of 80 pages ("much too long … really ought to be cut down," advised Rutherford) in *Philosophical Magazine* 26 (1913), with the title "On the Constitution of Atoms and Molecules." At the end of the article, Bohr summarizes his initial conditions, or rather postulates. The selection rule for the orbits appears there as follows:

> The different stationary states of a simple system consisting of an electron rotation round a positive nucleus are determined by the conditions that the ration between the total energy, emitted during the formation of the configuration, and the frequency of revolution of the electron is an entire multiple of $h/2$. Assuming that the orbit of the electron is circular, this assumption is equivalent with the assumption that the angular momentum of the electron round the nucleus is equal to an entire multiple of $h/2\pi$. [Jammer 1966, p. 81]

Bohr's postulates cannot be reduced to simpler basic assumptions. At this stage we have to view them as axioms. With their help we want to explain more complex phenomena. Hence their correctness is not directly evident but is assured only by the agreement of the conclusions drawn from them with the experimental facts.

The most daring assumption of the Bohr theory and also the one that aroused the most intense opposition was that, although an electron in an atom under normal circumstances moves in a circular path, in conflict with classical physics, in this case it does not radiate. We should not be surprised that after its publication Bohr's theory was enthusiastically accepted by some and rejected by others (Quotation 5.24).

Bohr's theory had to give answers to two cardinal questions. First, what is its relationship to classical physics, the radiation laws of classical electrodynamics? And to make further progress, it had to answer the question of how quantization conditions can be generalized to systems with more degrees of freedom. These questions were answered primarily by Bohr and later by Ehrenfest and Sommerfeld (Figure 5.49). It was in the *Sommerfeld ellipse model* that Bohr's atomic theory found its culmination.

The most important accomplishment of the theory is that it is able to explain the existence of discrete energy levels as well as in the simplest cases to show a connection between the numerical values obtained from spectral analysis and the fundamental constants of physics (Figure 5.50). The theory paints a reassuring

picture of the electron orbits that is easily visualized: The electrons travel—like the planets of the solar system—in elliptical paths of varying eccentricity around a central nucleus. The iconic image of the atom in popular culture is this model.

The eccentricity is determined by two whole numbers: the principal quantum number n and the azimuthal quantum number l (Figure 5.51). For describing the state of an electron in an atom, another quantum number, the magnetic quantum number m, is needed. This quantum number provides us with information about the energy shift of the electron due to an external magnetic field.

We may arrive at a general quantization rule by reformulating the Bohr selection rule. The position of an electron in its orbit is characterized by the angle φ; then with each *generalized spatial coordinate*, as in Section 4.2.7, is associated the following *generalized momentum*:

$$p_\varphi = \frac{\partial T}{\partial \dot{\varphi}} = \frac{\partial\left[m(r\dot{\varphi})^2/2\right]}{\partial \dot{\varphi}} = rmr\dot{\varphi} = rmv,$$

which is the angular momentum. The orbit selection rule can also be written in the following form:

$$\oint p_\varphi \,\mathrm{d}\varphi = \int_0^{2\pi} p_\varphi \,\mathrm{d}\varphi = 2\pi rmv = nh.$$

This rule can now easily be generalized to several degrees of freedom. Suppose an electron is moving along a (cyclic) space curve, and suppose as well that its position is characterized by the quantities r, ϑ, ψ, its state of motion by p_r, p_ϑ, p_ψ. Then the selection rules are as follows:

$$\oint p_r \,\mathrm{d}r = n_r h, \quad \oint p_\vartheta \,\mathrm{d}\vartheta = n_\vartheta h, \quad \oint p_\psi \,\mathrm{d}\psi = n_\psi h.$$

Instead of the quantum numbers n_r, n_ϑ, n_ψ, the quantum numbers presented above are introduced and used—the principal quantum number $n = n_r + n_\vartheta + n_\psi$, the azimuthal quantum number $l = n_\vartheta + n_\psi - 1 \equiv n_\varphi - 1$, and the magnetic quantum number $|m| = n_\psi$.

This general rule of quantization therefore asserts that the so-called phase integrals must assume values that are integer multiples of Planck's constant h. A deeper basis for this rule can be given with the help of the *Ehrenfest adiabatic hypothesis*.

It is a particularly ironic turn of fate that the illustrative extension of this already illustrative model with a rotating electron, having its own angular momentum, or *spin*, and also a magnetic moment, and the extension of which answered a number of open questions, was first proposed (Goudsmit–Uhlenbeck, 1925) at a time when the basic idea of Bohr's theory itself was beginning to be called into question. In the same year (1925), before the introduction of electron spin, Pauli formulated his *exclusion principle*, which made it possible to explain the periodic system of the elements (Figure 5.52).

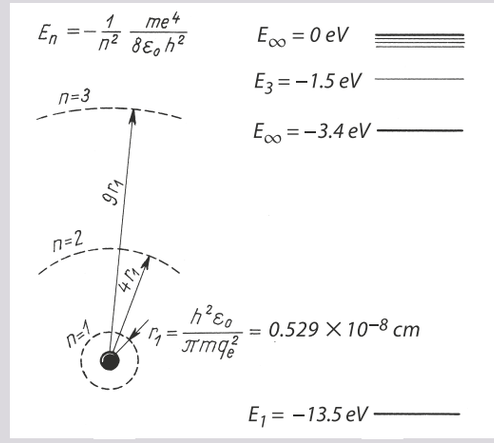
Thus, the state of an electron is characterized by four numbers: the principal quantum number n (its value may be 1, 2, 3, …, that is, any whole number), the azimuthal quantum number l (with possible values $l = 0, 1, 2, \ldots, n-1$, that, is any whole number less than n), the magnetic quantum number m (an arbitrary integer between $-l$ and l), and finally, the spin quantum number s (with the values $+\frac{1}{2}$, $-\frac{1}{2}$). In a state of equilibrium, the electrons assume the lowest energy levels but in such a way that the Pauli exclusion principle is still satisfied. This principle requires that within a system no two electrons can be in the same quantum state. For our purposes, this means that among the four quantum numbers, at least one has to be different for any two pairs of electrons.

5.3.6 *The Statistical Derivation of the Radiation Formula as Prelude to Quantum Electronics*

Einstein's derivation of the radiation formula mentioned above is based on the thesis of Bohr's atomic theory, according to which certain discrete energy states of a microsystem are stable, and an emission of



▲ **Figure 5.49** Arnold Sommerfeld (1868–1951): Studies in mathematics, then from 1900, professor of engineering mechanics at Aix-la-Chapelle (Aachen). 1906, called to the chair of theoretical physics at Munich. Sommerfeld's strength lay in recognizing the significance of new ideas, working for their recognition, and helping to work them out in detail, which he could do very effectively because of his outstanding mathematical ability. Together with Planck, he was among the first to support the theory of relativity; he gave the final form to Bohr's theory (Bohr–Sommerfeld model); and also solved a number of problems in wave mechanics. The books *Atombau und Spektrallinien* (Atomic structure and spectral lines; 1919) and *Wellenmechanische Ergänzungsband* (Wave-mechanical supplementary volume, 1929) were considered in the 1920s and 1930s to be the bible of this field. Pauli and Heisenberg were among his countless students.



▲ **Figure 5.50** Relationship between discrete energy levels and the permitted orbits in the Bohr model in the simplest case of the hydrogen atom.

From the Bohr selection rule

$$r_n m v_n = n\frac{h}{2\pi}$$

continued on next page

Figure 5.50 *continued*

and from the equation of motion

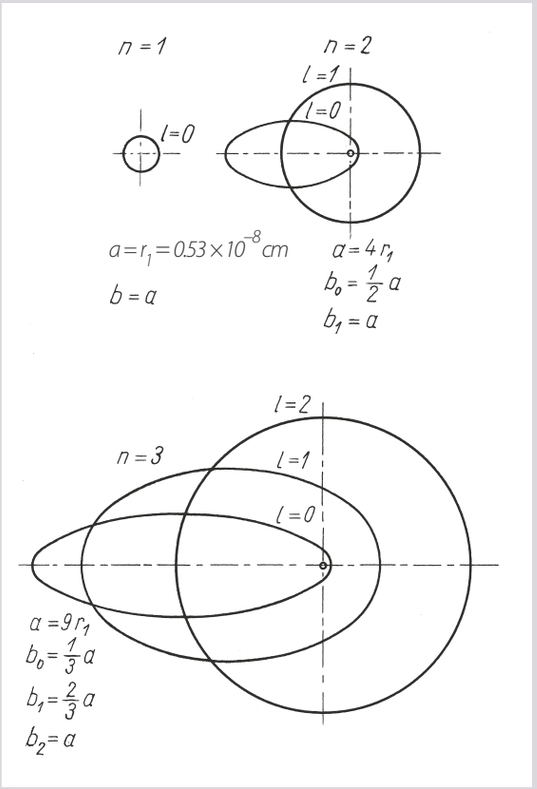
$$\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r_n^2}=\frac{mv_n^2}{r_n}$$

of an electron moving in a circular orbit, one obtains the total energy of the electron:

$$E_n=E_{\text{kinetic}}+E_{\text{potential}}=\frac{1}{2}mv_n^2-\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r_n}=-\frac{me^4}{8\varepsilon_0^2h^2}\frac{1}{n^2}.$$

The frequency condition $h\nu = E_{n2} - E_{n1}$ leads to the following expression for the wave number:

$$\frac{1}{\lambda}=\frac{\nu}{c}=\frac{me^4}{8\varepsilon_0^2h^3c}\left(\frac{1}{n_1^2}-\frac{1}{n_2^2}\right).$$



▲ **Figure 5.51** The elliptical orbits belonging to the various quantum numbers according to the Bohr–Sommerfeld model.

The complexity of the energy-level diagram of atoms with multiple electrons understandably grows with the number of electrons. The selection rules provide some clarity: For certain transitions, the quantum numbers *l* and *m* can change only at certain values. The complexity is increased in that a level that belongs to a certain principal quantum number is broken into several levels as a result of the diversity of interaction energies that belong to different *l* and *s* (fine structure of levels). To this are added more splittings that result from the interactions with the nuclear spin (hyperfine structure of the levels). All these phenomena were first dealt with systematically in the book *Linienspektren*, by F. Hund (1927).

continued on next page

radiation $E_2 - E_1 = h\nu$ occurs only during the transition from one energy state to another. Einstein's approach, published in 1916, was the following:

We consider for the sake of simplicity a microsystem (atom) with only two energy levels E_1 and E_2, where a transition from E_2 to E_1 leads to the emission of a photon with frequency ν. If we bring this system into a space filled with blackbody radiation, then a number of processes are possible (Figure 5.53).

(a) *Absorption*. The atom absorbs a quantum $h\nu$ and transitions from the lower-energy level E_1 to the higher level E_2. The number of absorption processes is proportional to the number of atoms N_1 in state E_1 as well as the energy density of the radiation field $\rho(\nu,T)$, that is, equal to $B_{12}N_1\rho(\nu,T)$.

(b) *Spontaneous emission*. The excited atoms (energy level E_2) return to the state E_1 without external influence with the emission of a photon $h\nu$. This emission is independent of the energy density of the radiation field, so that the number of spontaneous emissions is proportional only to the number of excited atoms N_2, that is, equal to AN_2.

(c) *Induced emission*. The excited atoms are induced by the radiation field to emit a photon $h\nu$ and return to the state E_1. The number of these emissions is proportional to N_2 and to $\rho(\nu,T)$, that is, equal to $B_{12}N_2\rho(\nu,T)$.

In a state of equilibrium (Figure 5.54), the number of absorption processes must be equal to the number of emission processes:

$$B_{12}N_1\rho(\nu,T)=AN_2+B_{21}N_2\rho(\nu,T).$$

If we consider that in thermal equilibrium, the number of atoms that are in the excited state E_2 is given by the Boltzmann distribution as

$$N_2=N_1\,e^{-\frac{E_2-E_1}{kT}}=N_1\,e^{-\frac{h\nu}{kT}},$$

we obtain for the energy density the expression

$$\rho(\nu,T)=\frac{A}{B_{12}\,e^{h\nu/kT}-B_{21}}.$$

We now write this relation in the form

$$\rho(\nu,T)=\frac{A/B_{12}}{e^{h\nu/kT}-B_{21}/B_{12}}$$

and require that in passing to an ever higher temperature, $\rho(\nu,T)$ must also increase without bound. From this requirement, it follows that $B_{21}/B_{12} = 1$, or $B_{12} = B_{21}$. If we finally also determine the quotient A/B_{12} from Wien's law,

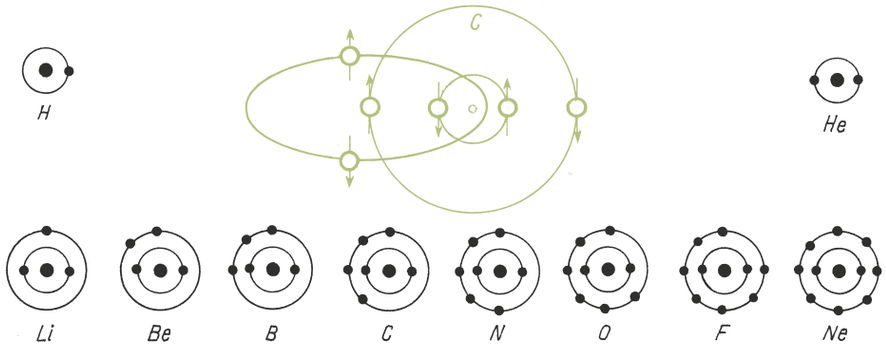
$$A/B_{12}\propto\nu^3,$$

we obtain Planck's formula.

5.3.7 *Heisenberg's Matrix Mechanics*

Despite achieving great success, and in parallel with this success, the limits of Bohr's theory were becoming ever more apparent. Shortcomings appeared mostly in the form that the results of the theory had to be corrected to achieve satisfactory agreement with experiment. Among these corrections, quite surprisingly, many followed the same pattern: For example, instead of l^2 one had to write $l(l+1)$, or some quantum numbers had to be reduced by 1.

At the beginning of the 1920s, it became apparent that further progress was impossible along these lines. The situation was summarized by Bohr in a lecture in 1922 in Göttingen (*Bohr festival*). It is worth saying a few words here concerning the context in which Bohr's lecture took place. In 1908, a mathematician named Wolfskehl established a prize of 100,000 marks for the solution of the famous problem known as Fermat's last theorem. (Fermat has been mentioned earlier in a



▲ **Figure 5.52** The interpretation of the periodic table of elements using the Pauli principle. All electrons must differ within the atom by at least one of the four quantum numbers that characterize their state. This means that in the first orbit there can be at most 2 electrons, in the second at most $8 = 2 \times 2^2$, and in the third at most $18 = 2 \times 3^2$. From these rules follow all the particular qualities of the elements that are reflected in the periodic table.

This model does not offer an accurate picture of the symmetry relationships in the arrangement of electrons. The elliptical model (shown in color) of the carbon atom is a better approximation, but even here, the pictorial representation should not be taken literally.

In general, for each quadruplet of numbers (n, l, m, s), there is exactly one possible state. For each triple (n, l, m), there are two associated states corresponding to the two possible spin alignments. For each pair (n, l) we have also the $(2l + 1)$ states that belong to the possible values of m (from $-l$ to $+l$). Thus, for each given n there are $2\sum_{i=0}^{n-1}(2l + 1) = 2n^2$ different states.

different context.) According to Fermat's conjecture, the equation

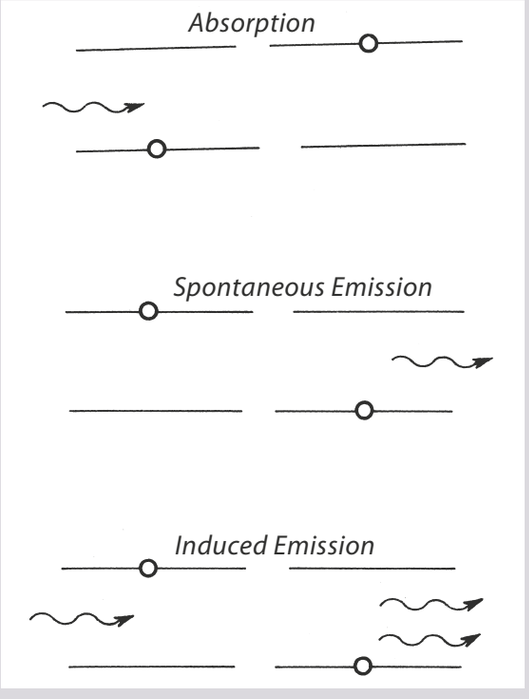
$$x^n + y^n = z^n$$

has no nontrivial solutions in integers for $n > 2$. The ancient Egyptians and Babylonians knew (Section 1.1.2) that for $n = 2$, one can find infinitely many *Pythagorean triples*, groups of numbers such as 3, 4, 5 and 5, 12, 13 that satisfy this equation. However, for $n > 2$, there are no known such triples. Fermat made a notation in the book that he was reading that he had an elegant and simple proof of his conjecture, but that there was insufficient space in the margin to write it down. This mathematical problem is notable for two reasons. First, among the many problems that remained unsolved for a long time (it was finally solved in 1995 by Andrew Wiles by using complex theorems from algebraic geometry), it is one of the few whose statement is simple enough to be understood by the general public. Second, there is scarcely a mathematical theorem anywhere for which more incorrect proofs have been proposed. While the prize of 100,000 marks remained unclaimed, the trustees of the endowment used the interest very wisely. Every year, outstanding lecturers were invited to Göttingen to present lectures on the state of scientific development and on current problems. Thus, in 1909, a year after the establishment of the prize, Poincaré gave a series of lectures, including the one on the theory of relativity in which the name Einstein was not mentioned! In 1910, Lorentz took the podium, and Bohr made an appearance in the series in 1922. This particular lecture by Bohr became famous because in the discussion that followed, a 21-year-old student by the name of Werner Heisenberg (Figure 5.55) took the floor and dared to contradict the internationally acclaimed and Nobel Laureate Bohr. Heisenberg at this time was working with Sommerfeld in Munich; in fact, it was Sommerfeld who brought him to this lecture. It was on this occasion that the collaboration between Bohr and Heisenberg, which later developed into a close friendship, got its start. It was also here that Pauli first met Bohr.

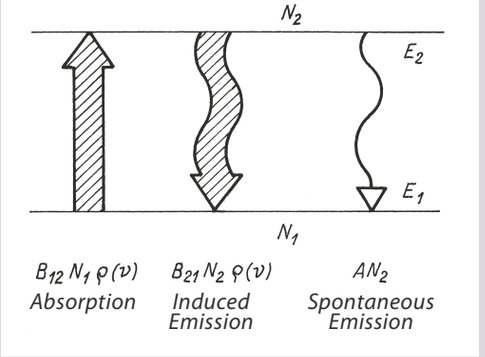
Heisenberg understood the fundamental theoretical and—one is tempted to say—philosophical error in Bohr's theory in that it explains the emission of discrete frequencies with moving particles and orbital paths that are defined and described in detail by using classical notions. All this detailed data has nothing to

Figure 5.51 *continued*

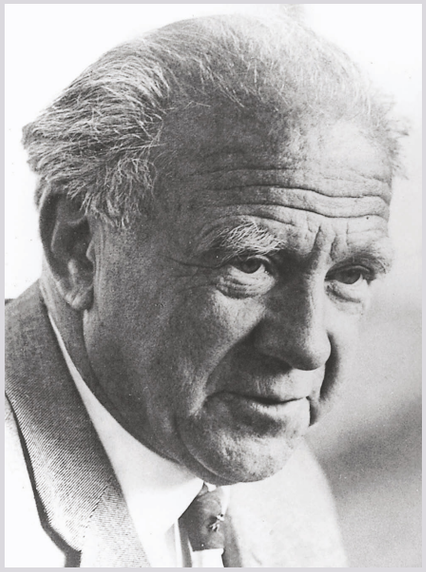
The modern theory of molecular spectra began in 1927 with the work of Born and Oppenheimer recognizing that the energy of a (diatomic) molecule can be represented as the sum of the rotational, vibrational, and electron energies. This yields the characteristic band structure of the molecular spectra. Hund also did groundbreaking work in this area. The most successful centers of molecular spectroscopy were those established by R. S. Mulliken (1895–1986) in the United States and by G. Herzberg (1904–1999) in Canada. Both these men were awarded the Nobel Prize in chemistry for their work. With their spectroscopic methods, they revealed the geometric structure of a large number of molecules and free radicals and thereby determined the consequent influence of the structure on chemical reactions.



▲ **Figure 5.53** Fundamental interaction types between a particle and a radiation field.



▲ **Figure 5.54** Thermal equilibrium between the radiation field and the particles.



◀ **Figure 5.55** Werner Heisenberg (1901–1976): Son of the well-known scholar of Byzantium August Heisenberg. Studies with Sommerfeld in Munich; 1923: dissertation on the theory of turbulent flow; 1924: private lecturer and assistant to Born in Göttingen; 1927: professor at Leipzig; 1941: successor to Sommerfeld in Munich. During the war, participated in the German nuclear program; then, after a brief internment, director of the Max Planck Institute for Physics in Göttingen, and from 1956 active in research and teaching in Munich. Chronology of his work: 1925, *Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen* (A quantum-theoretical reinterpretation of kinematic and mechanical relationships); 1927; Heisenberg uncertainty principle; 1940; theory of nuclear reactors; from 1956; unified theory of elementary particles. Although his work in this area was not entirely without success, he failed to achieve the goal he was seeking.

Heisenberg is one of the leading theoretical physicists of the twentieth century. It was granted to him to remain in the first rank of creative physicists until the very end of a long life full of success. He made significant contributions to the physics of the atomic shell, the atomic nucleus, and elementary particles; in addition, his natural philosophy is considered one of the great achievements of this century. One episode from his life: Sommerfeld described Heisenberg at his doctoral oral examination in 1923 as an incomparable talent, while the examining experimental physicist, Wilhelm Wien, thought that the candidate had shown "profound ignorance" of experimental physics.

> Only the true conservative can be a true revolutionary.
>
> —Heisenberg, quoted by C. F. von Weizsäcker

do with what is measurable. Therefore, we should try to construct the equations describing the behavior of atoms in such a way that only measurable quantities appear in them (Figure 5.56). This was the goal that was expressed by Heisenberg in 1925 in his groundbreaking work (Quotation 5.25, Plate XXIV). It is interesting to note that the legitimacy of this approach had been called into question by Einstein. In their discussion, Heisenberg reminds Einstein (see the following section) that he himself had also used such arguments when he began to reformulate the concepts of time and space.

Heisenberg's ideas were cast into mathematical form with help primarily from Born and Jordan. The basics of the matrix mechanics that they worked out can be summarized as follows.

Experiments show that an excited microsystem can emit radiation of several frequencies. It is well known (Section 4.5.4) that a periodic function can be represented as a Fourier series. So if we classically treat the motion of the electron as such a periodic function, then this will physically mean that in addition to the fundamental frequency, the integer multiples of that frequency appear in the emitted radiation:

$$\nu_1 = \nu_0, \quad \nu_2 = 2\nu_0, \quad \nu_3 = 3\nu_0, \quad \ldots, \quad \nu_n = n\nu_0, \quad \ldots.$$

In a classical treatment, we can even give the amplitudes associated with the individual frequencies as the coefficients of the Fourier series. However, a microsystem emits frequencies that can be characterized by two integral indices: a system transitioning from state m to state n will emit frequency ν_{mn}. It would therefore be better to replace the classical arrangement of frequencies

$$\nu_1, \nu_0, \ldots, \nu_n, \ldots$$

by a rectangular array, that is, a matrix of the form

$$\begin{bmatrix} \nu_{11} & \nu_{12} & \cdots & \nu_{1n} & \cdots \\ \nu_{21} & \nu_{22} & \cdots & \nu_{2n} & \cdots \\ \vdots & \vdots & & \vdots & \\ \nu_{n1} & \nu_{n2} & \cdots & \nu_{nn} & \cdots \\ \vdots & \vdots & & \vdots & \end{bmatrix}.$$

Analogously, the classical Fourier coefficients

$$q_1, q_0, \ldots, q_n, \ldots$$

should be replaced by the matrix

$$\begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} & \cdots \\ q_{21} & q_{22} & \cdots & q_{2n} & \cdots \\ \vdots & \vdots & & \vdots & \\ q_{n1} & q_{n2} & \cdots & q_{nn} & \cdots \\ \vdots & \vdots & & \vdots & \end{bmatrix},$$

which characterizes the position of the electron. To describe the state of motion, the momentum matrix

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2n} & \cdots \\ \vdots & \vdots & & \vdots & \\ p_{n1} & p_{n2} & \cdots & p_{nn} & \cdots \\ \vdots & \vdots & & \vdots & \end{bmatrix}$$

is introduced. The laws of microphysics can now be described as relationships among the newly introduced quantities. It turns out that the *Hamiltonian equations of motion* are valid, at least formally, for them. Conversely, some very simple rules that hold for the classical quantities q and p lose their validity for the matrices **q** and **p**. In particular, the matrix product **qp** is different from the product **pq**. One of the most important relationships of matrix mechanics is given by the difference of these products:

$$\mathbf{pq} - \mathbf{qp} = \frac{h}{2\pi i}\mathbf{1}.$$

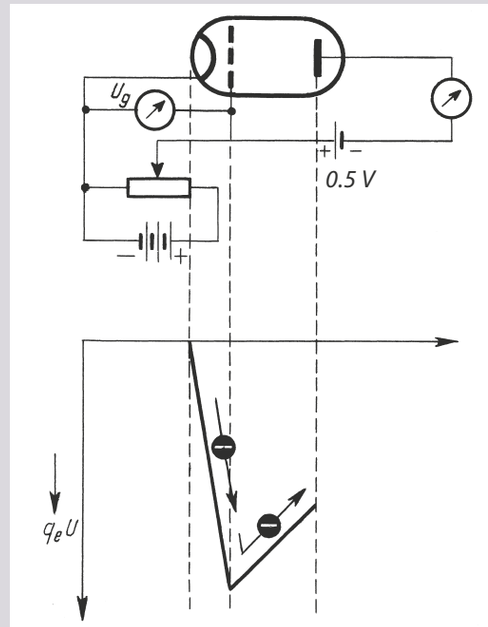
This is the famous commutation relation, in which the difference between matrix mechanics and classical mechanics is vividly expressed. At the same time, it also makes clear the crucial role played by the constant h. Were h equal to zero, we would be back in the classical realm.

Readers who feel that they have not quite grasped the secrets of matrix mechanics should not let that bother them, for they have eminent forebears in this regard (Quotations 5.26, 5.27, 5.28; Figure 5.57).

For the scientist, Table 5.1 gives a timeline of the development of quantum mechanics, showing who made the most significant contributions to the theory and when.

Table 5.1 The fundamental publications in quantum mechanics.

1. L. de Broglie:	*"Ondes et quanta," Comptes Rendus* 177 (1923), pp. 507–510, submitted September 10, 1923; two additional articles appeared in the same volume of *Comptes Rendus* (1923) and were presented on September 24 and October 8, 1923. $$\delta\int\frac{ds}{\lambda}=\delta\int\frac{m_0 v}{\sqrt{1-\beta^2}}ds=\delta\int p\,ds=0$$ $$\lambda=\frac{h}{p}$$
2. W. Heisenberg:	*"Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen," Zeitschrift für Physik* 33 (1925), pp. 879–893. Submitted July 29, 1925. $$h=8\pi^2 m\sum_{\tau=0}^{\infty}\left\{\left\lvert x_{n+\tau,n}\right\rvert^2 v_{n+\tau,n}-\left\lvert x_{n,n-\tau}\right\rvert^2 v_{n,n-\tau}\right\}$$ (Thomas Kuhn sum rule) Oscillator: $$E=\frac{h\omega_0}{2\pi}\left(n+\frac{1}{2}\right)$$ Rotator: $$E=\frac{h^2}{8\pi^2 ma^2}\left(n^2+n+\frac{1}{2}\right)$$



▲ **Figure 5.56** The discrete states of the atom can be studied with the Franck–Hertz experiment in addition to the line spectra.

The positive voltage on the grid of an electron tube filled with the gas under investigation is increased; the accelerated electrons can still reach the anode that is at a small negative potential. But when their energy is sufficient to bring the gas atoms into an excited state, their collisions will no longer be elastic, they will lose their energy, and then they cannot overcome the repulsion of the anode. Consequently, the arrival at the first excitation level will be signaled by a decrease of the anode current.

James Franck (1882–1964): Studies in Heidelberg and Berlin. Worked in Berlin, then, on Born's invitation, directed experimental work in Göttingen in a period considered the heyday of research there. After Hitler seized power, Franck emigrated to the United States, where he later took part in the atomic bomb program. In June 1945, he opposed the use of the atomic bomb in the famous Franck report.

Gustav Hertz (1887–1975): Nephew of Heinrich Hertz. Worked both in industry and university research; 1920–1925: worked at the Philips factory. 1935–1945: director of the research laboratory at Siemens. After the Second World War, taken to work in the Soviet Union. 1954–1961: director of the physics institute of Karl Marx University, Leipzig.

To study the hyperfine structure of spectra determined by the influence of the atomic nuclei, Hertz required a gas that was a pure isotope and was for this reason highly motivated to find methods for the separation of isotopes. With the discovery of nuclear fission, isotope separation became the key problem in the development of atomic bombs.

Quotation 5.25

The present paper seeks to establish a basis for theoretical quantum mechanics founded exclusively upon relationships between quantities which in principle are observable.

It is well known that the formal rules which are used in quantum theory for calculating observable quantities such as the energy of the hydrogen atom may be seriously criticized on the grounds that they contain, as basic element, relationships between quantities that are apparently unobservable in principle, e.g., position and period of revolution of the electron. Thus these rules lack an evident physical foundation, unless one still wants to retain the hope that the hitherto unobservable quantities may later come within the realm of experimental determination. This hope might be regarded as justified if the above-mentioned rules were internally consistent and applicable to a clearly defined range of quantum mechanical problems. Experience however shows that only the hydrogen atom and its STARK effect are amenable to treatment by these formal rules of quantum theory. Fundamental difficulties already arise in the problem of 'crossed fields' (hydrogen atom in electric and magnetic fields of differing directions). Also, the reaction of atoms to periodically varying fields cannot be described by these rules. Finally, the extension of the quantum rules to the treatment of atoms having several electrons has proved unfeasible.

It has become the practice to characterize this failure of the quantum-theoretical rules as deviation from classical mechanics, since the rules themselves were essentially derived from classical mechanics, since the rules themselves were essentially derived from classical mechanics. This characterization has, however, little meaning when one realizes that the Einstein-Bohr frequency condition (which is valid in all cases) already represents such a complete departure from classical mechanics, or rather (using the viewpoint of wave theory) from the kinematics underlying this mechanics, that even for the simplest quantum-theoretical problems the validity of classical mechanics simply cannot be maintained. In this situation it seems sensible to discard all hope of observing hitherto unobservable quantities, such as the position and period of the electron, and to concede that the partial agreement of the quantum rules with experience is more or less fortuitous. Instead it seems more reasonable to try to establish a theoretical quantum mechanics, analogous to classical mechanics, but in which only relations between unobservable quantities occur. One can regard the frequency condition and the dispersion theory of KRAMERS together with its extensions in recent papers as the most important first steps toward such a quantum-theoretical

continued on next page

Table 5.1 The fundamental publications in quantum mechanics, *continued*.

3. M. BORN–P. JORDAN: *"Zur Quantenmechanik,"* *Zeitschrift für Physik* 34 (1925), pp. 858–888. Submitted September 27, 1925.

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}$$

$$pq - qp = \frac{h}{2\pi i}1$$

4. BORN–HEISENBERG–JORDAN (Dreimänner Arbeit): *"Zur Quantenmechanik II,"* *Zeitschrift für Physik* 35 (1926), pp. 557–615. Submitted November 15, 1925.

$$\dot{f} = -\frac{2\pi i}{h}(fH - Hf)$$

$$P = U^{-1}pU, \quad Q = U^{-1}qU, \quad H(P,Q) = W \textit{ (diagonal)}$$

5. E. SCHRÖDINGER: *"Quantisierung als Eigenwertproblem I,"* *Annalen der Physik* 79 (1926), pp. 361–376. Submitted January 27, 1926.

$$\Delta\Psi + \frac{2m}{K^2}\left(E + \frac{e^2}{r}\right)\Psi = 0$$

$$E = -\frac{me^4}{2K^2m^2}, \quad K = \frac{h}{2\pi}$$

II. Submitted February 23, 1926.

$$\frac{\partial W}{\partial t} + T\left(q, \frac{\partial W}{\partial q}\right) + U(q) = 0$$

$$\Delta\Psi + \frac{8\pi^2 m}{h^2}(E - U)\Psi = 0$$

III. Submitted May 10, 1926.

$$L\Psi + E\Psi = 0$$

$$L\Psi + E\Psi = \lambda P\Psi$$

$$E = E_k + \lambda E_k$$

$$\Psi = \Psi_k + \lambda\varphi_k$$

IV. Submitted June 21, 1926.

$$-\frac{h^2}{8\pi^2 m}\Delta\Psi + U\Psi = \frac{h}{2\pi i}\frac{\partial\Psi}{\partial t}$$

6. E. SCHRÖDINGER: *"Über das Verhältnis der Heisenberg–Born–Jordanschen Quantenmechanik zu der meinen,"* *Annalen der Physik* 79 (1926) pp. 734–576. Submitted March 18, 1926.

$$pq - qp = \frac{h}{2\pi i}1$$

$$\left(\frac{h}{2\pi i}\frac{\partial}{\partial q}\right)q\Psi - q\left(\frac{h}{2\pi i}\frac{\partial}{\partial q}\right)\Psi = \frac{h}{2\pi i}\Psi$$

$$F_{jk} = \int u_j^* F\left(\frac{h}{2\pi i}\frac{\partial}{\partial q}, q\right)u_k \, dq$$

7. P. A. M DIRAC: *"The Fundamental Equations of Quantum Mechanics,"* *Proceedings of the Royal Society* (A) 109 (1925), pp. 642–653. Submitted November 7, 1925.

$$xy - yx \Leftrightarrow \frac{jh}{2\pi}\begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w} \\ \frac{\partial x}{\partial J} & \frac{\partial y}{\partial J} \end{vmatrix}$$

$$\frac{df}{dt} = \frac{2\pi}{jh}(fH - Hf)$$

Table 5.1 The fundamental publications in quantum mechanics, *continued*.

8. P. A. M Dirac:	*"Quantum Mechanics and a Preliminary Investigation of the Hydrogen Atom," Proceedings of the Royal Society* (A) 110 (1926), pp. 561–579. Submitted January 22, 1926.	*Spectrum of the hydrogen atom*
9. W. Pauli:	*"Über das Wasserstoffspektrum vom Standpunkt der neuen Quantenmechanik," Zeitschrift für Physik* 36 (1926), pp. 336–363. Completed in October 1925.	*Spectrum of the hydrogen atom; Stark effect; perturbations in crossed electric and magnetic fields*
10. M. Born:	*"Quantenmechanik der Stoßvorgänge," Zeitschrift für Physik* 38 (1926), pp. 803–827.	*The motion of particles is described through the laws of probability; however, the probability itself propagates in accord with the principle of causality.*
11. J. von Neumann:	*"Mathematische Begründung der Quantenmechanik," Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1927), pp. 1–57.	*Axioms of abstract Hilbert space; operators in Hilbert space; the eigenvalue problem*
12. P. A. M. Dirac:	*"The Quantum Theory of the Emission and Absorption of Radiation," Proceedings of the Royal Society* 114 (1927), pp. 243–265.	*Introduction to creation and annihilation operators*
13. P. A. M. Dirac:	*"The Quantum Theory of the Electron," Proceedings of the Royal Society* 117 (1928), pp. 610–624.	*Derivation of the relativistically invariant equation; electron spin; states with negative energy (as open problem)*

5.3.8 *Einstein and Heisenberg*

"What you have told us sounds extremely strange. You assume the existence of electrons inside the atom, and you are probably quite right to do so. But you refuse to consider their orbits, even though we can observe electron tracks in a cloud chamber. I should very much like to hear more about your reasons for making such strange assumptions."

"We cannot observe electron orbits inside the atom," I must have replied, "but the radiation which an atom emits during discharges enables us to deduce the frequencies and corresponding amplitudes of its electrons. After all, even in the older physics wave numbers and amplitudes could be considered substitutes for electron orbits. Now, since a good theory must be based on directly observable magnitudes, I thought it more fitting to restrict myself to these, treating them, as it were, as representatives of the electron orbits."

"But you don't seriously believe," Einstein protested, "that none but observable magnitudes must go into a physical theory?"

"Isn't that precisely what you have done with relativity?" I asked in some surprise. "After all, you did stress the fact that it is impermissible to speak of absolute time, simply because absolute time cannot be observed; that only clock readings, be it in the moving reference system or the system at rest, are relevant to the determination of time."

Quotation 5.25, ***continued***

mechanics. In this paper, we shall seek to establish some new quantum-mechanical relations and apply these to detailed treatment of a few specials programs. We shall restrict ourselves to problems involving one degree of freedom.

—Heisenberg, "Quantum Theoretical Re-interpretation of Kinematic and Mechanical Relations

Quotation 5.26

Hilbert was having a great laugh on Born and Heisenberg and the Göttingen theoretical physicists because when they first discovered matrix mechanics they were having, of course, the same kind of trouble that everybody else had in trying to solve problems and to manipulate and really do things with matrices. So they went to Hilbert for help, and Hilbert said the only times that he had ever anything to do with matrices was when they came up as a sort of by-product of the eigenvalues of the boundary-value problem of a differential equation. So if you look for the differential equation which has these matrices you can probably do more with that. They had thought it was a goofy idea and that Hilbert did not know what he was talking about. So he was having a lot of fun pointing out to them that they could have discovered Schrödinger's wave mechanics six months earlier if they had paid a little more attention to him.

—Edward U. Condon, "60 Years of Quantum Physics"

Quotation 5.27

[T]he fact that *xy* was not equal to *yx* was very disagreeable to me. I felt this was the only point of difficulty in the whole scheme, otherwise I would be perfectly happy.... I had written down, as the quantization rule the Thomas-Kuhn sum rule, but I had not realized that this was just $pq - qp$.

—Heisenberg, interviewed on February 15, 1963

Quotation 5.28

Heisenberg's great paper on matrix mechanics of 1925 did not appear sufficiently clear to Fermi, who reached a full understanding of quantum mechanics only later through Schrödinger's wave mechanics. I want to emphasize that this attitude of Fermi was certainly not due to the mathematical difficulties and novelty of matrix algebra; for him, such difficulties were minor obstacles; it was rather the physical ideas underlying these papers which were alien to him.

—Segrè, Biographical Introduction to *Collected Papers of Enrico Fermi*



▲ **Figure 5.57** David Hilbert (1862–1943): Studies in mathematics at Königsberg under Ferdinand Lindemann, who proved the transcendence of π. Began university career in Königsberg, then from 1895 to his death in Göttingen, which he made into a center of mathematical research during the first third of the twentieth century. Investigation of the foundations of geometry (1898–1902), discussed in Section 5.2.6. His most important work from the standpoint of physics is related to integral equations: foundations of all of functional analysis, Hilbert space, spectral theory of linear operators. Hilbert's direct contributions to physics are also significant. For instance, he succeeded in deriving Einstein's equations of the general theory of relativity from a variational principle (1915), provided a firmer foundation for Kirchhoff's formula (1912), and finally, clarified the mathematical foundations of quantum mechanics (Hilbert, von Neumann, Nordheim: *Über die Grundlagen der Quantenmechanik*, 1927).

Of the greatest importance for the development of modern physics, however, is Hilbert's book (written jointly with Courant) *Methods of Mathematical Physics*, the first volume of which appeared in 1924, presenting all the mathematical tools needed by the theoretical physicists of the time.

"Possibly I did use this kind of reasoning," Einstein admitted, "but it is nonsense all the same. Perhaps I could put it more diplomatically by saying that it may be heuristically useful to keep in mind what one has actually observed. But on principle, it is quite wrong to try founding a theory on observable magnitudes alone. In reality the very opposite happens. It is the theory which decides what we can observe. You must appreciate that observation is a very complicated process. The phenomenon under observation produces certain events in our measuring apparatus. As a result, further processes take place in the apparatus, which eventually and by complicated paths produce sense impressions and help us to fix the effects in our consciousness. Along this whole path—from the phenomenon to its fixation in our consciousness—we must be able to tell how nature functions, must know the natural laws at least in practical terms, before we can claim to have observed anything at all. Only theory, that is, knowledge of natural laws, enables us to deduce the underlying phenomena from our sense impressions. When we claim that we can observe something new, we ought really to be saying that, although we are about to formulate new natural laws that do not agree with the old ones, we nevertheless assume that the existing laws—covering the whole path from the phenomenon to our consciousness—function in such a way that we can rely upon them and hence speak of 'observations.'

"In the theory of relativity, for instance, we presuppose that, even in the moving reference system, the light rays traveling from the clock to the observer's eye behave more or less as we have always expected them to behave. And in your theory, you quite obviously assume that the whole mechanism of light transmission from the vibrating atom to the spectroscope or to the eye works just as one has always supposed it does, that is, essentially according to Maxwell's laws. If that were no longer the case, you could not possibly observe any of the magnitudes you call observable. Your claim that you are introducing none but observable magnitudes is therefore an assumption about a property of the theory that you are trying to formulate. You are, in fact, assuming that your theory does not clash with the old description of radiation phenomena in the essential points. You may well be right, of course, but you cannot be certain."

I was completely taken aback by Einstein's attitude, though I found his arguments convincing. Hence I said: "The idea that a good theory is no more than a condensation of observations in accordance with the principle of thought economy surely goes back to Mach, and it has, in fact, been said that your relativity theory makes decisive use of Machian concepts. But what you have just told me seems to indicate the very opposite. What am I to make of all this, or rather what do you yourself think about it?"

"It's a very long story, but we can go into it if you like. Mach's concept of thought economy probably contains part of the truth, but it strikes me as being just a bit too trivial."

—Werner Heisenberg, *Physics and Beyond* [pp. 62–65]

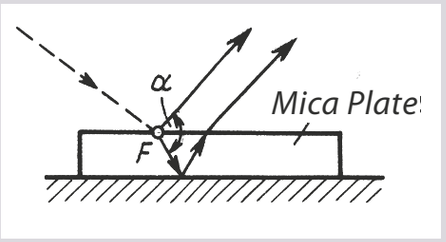
5.3.9 *Schrödinger's Wave Mechanics*

At about the same time, but following a completely different train of thought, Erwin Schrödinger arrived at a fundamental equation of a completely different kind.

During the first two decades of quantum theory, the question of the nature of light—specifically, its dual nature—regularly appears either intertwined with the major line of development or as an independent problem. At the end of the nine-

teenth century, this problem seemed to have been completely resolved: Light is a vibrational state of the ether, or more concretely, an electromagnetic wave.

However, the photoelectric effect again raised the issue of the particle nature of light. From this time on, the question—always posed as "either–or"—seemed to be decided with alternating fortunes in favor of one side or the other. We wish to mention here Pal Selényi, who proved with his wide-angle interference experiments that Einstein's "needle radiation" hypothesis could not be correct (Figure 5.58).



▲ **Figure 5.58** Principle of the wide-angle interference experiment by Selényi to refute Einstein's needle-radiation hypothesis.

Interference phenomena speak in favor of a wave nature of light, while the photoelectric and Compton (1923) effects suggest a particle nature. De Broglie (Figure 5.59) came to the following idea: We know that in addition to its wave nature, light also has corpuscular properties because, like particles with mass, it can collide with other particles satisfying the momentum- and energy-balance theorems. In the past, too much attention was given to the wave nature of light, and the other form was neglected, which is why we had difficulty in explaining the photoelectric and Compton effects. Are we not now making the same error with material particles but in the opposite direction? Are we not overemphasizing their particle nature as their only form at the expense of their wave nature? Perhaps it is because of our one-sided approach that we cannot explain the microphysical phenomena correctly. We should therefore grant wave properties to electrons and atoms as well.



▲ **Figure 5.59** Louis Victor, Duc de Broglie (1892–1981): Studies at the Sorbonne, where following the family tradition, he at first prepared for a career in the civil service by studying history and allied subjects. Under the influence of his elder brother, Maurice de Broglie, who ran a private laboratory in the family's ancestral castle, turned to the study of physics (see the quotation below). During the First World War, he worked on military applications of telecommunications technology. In his doctoral dissertation *Recherches sur la Théorie des Quanta*, submitted in 1924, he developed the theory of matter waves. It is reported that Schrödinger, after a brief glance at this work, set it aside with the remark, "Nonsense." He picked it up again only on the urging of Langevin and then developed the ideas in it further. In the interpretation of quantum mechanics, de Broglie was aligned with the conservative side. De Broglie's works on the history of science are also of significance. For example, in the introduction to his original paper, he dealt extensively with historical preconditions. He also made a suggestion on how to detect matter waves to a colleague of his brother but that person was busy with other projects. It was then by chance that the American physicist Clinton Joseph Davisson found the first evidence. Davisson was interested in secondary

continued on next page

With light, the wavelength was given, and we have been looking for the momentum. With the electron as a particle, the momentum is determined according to the formula $p = mv$; how can we find the characteristics of the corresponding wave? The answer suggests itself that the same relationship holds between the wavelength and momentum in the electron as for light. Using the relationship for light

$$p = mc = \frac{mc^2}{c} = \frac{h\nu}{c} = \frac{h}{\lambda},$$

we therefore associate with an electron having momentum p the wavelength

$$\lambda = \frac{h}{p} = \frac{h}{mv}.$$

The result of this new perspective that was most immediate theoretically and also historically was the simple explanation of the selection rule for stationary electron orbits in the Bohr atomic model. If we take this model's constraint equation

$$mv_n r_n = n\frac{h}{2\pi},$$

then, after a simple rearrangement, we obtain

$$2\pi r_n = n\frac{h}{mv_n}.$$

Substituting the wavelength of the electron yields

$$2\pi r_n = n\lambda.$$

This already has a very simple physical meaning: In the atom, the only stationary orbits possible are those whose circumference is an integer multiple of the electron's wavelength. We find this quite natural because this is also how we get a stationary vibrational state in mechanics (Figure 5.60).

The fact that with the help of the wave associated with the electron and its wavelength $\lambda = h/p$ we can interpret, or rather illustrate, a relationship that is not ex-

Figure 5.59 *continued*

emissions of electrons from ordinary polycrystalline nickel electrodes in vacuum tubes and made measurements of the angular distribution of electron scattering. One of the tubes accidently broke and the exposed electrode was oxidized. The oxide layer could be removed only after extensive heat treatment, which also turned the electrode into a single crystal. This completely changed the distribution curves. It took until 1927 for Davisson to obtain precise quantitative results that could be considered proof for matter waves. But in that year, George Paget Thomson, in Aberdeen, would also publish electron diffraction pictures from the passage of electrons through thin foils.

My brother Maurice, at the time a young physicist, was charged with producing the minutes of the first Solvay conference, which took place in Brussels in October 1911. The conference dealt with problems of quantum theory, and the greatest researchers in mathematical physics of the time—Max Planck, Einstein, Henri Poincaré, Lorentz, and many others—all took part. Soon afterward, my brother gave me a draft of the minutes that he had put together. I read the difficult text with great enthusiasm; I sensed the importance of quantum theory, which Planck had initiated in 1900 with his brilliant insight and which Einstein brought into a new form in 1905 with his light quantum hypothesis. I saw clearly the deeper meaning of this great revolution in theoretical physics: It meant that the wave and particle points of view that had been applied by theoreticians in physics in various fields—in particular the wave theory for describing light and wave processes, the particle theory for describing matter and the structure of matter—must be taken into account in both areas. But how can this synthesis be accomplished? Clearly, the "mysterious" Planck constant—whose secret has not been completely revealed even today, a half century later—must play an important role; but there were a few points that immediately caught my attention at the time: Around the middle of the nineteenth century, first Hamilton, and then later Jacobi, found many similarities between the laws of mechanics that determine the motion of particles and the wave-theoretic laws that describe the propagation of wave processes. According to the prevailing view, this similarity was purely formal; but my youthful imagination came into play, and I asked myself whether this similarity might have an underlying physical meaning and whether it could perhaps lead us to a synthesis of the wave and particle points of view, a synthesis whose necessity had been shown in physics by the appearance of the puzzling quantum. After receiving my degree, I decided that I would begin to investigate this problem.

—Louis de Broglie, *Certitude et incertitude de la science*, Paris, 1966

perimentally demonstrable anyway does not yet amount to a proof of the electron's wave nature. It was only in 1927 that Davisson and Germer directly determined the wavelength of the electron using the methods developed to measure the wavelength of x-rays (electron diffraction).

Let us now turn to Schrödinger's train of thought. There is a tight analogy between the path of a light ray as determined by the laws of geometric optics and the path of an electron in an electric field: The paths will be identical when there is a certain relationship between the index of refraction in the optical space and the electric potential in the electric field (Figure 5.61).

If we take the light ray optics to be the limiting case of the wave optics for small wavelengths, perhaps classical mechanics could then be the limiting case of a new mechanics, namely, a "wave mechanics." We need to find the fundamental equation of this new mechanics. It should be similar to the equation that describes light waves. This is how, in 1926, Schrödinger (Figure 5.62) arrived at the wave equation that bears his name.

Let us now look at Schrödinger's ideas in greater detail.

In today's textbooks on quantum mechanics, the student is introduced in a simple and straightforward way to the Schrödinger equation. However, the path cleared for us by Schrödinger was initially filled with obstacles. To start, Schrödinger had to reach back to the mechanics of Hamilton (Section 4.2.7).

Let $H(q,p)$ be the customary Hamilton function of the system representing the total energy of the system in terms of the spatial and momentum coordinates. We write the Hamiltonian action function, or *action integral*, as

$$\int_{t_0}^{t} (T-U)\,\mathrm{d}t,$$

which is nothing other than the time integral of the *Lagrange function* $L = T - U$. The integral is taken along the path traversed by the system, which we view as a function of time and of the endpoints of the path. The function $W(t)$ satisfies the Hamilton–Jacobi equation

$$\frac{\partial W}{\partial t} + H\left(q, \frac{\partial W}{\partial q}\right) = 0. \tag{5}$$

If we determine from this equation the action function W, then we can easily obtain the solution to the equation of motion. If we write this in the usual form

$$W = -Et + S(q),$$

Then equation (5) is transformed into the simpler

$$H\left(q, \frac{\partial S}{\partial q}\right) = E,$$

and moreover, we obtain

$$\left|\operatorname{grad} W\right|^2 = 2m(E-U), \quad \frac{\partial W}{\partial t} = -E.$$

From this it follows that the surfaces W = constant, the *equiphase surfaces*, move with phase velocity

$$v_f = \frac{E}{\left[2m(E-U)\right]^{1/2}},$$

in the direction specified by grad W. If we represent the waves propagating in an inhomogeneous

medium in the form

$$\varphi(x,y,z,t) = ae^{i[-\omega t + S(x,y,z)]},$$

then we see that we can write down a relationship for the phase of this function similar to that for W, and we see as well that E and v, or E and ω, must be proportional. Schrödinger naturally set the proportionality factor to h. We therefore obtain for λ the relationship

$$\lambda = \frac{v_f}{v} = \frac{1}{v}\frac{E}{\left[2m(E-U)\right]^{1/2}} = \frac{h}{\left[2m(E-U)\right]^{1/2}}. \qquad (6)$$

In the interest of historical accuracy, we should mention that Schrödinger only took this approach in his second article, which was submitted to the *Annalen der Physik* on February 23, 1936. In his first article (January 27, 1926), he begins with

$$H\left(q, \frac{\partial S}{\partial q}\right) = E,$$

where for S he substitutes $S = K \ln \Psi$. With this trial insertion (ansatz), he obtains

$$H\left(q, \frac{K}{\Psi}\frac{\partial \Psi}{\partial q}\right) = E, \qquad (7)$$

where the left-hand side of this equation is quadratic in both Ψ and $\partial\Psi/\partial q$. Schrödinger then sets the following conditions for Ψ: it must be a real, single-valued, twice continuously differentiable function, and the integral of the quadratic form resulting from (7) taken over the entire configuration space must be an extreme value. This is how he found the equation describing the hydrogen atom:

$$\Delta\Psi + \frac{2m}{K^2}\left(E + \frac{e^2}{r}\right)\Psi = 0.$$

Namely, for the hydrogen atom, we have

$$H(q,p) = \frac{1}{2m}\left(p_x^2 + p_y^2 + p_z^2\right) - \frac{e^2}{r} = E,$$

and from equation (7), we obtain

$$\frac{1}{2m}\left[\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2\right] - \frac{e^2}{r} - E = 0.$$

If we now set $S = K \ln \Psi$, then we obtain at once

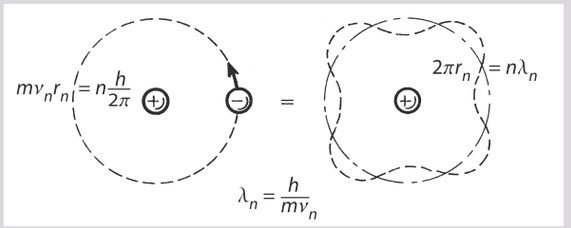
$$\frac{K^2}{2m}\frac{1}{\Psi^2}\left[\left(\frac{\partial \Psi}{\partial x}\right)^2 + \left(\frac{\partial \Psi}{\partial y}\right)^2 + \left(\frac{\partial \Psi}{\partial z}\right)^2\right] - \left(E + \frac{e^2}{r}\right) = 0,$$

and from this, the quadratic form

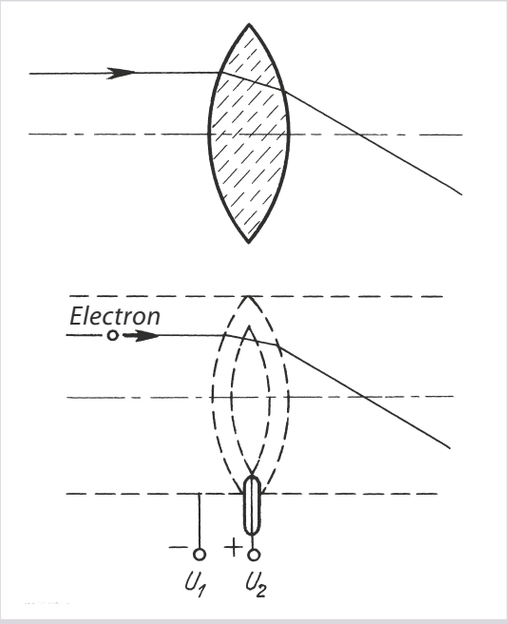
$$F = (\nabla\Psi)^2 - \frac{2m}{K^2}\left(E + \frac{e^2}{r}\right)\Psi^2.$$

Given the variational problem $\delta\iiint F\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 0$, the corresponding Euler–Lagrange differential equation is

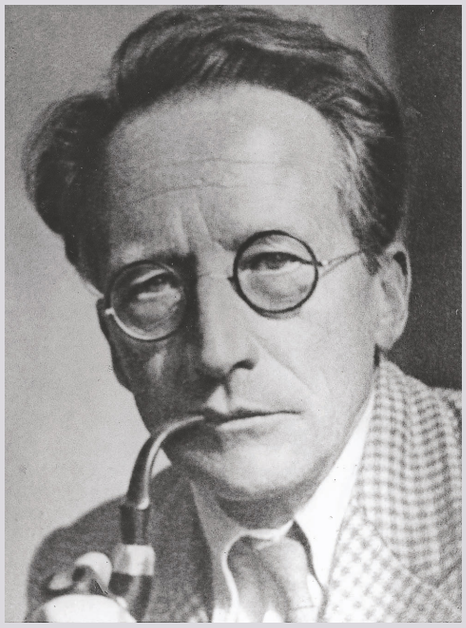
$$\frac{\partial F}{\partial \Psi} - \left(\frac{\partial}{\partial x}\frac{\partial F}{\partial(\partial\Psi/\partial x)} + \frac{\partial}{\partial y}\frac{\partial F}{\partial(\partial\Psi/\partial y)} + \frac{\partial}{\partial z}\frac{\partial F}{\partial(\partial\Psi/\partial z)}\right) = 0,$$



▲ **Figure 5.60** Selection rules for the permitted electron orbits according to Bohr and de Broglie.



▲ **Figure 5.61** Analogy between beam optics and classical mechanics.



▲ **Figure 5.62** Erwin Schrödinger (1887–1961): Physics studies in Vienna under Hasenöhrl, then in 1910 his assistant. One-semester appointments at several German universities; from 1921: professor for six years at the University of Zurich; 1927: called to Berlin; 1933: emigration first to Oxford, then return to Graz (1936). After the Austrian annexation, removed from his position, but was able to resume work in the peaceful atmosphere of Dublin. 1956, returned to Austria. Schrödinger was active in a great number of areas. He first worked on statistical thermodynamics, the theory of specific heat, and the theory of color vision, with lasting contributions even there. In his work on quantum mechanics, he first obtained the fundamental equation of wave mechanics in its relativistically invariant form (which is today called the Klein–Gordon equation); however, his explanation of the fine structure of spectra was faulty. In 1926, his seminal work *Quantisierung als Eigenwertproblem* (Quantization as eigenvalue problem) appeared. From 1927, when Max Born gave a statistical interpretation of the wave function, Schrödinger no longer supported the generally accepted Bohr–Heisenberg conception of quantum mechanics. In his last years, he published works on philosophical–ideological topics including a volume of poetry.

which, for the problem under consideration, will take the form

$$\Delta\Psi + \frac{2m}{K^2}\left(E + \frac{e^2}{r}\right)\Psi = 0.$$

Returning to the chronologically later, but following logically perhaps the more basic train of thought, let us take a second look at equation (6). If the most important characteristic of the wave, namely, the wavelength, is given, then we can find the transition from point mechanics to wave mechanics by analogy with the transition from geometric optics into wave optics: We need to insert the defined λ into a wave equation.

The wave equation for light has long been known as

$$\Delta\varphi - \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} = 0,$$

where the Laplacian $\Delta\varphi$ can be written in Cartesian coordinates as

$$\Delta\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2}.$$

In this wave equation, φ can be an arbitrary component of the electric or magnetic field strength of the electromagnetic field that forms the light wave.

If we are considering quantities that vary harmonically, that is, according to the sine function, then the wave equation can be further simplified. Namely, if we set

$$\varphi(x, y, z, t) = \varphi(x, y, z)\,\mathrm{e}^{-i2\pi\nu t},$$

we then obtain

$$\frac{\partial^2\varphi}{\partial t^2} = -(2\pi\nu)^2\varphi(x, y, z)\,\mathrm{e}^{-i2\pi\nu t} = -(2\pi\nu)^2\varphi(x, y, z, t).$$

Now the time-dependent factor $\mathrm{e}^{-2i\pi\nu t}$ can be canceled on both sides, yielding

$$\Delta\varphi = -\left(\frac{2\pi\nu}{c}\right)^2\varphi,$$

where φ depends only on the spatial coordinates. If we now take into account that $c/\nu = \lambda$, then the time-independent wave equation takes the form

$$\Delta\varphi + \left(\frac{2\pi}{\lambda}\right)^2\varphi = 0. \qquad (8)$$

We now look for the mechanical analogue of this equation. A wave associated with an electron is also described by a wave equation, and therefore, an electron must have some sort of wave function

$$\Psi(x, y, z)\,\mathrm{e}^{-i2\pi\nu t}$$

associated with it that varies harmonically with time and whose dependence on the spatial coordinates is yet to be determined. The wave function must satisfy a wave equation obtained from equation (8) by replacing λ with the expression given in equation (6):

$$\Delta\Psi + \frac{8\pi^2 m}{h^2}(E - U)\Psi = 0.$$

This equation is one of the fundamental equations of wave mechanics, the *time-independent Schrödinger equation*.

If, for a given problem, we now substitute the potential function $U(x, y, z)$ into the wave equation, we can then determine the function Ψ that depends on the spatial coordinates, and we obtain the time-dependent solution of the problem if we multiply Ψ by $\mathrm{e}^{-2i\pi\nu t}$. We obtain the energy E from the relationship

$$E = h\nu,$$

and the complete solution is then

$$\Psi(x,y,z,t) = \Psi(x,y,z)\mathrm{e}^{-i(2\pi/h)Et}.$$

It must be emphasized that the path that led SCHRÖDINGER to his equation cannot be thought of as a derivation based on logical deduction. At this level, we have to simply accept the Schrödinger equation as a fundamental law that has no further justification. Here we just sketched the motifs that helped SCHRÖDINGER to find his equation.

The significance of the Schrödinger equation is the following. Let's not worry for the time being about the physical interpretation of the quantity Ψ. Whatever physical interpretation it may have, we can in any case set certain general mathematical requirements for it, for instance, that it should be limited at finite distances and vanish at infinity. Physicists, and especially mathematicians, have long known that equations such as the wave equation of optics and the Schrödinger equation, given certain boundary conditions, can be solved only when the parameters in the equations have particular values. Just consider a string fixed at its endpoints, or a membrane, or even a microwave cavity resonator, where we will have a solution only when the constant in the wave equation—the wavelength—and the geometric size of the system are related in a very specific way, and in this relationship the whole numbers have distinguished roles. SCHRÖDINGER showed that in his wave equation, the natural conditions required of the function Ψ also have the characteristics of boundary conditions. Therefore:

> *For any given problem, the parameter E that represents the total energy of the system must be in a definite set of values for the wave equation to have a solution* (Quotation 5.29).

This set, in the most interesting cases, consists of discrete values. In other words, a given microsystem has a discrete set of total energies that can be determined from the Schrödinger equation. In mathematics, the parameter values for which a solution exists are called eigenvalues (or characteristic values), and the corresponding solutions are called eigenfunctions (characteristic functions). A physical problem, with the insertion of the potential function $U(x, y, z)$ into the SCHRÖDINGER equation, is reduced to a mathematical problem: One has to find discrete values for the total energy E, that is, the eigenvalues

$$E_1, E_2, \ldots, E_n, \ldots$$

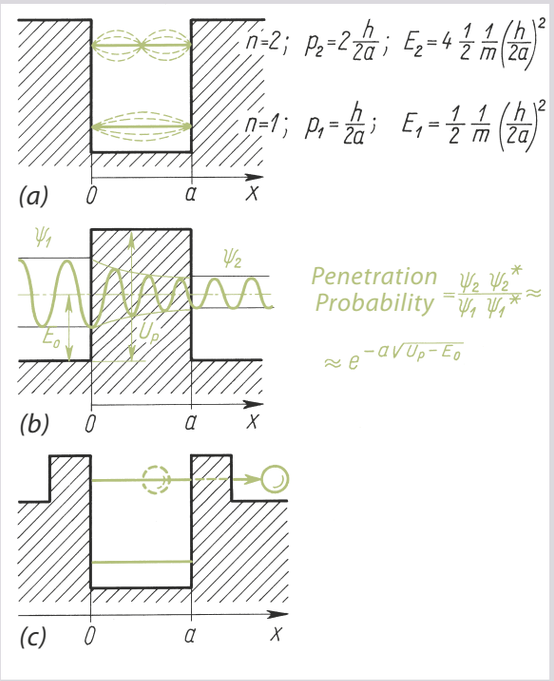
for which the Schrödinger equation is solvable. The associated wave functions

$$\Psi_1(x,y,z)\mathrm{e}^{-i2\pi(E_1/h)t}, \quad \Psi_2(x,y,z)\mathrm{e}^{-i2\pi(E_2/h)t}, \quad \ldots, \quad \Psi_n(x,y,z)\mathrm{e}^{-i2\pi(E_n/h)t}$$

are eigenfunctions, which are now of only secondary importance because we are mainly curious about the allowed—that is, possible—energy values.

Quantization therefore reduces to the mathematical problem of determining eigenvalues. This fact is forcefully reflected by the title of SCHRÖDINGER's article containing his famous equation: *Quantisierung als Eigenwertproblem*, that is, "quantization as an eigenvalue problem."

In Figure 5.63, we find the simplest—but still, in practice, very important—solutions to the Schrödinger equation. We have already mentioned that every new theory developed in the twentieth century has had to prove its viability by determining the energy levels of the hydrogen atom ("the hydrogen atom is invariant over all theories"). Let us now solve the Schrödinger equation for the hydrogen atom in the simplest case. Let Ψ be spherically symmetric; that is, Ψ is dependent only on r: $\Psi = \Psi(r)$. The expression $\Delta\Psi$ now



▲ **Figure 5.63** Application of the Schrödinger equation to three simple models: The equation is solvable in these cases using only elementary mathematics. It turns out that even from these models, which certainly provide an overly simplified view of reality, significant insights can be obtained. The general schema for solving problems in SCHRÖDINGER's wave mechanics is the following:

1. In analogy to the procedure in classical mechanics, where one has to substitute the characteristic force law for a concrete problem into Newton's equation, for every problem in wave mechanics, the corresponding expression U_p for the potential energy has to be written out in its classical form and substituted into the Schrödinger equation.
2. Based on physical considerations, we specify reasonable boundary conditions for the expected solution Ψ; in other words, we state how Ψ behaves at the boundaries of the area under investigation, for example, at infinity or at discontinuous points of U_p.
3. We find the function Ψ that satisfies both the partial differential equation from point 1 above and the boundary conditions of point 2.

The simplicity of the solution of the three problems depicted in the figure stems from the facts that they are one-dimensional (Ψ depends only on x) and that the potential energy is constant by region. Thus, the Schrödinger equation is reduced in each region to a differential equation that is not even partial but ordinary, or even simpler than that: it merely says that the desired $\Psi(x)$ is a function whose second derivative is equal to the function itself, disregarding only a constant factor. Such a function can be given at once; for example, the sine function. In addition, the boundary conditions are also simple: The problem illustrated in diagram (a) only requires that the function Ψ vanishes at the points $x = 0$ and $x = a$ (if the potential well is sufficiently deep).

continued on next page

Figure 5.63 *continued*

Diagram (a) is a model of a particle that is enclosed in a one-dimensional box. Very roughly, this is the situation in the nucleus of an atom. For the various possible wave functions, we obtain a familiar picture: the image of the fundamental vibration and harmonics of a stretched string fixed at both ends. From this analogy, SCHRÖDINGER believed that he could reduce the mysterious appearance of whole numbers to classical concepts. Here one obtains discrete values for the momentum and for the energy as well, due to $E_n = (1/2)p_n^2 / m$, which can be characterized by the quantum number n. We have a concrete manifestation of the general rule that a particle in a bound stationary state has discrete values of energy.

We also note the following fact: We know that a particle is located somewhere in the potential well of width a and that its minimal momentum is $p_1 = h/(2a)$. So we have $p_n a \geq h/2$, which is a concrete case of the Heisenberg uncertainty principle.

Whereas in diagram (a) we were dealing with discrete energy states of a stationary system, in the problem in diagram (b), we consider the behavior of a free particle when it meets a potential barrier. This problem can be illustrated classically and simply described: A ball can pass over a barrier only if its kinetic energy is greater than the potential energy at the highest point of the barrier; if this condition is not satisfied, then the ball rolls up (a slope) to the point determined by its energy, and then rolls back. However, the quantum-mechanical particle behaves completely differently from the classical one. According to quantum mechanics, there is a finite probability that we will find the particle on the other side of the barrier even if its kinetic energy is insufficient to pass over it. This surprising behavior is called the *tunnel effect* because the particle in effect tunnels its ways through the barrier. Of course, this picture is not to be taken too seriously. We approach reality more closely if we view the region of higher potential as an absorbing medium which the plane wave of constant amplitude corresponding to the particle (the electron) enters, is damped by, and comes out the other side of, with a smaller (but nonzero) amplitude. We should not forget here that the Ψ wave is a probability wave and so only has a physical meaning indirectly.

The tunnel effect plays an important role in a very large number of practically important electronic components (field electron emission, tunnel diodes, Josephson effect, etc.).

Diagram (c) shows a combination of the models in diagrams (a) and (b) that provides a better description of the relationships in the atomic nucleus. A bound particle located at a (quasi) stationary discrete energy level can leave the energy well in which it is trapped by way of the tunnel effect. Beginning with this picture, we can arrive, for example, at a theory of alpha decay.

Quotation 5.29

[The Schrödinger equation] has provided the basis of modern quantum mechanics, in which it seems to play the same part as do the equations established by NEWTON, LAGRANGE and HAMILTON in classical mechanics. Nevertheless there is an important distinction between these equations, consisting in the fact that in the latter equations the coordinates of the configuration point are not functions of time, but independent variables. Accordingly, while for any given system the classical equations of motion were more or less numerous and corresponded to

continued on next page

assumes the following form:

$$\Delta\Psi = \frac{\mathrm{d}^2\Psi}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}\Psi}{\mathrm{d}r}.$$

The potential energy of an electron with charge $-e$ in the field of the nucleus (here thought of as stationary) with charge $+e$ is

$$U = -\frac{e^2}{4\pi\varepsilon_0 r}.$$

The Schrödinger equation now reads

$$\frac{\mathrm{d}^2\Psi}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}\Psi}{\mathrm{d}r} + \frac{8\pi^2 m}{h^2}\left(E + \frac{e^2}{4\pi\varepsilon_0 r}\right)\Psi = 0.$$

We now prescribe the following conditions on Ψ: It is finite for finite values and approaches zero at infinity. The function $\Psi = e^{-ar}$ has these properties. Let us choose it then as a test function and investigate whether it can, in fact, satisfy the Schrödinger equation. After carrying out the indicated operations and rearranging, we obtain the following equation:

$$a^2 + \frac{8\pi^2 m}{h^2}E + \frac{1}{r}\left(\frac{8\pi^2 m}{h^2}\frac{e^2}{4\pi\varepsilon_0} - 2a\right) = 0.$$

This equation can work for all values of the independent variable r only if we have

$$a^2 + \frac{8\pi^2 m}{h^2}E = 0 \quad \text{and} \quad \frac{8\pi^2 m}{h^2}\frac{e^2}{4\pi\varepsilon_0} - 2a = 0.$$

If we now substitute the value of a from the right-hand equation into the left-hand equation, we obtain for the energy

$$E = \frac{me^4}{8e_0^2 h^2}.$$

This, in turn, is identical to the value given by Bohr's theory for the first energy level of the hydrogen atom (Figure 5.50).

There were many lengthy debates over what the proper interpretation of the state function Ψ should be, not always free of passion, and in some sense the debate continues even to this day. At the outset, there was a faint hope that quantum theory might somehow fit into the framework of classical physics: Physicists had frequently encountered differential equations similar to the Schrödinger equation in classical physics. Moreover, the Ψ function—in the case of a single particle—could be interpreted in three-dimensional space as a vibrational state. No wonder that Schrödinger's wave mechanics was much more popular with physicists than matrix mechanics, whose abstract mathematics alone already called attention to its revolutionary new character. SCHRÖDINGER himself was of the opinion that his theory would make a return to classical physics possible on a deterministic basis operating with continuous processes. Instead of a point-like electron, he envisaged a continuously distributed charge, with the function $e|\Psi|^2$ interpreted as the distribution of the charge -density.

It is primarily to the credit of MAX BORN (1882–1970) that a statistical interpretation was given to the Ψ function, appearing in his 1926 article jointly written with NORBERT WIENER. With this the debate began, reaching even beyond the boundaries of physics, about the determinacy or indeterminacy, causality or acausality, of natural phenomena. What emerged here requires an even more revolutionary rethinking of our worldview than did relativity.

Born speaks thus of his work in his Nobel lecture:

> It appeared to me that it was not possible to arrive at a clear interpretation of the Ψ-function by considering bound electrons. I had therefore been at pains, as early as the end of 1925, to extend the matrix method, which obviously covered only oscillatory processes, in such a way as to be applicable to aperiodic processes. I was at that time the guest of the Massachusetts Institute of Technology in the U.S.A., and there I found in Norbert Wiener a distinguished collaborator. In our joint paper we replaced the matrix by the general concept of the operator and, in this way, made possible the description of aperiodic processes. Yet we missed the true approach, which was reserved for Schrödinger; and I immediately took up his method, since it promised to lead to an interpretation of the ψ-function. Once more an idea of Einstein's gave the lead. He had sought to make the duality of particles (light quanta or photons) and waves comprehensible by interpreting the square of the optical wave amplitudes as probability density for the occurrence of photons. This idea could at once be extended to the Ψ-function: Ψ^2 must represent the probability density for electrons (or other particles). To assert this was easy; but how was it to be proved? ...
>
> But the factor that contributed more than these successes to the speedy acceptance of the statistical interpretation of the Ψ-function was a paper by Heisenberg that contained his celebrated uncertainty relationship, through which the revolutionary character of the new conception was first made clear.
>
> —Max Born, "Statistical Interpretation of Quantum Mechanics," Nobel Lecture

Even though the formalisms of Heisenberg's and Schrödinger's theories were so fundamentally different, Schrödinger himself pointed out that they are mathematically equivalent, despite the initial aversion the two scientists had to each other's ideas. Heisenberg's opinion:

> The more I work with the physical parts of the Schrödinger theory, the more disgusting I find it.

Schrödinger, in turn, wrote thus:

> I found the complicated methods of transcendental algebra, which make any visualization impossible, annoying, almost repulsive.

In their philosophical conclusions they differed as well: Schrödinger—like Einstein (Quotations 5.30, 5.31)—did not want to accept the fact that a return to a description of natural phenomena based on causal space-time relations is untenable even as an eventual future hope. However, the majority of physicists accepted the "Copenhagen interpretation" (Figure 5.64) worked out by Heisenberg and Bohr. See also Figures 5.65 through 5.70 for further illustrations of the times and the issues.

5.3.10 *Heisenberg: The Copenhagen Interpretation of Quantum Theory*

> The Copenhagen interpretation of quantum theory starts from a paradox. Any experiment in physics, whether it refers to the phenomena of daily life or to atomic events, is to be described in the terms of classical physics. The concepts of classical physics form the language by which we describe the arrangements of our experiments and state the results. We cannot and should

Quotation 5.29, ***continued***

> the number of degrees of freedom of the system, there can be only one single quantum equation for each system. In course of time the configuration point of classical theory describes a definite curve; on the other hand, the configuration point of the material wave fills at any given time the whole of infinite space, including those parts of space where potential energy is greater than the total energy, so that according to the classical theory, kinetic energy would become negative in these parts of space, and the momentum imaginary. ...
>
> From the discrete characteristic energy values, discrete characteristic values of the period of oscillation may be derived. The latter are determined according to the quantum postulate, in a similar manner to that of a stretched cord with fixed ends; with this distinction that the latter quantization is determined by an external condition, viz. the length of the cord, whereas in the present instance it depends upon the quantum of action, which in turn depends directly upon the differential equation.
>
> To each characteristic vibration there corresponds a particular wave function (ψ); this is the solution of the wave equation; and all these different characteristic functions form the component elements for the description of any movement in terms of wave mechanics.
>
> —Planck, *The Universe in the Light of Modern Physics* [pp. 28–30]

Quotation 5.30

> The generation to which Einstein, Bohr and I belong, was taught that there exists an objective physical world, which unfolds itself according to immutable laws independent of us; we are watching this process as the audience watches a play in a theatre. Einstein still believes that this should be the relation between the scientific observer and his subject. Quantum mechanics, however, interprets the experience gained in atomic physics in a different way. We may compare the observer of a physical phenomenon not with the audience of a theatrical performance, but with that of a football game where the act of watching, accompanied by applauding or hissing, has a marked influence on the speed and concentration of the players, and thus on what is watched. In fact, a better simile is life itself, where audience and actors are the same persons.
>
> —Born, "Physics and Relativity" [pp. 104–105]

Quotation 5.31

Quantum mechanics is certainly imposing. But an inner voice tells me that it is not yet the real thing. The theory says a lot, but does not really bring us any closer to the secret of the "old one". I, at any rate, am convinced that *He* is not playing at dice ... I am working very hard at deducing the equations of motion of material points regarded as singularities, given the differential equations of general relativity.

—Einstein, letter to Born, 1926 [Born 1971, p. 91]

It is my opinion that the contemporary quantum theory by means of certain definitely laid down basic concepts, which on the whole have been taken over from classical mechanics, constitutes an optimum formulation of the connections. I believe, however, that this theory offers no useful point of departure for future development. This is the point at which my expectation departs most widely from that of contemporary physicists.

—Einstein, "Autobiographical Notes" [p. 87]

not replace these concepts by any others. Still the application of these concepts is limited by the relations of uncertainty. We must keep in mind this limited range of applicability of the classical concepts while using them, but we cannot and should not try to improve them.

For a better understanding of this paradox it is useful to compare the procedure for the theoretical interpretation of an experiment in classical physics and in quantum theory. In Newton's mechanics, for instance, we may start by measuring the position and the velocity of the planet whose motion we are going to study. The result of the observation is translated into mathematics by deriving numbers for the co-ordinates and the momenta of the planet from the observation. Then the equations of motion are used to derive from these values of the co-ordinates and momenta at a given time the values of these co-ordinates or any other properties of the system at a later time, and in this way the astronomer can predict the properties of the system at a later time. He can, for instance, predict the exact time of an eclipse of the moon.

In quantitative theory the procedure is slightly different. We could for instance be interested in the motion of an electron through a cloud chamber and could determine by some kind of observation the initial position and velocity of the electron. But this determination will not be accurate; it will at least contain the inaccuracies following from the uncertainty relations and will probably contain still larger errors due to the difficulty of the experiment. It is the first of these inaccuracies which allows us to translate the result of the observation into the mathematical scheme of quantum theory. A probability function is written down which represents the experimental situation at the time of the measurement. ...

*

When the probability function in quantum theory has been determined at the initial time from the observation, one can from the laws of quantum theory calculate the probability function at any later time and can thereby determine the probability for a measurement giving a specified value of the measured quantity. We can, for instance, predict the probability for finding the electron at a later time at a given point in the cloud chamber. It should be emphasized, however, that the probability function does not in itself represent a course of events in the course of time. It represents a tendency for events and our knowledge of events. The probability function can be connected with reality only if one essential condition is fulfilled: if a new measurement is made to determine a certain property of the system. Only then does the probability function allow us to calculate the probable result of the new measurement. The result of the measurement again will be stated in terms of classical physics.

Therefore, the theoretical interpretation of an experiment requires three distinct steps: (1) the translation of the initial experimental situation into a probability function; (2) the following up of this function in the course of time; (3) the statement of a new measurement to be made of the system, the result of which can then be calculated from the probability function. For the first step the fulfillment of the uncertainty relations is a necessary condition. The second step cannot be described in terms of the classical concepts; there is no description of what happens to the system between the initial observation and the next measurement. It is only in the third step that we change over again from the "possible" to the "actual." ...

*

Bohr uses the concept of "complementarity" at several places in the interpretation of quantum theory. The knowledge of the position of a particle is complementary to the knowledge of its velocity or momentum. If we know the one

▶ **Figure 5.64** The symbol of Copenhagen is the mermaid, a figure from one of Andersen's fairy tales; the physicist, however, associates Copenhagen primarily with Bohr and his school, to which we may also count Heisenberg, Pauli, Dirac, Ehrenfest, Gamow, Landau, Kramers, and Klein. The atmosphere was familiar, but opponents were not spared. Heisenberg reports on a visit by Schrödinger in *Physics and Beyond* [pp. 73–76]:

> [A]lthough Bohr was normally most considerate and friendly in his dealings with people, he now struck me as an almost remorseless fanatic, one who was not prepared to make the least concession or grant that he could ever be mistaken.

On this occasion, Schrödinger offered the following bitter, and frequently quoted, opinion:

> If all this damned quantum jumping were really here to stay, I should be sorry I ever got involved with quantum theory.

Heisenberg recalls that the discussion became more and more violent:

> And so discussions continued day and night. After a few days Schrödinger fell ill, perhaps as a result of his enormous effort; in any case, he was forced to keep to his bed with a feverish cold. While Mrs. Bohr nursed him and brought in tea and cake, Niels Bohr kept sitting on the edge of the bed talking to Schrödinger: "But you must surely admit that…" No real understanding could be expected since, at the time, neither side was able to offer a complete and coherent interpretation of quantum mechanics. For all that, we in Copenhagen felt convinced toward the end of Schrödinger's visit that we were on the right track, though we fully realized how difficult it would be to convince even leading physicists that they must abandon all attempts to construct perceptual models of atomic processes.

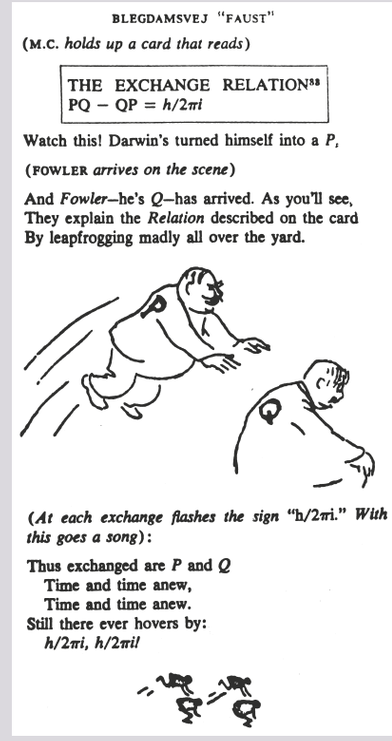
Some significant elements of the Copenhagen interpretation may have come from Bohr's upbringing. Jammer (1974) in his analysis of Bohr's philosophical background highlights Kierkegaard, whose student Høffding was a university colleague of Bohr's father. According to Kierkegaard, the creator of a philosophical system is himself a part of that system that he wishes to explain:

> One cannot view oneself, without self-deception, as an indifferent onlooker or impersonal observer; one is necessarily always a participant. The determination of the boundary between the objective and the subjective is an arbitrary act, and human life consists in a series of decisions. Science is nothing other than a well-determined action; and truth: a work of man, and indeed, not only because it is man who created knowledge, but because the significant object of knowledge is not something that has been standing complete since the beginning of time.

The psychology of William James also emphasizes the arbitrariness in demarcation and the uncontrollable influence of observation on the observed phenomenon. Finally, the philosophers of the Viennese School, who were greatly influenced by Mach, had similar ideas ("an assertion makes sense only if it can be empirically verified"). The motto of Wittgenstein, who came from this circle, was this: "Whereof one cannot speak thereof one must be silent."

> with high accuracy we cannot know the other with high accuracy; still we must know both for determining the behavior of the system. The space-time distribution of the atomic events is complementary to their deterministic description. The probability function obeys an equation of motion as the co-ordinates did in Newtonian mechanics; its change in the course of time is completely determined by the quantum mechanical equation, but it does not allow a description in space and time. The observation, on the other hand, enforces the description in space and time but breaks the determined continuity of the probability function by changing our knowledge of the system. …
>
> *
>
> A real difficulty in the understanding of this interpretation arises, however, when one asks the famous question: But what happens "really" in an atomic event? It has been said before that the mechanism and the results of an observation can always be stated in terms of the classical concepts. But what one deduces from an observation is a probability function, a mathematical expression that combines statements about possibilities or tendencies with statements about our knowledge of facts. So we cannot completely objectify the result of an observation, we cannot describe what "happens" between this observation and the next. This looks as if we had introduced an element of subjectivism into the theory, as if we meant to say: what happens depends





BLEGDAMSVEJ "FAUST"

(M.C. *holds up a card that reads*)

THE EXCHANGE RELATION[88]
PQ − QP = $h/2\pi i$

Watch this! Darwin's turned himself into a *P*,

(FOWLER *arrives on the scene*)

And *Fowler*—he's *Q*—has arrived. As you'll see,
They explain the *Relation* described on the card
By leapfrogging madly all over the yard.

(*At each exchange flashes the sign "h/2πi." With this goes a song*):

Thus exchanged are *P* and *Q*
Time and time anew,
Time and time anew.
Still there ever hovers by:
$h/2\pi i$, $h/2\pi il$

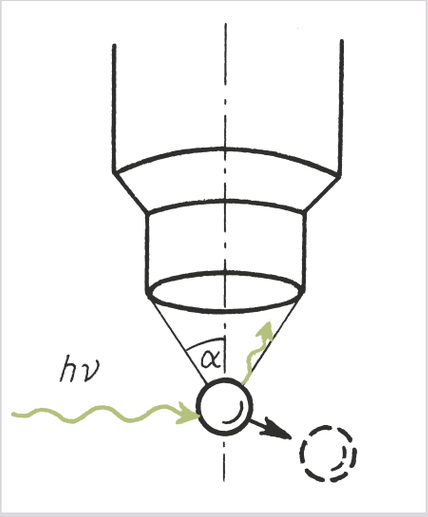
Figure 5.65 The collaborators in Bohr's school could not free themselves of science even in their amusements. We show here a facsimile page of a parody of Faust in which the distribution of roles is telling: Faust is Paul Ehrenfest; Mephisto, Wolfgang Pauli; and God, Niels Bohr.

Paul Ehrenfest (1880–1933): Student of Boltzmann in Vienna; 1912: successor to Lorentz in the chair of theoretical physics in Leyden. The adiabatic principle established by him in 1913 and worked out in detail in subsequent years bears his name. Ehrenfest showed in 1927 that the equations of motion of classical mechanics follow from wave mechanics as a limiting case. With his critical and pedagogically valuable contributions to discussions, he contributed to the development of physics far beyond his concrete results. This contradiction between his important position and his creative capacity may have contributed to his depression and suicide.

continued on next page

Figure 5.65 *continued*

In our picture, Darwin and Fowler are playing leapfrog, which is supposed to represent the commutation relation. C. Darwin participated in the solution of the subproblems in quantum theory, in particular the incorporation of spin into quantum mechanics. The name R. H. Fowler is familiar to telecommunications engineers from the Fowler–Nordheim formula for field electron emission.



▲ **Figure 5.66** It was already known in classical physics that a measurement is always connected with a disturbance of the system being measured, so that it can never be made with infinite precision. Imagine, for example, that we wished to measure the temperature of a thimble full of warm water using an ordinary thermometer. Placing the thermometer in the water will cool the water, and we will measure this cooler temperature. But classical physics was firmly convinced that either the disturbance could be made arbitrarily small (in our example, we could use smaller and smaller thermometers) or the result of the measurement could be corrected by taking the disturbing effect into account (in our example, we could consider the thermal capacity of the thermometer).

In quantum mechanics, the interaction energy can never be made zero due to the energy quanta, and the fundamentally probabilistic nature of the elementary processes makes a precise correction impossible. More accurately: the possibility of reducing the influence and the possibility of correction work against each other.

We can make all of this more concrete by considering the measurement of the state of motion of an electron: its location and its momentum. To locate the electron within the viewing field of a microscope, we illuminate it with light. The elementary process leading to an observation will be a collision of a photon of energy *hv* with the electron, which changes the photon's path (classically, we would speak of light scattering; in quantum mechanics, this is called the *Compton effect*).

If we don't want the photon to collide too violently with the electron and thereby change the electron's momentum in a way that is not precisely correctable due to the statistical nature of the interaction, then we must use light of a lower frequency, and hence a greater wavelength. But from optics we know that the resolution capacity of the microscope and

continued on next page

on our way of observing it or on the fact that we observe it. Before discussing this problem of subjectivism it is necessary to explain quite clearly why one would get into hopeless difficulties if one tried to describe what happens between two consecutive observations.

For this purpose it is convenient to discuss the following ideal experiment: We assume that a small source of monochromatic light radiates toward a black screen with two small holes in it. The diameter of the holes may not be much bigger than the wave length of the light, but their distance will be very much bigger. At some distance behind the screen a photographic plate registers the incident light. If one describes this experiment in terms of the wave picture, one says that the primary wave penetrates through the two holes; there will be secondary spherical waves starting from the holes that interfere with one another, and the interference will produce a pattern of varying intensity on the photographic plate. [See Figure 5.67.]

The blackening of the photographic plate is a quantum process, a chemical reaction produced by single light quanta. Therefore, it must also be possible to describe the experiment in terms of light quanta. If it would be permissible to say what happens to the single light quantum between its emission from the light source and its absorption in the photographic plate, one could argue as follows: The single light quantum can come through the first hole or through the second one. If it goes through the first hole and is scattered there, its probability for being absorbed at a certain point of the photographic plate cannot depend upon whether the second hole is closed or open. The probability distribution on the plate will be the same as if only the first hole was open. If the experiment is repeated many times and one takes together all cases in which the light quantum has gone through the first hole, the blackening of the plate due to these cases will correspond to the probability distribution. If one considers only those light quanta that go through the second hole, the blackening should correspond to a probability distribution derived from the assumption that only the second hole is open. The total blackening, therefore, should just be the sum of the blackenings in the two cases; in other words, there should be no interference pattern. Therefore, the statement that any light quantum must have gone *either* through the first *or* through the second hole is problematic and leads to contradictions. This example shows clearly that the concept of the probability function does not allow a description of what happens between two observations. Any attempt to find such a description would lead to contradictions; this must mean that the term "happens" is restricted to the observation.

Now, this is a very strange result, since it seems to indicate that the observation plays a decisive role in the event and that the reality varies, depending upon whether we observe it or not. To make this point clearer we have to analyze the process of observation more closely.

To begin with, it is important to remember that in natural science we are not interested in the universe as a whole, including ourselves, but we direct our attention to some part of the universe and make that the object of our studies. In atomic physics this part is usually a very small object, an atomic particle or a group of atomic particles, sometimes much larger—the size does not matter; but it is important that a large part of the universe, including ourselves, does *not* belong to the object.

Now, the theoretical interpretation of an experiment starts with the two steps that have been discussed. In the first step we have to describe the arrangement of the experiment, eventually combined with a first observation, in terms of classical physics and translate this description into a probability function.

This probability function follows the laws of quantum theory, and its change in the course of time, which is continuous, can be calculated from the initial conditions; this is the second step. The probability function combines objective and subjective elements. It contains statements about possibilities or better tendencies ("potentia" in Aristotelian philosophy), and these statements are completely objective, they do not depend on any observer; and it contains statements about our knowledge of the system, which of course are subjective insofar as they may be different for different observers. In ideal cases the subjective element in the probability function may be practically negligible as compared with the objective one. The physicists then speak of a "pure case."

When we now come to the next observation, the result of which should be predicted from the theory, it is very important to realize that our object has to be in contact with the other part of the world, namely, the experimental arrangement, the measuring rod, etc., before or at least at the moment of observation. This means that the equation of motion for the probability function does now contain the influence of the interaction with the measuring device. This influence introduces a new element of uncertainty, since the measuring device is necessarily described in the terms of classical physics; such a description contains all the uncertainties concerning the microscopic structure of the whole world. These uncertainties may be called objective insofar as they are simply a consequence of the description in the terms of classical physics and do not depend on any observer. They may be called subjective insofar as they refer to our incomplete knowledge of the world.

After this interaction has taken place, the probability function contains the objective element of tendency and the subjective element of incomplete knowledge, even if it has been a "pure case" before. It is for this reason that the result of the observation cannot generally be predicted with certainty; what can be predicted is the probability of a certain result of the observation, and this statement about the probability can be checked by repeating the experiment many times. The probability function does—unlike the common procedure in Newtonian mechanics—not describe a certain event but, at least during the process of observation, a whole ensemble of possible events.

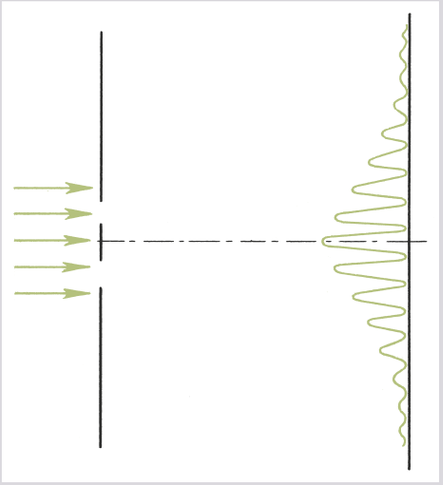
The observation itself changes the probability function discontinuously; it selects of all possible events the actual one that has taken place. Since through the observation our knowledge of the system has changed discontinuously, its mathematical representation also has undergone the discontinuous change and we speak of a "quantum jump." When the old adage "Natura non facit saltus" is used as the basis for criticism of quantum theory, we can reply that certainly our knowledge can change suddenly and that this fact justifies the use of the term "quantum jump."

Therefore, the transition from the "possible" to the "actual" takes place during the act of observation. If we want to describe what happens in an atomic event, we have to realize that the word "happens" can apply only to an observation, not to the state of affairs between two observations. It applies to the physical, not to the psychical act of observation, and we may say that the transition from the "possible" to the "actual" takes place as soon as the interaction of the object with the measuring device, and thereby with the rest of the world, has come into play; it is not connected with the act of registration of the result by the mind of the observer. The discontinuous change in the probability function, however, takes place with the act of registration, because it is the discontinuous change of our knowledge in the instant of registration that has its image in the discontinuous change of the probability function.

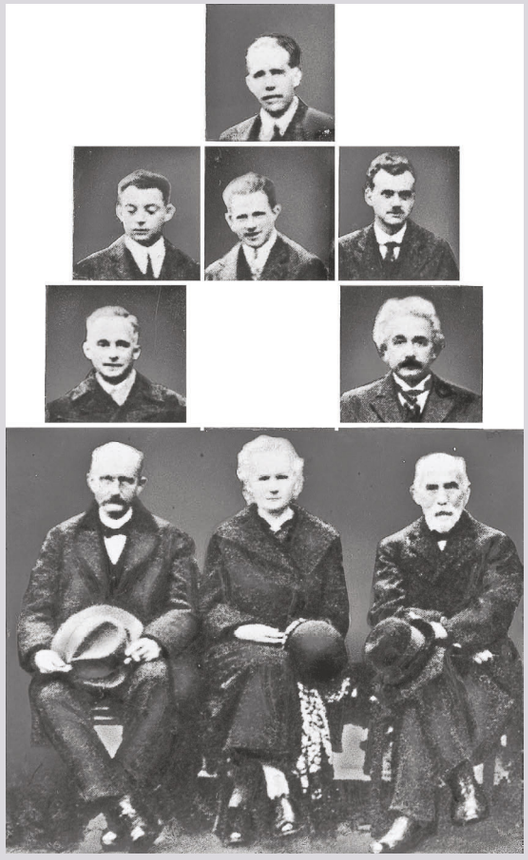
Figure 5.66 *continued*

hence the precision of the determination of location of the electron is reduced with increasing wavelength λ. According to Abbe's law, the uncertainty of the location is given by $\Delta x = \lambda/\sin\alpha$. (We cannot resist mentioning the odd fact that Heisenberg, at his doctoral oral examination in 1923, was unable to answer a question posed by Wien about the resolution capability of a microscope.) From the point of view of determining the location, it is desirable to have a smaller value of λ that has a greater frequency ν. For a given λ, the momentum of a light quantum of energy $h\nu$ is h/λ, and after the collision, the horizontal component of the photon's momentum must lie between $-(h/\lambda)\sin\alpha$ and $+(h/\lambda)\sin\alpha$ so that the photon can be scattered into the field of the microscope. But from this follows the uncertainty of the electron's momentum at $\Delta p_x = (h/\lambda)\sin\alpha$. No matter how we choose λ, the following relationship between the uncertainty of the location measurement Δx and the momentum measurement Δp_x will hold:

$$\Delta p_x \cdot \Delta x \approx \frac{\lambda}{\sin\alpha}\frac{h}{\lambda}\sin\alpha = h.$$



▲ **Figure 5.67** The interference experiment described by Heisenberg (Section 5.3.10): The plane wave is diffracted at the two slits, and the interference can be seen on the screen. These experiments are of particular interest in the interpretation of quantum theory when the light intensities are so weak that one can be certain that at any given moment, only one photon is interacting with the system of slits and in effect it interferes "with itself."



▲ **Figure 5.68** Portraits of the participants in the Solvay Congress of 1927: The leading participants were Bohr and the "youngsters" (Pauli, Heisenberg, Dirac); two representatives of the middle generation, one of whom (Born) was their supporter and the other (Einstein) an opponent; and finally, physicists of the older generation (Planck, Mme. Curie, Lorentz), who looked on with a mixture of estrangement and affection at what their ideas had grown into. Seeing the picture of these three greats of early twentieth-century physics does not just evoke awe for their towering intellects but, because of their frailty, also our empathy. They were still creative. Lorentz was still at work replying at length to Schrödinger's letters and to the questions posed by Goudsmit and Uhlenbeck related to spin.

The official topics of the conference: electrons and photons. The battle lines formed around the "conservatives" (Einstein, Schrödinger, de Broglie) on one side and representatives of the Copenhagen school (Bohr, Heisenberg, Pauli, Born) on the other side. The conservatives, led by Einstein, still hoped that the fundamental equations of quantum mechanics could be interpreted within the schema of the classical equations and within the framework of a deterministic physics, whereas the "Copenhageners" had argued for the fundamentally statistical nature of physical phenomena and for all that followed in consequence.

The 74-year-old Lorentz was the president of the congress. After lectures by Bragg, Compton, and de Broglie, Born and Heisenberg spoke on matrix mechanics, transformation theory, and the statistical interpretation. They provocatively summarized their position at the end of their lecture:

continued on next page

To what extent, then, have we finally come to an objective description of the world, especially of the atomic world? In classical physics, science started from the belief—or should one say from the illusion?—that we could describe the world or at least parts of the world without any reference to ourselves. This is actually possible to a large extent. We know that the city of London exists whether we see it or not. It may be said that classical physics is just that idealization in which we can speak about parts of the world without any reference to ourselves. Its success has led to the general idea of an objective description of the world. Objectivity has become the first criterion for the value of any scientific result. Does the Copenhagen interpretation of quantum theory still comply with this ideal? One may perhaps say that quantum theory corresponds to this ideal as far as possible. Certainly quantum theory does not contain genuine subjective features, it does not introduce the mind of the physicist as a part of the atomic event. But it starts from the division of the world into the "object" and the rest of the world, and from the fact that at least for the rest of the world we use the classical concepts in our description. This division is arbitrary and historically a consequence of the general human way of thinking. But this is already a reference to ourselves and insofar as our description is not completely objective.

It has been stated in the beginning that the Copenhagen interpretation of quantum theory starts with a paradox. It starts from the fact that we describe our experiments in the terms of classical physics and at the same time from the knowledge that these concepts do not fit nature accurately. The tension between these two starting points is the root of the statistical character of quantum theory. Therefore, it has sometimes been suggested that one should depart from the classical concepts altogether and that a radical change in the concepts used for describing the experiments might possibly lead back to a nonstatistical, completely objective description of nature.

This suggestion, however, rests upon a misunderstanding. The concepts of classical physics are just a refinement of the concepts of daily life and are an essential part of the language which forms the basis of all natural science. Our actual situation in science is such that we *do* use the classical concepts for the description of the experiments, and it was the problem of quantum theory to find theoretical interpretation of the experiments on this basis. There is no use in discussing what could be done if we were other beings than we were. At this point we have to realize, as von Weizsäcker has put it, that "Nature is earlier than man, but man is earlier than natural science." The first part of the sentence justifies classical physics, with its ideal of complete objectivity. The second part tells us why we cannot escape the paradox of quantum theory, namely, the necessity of using the classical concepts.

—Werner Heisenberg, *Physics and Philosophy* [pp. 46–56]

5.3.11 *Operators. Quantum Electrodynamics*

A glance at Table 5.1 should be sufficient to convince oneself that the formative years of quantum mechanics were witness to the most tempestuous development in the history of physics. In the course of a few years, classical Newtonian mechanics was replaced in its entirety by a new mechanics. The young scientists effecting this transition were Heisenberg, Schrödinger, Pauli (Figure 5.71), and Dirac (Figure 5.72). The physicists of the older generation, with a mixture of admiration and mockery, dubbed quantum mechanics the "physics of the twenty-year-olds." Among the senior professors, Born managed to keep pace with the younger gen-

eration, among them his assistant Heisenberg, but under the stress of doing so he suffered a nervous breakdown in 1928 and stopped working for several years. From the middle generation, Bohr allied himself with the young physicists from the start, and soon it was he who set the pace. Sommerfeld was averse to the new ideas, yet because of his familiarity and skill with the mathematical methods, was able to contribute to the pedagogical refinement of the quantum theory and, with the solution of concrete problems, to the formation of its future. Einstein's role was limited to that of a skeptical but constructive critic.

Heisenberg's matrix mechanics, Schrödinger's wave mechanics, the proof of the equivalence of the two theories, Dirac's operator representation, the establishment of the mathematical basis of the theory by John von Neumann, and finally the pinnacle, the formulation of the relativistically invariant fundamental equation of Dirac—these were the steps that led to the form of quantum mechanics that we learn and teach today at our universities.

For an axiomatic formulation of quantum theory, we begin with assuming that every physical quantity is represented by an operator, that is, a rule by which it is computed. (Of course, the operators must satisfy certain mathematical conditions.) The state of a physical system, in turn, is described by a state function. The computational rule refers to this state function, or, as the physicists put it, the operators are applied to the state function. The so-called characteristic states or eigenstates of the system have particular significance. These states are described by well-determined functions (called eigenfunctions) to which fixed values of the physical quantity in question (called the eigenvalues) belong. To determine the eigenvalues and eigenstates of a physical quantity, an eigenvalue equation similar to the Schrödinger equation needs to be found.

Now, if we take an arbitrary state function and apply the operator for a given physical quantity to it, that is, if we perform the computational rule, we will, of course, get another function. The fundamental question in the search for eigenfunctions is then the following: Can we find a function such that if we apply the operator to it, we get the same function back multiplied by a constant? At the same time, we will also require of a state function that it satisfy some reasonable mathematical conditions, as we have already seen in the discussion of the Schrödinger equation. Such, for instance, is the requirement that the function be single-valued, which is to say that at any particular point in space it should have just one definite value. It can be shown in general that such an eigenvalue equation at definite values of the above-mentioned multiplier constant has definite solutions: These are just the eigenvalues and the eigenfunctions.

Given all this highly formal mathematics, the question remains how to connect it with physical reality. This connection is given by the other basic feature of quantum mechanics, namely, that it determines the expected, or average, value over a series of repeated measurements of the physical quantity in question, given the state function that describes the state of the system. *Compared with classical mechanics, the differences are twofold: First, the concept of a physical quantity itself has become more abstract; and second, the relationships of quantum mechanics record only expected values, and do not predict the concrete result of any given actual measurement.*

We will now translate the above description into the technical language of physics. Let us begin by writing down the Schrödinger equation:

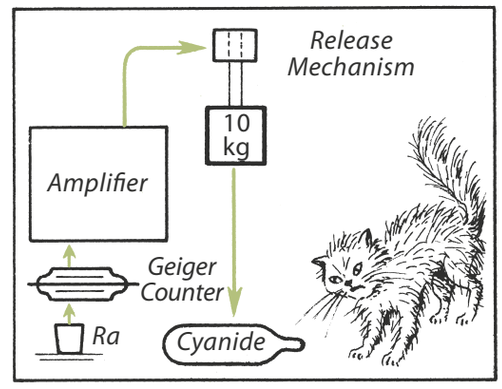
$$\Delta\Psi + \frac{8\pi^2 m}{h^2}(E - U)\Psi = 0.$$

Figure 5.68 *continued*

> Quantum mechanics is fundamentally a statistical theory and as such a closed theory that permits no modification of its physical and mathematical postulates.

After Schrödinger's lecture, Bohr spoke on epistemological implications. Heisenberg described the tense atmosphere as follows:

> Bohr and Einstein were in the thick of it all. Einstein was quite unwilling to accept the fundamentally statistical character of the new quantum theory. Needless to say, he had no objections against probability statements whenever a particular system was not known in every last detail—after all, the old statistical mechanics and thermodynamics had been based on just such statements. However, Einstein would not admit that it was impossible, even in principle, to discover all the partial facts needed for the complete description of a physical process. "God does not throw dice" was a phrase we often heard from his lips in these discussions. And so he refused point blank to accept the uncertainty principle, and tried to think up cases in which the principle would not hold. … Then, in the course of the day, we would have further discussions on the matter, and, as a rule, by suppertime we would have reached the point where Niels Bohr could prove to Einstein that even his latest experiment failed to shake the uncertainty principle. … Einstein's friend Paul Ehrenfest, a physicist from Leyden in Holland, said: "Einstein, I am ashamed of you; you are arguing against the new quantum theory just as your opponents argue about relativity theory." But even this friendly admonition went unheard.
>
> —*Physics and Beyond* [p. 80]

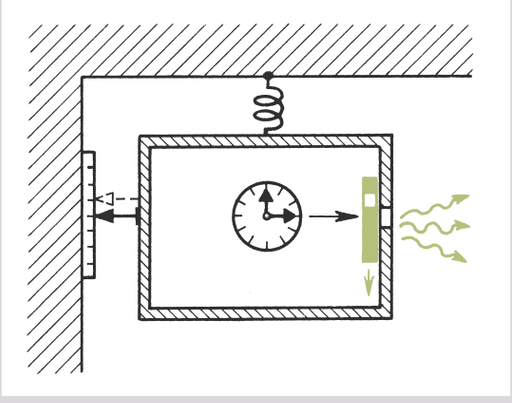


▲ **Figure 5.69** A much-quoted example put forth by Schrödinger to explain the inadequacy of the quantum-mechanical formulation: A closed container contains a radioactive atom that will decay with 50% probability within the next 60 minutes. The particle emitted when the atom decays is detected by a Geiger counter, whose signal is amplified, resulting in the release of a weight that falls, breaking open a container of cyanide, killing the cat.

continued on next page

Figure 5.69 *continued*

The question now to be answered is whether the cat is dead or alive 60 minutes after the end of the experiment. According to the probabilistic interpretation of quantum mechanics, the state of the cat is described by a Ψ function that results from a combination of the two possible states of the cat (alive or dead) in equal portions. However, if we open the container and look in, we would be able to see whether the cat is alive or dead, and it is obvious that it would not be due to our looking into the box because the state of the cat had been well determined, independent of us or our observation, even before the container was opened. This example can be much less dramatically and much more simply presented thus: Radioactive decay causes a mechanical counter to advance by one unit, and the entire technology of particle counting relies, of course, on the fact that the counting mechanism can operate on its own without any observations, that is, it objectively records the particles independent of us—there is no need to execute a cat in such a complicated way every time a decay takes place. BOHR commented that the experimental situation and the experimental result are two parts of a single phenomenon that join our macroworld and so we cannot limit ourselves to describing the intervening processes that are dominated by quantum laws.



▲ **Figure 5.70** According to HEISENBERG, the opponents of the Copenhagen interpretation can be divided into three groups. The first is in complete agreement with all the physical predictions of quantum mechanics, with the mathematical formalism, and with the limitations on the precisions of the measurements, but is not comfortable with the philosophical foundations. The second group agrees that the current formalism admits *only* the Copenhagen interpretation, but they conclude that the problem is then with the theory itself. The third group simply does not like quantum mechanics *in general*, even though they cannot find any logical problems with it and do not offer any constructive alternatives. HEISENBERG lists EINSTEIN, together with SCHRÖDINGER and LAUE, in the latter group. EINSTEIN frequently constructed concrete counterexamples to illustrate his misgivings. One of his famous counterexamples, with which he surprised BOHR at the 1930 Solvay Congress (although this topic was not on the conference's agenda), is shown here.

continued on next page

We now do a bit of rearrangement, all within the confines of high-school algebra:

$$-\frac{h^2}{8\pi^2 m}\Delta\Psi + U\Psi = E\Psi.$$

Advancing to college-level mathematics, if we bring this equation into the form

$$\left(-\frac{h^2}{8\pi^2 m}\Delta + U\right)\Psi = E\Psi, \qquad (9)$$

we can recast the problem of solving the Schrödinger equation as follows. The expression in parentheses on the left-hand side of equation (9) amounts to the following rule: Partially differentiate Ψ twice with respect to x, y, and z; add the second derivatives; and multiply this sum by $(-h^2/8\pi^2 m)$. Finally, multiply Ψ by U and add the product to the expression just calculated. Put more briefly, apply the operator **H** to the state function Ψ, where, by definition, **H** has the following form:

$$\mathbf{H} = -\frac{h^2}{8\pi^2 m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + U(x, y, z).$$

Using the **H** operator, the Schrödinger equation can be written in the simple form

$$\mathbf{H}\Psi = E\Psi.$$

The letter **H** refers to the name HAMILTON, and the operator is called the Hamiltonian operator. As we shall soon see, this operator is very closely connected with the classical Hamiltonian function.

Thus we have transformed the Schrödinger equation into an eigenvalue equation: We look for functions Ψ that remain unchanged up to a well-defined constant after execution of the mathematical operations prescribed by the operator **H**. In this case, the constants signify the possible energy values of the system. Through the solution of the eigenvalue problem, we have obtained the set $\Psi_1, \Psi_2, \ldots, \Psi_v, \ldots$ of eigenfunctions with the associated set $E_1, E_2, \ldots, E_v, \ldots$ of eigenvalues.

From this point of view, the Schrödinger equation is nothing other than the eigenvalue equation associated with the Hamiltonian operator. But what does HAMILTON, who lived one hundred years earlier, have to do with quantum mechanics? The answer to this question sheds light on the close connection between quantum mechanics and classical physics. Let us once more write down the Hamiltonian operator and the classical Hamiltonian function, termwise, one below the other:

$$\mathbf{H} = \frac{1}{2m}\left[\left(\frac{h}{2\pi i}\right)^2\frac{\partial^2}{\partial x^2} + \left(\frac{h}{2\pi i}\right)^2\frac{\partial^2}{\partial y^2} + \left(\frac{h}{2\pi i}\right)^2\frac{\partial^2}{\partial z^2}\right] + U(x, y, z),$$

$$\mathrm{H} = \frac{1}{2m}\left[\mathrm{p}_x^2 + \mathrm{p}_y^2 + \mathrm{p}_z^2\right] + U(x, y, z).$$

We see that one can pass from the classical operator H to the operator **H** if one writes, in place of the classical quantities, the so-called momentum operators

$$\mathbf{p}_x = \frac{h}{2\pi i}\frac{\partial}{\partial x}, \quad \mathbf{p}_y = \frac{h}{2\pi i}\frac{\partial}{\partial y}, \quad \mathbf{p}_z = \frac{h}{2\pi i}\frac{\partial}{\partial z}.$$

In this case, we have

$$\mathbf{p}_x^2 = \mathbf{p}_x\mathbf{p}_x = \left(\frac{h}{2\pi i}\frac{\partial}{\partial x}\right)\left(\frac{h}{2\pi i}\frac{\partial}{\partial x}\right) = -\frac{h^2}{4\pi^2}\frac{\partial^2}{\partial x^2},$$

and so forth. We have now come a step further: We have found the operators associated with the *classical* momentum components. And even more: We have found a general rule—or better, a guideline—for finding the quantum-mechanical operator for a given physical quantity. To wit, we write down the *classical* formula for the quantity as a function of the position and the momentum, and then in the formula we substitute the operators for the position and the momentum. An example should help to clarify this

process. The components of the angular momentum in classical physics are

$$L_x = [\mathbf{r} \times \mathbf{p}]_x = yp_z - zp_y, \quad L_y = [\mathbf{r} \times \mathbf{p}]_y = zp_x - xp_z, \quad L_z = [\mathbf{r} \times \mathbf{p}]_z = xp_y - yp_x.$$

Following the instructions given above, we obtain the operator, for example, of the z component of the angular momentum

$$\mathbf{L}_z = \frac{h}{2\pi i}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right).$$

Starting from this relationship, we can study a particularly simple and transparent but very instructive case of finding eigenvalues: If we transform the formula for spherical coordinates, the expression becomes

$$\mathbf{L}_z = \frac{h}{2\pi i}\frac{\partial}{\partial \varphi}.$$

The eigenvalue equation is now

$$\mathbf{L}_z \Psi = l_z \Psi \rightarrow \frac{h}{2\pi i}\frac{\partial}{\partial \varphi}\Psi = l_z \Psi.$$

A substitution shows immediately that the function

$$\Psi = A\mathrm{e}^{i(2\pi/h)l_z\varphi}$$

satisfies the equation. The case is instructive because it shows how the natural requirement that the function be *single-valued* leads to the discrete eigenvalues. In a function of the form $e^{ia\varphi}$, if we replace φ with $\varphi + 2\pi$, we return to the same point in space, so the values of the function must be the same. In other words,

$$A\mathrm{e}^{i(2\pi/h)l_z\varphi} = A\mathrm{e}^{i(2\pi/h)l_z(\varphi+2\pi)} = A\mathrm{e}^{i(2\pi/h)l_z\varphi}\mathrm{e}^{i(2\pi/h)l_z 2\pi}.$$

From this, we immediately obtain the equation

$$\mathrm{e}^{i(2\pi/h)l_z 2\pi} = 1.$$

We know that an equation of the form $\mathrm{e}^{im2\pi} = 1$ can be satisfied by the values $m = 0, \pm 1, \pm 2, \ldots$, from which follows $(2\pi/h)l_z = m$, or

$$l_z = m\frac{h}{2\pi}, \quad m = 0, \pm 1, \pm 2, \ldots.$$

We have thus found the eigenvalues of the eigenvalue equation $\mathbf{L}_z = l_z\Psi$, which represent the possible values of the z-component of the angular momentum that can occur as the results of a measurement. We see that these correspond exactly to the Bohr requirement: The possible values of the angular momentum are integer multiples of $h/2\pi$. The systems of eigenvalues and eigenfunctions are therefore

$$l_z = \frac{h}{2\pi}, 2\frac{h}{2\pi}, 3\frac{h}{2\pi}, \ldots, m\frac{h}{2\pi}, \ldots, \quad \Psi = A\mathrm{e}^{i\varphi}, A\mathrm{e}^{i2\varphi}, A\mathrm{e}^{i3\varphi}, \ldots, A\mathrm{e}^{im\varphi}.$$

Let us return to making the above general schema and the connection between the mathematics and physical reality more precise. We can first assert that *a single measurement always returns an eigenvalue*. However, if a system is characterized by the state function Ψ and we measure the physical quantity associated with the operator $\mathbf{M}$ repeatedly, then we obtain a collection of eigenvalues, each with a certain probability. Quantum mechanics asserts that this expected, or average, value of these quantities is given by

$$M_{\text{expected}} \equiv \langle M \rangle = \int_V \Psi \cdot \mathbf{M}\Psi \, \mathrm{d}V.$$

The integral must be taken over the entire configuration space. It now becomes useful to introduce a

Figure 5.70 *continued*

A clock placed in a closed system activates a slotted shutter that allows energy to escape the system in the form of radiation. Before the shutter is opened, we measure the mass of the system, for example, by suspending the system from a spring balance. After energy has been emitted, we again weigh the mass, and, of course, we register a loss in mass of E/c^2. The determination of mass both before and after the emission of energy can be carried out over an arbitrarily long period of time, so we may take our measurements to any desired degree of accuracy, and in this way the imprecision ΔE of the energy measurement can be made to approach zero. There are also no theoretical limits on the precision of the time measurement, and so we obtain $\Delta E \times \Delta t \rightarrow 0$, in contradiction to the HEISENBERG uncertainty principle, which places a limit on the precision of simultaneous measurement of E and t.

The problem kept BOHR awake all night before he arrived at a solution: The key is provided by no less than Einstein's general theory of relativity, for the emission of energy causes the entire system to move in the gravitational field. The change in gravitational potential leads to a change in the speed of the clock, which leads, although indirectly, to the relationship $\Delta E \times \Delta t \approx h$.



▲ **Figure 5.71** Wolfgang Pauli (1900–1958): Studies with Sommerfeld in Munich; worked briefly with Born in Göttingen and Bohr in Copenhagen. 1923–1928: Hamburg; from 1928 to his death at the University of Zurich. During the Second World War (1940–1946) in the United States, where he was one of the few leading physicists who *did not* participate in the creation of the atomic bomb.

While still a student, Pauli, at Sommerfeld's behest, wrote a survey article on the theory of relativity for the *Enzyklopädie der mathematischen Wissenschaften*, which remains today one of the best summaries of this subject.

In physics, Pauli is remembered most for the exclusion principle, to which he was led by two sets of problems: the splitting of spectral lines in very strong magnetic fields, which suggested that the state of the electron is characterized by four quantum numbers, and the question of why all the electrons of the atomic shell were not in a state of least energy. In his first published work, in 1925, the introduction of the fourth quantum number j is purely formal, although he already determined that $j = l + s$ (where l is the angular momentum and $s = \pm ½$) and even that the interaction energy with the magnetic field is $m_l + 2m_s$. It could almost be said that Pauli already expressed in his theory the properties of an electron as having a spin and a magnetic moment (spin = $½h/2\pi$, magnetic moment = $2m_s\mu_B$); the physical interpretation, however, had to wait for G. E. Uhlenbeck and S. Goudsmit (1925).

The work of Heisenberg, Dirac, Slater, and Pauli in the years 1926–1929 made it clear that the Pauli exclusion principle, in the language of quantum-mechanics, is equivalent to the requirement that the state function of the electron (including its

continued on next page

simplifying notation. If we define the scalar product of the functions Ψ_a and Ψ_b by the equation

$$(\Psi_a, \Psi_b) = \int_V \Psi_a \cdot \Psi_b \, \mathrm{d}V,$$

then the expected value can be written in the following simple form:

$$\langle M \rangle = (\Psi, \mathbf{M}\Psi).$$

From this point of view, the state function can be conceptualized as a vector in abstract Hilbert space—or to simplify this a bit, in the space of square-integrable functions. In this way, concepts known from elementary vector algebra—such as, for example, "parallel," "vertical," and "projection"—can be used with some necessary caution to great heuristic effect.

We have not yet discussed the time dependence of Ψ, apart from the steady-state function

$$\mathrm{e}^{-i2\pi\nu t} = \mathrm{e}^{-i2\pi(E/h)t}.$$

If we differentiate with respect to time, we obtain

$$\frac{\partial}{\partial t}\mathrm{e}^{-i2\pi(E/h)t} = -i2\pi\frac{E}{h}\mathrm{e}^{-i2\pi(E/h)t}$$

and therefore also the equation

$$-\frac{h}{2\pi i}\frac{\partial}{\partial t}\mathrm{e}^{-i2\pi(E/h)t} = E\,\mathrm{e}^{-i2\pi(E/h)t}.$$

We can therefore see that in this simple case, multiplication by E corresponds to applying the operator $-(h/2\pi i)\partial/\partial t$. Hence we can write the Schrödinger equation in the following form:

$$\mathbf{H}\Psi = -\frac{h}{2\pi i}\frac{\partial \Psi}{\partial t}.$$

This equation, postulated to have general validity, now giving us the time-dependencies of the state function Ψ as well, is the most important formula of quantum mechanics.

For the physical quantities that have been abstracted into operators, the usual calculation rules are not valid. So, in general, the order in which we apply two operators to the state function is not irrelevant. In other words, the usual commutative rule of multiplication does not hold for operators. In general, $\mathbf{M}(\mathbf{N}\Psi)$ is different from $\mathbf{N}(\mathbf{M}\Psi)$, or, using a more compact notation,

$$\mathbf{MN} \neq \mathbf{NM}, \quad \mathbf{MN} - \mathbf{NM} \neq 0.$$

An important example of a pair of such noncommutative operators is that of the position and momentum operators x and $\mathbf{p}_x$. If we calculate $x\mathbf{p}_x$ and then $\mathbf{p}_x x$, we obtain

$$(x\mathbf{p}_x)\Psi = x(\mathbf{p}_x\Psi) = x\frac{h}{2\pi i}\frac{\partial \Psi}{\partial x} = \frac{h}{2\pi i}x\frac{\partial \Psi}{\partial x},$$

$$(\mathbf{p}_x x)\Psi = \mathbf{p}_x(x\Psi) = \frac{h}{2\pi i}\frac{\partial}{\partial x}(x\Psi) = \frac{h}{2\pi i}\left[\Psi + x\frac{\partial \Psi}{\partial x}\right],$$

and we have

$$(\mathbf{p}_x x)\Psi - (x\mathbf{p}_x)\Psi = \frac{h}{2\pi i}\left[\Psi + x\frac{\partial \Psi}{\partial x} - x\frac{\partial \Psi}{\partial x}\right] = \frac{h}{2\pi i}\Psi.$$

We thus obtain the following commutation rule:

$$\mathbf{p}_x x - x\mathbf{p}_x = \frac{h}{2\pi i}.$$

The position and momentum operators are thus noncommutative. This fact has broad ramifications. Namely, the only physical quantities that can be measured simultaneously in a physical system with arbitrary precision are those associated with commutative operators. For quantities with noncommutative operators, the Heisenberg uncertainty principle holds. In our case, it takes the following form:

$$\Delta x \Delta p_x \geq \frac{1}{2}\frac{h}{2\pi}.$$

A further important role is assumed by the expression **MN – NM**, the *commutator*. Dirac pointed out that $(2\pi i/h)$(**MN – NM**) assumes the role in quantum mechanics of the Poisson–Jacobi bracket in classical mechanics. Correspondingly, we obtain, as in the last equation in Section 4.2.7, the rule for the derivative of an operator with respect to time:

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \frac{\partial \mathbf{L}}{\partial t} + \frac{2\pi i}{h}[\mathbf{HL} - \mathbf{LH}].$$

This yields, among other things, the important theorem that if it does not depend explicitly on time, an operator that commutes with the Hamiltonian operator is invariant with respect to time. From this follow, for example, the conservation laws of quantum mechanics.

We have not yet mentioned the greatest flaw in the Schrödinger equation: It is not invariant with respect to the Lorentz transformation. Dirac's great accomplishment was removing this flaw for the electron's energy equation.

Let us attempt to construct the wave equation according to the usual method, by first writing down the classical Hamiltonian function, bringing it into an operator form, and applying it to the state function, where the right-hand side of the equation is expressed by $E\Psi$ (with $E = (h/2\pi i)\partial/\partial t$ as energy operator). But then, in the relativistic case, we run into difficulties: The equation of motion

$$\frac{\mathrm{d}(m\mathbf{v})}{\mathrm{d}t} = q_e\mathbf{E} + q_e\mathbf{v}\times\mathbf{B} \quad \left(\text{with } m = \frac{m_0}{\sqrt{1 - v^2/c^2}}\right)$$

is associated with the Hamiltonian function $\mathrm{H} = q_e U + \sqrt{m_0^2c^4 + c^2\left(\mathbf{p} + q_e\mathbf{A}\right)^2}$, and if we insert **p** as an operator, we end up, because of the square root, with an undefined operator expression. From 1926 on, a number of physicists, among them Schrödinger himself, as well as Klein, Gordon, and Fock, attempted to circumvent this difficulty by setting the operator $(\mathbf{H} - q_e U)^2$ equal to the operator $(E - q_e U)^2$. It later turned out that this results in an equation describing the behavior of particles with zero spin.

Dirac took a different course: He looked for a quantity that, multiplied by itself, gives precisely the expression under the square root because in this way, at least formally, one can then extract the square root. To this end, however, matrices with particular commutation properties have to be introduced. One can see by performing the operations that the factor

$$\left[c\rho_1\left(\sigma_x p_x + \sigma_y p_y + \sigma_z p_z\right) + \rho_3 m_0 c^2\right]\left[c\rho_1\left(\sigma_x p_x + \sigma_y p_y + \sigma_z p_z\right) + \rho_3 m_0 c^2\right]$$

is just the desired expression $m_0^2c^4 + c^2(p_x^2 + p_y^2 + p_z^2)$ (where for simplicity we have left out the vector potential **A**) if the quantities ρ and σ are given by the matrices

$$\sigma_x = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

$$\rho_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \rho_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

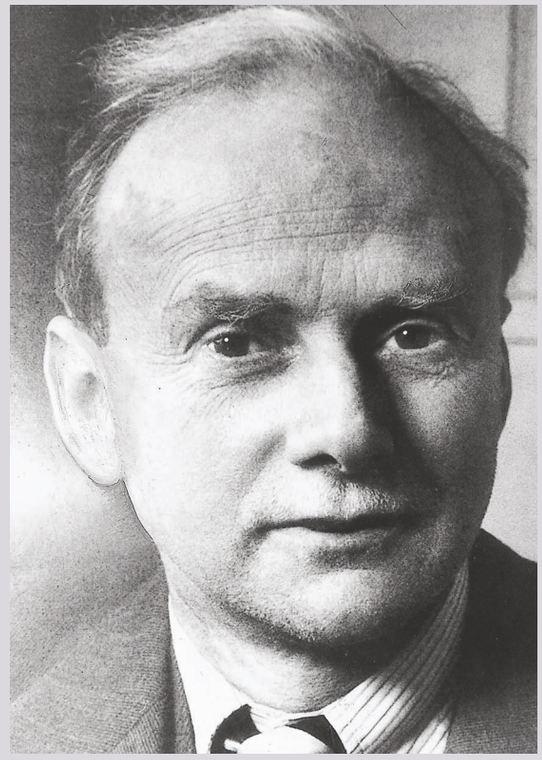
Figure 5.71 *continued*

spin) must be antisymmetric. In his 1940 paper "The Connection between Spin and Statistics," Pauli gives a derivation of his principle based on the relativistic invariance requirement.

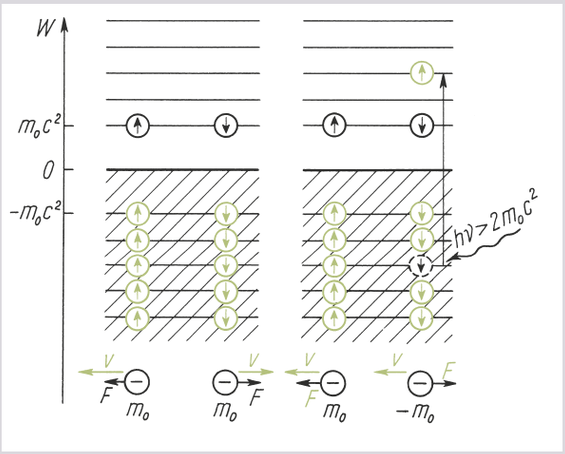
With his matrix-mechanical treatment of the states of the hydrogen atom, Pauli made a crucial contribution to the triumph of matrix mechanics.

In 1927, using spin matrices and two-component wave functions, Pauli was able to incorporate spin into the Schrödinger equation. This method remained useful for solving problems to the first approximation even after the more general Dirac equation had been established. In the area of nuclear physics, Pauli postulated the existence of a particle in 1930, which was later called the neutrino by Fermi, to explain beta decay, although Pauli still thought that the neutrino was a part of the nucleus (the correct theory of beta decay was finally worked out by Fermi in 1934).

Pauli was considered one of the most important figures in physics, contributing decisively to the advance of knowledge not only with his concrete results, but also through his work at international conferences. Together with Einstein, Bohr, and Heisenberg, he is among the leading personalities of the first half of the twentieth century.



▲ **Figure 5.72** Paul Adrien Maurice Dirac (1902–1984): Studied at Bristol, where he was born, then at Cambridge. Planned on becoming an electrical engineer, but his dissertation was devoted to problems in theoretical physics. From 1932, Lucasian chair at Cambridge, which had been held by Newton; 1926: connection between the symmetry properties of the wave function and statistics; 1926–1927: Dirac formalism of quantum mechanics; 1927: field quantization, second quantization; 1928: relativistic wave equation; 1948; magnetic monopole.



▲ **Figure 5.73** The Dirac equation has solutions for free electrons with kinetic energies not only greater than m_0c^2 but also less than $-m_0c^2$. The latter states can be interpreted as electrons with negative mass. An electron with negative mass would interact with one of positive mass in a special way: By Coulomb's law, the electrons would repel each other, but the acceleration of the particle with a negative mass is opposite to the direction of the force, so that the two particles would end up "playing tag." Such phenomena are not found in nature, and Dirac was forced to assume that all states with a negative energy are occupied and that is why no effects are seen. But, if with the help of a photon with an energy $h\nu > 2m_0c^2$, we bring an electron to the positive energy level, we get a "normally behaving" electron with rest mass m_0 and charge $-e$. The remaining empty place should now become detectable. The missing particle in the sea of particles of negative charge and negative mass is expressed as a particle of positive charge and positive mass. Because the proton has a positive charge and at the time no other elementary particles were known, it is understandable that Dirac interpreted the "hole" as a proton. Only with the experimental discovery of the positron could the hole be identified with the electron's antiparticle. The pair-creation process can now be readily interpreted: Under the influence of a gamma ray, an electron is raised to a positive energy level, so that a *previously unobserved* electron appears and the empty place can be observed as a positron. In the annihilation process, in turn, an electron jumps from a positive energy level to an empty place with a negative energy, whereupon a gamma quantum is emitted, and the electron and the positron vanish, or more accurately, become unobservable.

The Hamiltonian operator can then be written as

$$\mathbf{H} = c\rho_1\left(\sigma_x p_x + \sigma_y p_y + \sigma_z p_z\right) + \rho_3 m_0 c^2 + q_e U,$$

and the relativistic equation of the electron, the Dirac equation, assumes the form

$$\left[\frac{hc}{2\pi i}\rho_1\left(\sigma_x\frac{\partial}{\partial x} + \sigma_y\frac{\partial}{\partial y} + \sigma_z\frac{\partial}{\partial z}\right) + \rho_3 m_0 c^2 + q_e U\right]\Psi = -\frac{h}{2\pi i}\frac{\partial\Psi}{\partial t}.$$

The state function Ψ appearing in this equation incorporates the four space-time functions.

The Dirac equation effortlessly yields the spin and the magnetic moment of the electron. By 1928, Gordon and Darwin had derived the Sommerfeld fine structure constant from it, and Klein and Nishina, also in 1928, could apply it to the scattering of light by free electrons. This work helped the Dirac equation gain general acceptance. From this equation, one can also infer the existence of a particle with positive charge. In 1929, Dirac still thought that this particle would be the proton, but in 1930, Oppenheimer, and later Dirac and Tamm as well, noticed that the asymmetry between electron and proton leads to new difficulties. Finally, in 1931, Dirac suggested the existence of antielectrons (positrons), which were then found by Anderson the following year. Dirac even conjectured that antiprotons might also exist (Figure 5.73).

With this axiomatic formulation, quantum mechanics achieved a certain closure, at least in the description of systems with a finite number of particles. But already in the first years of quantum mechanics, the question had already arisen as to how the theory could be generalized to systems with an infinite number of degrees of freedom or, more simply stated, the idea of how to establish quantization rules for *fields*. When it looks for operators to represent physical quantities, quantum mechanics relies on analogies with *classical* mechanics in terms of generalized position and momentum coordinates. An electromagnetic field is as much a physical reality as a system made up of point masses. However, this physical reality is described not by the fundamental equations of classical mechanics but by the Maxwell equations. How, then, can an electromagnetic field be quantized within the framework of quantum mechanics? The answer is provided by *quantum electrodynamics* (Figures 5.74, 5.75).

The path to field quantization is of great interest in itself, and it provides a deeper insight into the fundamental relationships of classical electrodynamics. We proceed according to a simple schema: We already know the quantization rules for the phenomena described by classical mechanics, so we need to clothe Maxwell's equations in the formal garb of classical mechanics. This is indeed possible: One can introduce quantities for the electromagnetic field that correspond to the conjugate position and momentum coordinates satisfying equations similar to those of classical mechanics. The quantization rules for these quantities are then practically the same as those for mechanical systems.

When this schema is implemented in detail, we get back the same results we have already seen: The energy of the electromagnetic field is composed of the $h\nu$ energies of quanta, that is, of photons, and each such photon has a momentum h/λ as demanded by de Broglie's theory. But there is also a host of new and surprising results. For example, it turns out that the operators associated with electric and magnetic fields are not interchangeable, and therefore their fields cannot be determined simultaneously with arbitrary precision. This also means that a state with $\mathbf{E} = 0$ and $\mathbf{B} = 0$ cannot be realized, and therefore even a radiation-free space, a photon vacuum, possesses an energy: the zero-point energy.

In following years, quantum electrodynamics attained great importance as a successful example of field quantization. In the formalism of quantum electrodynamics, it was possible and necessary to intro-

duce at the outset operators for interaction processes where particles are created (emitted) or annihilated (absorbed); these are the creation and annihilation operators. A general theory of elementary particles requires these operators because the interactions of elementary particles typically involve the transformation of particles, that is, the disappearance of one and the appearance of another. In analogy to quantum electrodynamics, the theory of the electron–positron field was developed. The quanta of this field are electrons and positrons, just as photons are the quanta of an electromagnetic field.

These quanta can be obtained naturally only with the help of the quantization method. This quantization is called a second quantization. The Dirac equation here assumes the role of the Maxwell equation, where the state function in the Dirac equation is to be similarly quantized. The Dirac equation had already been obtained by a quantization process, so the term "second quantization" makes sense. The investigation of the coupling between the electromagnetic field and the electron–positron field led to an extension of the concept of interaction: Photons mediate the electromagnetic interaction.

The first phase of the development of quantum electrodynamics is marked by the work of Dirac, Jordan, and Pauli (1927), Heisenberg and Pauli (1929), Fock (2932), and Pauli and Weisskopf (1934). The second phase begins at the start of the 1940s with a 1943 work of Sin-Itiro Tomonaga (1906–1979). Due to his isolation caused by the war, his discoveries had to be derived again in the West. The impetus for this development came from experiments. By the 1930s it was already suspected that the $2^2S_{1/2}$ and $2^2P_{1/2}$ energy levels of the hydrogen atom were not exactly the same, although according to Dirac's theory they should be. In 1947, Willis Lamb and Robert Retherford proved that there is a difference between the two levels, but it is so small that the associated beat frequency lies in the radio frequency range. Microwave spectroscopy was now able to extend optical spectroscopy in the investigation of atomic spectra. We need to mention yet another surprising measurement: According to Dirac's theory, the magnetic moment of the electron should be equal to a Bohr magneton, but in 1941, Polykarp Kusch found a value that deviated from this, although only slightly (about 0.1%). The theoretical explanation for all anomalies was given independently by Julian Schwinger (1918–1994), Richard Phillips Feynman (1918–1988), and Freeman Dyson (1923–) and also by Tomonaga by a different method using the further developed quantum electrodynamics of 1947/48.

5.3.12 *The Causality Problem*

Every physicist who writes the phrase "causality problem" or reader who encounters it in a physics book should bear Schrödinger's warning in mind:

> Whether what I am writing about is truly the causality problem in the philosophical sense is something that I do not wish to decide in advance simply by having used the term.

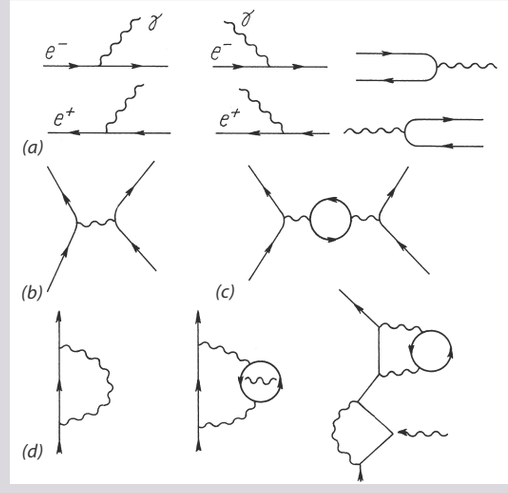
The debate between the physicists and the philosophers, as well as the debates among philosophers and physicists separately, have often led nowhere because they were not talking about the same thing (Quotations 5.32, 5.33, 5.34, 5.35). The goal of the physicists in the debate is quite clear. They want to move the problem to their own turf—in other words, to restate it so that the verification methods already acceptable to them can be used. At the same time, the philosophers, seeing the causality principle as something much more general, feel that the focus of the physicists is too narrow.

We would first like to try to summarize the formulations of the causality principle as given by certain philosophers—*not philosophy itself*—so that we can then consider the reservations brought up by the philosophers and especially by the physicists regarding its validity. Finally, we shall attempt to give a formulation that necessarily limits the scope of the problem but whose validity can be established with physical methods.

Among philosophers, it is quite generally believed that causality, or the law of cause and effect, was recognized by mankind through human activity and through the consequential effects of that activity. According to a very early formulation given by Anaximander, a cause results in an effect just as guilt results in a punishment. So it is by no means a coincidence that the Greek word for "cause" (αἰτία) also has the meaning of "guilt." This anthropomorphic origin of the principle was to reappear again and again. But already in classical Greek philosophy it is seen in



▲ **Figure 5.74** Richard Philips Feynman (1918–1988). 1950: Professor at the California Institute of Technology. Many physicists associate his name with Feynman diagrams, which serve as a pictorial representation of interaction processes. Moreover, Feynman became widely known for his textbooks *Lectures in Physics* (1963–1965), which have been translated into many languages and are distinguished by an original pedagogical approach.



▲ **Figure 5.75** Pictorial representation of the fundamental phenomena of quantum electrodynamics: the Feynman diagrams.

(a) Elementary actions: The electron emits a photon, absorbs a photon; the positron and electron generate an electron and a positron (pair formation). These elementary actions are, in and of themselves, impossible because

continued on next page

Figure 5.75 *continued*

they cannot simultaneously satisfy the laws of conservation of momentum and conservation of energy (see also Figure 5.138).

(b) A real process: The interaction between two electrons is mediated by photons; the photon emitted from one electron is absorbed by the other.

The question arises, how can the force resulting from emission and absorption be modeled in classical mechanics? Consider two boats close to each other with two persons, one in each boat, tossing balls back and forth: we get an impulse-like repulsive force. With the people tossing boomerangs instead, we could even model an attractive force. Such visualizations can be pedagogically useful, but one shouldn't take them too seriously.

(c) The Heisenberg relation in the form $\Delta E \times \Delta t \approx h$ makes it possible for phenomena with large energies to exist for a short time. Here the mediator photon creates an electron–positron pair that again unites into a photon. In calculating the interaction, one must take into consideration the effects of these "virtual" particles. Fortunately, because the coupling factor ($e^2 / \hbar c = 1/137$) is small, the influence of more complex processes becomes less and less.

(d) Reaction of the electron on itself: This was a problem that classical electrodynamics could not solve. What holds the electron together given that the negative charge exerts a repulsive force on itself? If we looked at the electron as a point-like particle, a difficulty would arise: the self-energy of the electron would be infinite.

In quantum electrodynamics, the electron emits (virtual) photons and absorbs them again. The photons can evoke very complicated processes such as, for example, the illustrated pair formation and annihilation, even with an exchange of a photon. Every such loop—with its uncountably many combinations—makes an infinitely large contribution to the self-energy; and because there is an infinite number of such initial loops, we get an infinite sequence of infinite quantities. So the problem of the self-energy of the electron seemed unsolvable in quantum electrodynamics as well. The brilliant insight independently made by Tomonaga, Schwinger, and Feynman was to recognize that this many-times infinite quantity can be reduced to a finite one by the subtraction of a single infinity. Their method is called renormalization. One can picture it by imagining that the zero point on the energy scale is shifted to infinity, so the small differences appear as well-defined finite values.

The third diagram shows how the anomalous magnetic moment of the electron arises as the result of many complicated processes (the arrow coming in from the right represents the photon that mediates the effect of the external magnetic field). If one takes more and more processes into consideration, then one obtains, instead of 1 magneton (which would follow from the Dirac equation), the theoretical value 1.00115965246. The measured value is 1.00115965221. Thus, quantum electrodynamics became a branch of physics in which the theory could be verified experimentally with the greatest precision.

The success of quantum electrodynamics follows from its renormalizability, and this, in turn, follows from the fact that the electromagnetic field is gauge-invariant. In the simplest case, gauge invariance means that the values of the physically relevant quantities remain unchanged when the potential is changed by an arbitrary amount (see Figures 5.167, 5.168, 5.169).

its modern form: everything that happens has some cause, after which it necessarily follows as an effect. "Like causes at all times and all places have like effects." The prevailing sentiment, consciously or tacitly, was that this principle belongs among the fundamental assumptions of all scientific research.

The representatives of the two extreme points of view on causality were Hume and Kant. Although Hume did not go so far as to deny the existence of general relationships, he called into question their logical necessity: Between things there exist no inner connections; instead, we have no other justification for concluding the existence of one thing from the existence of another other than the habit that influences our imagination.

As we have already seen, Kant, in contrast, conceived of causality as a category—an organizational principle of the mind—whose validity is ensured a priori, that is, independent of our experience. The basic thesis of the causal relations between events is therefore valid prior to any experience because it is precisely the causality principle that is the basis for the possibility of experience.

Many physicists, among them Helmholtz, espoused the role of causality in scientific work in the sense of Kant and believed in the validity of the causality principle (Quotation 5.36). Others, however, sensed anew a tinge of human influence in the word "cause," which they could not reconcile with the physical laws of the nineteenth century that expressed relationships in a functional way. At the very least, they doubted whether a division of phenomena into causes and effects could be helpful for understanding natural phenomena or in their research. In this regard we quote Kirchhoff [1876]:

> One generally defines mechanics as the science of *forces*, and forces as the *causes* that result in motion or that *strive* to bring forth motion. Certainly this definition has been most useful in the development of mechanics and also in learning the science when it is explicated through examples of forces that are taken from the experience of our daily lives. But there is an associated lack of clarity from which the notions of cause and striving cannot free themselves. This lack of clarity is revealed, for example, in the differing views over whether the law of inertia and the parallelogram law of forces are to be seen as the result of experience, as axioms, or as theorems that can be logically proved and that thus have to be proved. Given the clarity that the laws of mechanics otherwise allow, it seems to me desirable to remove such murky points from it, even if such is possible only by limiting its scope. For this reason, I represent as the task of mechanics to *describe* the motions occurring in nature, *and indeed to do so completely and in the simplest possible manner*. I wish thereby to say that what should be at stake is only to specify *what* the phenomena are that occur, not to determine their *causes*.

The same attitude can be found in Mach, even more clearly:

> In speaking of cause and effect we arbitrarily give relief to those elements to whose connection we have to attend in the reproduction of a fact in the respect to which it is important to us. There is no cause nor effect in nature; nature has but an individual existence; nature simply *is*. Recurrences of like causes in which *A* is always connected with *B*, that is, like results under like circumstances, that is again, the essence of the connection of cause and effect, exist but in the abstraction which we perform for the purpose of mentally reproducing the facts. Let a fact become familiar, and we no longer require this putting into relief of its connecting marks, our attention is no longer attracted

to the new and surprising, and we cease to speak of cause and effect. Heat is said to be the cause of the tension of steam; but when the phenomenon becomes familiar we think of the steam at once with the tension proper to its temperature. Acid is said to be the cause of the reddening of tincture of litmus; but later we think of the reddening as a property of the acid. ...

Cause and effect, therefore, are things of thought, having an economical office. [1893, pp. 580–581]

At another point, Mach expresses the hope that the notions of cause and effect, which reek of fetishism and are obscure in their formalism, will be banished from the vocabulary of the natural sciences in the future.

It would be of interest to investigate the extent to which Mach's prognosis has been borne out. We could, for example, count how often the words "cause" and "effect" appear in a current textbook on theoretical physics, or better yet, in the 55 volumes of the *Handbook of Physics*.

Let us now examine to what a narrow extent the causality principle can serve the physicist as a worthwhile methodological tool.

1. In the introduction to this book, we stated that the task of physics is to interpret natural phenomena. We also proposed a schema according to which this task could be accomplished: characterization of concrete situations and provision of laws (*explanans*), with the help of which the quantitative characteristics of the phenomenon under investigation (at least in principle) can be determined.

   With this formulation, the causality principle is identical to the requirement that *the physical phenomena be amenable to interpretation in the sense given above*. This, as a methodological schema, is generally accepted by physicists.
2. We can state the principle of cause and effect as the most general principle of physics that sets the boundaries within which physical events can affect one another. Thus, it excludes the effect of an event on what already has happened. (The future cannot act on the past; a cause event precedes the effect event in time.) Moreover, the principle excludes the interaction of events to which the words *earlier* and *later* cannot be applied. (These notions are defined precisely in the theory of relativity.)

   In this formulation, the principle can serve physicists as a useful and concrete tool. Thus, it prescribes the boundary conditions that ensure the unique solution of a problem and gives a means of excluding certain solutions (e.g., advanced potential) as physically impossible. One can even obtain concrete statements about the properties of a particular physical system. Thus, for example, in the theory of electrical circuits, if we set the obvious requirement that the reaction of a network to some external excitation can occur only after the excitation, we can immediately make some conclusions about the circuit elements and exclude some models. In general field theory as well, three fundamental requirements are used as a basis in the search for general laws: relativistic invariance, the validity of the laws of quantum mechanics, and the causality principle.

   Starting with the causality principle, one can require that operators that are associated with space-time points among which an interaction is impossible be interchangeable. For instance, the CPT invariance can be supported with such an argument (Section 5.5).

Quotation 5.32

The use of the concept of causality for describing the law of cause and effect is of relatively recent origin. In previous philosophies the word *causa* had a very much more general significance than it has today. Thus scholasticism, following Aristotle, spoke of four kinds of "causes": the *causa formalis* which might be considered as the form or the spiritual essence of a thing, the *causa materialis* which referred to the matter of which the thing consisted, the *causa finalis* or the purpose for which the thing was created, and finally, the *causa efficiens*. Only the *causa efficiens* corresponds to what is meant by the word "cause" today.

—Werner Heisenberg, *The Physicist's Concept of Nature* [p. 33]

Quotation 5.33

This definition of determinism, relying on the exact predictability of events, is perhaps the only one that a physicist could accept, since only this one can be subject to actual verification. However, one should not conceal the fact that this physical definition of determinism brings with it some difficulties. Above all, that in nature there reigns a universal interaction and the motion of the smallest atom can be influenced by the farthest star, the exact prediction of an arbitrary future event would thus in principle require the complete knowledge of the current state of the entire universe, which of course is unrealizable. Clearly, however, this is only a theoretical objection, since in general, the advance calculation of a future process can be realized in practice using a finite amount of information regarding the current state. Much more difficult is the objection that follows from the necessarily approximative character of our observations and measurements.

The data obtained through observation and measurement are always subject to experimental errors, and so the foresight obtained with them is also possible only with some degree of uncertainty, so that the verification of the strict predictability of events and hence determinism in the sense described above, always possess only an approximate validity. Nevertheless, this new objection also seems irrefutable: The precision of our observations and measurements can always be improved, whether through the refinement of methods or through the perfection of the experimental procedure. To the same degree as the precision of our observations increases, we obtain an ever stricter predictability, and thus we can consider determinism confirmed as the result of a crossing a boundary. In classical physics, nothing

continued on next page

Quotation 5.33, ***continued***

seems to speak against the idea that our prediction of future processes could become ever more precise as a result of improved observational processes. This meant that our trust in physical determinism up through the deepening of our knowledge of quantum processes was not unjustified.

—De Broglie, *Continu at discontinu en Physique moderne* [pp. 59–60]

In conclusion we may therefore say: The law of causality is neither true nor false. It is rather a heuristic principle, a signpost—and in my opinion, our most valuable signpost—to help us find our bearings in a bewildering maze of occurrences, and to show us the direction in which scientific research must advance in order to achieve fertile results. The law of causality, which immediately impresses the awakening soul of the child and plants the untiring question "Why?" into his mouth, remains a lifelong companion of the scientist and confronts him incessantly with new problems. For science is not a contemplative repose amidst knowledge already gained, but is indefatigable work and an ever progressive development.

—Planck, *Scientific Autobiography...* [pp. 149–150]

Quotation 5.34

We shall now say a few words about the relationship of the notions of determinism and causality. The relation between these two concepts seems not to be very precise; moreover, it depends in great measure on the definition that we give to the one and to the other. Thus some authors, for example, view the concept of causality as more restricted than that of determinism and maintain that determinism remains valid in quantum mechanics, while causality breaks down. In contrast to this, it seems to us much more natural to say that in quantum mechanics, there is no determinism in the sense presented above; however, causality exists in a somewhat wider sense, about which we shall speak now.

Let us consider a phenomenon *A* following which certain phenomena B_1, B_2, B_3, ... always occur. If now the phenomena B_1, B_2, B_3, ... do not occur when *A* does not occur, we may say—and indeed based on the expanded definition of causality—that *A* represents the cause for phenomena B_1, B_2, B_3, ...; this defintion will be in complete agreement with the old maxim *Sublata causa, tollitur effectus*. According to this definition, there exists a causal relationship between the phenomenon *A* and the phenomena B_1, B_2, B_3, ..., but we cannot speak of determinism in the sense given above if we have no possibility of predicting *which* of the phenomena B_1, B_2, B_3, ...

continued on next page

3. Planck saw a parallel between the causality principle and the possibility of predicting the future:

   An event is causally *determined if it can be predicted with certainty.*

   In classical physics, Laplace's demon is most often invoked to characterize deterministic laws (see the epigraph to Chapter 4).

These ideas are usually expressed as follows: The laws of classical physics can be described by differential equations that, given present conditions, unambiguously determine the future. However, Planck had already pointed out that the statement about predetermination of the future cannot be sustained down to the last implication, even in classical physics:

In not one single case is it possible to predict a physical event with certainty.

We find a more precise statement with de Broglie and others (Quotations 5.33, 5.34, 5.35).

In looking more closely at the problem, we meet additional difficulties. Both in the search for natural laws and in their application we have to specify identical causes—in the language of physics, identical boundary conditions—and then expect to observe identical effects states with identical characteristic quantities.

For this, however, we must assume that the history of the entire universe can be decomposed into histories that occur in parallel with and independently of one another. However, this assumption harbors within itself a host of difficulties. First, how are laws to be formulated for events that occur only once? An example of this is the evolution of the universe, for which we certainly cannot carry out a controlled experiment. In this case, we also cannot satisfy the requirement discussed in the introduction that a physicist make only such "intersubjective" assertions that can be experimentally checked by anyone. And now to the second difficulty: When we separate the partial phenomena, the subject of our investigation, the connection to the other parts of the world is represented by the boundary conditions as specified by us; here the possibility of the exclusion of the subject must be carefully ascertained. If we wish to prove determinacy in classical electrodynamics in this sense, then have to assume either that at the edges of the selected group of phenomena not only the present but also future effects are known, or else we must again include the entire universe in our investigations.

In this limited sense, physicists can give an answer to much disputed questions about causality–acausality or determinism–indeterminism. Heisenberg said that in the sentence, "If we know the initial state, then the future is unambiguously determined," it is not that the conclusion is false, but that the premise can never be fulfilled. More precisely, in quantum mechanics, the state function satisfies a differential equation that is not different in character from the equations for arbitrary physical quantities in classical physics. In other words, if we know the state function now, the fundamental equations of quantum mechanics make it possible to determine this state function uniquely at any later time. The difference from classical mechanics is that there the state characteristics were identical with the quantities that resulted from the measurements. The quantum-mechanical state function, in contrast, describes only the expected values from the measurements, and in this sense can give only a probabilistic answer.

This takes us directly to the question of hidden parameters. One obvious idea is that the statistical results of quantum mechanics should be viewed in the same way as the statistical results of the classical kinetic theory of gases, that is, as average values of phenomena that satisfy strictly deterministic laws.

Let us assume that the state function $\Psi(\mathrm{r},t)$ depends as well on an additional quantity, the *hidden parameter* λ, and we can detect only the average

$$\Psi(\mathbf{r},t) = \overline{\Psi(\mathbf{r},t,\lambda)}^2$$

over the values of the hidden parameter.

John von Neumann (Figure 5.76) proved under very general assumptions that the problem cannot be solved with the introduction of such a hidden parameter because doing so would lead to contradictions with experimentally proved laws of quantum mechanics. The postulates used in proof are these:

1. Let us assume that
$$\left(\Psi, \mathbf{M}\Psi\right) = \int_{\lambda} M(\lambda,\Psi)\rho(\lambda)\mathrm{d}\lambda.$$
Here, **M** is an operator that is associated with an arbitrary physical quantity, and so the left-hand side of the equation denotes the expected value of a measurement of this physical quantity when the system is in the state given by the function Ψ. On the right-hand side of the equation, $M(\lambda, \Psi)$ is the quantity's now uniquely determined value, given Ψ and λ, and $\rho(\lambda)$ is the distribution function that is a measure of the frequency of the measurement results that correspond to different values of λ.

2. Every value $M(\lambda_i, \Psi)$ measured for an arbitrarily chosen parameter λ_i is one of the eigenvalues.

During the following years, von Neumann's theory that the fundamental equations of quantum mechanics cannot be recast into classically deterministic form even with the assumption of hidden parameters was proven under more and more general conditions. The counter arguments and counterexamples ran into such difficulties that most physicists find von Neumann's formulation simpler and more reassuring and accept that phenomena in the microworld have a fundamentally probabilistic character (Quotations 5.37, 5.38).

5.3.13 *John von Neumann on Causality and Hidden Parameters*

Let us forget the whole of quantum mechanics but retain the following. Suppose a system **S** is given, which is characterized for the experimenter by the enumeration of all the effectively measurable quantities in it and their functional relations with one another. With each quantity we include the directions as to how it is to be measured—and how its value is to be read or calculated from the indicator positions on the measuring instruments. ...

... The investigation of the physical quantities related to a single object *S* is not the only thing which can be done—especially if doubts exist relative to the simultaneous measurability of several quantities. In such cases it is also possible to observe great statistical ensembles which consist of many systems $\mathbf{S}_1, ..., \mathbf{S}_N$ (i.e., *N* models of **S**, *N* large). In such an ensemble $[\mathbf{S}_1, ..., \mathbf{S}_N]$ we do not measure the "value" of a quantity *R* but its distribution of values: i.e., for each interval $a' < a$ ffi a'' (a', a given, a' ffi a'') the number of those among the $\mathbf{S}_1, ..., \mathbf{S}_N$ for which the value of *R* lies in the interval—dividing this number by *N* we obtain the probability function $w(a', a'') = w(a'') - w(a')$ ....

For such ensembles, it is not surprising that a physical quantity *R* does not have a sharp value, i.e., that its distribution function does not consist of a single value a_o, but that several values or intervals of values are possible, and that a positive dispersion exists. However, two different reasons for this behavior are a priori conceivable:

I. The individual systems $\mathbf{S}_1, ..., \mathbf{S}_N$ of our ensemble can be in different states, so that the ensemble $[\mathbf{S}_1, ..., \mathbf{S}_N]$ is defined by their relative frequencies. The fact that we do not obtain sharp values for the physical quantities in this case is caused by our lack of information: we do not know in which state we are measuring, and therefore we cannot predict the results.

II. All individual systems $\mathbf{S}_1, ..., \mathbf{S}_N$ are in the same state, but the laws of nature are not causal. Then the cause of the dispersions is not our lack of information, but is nature itself, which has disregarded the "principle of sufficient cause."

Quotation 5.34, ***continued***

appears when *A* occurs. Determinism is present only when there is only a single phenomenon B_1. In other words, we are convinced that we are looking at such causality without determinism in quantum mechanics, where precise predictions are possible only in exceptional cases, namely in those cases that the theoreticians of the new mechanics call "pure cases."

—De Broglie, *Continu at discontinu en Physique moderne* [pp. 64–65]

Quotation 5.35

Luck and *chance*, too, are said to be causes, and many things re said to exist and to come to be through luck or *chance*. Accordingly, we must inquire in what manner luck and *chance* are causes among those given, whether luck and *chance* are the same or distinct, and, as a whole, what luck is and what *chance* is.

Some thinkers even raise the problem of whether luck and *chance* exist or not; for they say that nothing comes by luck, but that in every case in which we say that a thing comes to be by *chance* or luck there is a definite cause ... and this is not luck; for it would indeed appear strange if luck were something, and one might even raise the question as to why not even one of the ancient wise men, in speaking of the causes of generation and of destruction, said anything definite about luck.

—Aristotle, *Physics,* II.4 [1969, p. 32]

Quotation 5.36

Were therefore the law of causation a law of experience, its inductive proof would seem to be in very bad shape. The best we could say is that it was not any more valid than rules of meteorology ...

... the law of causation bears on its face the character of a purely logical law, chiefly because the conclusions derived from it do not concern actual experience, but its interpretation. Hence it cannot be refuted by any possible experience. For if we founder anywhere in applying the law of causation, we do not conclude that it is false, but simply that we do not yet completely understand the complex of causes mutually interacting in the given phenomenon.

—Helmholtz, *Treatise on Physiological Optics* [Vol. III, pp. 32–33]

Quotation 5.37

Indeterminism, which up to 1927 had been equated with obscurantism, became the ruling fashion; and some great scientists, such as Max Planck, Erwin Schrödinger, and Albert Einstein, who hesitated to abandon determinism, were considered old fogies, although they had been in the forefront of the development of quantum theory. I myself once heard a brilliant young physicist describe Einstein, who was then still alive and hard at work, as 'antediluvian'. The deluge that was supposed to have swept Einstein away was the new quantum theory, which had risen during the years from 1925 to 1927, and to whose advent at most seven people had made contributions comparable to those of Einstein.

—Karl Popper, *Objective Knowledge* [p. 214]

Quotation 5.38

[T]he statistical interpretation of quantum mechanics ... is not a simple matter, and demands not only knowledge of a complicated mathematical formalism, but a certain philosophical attitude: the willingness to sacrifice traditional concepts and accept new ones, like Bohr's principle of complementarity. I am far from saying that the present interpretation is perfect and final. I welcome Schrodinger's attack against the complacency of many physicists who are accepting the current interpretation because it works, without worrying about the soundness of its foundations. Yet I do not think that Schrodinger has made a positive contribution to the philosophical problems. ... I choose as my witness W. Pauli who is generally acknowledged to be the most critical, logically and mathematically exacting amongst the scholars who have contributed to quantum mechanics. I translate a few lines from a letter (in German) which I have recently received:

Against all retrograde efforts (Schrödinger, Bohm, etc., and in a certain sense, also Einstein) I am certain that the statistical character of the -function, and thus of the laws of nature—which you have, right from the beginning, strongly stressed in opposition to Schrodinger—will determine the style of the laws for at least some centuries. It is possible that later, for example in connection with the processes of life, something entirely new may be found, but to dream of a way back, back to the classical style of Newton-Maxwell (and it is nothing but dreams which those gentlemen indulge in), that seems to me hopeless, off the way, bad taste. And we could add "it is not even a lovely dream."

—Born, "Statistical Interpretation of Quantum Mechanics" [pp. 149–150]

Case I is generally well-known, while Case II is important and new. ...

One might object against [Case] II that nature cannot violate the "principle of sufficient cause," i.e., causality, at all, because this is merely a definition of identity. That is, the theorem that two identical objects $\mathbf{S}_1, \mathbf{S}_2$—i.e., two replicas of the system **S** which are in the same state—will remain identical in all conceivable interactions is true because it is tautological. For if $\mathbf{S}_1, \mathbf{S}_2$ could react differently to the same intervention in their interaction (i.e., if they gave different values in the measurement of a quantity *R*), then we would not have called them identical. Therefore, in an ensemble $[\mathbf{S}_1, ..., \mathbf{S}_N]$ which has dispersion relative to a quantity *R*, the individual systems $\mathbf{S}_1, ..., \mathbf{S}_N$ cannot (by definition) all be in the same state. (The application to quantum mechanics would be: Since one will obtain different values in the measurement of the same quantity *R* in several systems, which all are in the state with the wave function φ—if φ is not an eigenfunction of the operator R of *R*—therefore these systems are not equal to one another—i.e., the description by the wave function is not complete. Therefore other variables must exist, the "hidden parameters" ... We shall soon see what difficulties this presents.) In a large statistical ensemble therefore, as long as any quantity *R* possesses a dispersion in it, the possibility must exist of resolving it into several differently constituted parts (according to the various states of its elements). This is all the more plausible, since ordinarily a simple method of such a resolution seems to exist: namely, we can resolve it according to the various values which *R* has in the ensemble. After a subdivision or resolution relative to all quantities *R, S, T,...* which are present has been carried out, a truly homogeneous ensemble would then be obtained. At the end of the process, these quantities would have no further dispersion in any of the sub-ensembles.

First of all, the statements contained in the last sentence are incorrect because we did not consider the fact that measurement changes the measured system. ...

Therefore we have no method which would make it always possible to resolve further the dispersing ensembles (without a change of their elements) or to penetrate to those homogeneous ensembles which no longer have dispersion. The last ones are the ensembles we are accustomed to consider to be composed of individual particles, all identical, and all determined causally. Nevertheless, we could attempt to maintain the fiction that each dispersing ensemble can be divided into two (or more) parts, different from each other and from it, without a change in its elements. That is, the division would be such that the superposition of two resolved ensembles would again produce the original ensemble. As we see, the attempt to interpret causality as an equality definition led to a question of fact which can and must be answered, and which might conceivably be answered negatively. This is the question: is it really possible to represent each ensemble $[\mathbf{S}_1, ..., \mathbf{S}_N]$, in which there is a quantity *R* with dispersion, by the superposition of two (or more) ensembles different from one another and from it?

[Here follows the precise mathematical formulation and the solution of the problem raised with the conclusion that there are no ensembles which are free from dispersion and no further assumption of hidden parameters would help.]

To summarize, the position of causality in modern physics can therefore be characterized as follows: In the macroscopic case there is no experiment which supports it, and none can be devised because the apparent causal order of the world in the large (i.e., for objects visible to the naked eye) has certainly no other cause than the "law of large numbers" and it is com-

pletely independent of whether the natural laws governing the elementary processes are causal or not. That macroscopically identical objects exhibit identical behavior has little to do with causality: They are in fact not equal at all, since the coordinates which determine the states of their atoms almost never coincide exactly, and the macroscopic method of observation averages over these coordinates (here they are the "hidden parameters"). The number of these coordinates is, however, great (for 1 gram of matter, about 10^{25}) and therefore the above mentioned averaging process entails an extensive diminution of all dispersions, according to the well-known laws of the calculus of probability. (Naturally this is true only in the general case; in suitable special cases—such as Brownian motion, unstable states, among others—this apparent microscopic causality fails.) The question of causality could be put to a true test only in the atom, in the elementary processes themselves, and here everything in the present state of our knowledge militates against it. The only formal theory existing at the present time which orders and summarizes our experiences in this area in a half-way satisfactory manner, i.e., quantum mechanics, is in compelling logical contradiction with causality. Of course it would be an exaggeration to maintain that causality has thereby been done away with: quantum mechanics has, in its present form, several serious lacunae, and it may even be that it is false, although this latter possibility is highly unlikely, in the face of its startling capacity in the qualitative explanation of general problems, and in the quantitative calculation of special ones. In spite of the fact that quantum mechanics agrees well with experiment, and that it has opened up for us a qualitatively new side of the world, one can never say of the theory that it has been proved by experience, but only that it is the best summarization of experience. However, mindful of such precautions, we may still say that there is at present no occasion and no reason to speak of causality in nature—because no experiment indicates its presence, since the macroscopic are unsuitable in principle, and the only known theory which is compatible with our experiences relative to elementary processes, quantum mechanics, contradicts it.

To be sure, we are dealing with an age old way of thinking of all mankind, but not with a logical necessity (this follows from the fact that it was at all possible to build a statistical theory), and anyone who enters upon the subject without preconceived notions has no reason to adhere to it. Under such circumstances, is it sensible to sacrifice a reasonable physical theory for its sake?

—John Von Neumann, *Mathematical Foundations of Quantum Mechanics* [pp. 297–306, 323–324, 326–328]

5.3.14 *Quantum Mechanics as a Tool and as a Philosophy*

In the previous sections, we have discussed in some detail the philosophical background and interpretation of quantum mechanics and problems with the theory that have not been completely cleared up even today. At the same time, quantum mechanics is also an everyday tool that the physicist trusts as completely as an engineer trusts classical mechanics or Maxwell's electrodynamics. Where does the physicist get this certainty from? Here we can benefit from an analogy.

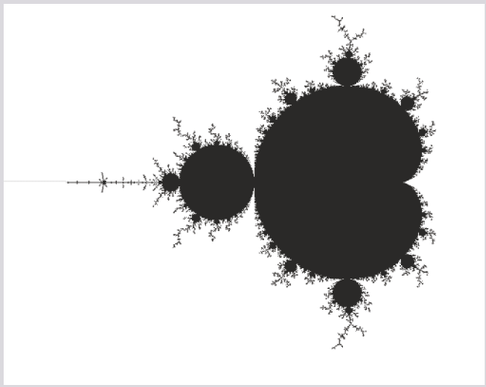
For radiologists, an x-ray machine is an everyday tool. To be able to fulfill their job, they must be familiar with the *use* of the apparatus. This means having knowledge of the steps that generally appear in the instruction manual that comes with the piece of equipment. Moreover, they must be able to evaluate the results (the screen image or x-ray film) diagnostically. In contrast, radiologists do not have to



▲ **Figure 5.76** John von Neumann (1903–1957): Born in Budapest. University studies in Berlin and Zurich, where he received a degree as a chemical engineer. In the same year, doctorate in Budapest with a dissertation on mathematical problems. 1926–1930: private lecturer in Berlin and Hamburg, from 1930 to his death, professor at the Institute for Advanced Study in Princeton. During the war, participated in the atomic bomb program, and after the war, in the development of the hydrogen bomb.

His work on quantum mechanics is collected in the book *Mathematical Foundations of Quantum Mechanics* (1932). His contributions to physics (more generally, to applied mathematics) were in a number of diverse areas: statistical physics, precise formulation and proof of the ergodic hypothesis, problems of aero- and hydrodynamics, theory of shock waves, meteorological problems, computer science.

The investigation of rotation groups connected to rings of operators (jointly with F. J. Murray) in the 1930s led him to the theory of spaces with continuous dimensionality. Such spaces, with fractional dimension, considered from another vantage point and with another set of physical objectives, gained currency in the 1960s (Figure 5.77).



▲ **Figure 5.77** One of the most popular fractals is the Mandelbrot set. Benoît Mandelbrot (1924–2010) investigated the trivial problem of determining the length of Britain's coastline. To be sure, the result depends very strongly on how one carries out the measurement, for example, the extent to which one takes into account every zig and zag along the shore. An apparently smooth stretch of coastline reveals on closer examination a further array of zigzags, so that in the final tally, the coastline, that is, the totality of points on the boundary between land and sea, is represented more as a strange surface than as a line. Its dimension falls between 1 and 2 (of course, there is a precise definition of dimension, which we do not require for this overview). Mandelbrot gave the name *fractal* to point sets characterized by a fractional dimension. The fractal pictured here is named in his honor. It is of interest for a number of reasons. First, it presents a compelling example of how a simple mathematical question can lead to an almost limitless treasure trove of knowledge. Second, it points to the relationship between fractal geometry and chaotic motion, and thereby this set of problems became an integral part of the problem of causality in physics.

Mandelbrot considers the discrete sequence of complex numbers determined by the simple recurrence formula

$$z_{n+1} = z_n^2 + c,$$

where c is a given constant and z_n is an element of the complex plane. If we choose the initial point $z_0 = 0$, then we obtain the sequence $z_1 = c$, $z_2 = c^2 + c$, $z_3 = (c^2 + c)^2 + c$, etc. These points may be construed as the discrete points of the path of the point z_0. Such paths will of course be different for different values of c. If one now asks for the location of the points (represented in the c-plane) that belong to such paths, those that go off to infinity and those that remain finite, we obtain the Mandelbrot set: The points shown in black belong to the values of c for which the path of the point ($z_0 = 0$) remains constant. For all other points, the path goes off to infinity. In the picture one can see the characteristic appearance of a fractal: The contours repeat themselves in smaller and smaller regions (self-similarity). And for the physicist, what is striking about this mathematical model is the sensitivity of the paths to the initial conditions.

understand the apparatus as an engineering creation or in its technical details, and need only a general familiarity with the underlying physical processes. They can still perform the routine work without faults.

However, when it comes to research and development of radiological equipment, the level of knowledge described above is far from adequate. Now the interactions between physical and biological processes must be considered, and one has to reach back to more fundamental phenomena of those sciences.

Finally, we could be discussing the social implications of radiological equipment: for example, the effectiveness and costs of mass screening, or the issues of radiation danger and protection of patients and technicians.

1. The routine task of a physicist with respect to quantum mechanics—or more accurately, the content of a college-level textbook or monograph on quantum mechanics—consists of the solution of concrete problems and in the evaluation of proposed solutions, that is, in the comparison of theoretically obtained results with the results of experimental measurements. To this end, quantum mechanics clearly has the necessary mathematical apparatus together with the associated instruction manual.

   The state function Ψ suffices to describe a microphysical system. For two functions Ψ_1 and Ψ_2, the superposition principle holds; that is, if Ψ_1 and Ψ_2 represent two possible states of the system, then $c_1\Psi_1 + c_2\Psi_2$ also represents a possible state.

   The mathematical conditions for the state functions are summarized by the requirement that they represent an element of Hilbert space.

   The temporal course of the state function Ψ is described by the time-dependent Schrödinger equation

$$\mathrm{H}\Psi = -\frac{h}{2\pi i}\frac{\partial \Psi}{\partial t}.$$

   If we repeatedly measure the physical quantity characterized by the operator Λ in a system in state Ψ, then each time we obtain as the measured value an eigenvalue Λ_v associated with the eigenfunction Ψ_v, and we obtain it with probability $|(\Psi_v, \Psi)|$.

   To find the operator associated with a given physical quantity, we are aided by the generalized correspondence principle.

2. Let us move into the second phase: We would like to look more deeply into the physics of processes. We have to forgo demonstrations with physical models at the outset, but with electrodynamics, we have gotten used to visualizing wavelike phenomena that disperse in space. We would like to imagine the Ψ function as something similar, especially because it satisfies a wave equation similar to that satisfied by an electric or magnetic field, and similarly satisfies that given some initial conditions, the Ψ function can be determined at an arbitrary later time. Still, we cannot model it as the fluctuation of some physical reality: Ψ does not have any physical meaning, and in regard to reality we can only elicit probabilistic statements with its help. In the same vein, we cannot give an illustrative answer to the question of what happens with the function Ψ during the measuring process: The measurement selects so to speak from the components of Ψ the eigenfunction Ψ_v that is associated with the quantity being measured. Is it possible to have an illustrative picture of transition $\Psi \rightarrow \Psi_v$?

   For this second set of problems, the Copenhagen interpretation provides an answer that the majority of physicists finds satisfactory. We will try to

summarize the concept by highlighting key passages from Heisenberg's work from Section 5.3.10.

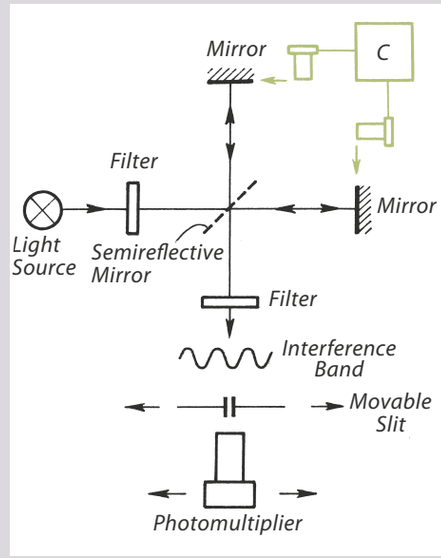
> If we want to describe what happens in an atomic event, we have to realize that the word "happens" can apply only to an observation, not to the state of affairs between two observations. It applies to the physical, not to the psychical act of observation, and we may say that the transition from the "possible" to the "actual" takes place as soon as the interaction of the object with the measuring device, and thereby with the rest of the world, has come into play; it is not connected with the act of registration of the result by the mind of the observer. The discontinuous change in the probability function, however, takes place with the act of registration, because it is the discontinuous change of our knowledge in the instant of registration that has its image in the discontinuous change of the probability function.

As an illustration of this, we discuss a variant of the famous "single photon" experiment that tellingly illustrates how quantum mechanics, concretely the Heisenberg uncertainty relation, resolves the wave–particle dualism of light. The experiment also illustrates the character of the wave function that cannot be given a direct physical interpretation and the related conceptual difficulties.

Let us consider a Michelson interferometer of a type similar to what we have already seen in the discussions of the theory of relativity: A ray of light from a light source enters a filter that can weaken the light's intensity; it falls on a half-silvered mirror that reflects part of the beam and lets the other half through. Then after reflection of the separated rays, they are made to interfere (Figure 5.78). From the distance between the interference lines, one can determine the wavelength of the light, or in the language of particles, the momentum of the photons. Thus, this arrangement can also be viewed as an apparatus to measure momentum.

Now let us replace the mirrors by photon counters. Using a coincidence circuit, the number of photons falling simultaneously on the two mirrors can be measured. With this modification, we are essentially measuring the positions of the photons (shown in color in Figure 5.78).

We now diminish the intensity of the light to such a point that we have ensured that at any time there is never more than a single photon in the interferometer. From a naive point of view, we may assume that in the first variant of our apparatus, that is, in measuring wavelength and momentum, this single photon is reflected in one or the other of the two mirrors: Thus, no interference would be expected. *Yet in such cases, one also observes interference lines*: A single photon somehow "senses" the presence of the two mirrors; one could again naively suppose that the photon is, so to speak, split in two at the mirror, and these "half-photons" are reflected and brought into interference. But if we move to the second variant of the experimental setup, we find that either one or the other counter will register the presence of the photon; no coincidences will occur. The Copenhagen interpretation holds these results to be self-evident: The interference experiment, as we already mentioned, is a measurement of the momentum of the photons. In such a situation, according to Heisenberg's uncertainty principle, we can say nothing at all about the position of the photon. Thus, the question of where the photon is located cannot be asked, and even less are we entitled to ask what its path is. In the



▲ **Figure 5.78** A variant of the "single photon" experiment. Such experiments testify to the correctness of the Copenhagen interpretation; but, at the same time, they also illustrate the difficulties.

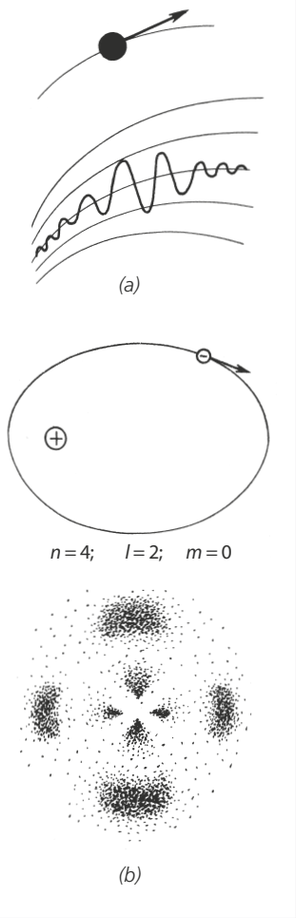
A topic of endless controversy, continuing to our day, is presented by the paradox published in 1935 by Einstein, Boris Podolsky, and Nathan Rosen (the EPR paradox). The essence of their thought experiment can be put as follows: Consider two systems that interact for a brief period—in the simplest case, we can assume the collision of two particles. After the interaction, we determine a physically relevant quantity of one system by using a measurement taken in the other system and applying a conservation law. For example, after the interaction of two electrons e_1 and e_2, if we measure the momentum of electron e_1 with the utmost accuracy, we lose precision regarding its location (which is for the moment irrelevant), but from the law of conservation of momentum, we can calculate the momentum of electron e_2. Then we measure the location of electron e_2 with utmost precision. Therefore, for electron e_2 we would *not* have the relationship $\Delta p \Delta q \approx h$. Bohr's answer was this: Momentum and location have no objective reality in and of themselves; they are realized only through the process of measurement. This fundamental quandary could be decided only experimentally, as was proposed by John Stuart Bell in 1964 (Bell inequalities). He began with a variant of the EPR paradox formulated by Bohm and expressed the relationship between measurements in the two systems if, following Einstein, one ascribes an objective reality to the physical parameters (momentum, location, spin). The relevant experiment, requiring extreme precise techniques, was carried out in 1982 by Alain Aspect. The results of the measurements speak unequivocally in favor of the Bohr conception.

Quotation 5.39

In conclusion, may I be permitted to express a wish? Suppose that in a few years from now these theories are subjected to new tests and come out triumphant, our secondary education will then run a great risk. Some teachers will no doubt wish to make room for the new theories. Novelties are so attractive, and it is so hard not to appear sufficiently advanced! At least they will wish to open up prospects to the children, who will be warned, before they are taught the ordinary mechanics, that it has had its day, and that at most it was only good for such an old fogey as Laplace. Then they will never become familiar with the ordinary mechanics.

Is it good to warn them that it is only approximate? Certainly, but not till later on; when they are steeped to the marrow in the old laws, when they have got into the way of thinking in them, and are no longer in danger of unlearning them, then they may safely be shown their limitations.

—Poincaré, *Science and Method* [pp. 249–250]



◀ **Figure 5.79** In describing microphysical processes, quantum mechanics abstracts physical quantities into operators, and states into Ψ functions.

(a) The electrically charged billiard ball traveling along a determined path as a model of the electron becomes an indistinct wave packet.

(b) The model of the atom as a solar system becomes a particular vibrational state of the descriptive Ψ function. The quantum numbers (n, l, m) now characterize this vibrational state: They are closely connected to the nodal surfaces of the vibrational states, which may be planes, spheres, and cones.

second experimental setup, when we measure the location, we cannot talk about light as a wave phenomenon at all.

We summarize once more the problems with the Copenhagen interpretation. According to Heisenberg, the opponents of this interpretation can be divided into three groups.

The first group agrees fully with all the assertions of a physical nature as well as with the mathematical formalism and also has no objections to its conclusions with respect to measurements and in particular to the theoretical limits to the measurement's precision. Yet they cannot come to terms with the philosophical consequences.

Representatives of the second group observe that the current formalism of quantum mechanics permits only the Copenhagen interpretation, and because of that, they would want to correct the theory itself.

The third group simply does not like the idea of quantum mechanics as such, although they recognize that logically it makes sense and there is no constructive counterproposal available. Heisenberg counts in this group, among others, Einstein, von Laue, and Schrödinger.

3. The assertions of the Copenhagen school lead us into the next sphere: the realm of pure philosophy. When we say that the state function Ψ gives the most complete possible description of a system, this amounts to saying that there is no possibility of extending it with such further (hidden) parameters so that we could definitely predict the result of a particular experiment, and therefore that there is no hope of returning to the old deterministic physics. This does not mean, of course, that quantum mechanics is a final, closed, and unchangeable system, but merely expresses the physicists' conviction that the progress is irreversible. Physicists believe that the processes controlling our world—the fundamental processes of the microworld—are fundamentally statistical in nature. *While classical physics imagined that behind the seemingly random phenomena of the macroworld stood the strictly deterministic laws of the microworld, today, in contrast, we believe that the seemingly deterministic laws of the macroworld are the result of the microworld's intrinsically probabilistic laws*.

   What this means for the practicing physicist is that understanding quantum mechanics consists of knowing, for a given situation, how to write down the relevant equation (such as the Schrödinger equation), solve this equation, and reconcile the solution with experimental findings. There are many who maintain that this is too superficial, and others believe that this is the most one can demand of a theory. Finally, there are those who go even further: One cannot indeed expect more, but in fact, one does not need anything more.

5.3.15 *What Remains of Classical Physics?*

Having concluded our discussion of the special and general theories of relativity and quantum theory, we would like to ask what of classical physics can survive and what must be left behind. Moreover, we would like to consider the extent to which our insight into the fundamental laws of nature has improved.

Let us examine the new physics.

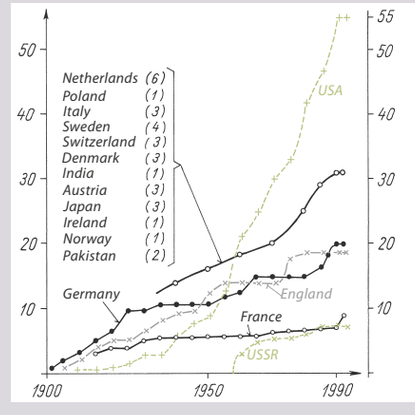
If we look at the assertions of relativity theory, such as the relativistic equation of motion, we observe that the departures from classical physics are of significance only for speeds on the order of magnitude of the speed of light. So we may say that classical mechanics is the limiting case of relativistic mechanics for velocities that are small in comparison to that of light. In other words, the laws of classical



▲ **Figure 5.80** The participants at the Solvay Congress of 1933 in Brussels. Next to the names of Nobel Laureates is the year in which their prize was awarded [Taton 1972]:

1. E. Henriot
2. E. Schrödinger, 1933
3. F. Perrin, 1926
4. F. Joliot, 1935
5. Irène Joliot-Curie, 1935
6. W. Heisenberg, 1932
7. H. A. Kramers
8. E. Stahel
9. N. Bohr, 1922
10. E. Fermi, 1938
11. A. Ioffe
12. E. T. S. Walton, 1951
13. P. A. M. Dirac, 1933
14. Mme. Curie, 1903, 1911
15. P. Debye, 1936
16. N. F. Mott, 1977
17. B. Cabrera
18. G. Gamow
19. P. Langevin
20. O. W. Richardson, 1928
21. M. S. Rosenblum
22. P. Blackett, 1948
23. Lord Rutherford, 1908
24. J. Errera
25. E. Bauer
26. T. de Donder
27. W. Pauli, 1945
28. J. E. Verschaffelt
29. M. Cosyns
30. M. de Broglie
31. E. Herzen
32. J. D. Cockcroft, 1951
33. C. D. Ellis
34. R. Peierls
35. L. de Broglie, 1929
36. Auguste Piccard
37. E. O. Lawrence, 1939
38. Lise Meitner
39. J. Chadwick, 1935
40. L. Rosenfeld
41. W. Bothe, 1954

mechanics retain their validity for describing phenomena that we encounter every day because the speeds attainable in practice by macroscopic bodies are very small compared to the speed of light. Thus, Newtonian mechanics is completely sufficient for the solution of most technical problems that arise in projects such as the engineering of roads, bridges, buildings, and machines. However, the question of the extent to which it is permissible to base the construction of an apparatus on Newtonian mechanics must be investigated precisely in cases in which the motion of microscopic particles, such as electrons or protons, plays a significant role. Apparatuses of this type include electron tubes used in high-frequency technology, x-ray tubes, and various nuclear particle accelerators. In these cases as well, a simple rule can be invoked to determine the applicability of Newtonian mechanics: Newtonian mechanics is applicable when the energy that a particle absorbs in the given apparatus is negligible in comparison with its rest energy; otherwise, relativistic equations must be used. In this sense, we may say that the theory of relativity has

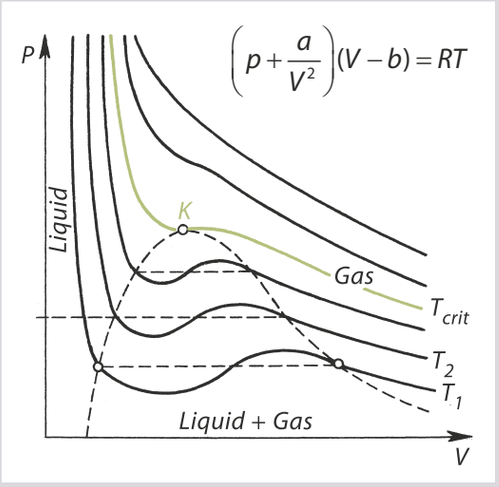


◀ **Figure 5.81** World events reflected in science: the total number of German, British, American, and Soviet Nobel Laureates as a function of time. The beginning of stagnation in

continued on next page

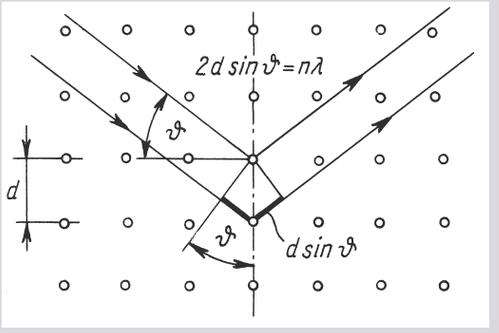
Figure 5.81 *continued*

physics in Germany clearly begins with Hitler's rise to power. After the Second World War, the effects of the "brain drain" become apparent as scientists move to the United States.



▲ **Figure 5.82** The van der Waals state equation and its graphical representation; here a/V^2 is the pressure-correction term due to the interaction of the molecules, and b accounts for the molecule's sizes.

Johannes Diderik van der Waals (1837–1923): Autodidact; 1877–1907, professor at the University of Amsterdam. In 1873, began to investigate the phase transformations liquid–gas and gas–liquid; he formulated his equation in 1881. In doing so, he noted the important fact that above a certain critical temperature, a gas will not turn into liquid no matter how high the pressure applied is. His research pointed the way for James Dewar and Heike Kamerlingh-Onnes's work on the liquefaction of hydrogen and helium. The name van der Waals is associated with the weak and short-range van der Waals force that acts between molecules and atoms.



▲ **Figure 5.83** Bragg reflection: x-rays are reflected on an array of lattice planes of atoms acting as a kind of mirror when the relationship $2d \sin \vartheta = n\lambda$ between the interplanar spacing, the wavelength, and the angle of incidence is satisfied, in which case, the waves scattered by the individual lattice points are mutually strengthened through interference. This method was also used by Davisson to demonstrate the wave nature of the electron and the validity of the de Broglie relationship $\lambda = h/mv$.

continued on next page

become a tool not only of the physicist, but also of the engineer. Nevertheless, we must continue to teach and learn classical Newtonian mechanics—and its mathematically elegant variants, the Lagrange and Hamilton equations—because they continue to offer the most adequate description of the motion of macroscopic bodies. But we have to know them also because they show how to arrive at the laws of microsystems and establish the fundamental relationships of quantum mechanics (Quotation 5.39).

The correlations of relativity theory are naturally more comprehensive than those of classical mechanics. *However, the theory of relativity does not tear down the edifice of classical mechanics, but rather contains it as a special case, and therefore each reinforces the credibility of the other.*

The special theory of relativity changed the backdrop and also the stage on which the story of physics unfolds. Instead of the Newtonian concepts of "absolute space" and "absolute time," it emphasizes that all inertial systems are equivalent, capitalizing on the fact that the speed of light is constant, and provides transformation formulas for the measurements of physical quantities and also, among other things, for such fundamental quantities as length and time in passing from one inertial system to another. The crucial requirement of relativity theory is the invariance of the form of natural laws for all coordinate systems that move linearly relative to one another.

With its synthesis of physics and geometry, the general theory of relativity represents a difficult challenge to our powers of imagination. Here, too, one can show that for gravitational fields of limited strength, we get back to Newton's law of gravitation, and therefore Newton's equations of motion for celestial bodies in gravitational fields of limited strength. In our solar system, deviations from Newtonian mechanics appear only in the precession of the perihelion of the orbit of Mercury, the closest planet to the Sun. This displacement of the perihelion can be established only with precise astronomical observations. For the universe as a whole, in contrast, the general theory of relativity is even today the only usable theory and the only one suitable for establishing quantitatively calculable models.

At first glance, at least, it would seem as though the equations of motion of classical mechanics have been completely displaced by quantum mechanics and are no longer in force. In fact, the postulates of quantum mechanics introduced in Section 5.3.11 do not have much in common with the fundamental laws of Newtonian mechanics. However, a connection can be reestablished without much difficulty. If we investigate the temporal change in the expected value of a momentum (that is, the average value of the momentum over a series of measurements), then according to the rules of quantum mechanics, we obtain an equation that allows this temporal change to be related to the average value of the spatial change of the potential energy. More precisely, we have the equation (Ehrenfest's theorem)

$$\frac{\mathrm{d}\langle p\rangle}{\mathrm{d}t}=-\left\langle\frac{\partial U}{\partial x}\right\rangle,$$

which can be recognized as $F = ma$, Newton's second law. For macroscopic bodies, Planck's constant h plays no role because of its small size, so that the commutation relations and the resulting Heisenberg uncertainty principle are insignificant. The precision attainable in the measurement of a physical quantity is then determined, as in classical mechanics, by the measurement process. We may therefore assume in this case that instead of the expected value or average value or probable value, we can talk about the measured value itself. In this way, we arrive once again at the

fundamental equations of classical mechanics. We have thus managed to fit classical mechanics into the framework of quantum mechanics and thereby ensure that in the investigation of macroscopic bodies, classical mechanics retains its validity. The limits of applicability of the classical theory are prescribed by the numerical value of Planck's constant h.

But quantum mechanics, both with its concrete results and with its philosophical background, allows us a deeper understanding of the fundamental structure of our world. It achieved this by completely giving up some of the concepts that we have abstracted from our everyday interaction with the macroworld, and thus it can be naturally visualized, such as the concept of path. Similarly, we have to modify our thinking about a concept from the macroworld that was primarily a result of our human experience: the causality principle (Figure 5.79).

Curved four-dimensional spaces, particles that appear sometimes as points, sometimes as extended waveforms—indeed, how can one grasp such things visually? We must accept as quite natural that for microphysics and the physics of the universe, there is no visually clear representation, nor can there be—at least not at present—because visualization at the same time implies a reduction to known sensory impressions. We must consider it a miracle that the human mind is at all capable of representing the array of physical processes that it has observed in a unified mathematical theory, even if an abstract one.

5.4 Nuclear Structure, Nuclear Energy

5.4.1 *A Backward Glance at the First Three Decades*

By the beginning of the 1930s, quantum mechanics, after spectacular successes and with its shortcomings and contradictions already resolved, had already become a routine theoretical tool for the physicists. The successful interpretation of many phenomena also reinforced the faith of physicists in the correctness of the conceptual basis, or, if you will, of the philosophical basis of the theory. With the 1930 publication of Heisenberg's book *The Physical Principles of the Quantum Mechanics* and with John von Neumann's *Mathematical Foundations of Quantum Mechanics*, published in 1932, quantum mechanics could finally be seen as having matured to the stage where textbooks could be published and where it could join alongside the recognized classical fields as a full-fledged new discipline. It was also true that, in the preface to his book, Heisenberg complained that physicists *believe* in quantum mechanics rather than *understand* it. In any case, as the disagreements over the interpretation of quantum mechanics had died down, the "greats" were now free to address open problems in other areas of physics with unabated energy.

However, the problems of the interpretation of the basic principles of quantum mechanics remained, and they reappear occasionally even today. At the time, the embers of doubt were kept smoldering by Einstein who was unwilling to believe that "God would play at dice" (Quotation 5.29). Even in the second half of the 1930s, Einstein continually found fresh arguments against the Bohr–Heisenberg interpretation of quantum mechanics. However, the focus of general interest had meanwhile shifted elsewhere.

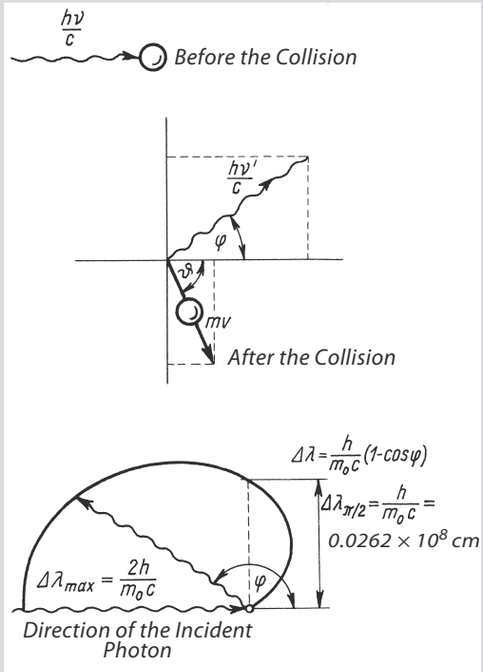
The reaching of a plateau in the understanding of the theoretical questions of quantum mechanics, and with it of the atomic shell, coincided with a significant turning point in the field of nuclear physics: In 1932, Chadwick discovered the neutron. The 1933 Solvay Conference in Brussels (Figure 5.80) was devoted to a

Figure 5.83 *continued*

William Henry Bragg (1862–1942): Studies at Cambridge; 1885–1907: professor at the University of Adelaide (Australia); 1912: ionization spectrometer and the first measurements of x-ray wavelengths and crystal data. His son, William Lawrence Bragg (1890–1971), also studied at Cambridge and at that time, together with his father—inspired by Laue's work—began work on crystals. The measuring apparatus was built by the father, and the son formulated the condition for reflection. Both played important roles in the scientific life of Great Britain.

W. L. Bragg was Rutherford's successor as director of the Cavendish Laboratory.

Together with his father, W. L. Bragg made a valuable contribution to physics education at the middle-school level: 20,000 students annually attended the lectures at the Royal Institution that they directed. They showed experiments at these lectures that would have been too expensive for any given school. They also promoted the continuing education of teachers.



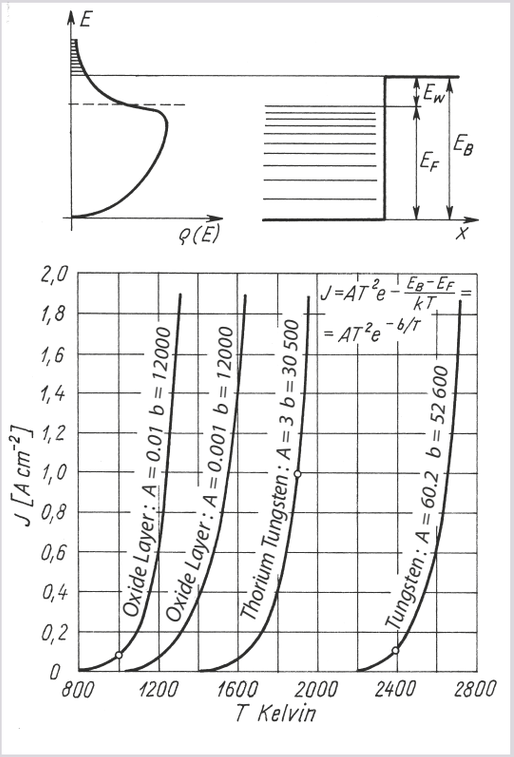
▲ **Figure 5.84** The Compton effect: In 1922, Compton became aware of the phenomenon that when x-rays are scattered on electrons, their wavelength changes. He gave an explanation for this in 1923: he described the scattering as a collision between a stationary free electron and a photon of energy $h\nu$ and of momentum $h\nu/c$. Using the conservation laws of momentum and energy, he was also able to correctly express the dependence of the wavelength on the direction. The Compton effect is primarily responsible for the absorption of x-rays at mid-range energies. At high energies, pair formation plays an increasing role, whereas at low energies, the photoelectric effect becomes significant.

As early as 1924, Bohr, Kramers, and Slater conjectured that, in microscopic processes, the energy and momentum conservation laws might be satisfied only in the statistical average. However, Hans Geiger and Walter Bothe showed, using the coincidence method they developed, that in Compton scattering, both conservation laws are also valid for every elementary process.

continued on next page

Figure 5.84 *continued*

Arthur Holly Compton (1892–1962): Doctorate at Princeton; 1923–1945: University of Chicago. Here he worked under Fermi's direction on the first realization of a sustained nuclear chain reaction.



▲ **Figure 5.85** Richardson's law (in the form given by Dushman) gives the current density of thermal electron emission from metals at temperature *T*. In the derivation, one pictures—based on the Sommerfeld model—the metal as a potential well in which, at temperatures close to absolute zero, the electrons completely fill all the energy levels below a particular one (called the Fermi level). At higher temperatures, the kinetic energy of a portion of the electrons increases enough so that they can escape the potential well. The difference in the height of the potential threshold—in other words, the depth of the potential well—and that of the Fermi level yields the work of escape. Along with the temperature, this material constant has the greatest influence on the values of the emission current density. The left part of the upper picture shows the electron density (horizontal axis) as a function of the energy (vertical axis). Electrons in the small shaded region have sufficient energy to be emitted.

discussion of the properties of the atomic nucleus. One-fourth of the physicists attending were already Nobel Laureates, but the greats of the future were also there: 20 current or future Nobel Laureates at the same conference! Alongside the veterans of nuclear physics and nuclear chemistry—Mme. Curie and Rutherford—we find, with Bohr at the head, all the pioneers of quantum mechanics: de Broglie, Heisenberg, Schrödinger, Dirac, Pauli, Fermi. Among these, Heisenberg, Bohr, and above all Fermi would play an eminent role in nuclear physics. Einstein and Planck are not in the photograph. We should not forget that we are talking about the year 1933, the year that Hitler seized power, which decisively affected the fate of many leading physicists (Figure 5.81). However, the absence of Planck and Einstein has also to do with physics: Planck was by now only an observer, and Einstein was in the opposition to quantum mechanics; furthermore, he is more interested in cosmological problems than in questions of nuclear physics.

In closing out one complex of topics and before launching into another, it is worth pausing for a moment to consider how the *physicists themselves* viewed and evaluated the achievements of the first three decades of the twentieth century. To a first approximation we can say that the Nobel Prize represents the best opinion poll of physicists: Great physicists get Nobel Prizes (Table 5.2, Figures 5.82–5.94); if we then look at the list of Nobel Laureates, we see that this hypothesis is more or less reversible: Nobel Laureates are great physicists.

Table 5.2 Nobel Prizes in physics up to 1991.

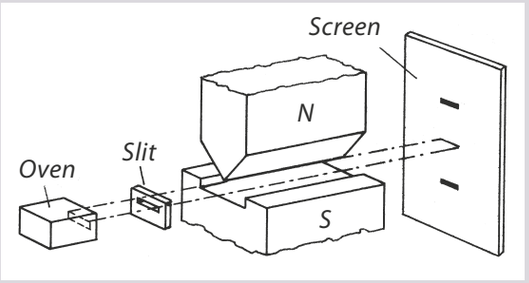
1901	W. Roentgen, 1845–1923 (Germany)	in recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him
1902	Hendrik Antoon Lorentz, 1853–1928 (Netherlands) Pieter Zeeman, 1865–1943 (Netherlands)	in recognition of the extraordinary service they rendered by their researches into the influence of magnetism upon radiation phenomena
1903	H. A. Becquerel, 1852–1908 (France)	in recognition of the extraordinary services he has rendered by his discovery of spontaneous radioactivity
	P. Curie, 1859–1906 (France) M. Curie née Skłodowska, 1867–1934 (Poland)	in recognition of the extraordinary services they have rendered by their joint researches on the radiation phenomena discovered by Professor Henri Becquerel
1904	Lord John William Strutt Rayleigh, 1842–1919 (England)	for his investigations of the densities of the most important gases and for his discovery of argon in connection with these studies
1905	Philipp Eduard Anton Lenard, 1862–1947 (Germany)	for his work on cathode rays
1906	Sir Joseph John Thomson, 1856–1940 (England)	in recognition of the great merits of his theoretical and experimental investigations on the conduction of electricity by gases
1907	Albert Abraham Michelson, 1852–1931 (USA)	for his optical precision instruments and the spectroscopic and metrological investigations carried out with their aid
1908	Gabriel Lippman, 1845–1921 (France)	for his method of reproducing colors photographically based on the phenomenon of interference
1909	Guglielmo Marconi, 1874–1937 (Italy) Karl Ferdinand Braun, 1850–1918 (Germany)	in recognition of their contributions to the development of wireless telegraphy
1910	Johannes Diderik van der Waals, 1837–1923 (Netherlands)	for his work on the equation of state for gases and liquids (Figure 5.82)

continued on next page

Table 5.2 Nobel Prizes in physics up to 1991, *continued*.

1911	Wilhelm Wien, 1864–1928 (Germany)	for his discoveries regarding the laws governing the radiation of heat
1912	Nils Gustaf Dalén, 1869–1937 (Sweden)	for his invention of automatic regulators for use in conjunction with gas accumulators for illuminating lighthouses and buoys
1913	Heike Kamerlingh-Onnes, 1853–1926 (Netherlands)	for his investigations on the properties of matter at low temperatures, which led, inter alia, to the production of liquid helium
1914	Max von Laue, 1879–1960 (Germany)	for his discovery of the diffraction of x-rays by crystals
1915	Sir William Henry Bragg, 1862–1942 (England) Sir William Lawrence Bragg, 1890–1971 (England)	for their services in the analysis of crystal structure by means of x-rays (Figure 5.83)
1916	—	
1917	Charles Glover Barkla, 1877–1944 (England)	for his discovery of the characteristic Roentgen radiation of the elements
1918	Max Karl Ernst Ludwig Planck, 1858–1947 (Germany)	in recognition of the services he rendered to the advancement of physics by his discovery of energy quanta
1919	Johannes Stark, 1874–1957 (Germany)	for his discovery of the Doppler effect in canal rays and the splitting of spectral lines in electric fields
1920	Charles Edouard Guillaume, 1861–1938 (Switzerland)	in recognition of the service he has rendered to precision measurements in physics by his discovery of anomalies in nickel steel alloys
1921	Albert Einstein, 1879–1955 (Germany)	for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect
1922	Niels Bohr, 1885–1962 (Denmark)	for his services in the investigation of the structure of atoms and of the radiation emanating from them
1923	Robert Andrews Millikan, 1868–1953 (USA)	for his work on the elementary charge of electricity and on the photoelectric effect
1924	Karl Manne Georg Siegbahn, 1886–1978 (Sweden)	for his discoveries and research in the field of x-ray spectroscopy
1925	James Franck, 1882–1964 (Germany) Gustav Hertz, 1887–1975 (Germany)	for their discovery of the laws governing the impact of an electron upon an atom
1926	Jean Baptiste Perrin, 1870–1942 (France)	for his work on the discontinuous structure of matter, and especially for his discovery of sedimentation equilibrium
1927	Arthur Holly Compton, 1892–1962 (USA)	for his discovery of the effect named after him (Figure 5.84)
	Charles Thomson Rees Wilson, 1869–1959 (Scotland)	for his method of making the paths of electrically charged particles visible by condensation of vapor
1928	Sir Owen Willans Richardson, 1879–1959 (England)	for his work on the thermionic phenomenon and especially for the discovery of the law named after him (Figure 5.85)
1929	Prince Louis-Victor de Broglie, 1892–1981 (France)	for his discovery of the wave nature of electrons
1930	Sir Chandrasekhara Venkata Raman, 1888–1970 (India)	for his work on the scattering of light and for the discovery of the effect named after him
1931	—	

continued on next page



▲ **Figure 5.86** The experiment of Stern and Gerlach.

It is to be expected from the Rutherford–Bohr–Sommerfeld model that the atom possesses a mechanical angular momentum and with it a tightly coupled magnetic moment; both of these follow at once from the idea that the electrons orbit the nucleus. An experiment by Einstein and W. J. de Haas measured the gyromagnetic ratio between the angular momentum and the magnetic moment for the first time. In this experiment, it was found that when a suspended iron rod was demagnetized, the reversal of "molecular currents" caused a change in the angular momentum that was compensated for by the observable turning of the rod (Einstein, de Haas: "Experimenteller Nachweis der ampereschen Molekularströme," 1915). The authors reported that their measurements were in good agreement with the theoretical values. However, more careful measurements made by E. Back in 1919 showed that the experiments, actually only give *half* of what is required by the theory. In searching for an explanation for the scattering of x-rays, it first occurred to Compton in 1921 that an electron could also have its own spin and magnetic moment. Put more vividly, not only does the electron orbit the nucleus, it also rotates on its own axis, which creates the magnetic moment. Goudsmit and Uhlenbeck gave definite evidence for this in 1925. The spin finally took its place in quantum theory in the Dirac equation.

In order to explain the hyperfine structure of the spectra, in 1924 Pauli proposed the idea of nuclear spin. Proton spin was required by D. M. Dennison in 1927 for the explanation of the existence of ortho- and parahydrogen.

Landé succeeded (in 1921–1923) in determining the magnetic moments of composite systems; through his work, he pointed out shortcomings in the Bohr theory (the absolute value of the angular momentum is not l but $\sqrt{l(l+1)}$) and also laid the groundwork for the introduction of electron spin.

The development of methods for direct measurement of the magnetic moment brought with it a closer connection between theory and experiment. In a homogeneous magnetic field, a particle with a magnetic moment experiences only a torque; the path of the particle can be altered only by an inhomogeneous field. The first experiment of this kind was carried out by Otto Stern and Walter Gerlach in 1922 (see illustration). They showed only directional quantization: the beam of evaporated silver atoms moving through an inhomogeneous magnetic field decomposed into two separate beams. The correct interpretation of this phenomenon was possible only after the introduction of electron spin.

Otto Stern (1888–1969): Assistant to Einstein in Prague; he followed him to Zurich and later even to America.

Isidor Isaac Rabi (1898–1988): Otto Stern's student, refined Stern's methods in the United States. With some modification Rabi's experimental setup serves as the standard for our measurement of time: the cesium atomic clock.

continued on next page

Figure 5.86 *continued*

Nuclear spins and the associated magnetic moments play a central role in the structural analysis of molecules (NMR spectroscopy) and in health care (MRI magnetic resonance imaging). As mentioned above, a torque acts on the magnetic moment in a homogeneous magnetic field. Viewed classically, the nucleus executes a precessional motion, like a gyroscope, with a given frequency. If one now overlays a linearly polarized high-frequency field on the *B*-field perpendicular to the *B* direction and varies its frequency, then at certain frequencies, resonance occurs that can be observed as an increased absorption of the high-frequency energy. Viewed quantum-mechanically, the absorbed energy leads to a change in the $2J + 1$ possible spin settings. Because the nuclear magnets of the atoms of a molecule mutually influence one another, one can draw conclusions from the resonance spectrum about the arrangement of the nuclei, that is, about the molecular structure.



▲ **Figure 5.87** The discoverers of the transistor.

John Bardeen (1908–1991): Doctorate at Princeton, after the Second World War until 1951: at Bell Laboratories; 1956: Nobel Prize for the discovery of the transistor; 1972: Nobel Prize for the theory of superconductivity (both times, prize awarded jointly).

Walter Houser Brattain (1902–1987): 1929–1967: Bell Laboratories.

William Bradford Shockley (1910–1989): From 1936, at Bell Laboratories; during the war, worked on technical problems in antisubmarine operations; from 1963, professor at Stanford University.

Table 5.2 Nobel Prizes in physics up to 1991, *continued*.

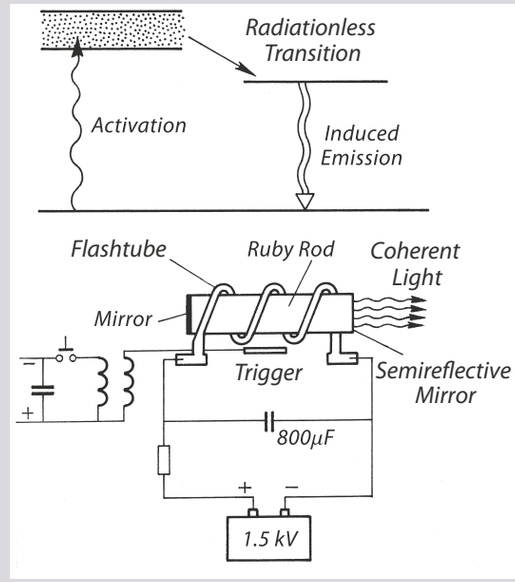
1932	Werner Heisenberg, 1901–1976 (Germany)	for the creation of quantum mechanics, the application of which has, inter alia, led to the discovery of the allotropic forms of hydrogen
1933	Erwin Schrödinger, 1887–1961 (Austria) Paul Adrien Maurice Dirac, 1902–1984 (England)	for the discovery of new productive forms of atomic theory
1934	—	
1935	Sir James Chadwick, 1891–1974 (England)	for the discovery of the neutron
1936	Victor Franz Hess, 1883–1964 (Austria)	for his discovery of cosmic radiation
	Carl David Anderson, 1905–1991 (USA)	for his discovery of the positron
1937	Clinton Joseph Davisson, 1881–1958 (USA) Sir George Paget Thomson, 1892–1975 (England)	for their experimental discovery of the diffraction of electrons by crystals
1938	Enrico Fermi, 1901–1954 (Italy)	for his demonstrations of the existence of new radioactive elements produced by neutron irradiation, and for his related discovery of nuclear reactions brought about by slow neutrons
1939	Ernest Orlando Lawrence, 1901–1958 (USA)	for the invention and development of the cyclotron and for results obtained with it, especially with regard to artificial radioactive elements
1940–1942	—	
1943	Otto Stern, 1888–1969 (USA)	for his contribution to the development of the molecular ray method and his discovery of the magnetic moment of the proton (Figure 5.86)
1944	Isidor Isaac Rabi, 1898–1988 (USA)	for his resonance method for recording the magnetic properties of atomic nuclei (Figure 5.87)
1945	Wolfgang Pauli, 1900–1958 (Austria)	for the discovery of the exclusion principle, also called the Pauli principle
1946	Percy Williams Bridgman, 1882–1961 (USA)	for the invention of an apparatus to produce extremely high pressures, and for the discoveries he made therewith in the field of high-pressure physics
1947	Sir Edward Victor Appleton, 1892–1965 (England)	for his investigations of the physics of the upper atmosphere especially for the discovery of the so-called Appleton layer
1948	Lord Patrick Maynard Stuart Blackett, 1897–1974 (England)	for his development of the Wilson cloud chamber method, and his discoveries therewith in the fields of nuclear physics and cosmic radiation
1949	Hideki Yukawa, 1907–1981 (Japan)	for his prediction of the existence of mesons on the basis of theoretical work on nuclear forces
1950	Cecil Frank Powell, 1903–1969 (England)	for his development of the photographic method of studying nuclear processes and his discoveries regarding mesons made with this method
1951	Sir John Douglas Cockcroft, 1897–1967 (England) Ernest Thomas Sinton Walton, 1903–1995 (Ireland)	for their pioneering work on the transmutation of atomic nuclei by artificially accelerated atomic particle
1952	Felix Bloch, 1905–1983 (USA) Edward Mills Purcell, 1912–1997 (USA)	for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith

continued on next page

Table 5.2 Nobel Prizes in physics up to 1991, *continued*.

1953	Frits (Frederik) Zernike, 1888–1966 (Netherlands)	for his demonstration of the phase contrast method, especially for his invention of the phase contrast microscope
1954	Max Born, 1882–1970 (Germany)	for his fundamental research in quantum mechanics, especially for his statistical interpretation of the wave function
	Walther Bothe, 1891–1957 (Germany)	for the coincidence method and his discoveries made therewith
1955	Willis Eugene Lamb, 1913–2008 (USA)	for his discoveries concerning the fine structure of the hydrogen spectrum
	Polykarp Kusch, 1911–1993 (USA)	for his precision determination of the magnetic moment of the electron
1956	William Shockley, 1910–1989 (USA) John Bardeen, 1908–1991 (USA) Walter Houser Brattain, 1902–1987 (USA)	for their researches on semiconductors and their discovery of the transistor effect (Figure 5.88)
1957	Chen Ning Yang, b. 1922 (USA) Tsung-Dao Lee, b. 1926 (USA)	for their penetrating investigation of the so-called parity laws, which has led to important discoveries regarding elementary particles
1958	Pavel Alekseevich Cherenkov, 1904–1990 (USSR) Ilya Mikhailovich Frank, 1908–1990 (USSR) Igor Yevgenevich Tamm, 1895–1971 (USSR)	for the discovery and the interpretation of the Cherenkov effect
1959	Emilio Gino Segrè, 1905–1989 (USA) Owen Chamberlain, 1920–2006 (USA)	for their discovery of the antiproton
1960	Donald A. Glaser, b. 1926 (USA)	for the invention of the bubble chamber
1961	Robert Hofstadter, 1915–1990 (USA)	for his pioneering studies of electron scattering in atomic nuclei and for his thereby achieved discoveries concerning the structure of the nucleons
	Rudolf Ludwig Mössbauer, b. 1929 (Germany)	for his researches concerning the resonance absorption of gamma radiation and his discovery in this connection of the effect that bears his name
1962	Lev Davidovich Landau, 1908–1969 (USSR)	for his pioneering theories for condensed matter, especially liquid helium
1963	Eugene P. Wigner, 1902–1995 (USA)	for his contributions to the theory of the atomic nucleus and the elementary particles, particularly through the discovery and application of fundamental symmetry principles
	Maria Goeppert-Mayer, 1906–1972 (USA) H. Hans D. Jensen, 1907–1973 (Germany)	for their discoveries concerning nuclear shell structure
1964	Charles H. Townes, b. 1915 (USA)	for fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser–laser principle (Figure 5.90)
	Nikolai Gennadievich Basov, 1922–2001 (USSR) Aleksandr Mikhailovich Prokhorov, 1916–2002 (USSR)	for fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser–laser principle (Figure 5.89)
1965	Sin-Itiro Tomonaga, 1906–1979 (Japan) Julian S. Schwinger, 1918–1994 (USA) Richard P. Feynman, 1918–1988 (USA)	for their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles
1966	Alfred Kastler, 1902–1983 (France)	for the discovery and development of optical methods for studying Hertzian resonances in atoms
1967	Hans Albrecht Bethe, 1906–2005 (USA)	for his contributions to the theory of nuclear reactions, especially his discoveries concerning the energy production in stars

continued on next page



▲ **Figure 5.88** Diagram of the pulsed ruby laser.

The operation of a laser is based on the following principle: When a microsystem passes from a higher energy state to a lower one due to the stimulation of an external excitatory wave, the emitted electromagnetic wave will be coherent with the incident wave (Figures 5.53, 5.54). In a ruby laser, the "pumping"—that is, the raising of the microsystem to a higher energy level—is accomplished with a flashtube. From the pumped energy states, which range over a relatively wide band, the microsystems first move into a sharply defined, somewhat lower-lying energy state, from which they move back into the ground state under the action of the excitatory wave. The excitation in the ruby laser is accomplished by an electromagnetic field that is maintained as a standing wave between two mirrors (one of which is semireflective).

The pioneers of quantum electronics:

Charles Hard Townes (b. 1915): 1939–1948: development of radar systems at Bell Laboratories; 1948–1959: Columbia University; from 1961: Massachusetts Institute of Technology.

Nikolai Gennadievich Basov (1922–2001): Doctorate in engineering physics (1950), then worked as the Lebedev Physical Institute of the Academy of Sciences of the USSR. From 1963, professor at the Moscow Engineering Physics Institute.

Aleksandr Mikhailovich Prokhorov (1916–2002): Studies in Leningrad; from 1946, work at the Physical Institute of the Academy of Sciences of the USSR; from 1959, professor at Moscow State University (MGU).

Table 5.2 Nobel Prizes in physics up to 1991, *continued*.

1968	Luis W. Alvarez, 1911–1981 (USA)	for his decisive contributions to elementary particle physics, in particular the discovery of a large number of resonance states, made possible through his development of the technique of using hydrogen bubble chamber and data analysis
1969	Murray Gell-Mann, b. 1929 (USA)	for his contributions and discoveries concerning the classification of elementary particles and their interactions
1970	Hannes Alfvén, 1908–1995 (Sweden)	for fundamental work and discoveries in magneto-hydrodynamics with fruitful applications in different parts of plasma physics
	Louis Néel, 1904–2000 (France)	for fundamental work and discoveries concerning antiferromagnetism and ferrimagnetism, which have led to important applications in solid state physics
1971	Dennis Gabor, 1900–1979 (England)	for his invention and development of the holographic method (Figures 5.93, 5.94)
1972	John Bardeen, 1908–1991 (USA) Leon N. Cooper, b. 1930 (USA) J. Robert Schrieffer, b. 1931 (USA)	for their jointly developed theory of superconductivity, usually called the BCS theory
1973	Leo Esaki, b. 1925 (Japan) Ivar Giaever, b. 1929 (Norway)	for their experimental discoveries regarding tunneling phenomena in semiconductors and superconductors, respectively
	Brian D. Josephson, b. 1940 (England)	for his theoretical predictions of the properties of a supercurrent through a tunnel barrier, in particular those phenomena that are generally known as the Josephson effects
1974	Sir Martin Ryle, 1918–1984 (England) Antony Hewish, b. 1924 (England)	for their pioneering research in radio astrophysics, Ryle for his observations and inventions, in particular, of the aperture synthesis technique; and Hewish for his decisive role in the discovery of pulsars
1975	Aage Bohr, b. 1922 (Denmark) Ben Mottelson, 1926–1981 (Denmark) James Rainwater, 1917–1986 (USA)	for the discovery of the connection between collective motion and particle motion in atomic nuclei and the development of the theory of the structure of the atomic nucleus based on this connection
1976	Burton Richter, b. 1931 (USA) Samuel C. C. Ting, b. 1936 (USA)	for their pioneering work in the discovery of a heavy elementary particle of a new kind
1977	Philip W. Anderson, b. 1923 (USA) Sir Nevill F. Mott, 1905–1996 (England) John H. van Vleck, 1899–1980 (USA)	for their fundamental theoretical investigations of the electronic structure of magnetic and disordered systems
1978	Pyotr Leonidovich Kapitsa, 1894–1984 (USSR)	for his basic inventions and discoveries in the area of low-temperature physics
	Arno A. Pensias, b. 1933 (USA) Robert W. Wilson, b. 1936 (USA)	for their discovery of cosmic microwave background radiation
1979	Sheldon L. Glashow, b. 1932 (USA) Abdus Salam, b. 1926 (Pakistan) Steven Weinberg, b. 1933 (USA)	for their contributions to the theory of the unified weak and electromagnetic interaction between elementary particles, including inter alia, the prediction of the weak neutral current
1980	James W. Cronin, b. 1931 (USA) Val L. Fitch, b. 1923 (USA)	for the discovery of violations of fundamental symmetry principles in the decay of neutral K mesons
1981	Nicolaas Bloembergen, b. 1920 (USA) Arthur L. Schawlow, 1921–1999 (USA)	for their contribution to the development of laser spectroscopy
	Kai M. B. Siegbahn, 1918–2007 (Sweden)	for his contribution to the development of high-resolution electron spectroscopy

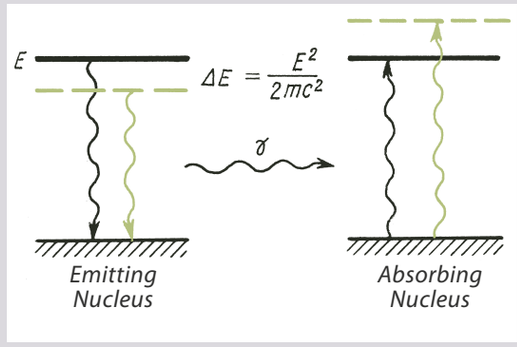
continued on next page

Table 5.2 Nobel Prizes in physics up to 1991, *continued*.

1982	Kenneth G. Wilson, b. 1936 (USA)	for his theory for critical phenomena in connection with phase transitions
1983	Subramanyan Chandrasekhar, 1910–1995 (Pakistan–England–USA)	for his theoretical studies of the physical processes of importance to the structure and evolution of the stars
	William A. Fowler, 1911–1995 (USA)	for his theoretical and experimental studies of the nuclear reactions of importance in the formation of the chemical elements in the universe
1984	Carlo Rubbia, b. 1934 (Italy) Simon van der Meer, b. 1925 (Netherlands)	for their decisive contributions to the large project that led to the discovery of the field particles W and Z, communicators of the weak interaction
1985	Klaus von Klitzing, b. 1943 (Germany)	for the discovery of the quantized Hall effect (Figure 5.91)
1986	Ernst Ruska, 1906–1988 (Germany)	for his fundamental work in electron optics, and for the design of the first electron microscope
	Gerd Binnig, b. 1947 (Germany) Heinrich Rohrer, b. 1933 (Switzerland)	for their design of the scanning tunneling microscope (Figure 5.92)
1987	J. Georg Bednorz, b. 1950 (Germany) K. Alexander Müller, b. 1927 (Switzerland)	for their important breakthrough in the discovery of superconductivity in ceramic materials
1988	Leon M. Lederman, b. 1922 (USA) Melvin Schwartz, b. 1932 (USA) Jack Steinberger, b. 1921 (USA)	for the neutrino beam method and the demonstration of the doublet structure of the leptons through the discovery of the muon neutrino
1989	Norman F. Ramsey, b. 1915 (USA)	for the invention of the separated oscillatory fields method and its use in the hydrogen maser and other atomic clocks
	Hans G. Dehmelt, b. 1922 (USA) Wolfgang Paul, 1913–1993 (Germany)	for the development of the ion trap technique
1990	Jerome I. Friedman, b. 1930 (USA) Henry W. Kendall, b. 1926 (USA) Richard E. Taylor, b. 1929 (Canada)	for their pioneering investigations concerning deep inelastic scattering of electrons on protons and bound neutrons, which have been of essential importance for the development of the quark model in particle physics
1991	Pierre-Gilles de Gennes, b. 1932 (France)	for discovering that methods developed for studying order phenomena in simple systems can be generalized to more complex forms of matter, in particular, to liquid crystals and polymers

We may note, with a measure of surprise, that in 1912, Nils Gustav Dalén received a Nobel in physics for the automatic gas regulator for lighthouses and buoy lights. This apparatus was no doubt of great practical importance for seafaring nations, but it advanced physics precious little. However, we can forgive the Royal Swedish Academy for a bit of subjectivity.

In the list we, of course, encounter the names Planck and Einstein. It nonetheless surprises us how late both received the Nobel Prize: Planck in 1918 and Einstein in 1921. It is also worth noting that Einstein received the Nobel Prize neither for his formulation of the theory of relativity nor for his discovery of the equivalence of mass and energy; in the award citation, Einstein's interpretation of the photoelectric effect was singled out for emphasis. From this follows the peculiar situation that the relationship $E = mc^2$, so incredibly important for the atomic age, was not listed as the outstanding individual achievement. We have seen in the previous section that the form of this formula that eventually crystallized was indeed the result of the collective work of a number of physicists.

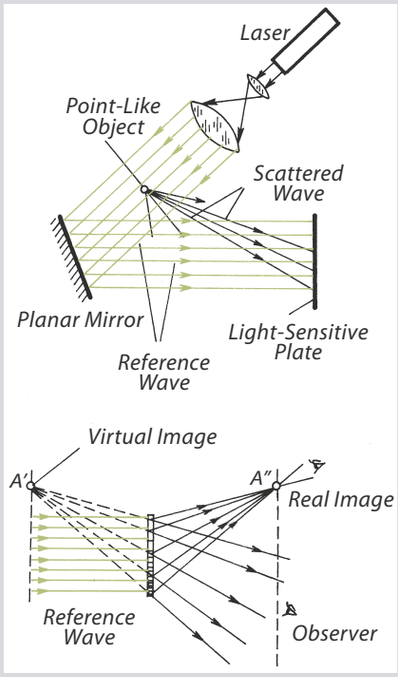


▲ **Figure 5.89** The Mössbauer effect.

If an atomic nucleus passes from an excited state into the ground state with the emission of a gamma quantum, then the nucleus can also move into the excited state through the absorption of the same quantum (resonance absorption). In reality, this simple phenomenon is complicated by the fact that as compensation for the momentum of the emitted gamma quantum, the radiating nucleus suffers a recoil and takes a portion of the energy with it. Given the available energy E, momentum $p = E/c$, and the mass of the nucleus m, the energy of the emitted gamma quantum will be reduced by

$$\Delta E = \frac{p^2}{2m} = \left(\frac{E}{c}\right)^{22} \times \frac{1}{2m}.$$

In contrast, during absorption, the nucleus will receive momentum, and if it is to acquire enough energy to move to the excited state, it must absorb a gamma quantum with ΔE to spare. With nuclei that are built into a crystal lattice, it is possible for the emitting or absorbing nuclei to lean, so to speak, on the entire lattice and thus give off or take up only a miniscule amount of energy when they emit or absorb the gamma quantum. In such cases, the resonance absorption proceeds with very sharply defined energy values, and therefore even the least effect resulting in a change in energy can be detected—for example, the state of motion of the emitting or absorbing nucleus, or the difference between the gravitational potentials at the places of emission and absorption.



▲ **Figure 5.90** Holography. In ordinary optical imaging (on a black-and-white photographic plate), only the intensity distribution of the waves coming from the object is preserved, even though these waves carry much more information—the total information being encoded in both the amplitude and the phase. The process of holography allows this phase information to be recorded also. This is done by bringing the light waves scattered from the object into interference with the waves of a reference beam, which must be coherent with the beam illuminating the object. The resulting interference pattern is recorded on the photographic plate as a hologram.

To reconstruct the image, a coherent beam identical to the reference beam illuminates the hologram, which lets the observer enjoy a complete spatial image from many viewpoints.

Dennis Gabor (1900–1979): Born in Budapest; received engineering degree. 1927–1933, research engineer at the Siemens factory in Berlin; from 1949, Imperial College, London; from 1958, professor there. The idea of holography goes back to the year 1947, while the practical realization had to wait until 1960 with the advent of laser technology.

Dennis Gabor was a multifaceted researcher who worked productively in a number of areas: communications theory, high-speed oscillography, television technology, and cryogenics (Nobel Prize 1971).

We should also mention that in 1928, that is, in one of the most exciting years of quantum mechanics, it was Richardson who was awarded the Nobel Prize for the theory of thermal electron emission.

If we look for recognition of the outstanding achievements in nuclear physics, then we encounter in the table only Becquerel and the Curies, who in 1903 received the Nobel Prize for the discovery of radioactivity and for their pioneering work in this field.

A bit belatedly, also in the years of ferment in quantum mechanics, Wilson was recognized with a Nobel Prize for his cloud chamber. Many discoveries in nuclear physics are the result of measurements made possible with this instrument.

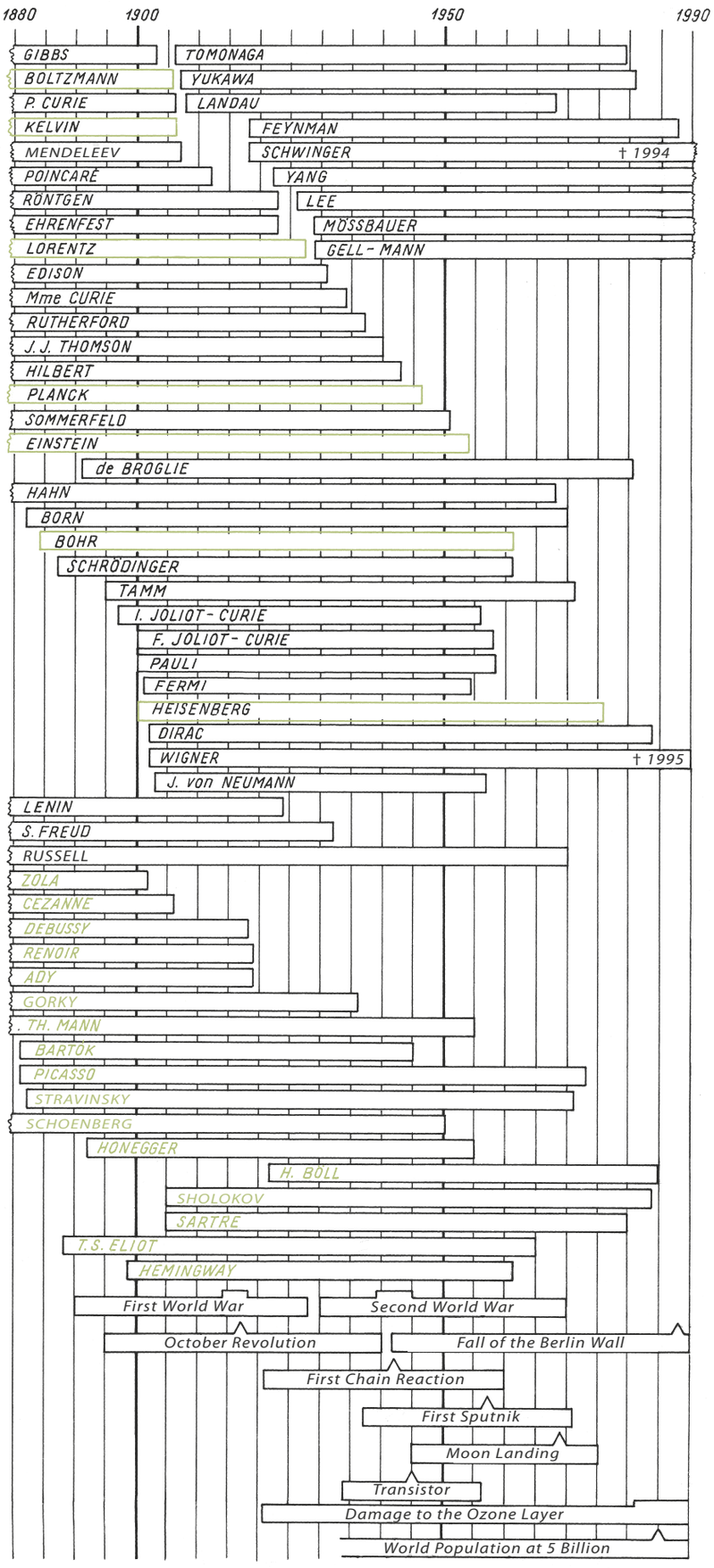
Surprisingly, we will look in the table in vain for Rutherford, the pioneer of nuclear physics and one of the most important experimental physicists of all time. He received his Nobel Prize in chemistry in 1908 for his work on natural radioactivity. We also encounter among the chemists the name Soddy, who in 1921 won the Nobel Prize for chemistry (the Nobel Prize for physics was awarded to Einstein in that year). At least Soddy was a chemist, not a physicist. Aston, too, received a Nobel Prize in chemistry, for his studies of isotopes (the Nobel Prize in physics went that year to Bohr). Aston's discoveries are important to both physics and chemistry; his research approach, however, was entirely physical in nature. We can thus understand why in his address given upon receiving the Nobel Prize, Rutherford spoke of a variety of transformations that he had encountered in his work. Some of these transformations took a comparatively long time, and others happened quite quickly. But none was as quick as the one that transformed him from a physicist into a chemist.

Table 5.3 presents the life spans of some of the physicists who were active at the turn of the century and into the twentieth century in the context of other events of the era.

5.4.2 *Stations of the Study of the Atomic Nucleus*

Let us now turn our attention to the most important stations marking the advance in our knowledge of the structure of the atomic nucleus. As we have already seen, the leading area of research, or more frivolously the most fashionable one, during the first third of the twentieth century was the problem of the atomic shell. In Plate XXI, we have attempted to give an impression of the intensity with which research was taking place in various areas and to illustrate how the center of interest shifted over the course of time. We see that the physics of the atomic nucleus came into vogue at the beginning of the 1930s, only to surrender its place at the end of the 1940s to the physics of elementary particles as the new headline-grabber. The same color plate differentiates between the experiments related to the atomic nucleus and the atomic shell from the beginning of this century, although we know from the previous sections that these areas were consciously separated only after 1911, when Rutherford created his atomic model, consisting of a central nucleus and a number of electrons orbiting the nucleus.

The year 1911, or rather the year 1913, in which the refined *Rutherford–Bohr model* was presented, marked the end of the first epoch of nuclear physics, but for the true Heroic Age we count the eight years following the first observation of a nuclear phenomenon, that is, the years 1896 to 1904. The leading figures of this period were Becquerel, the Curies, and Rutherford. It is almost beyond our comprehension today just how revolutionary their discoveries were. Their courage in reporting their results was buttressed by the most circumspect experimental work. To think that Lavoisier had come up with the notion of a chemical element



◀ **Table 5.3** Physicists of the twentieth century from the turn of the century through the 1980s.

To attempt a brief characterization of this century, one can gather courage only from ignorance. Considering the many spectacular achievements of *technology*, one might speak of the age of the atom, the age of space travel, or the age of the computer.

From a *philosophical* point of view, the following can be viewed as the most significant advances in basic physics research: the space-time concept of relativity theory, the probabilistic interpretation of quantum mechanics, the connection between symmetries and the conservation laws.

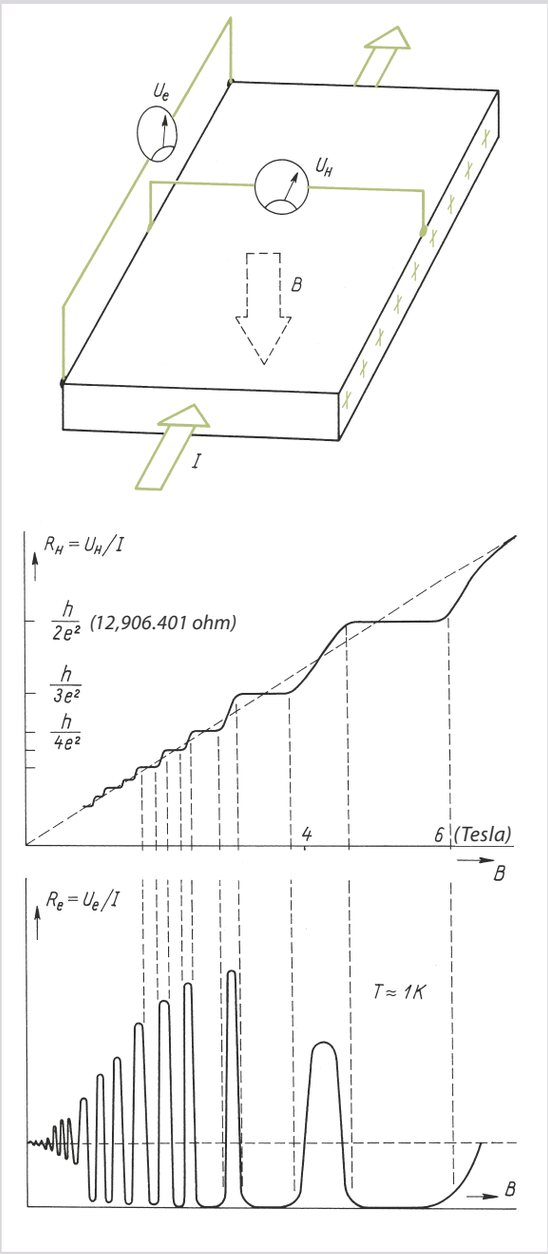
The physics of the nineteenth century did not make use of the idea of evolution, which was the leitmotif of biology and of the social sciences; rather, it painted, as we have seen, the vision of heat death, excluding all possible change. In contrast, the physics of today is based on the likelihood of a dynamic expanding universe.

The eighteenth century still lived under the belief in the omnipotence of the spirit and the knowledge created by it. In the nineteenth century, the social problems resulting from the practical application of scientific accomplishments—from the industrial revolution—could no longer be overlooked; but the hope remained that reason could overcome them.

The first half of this century was overshadowed by two world wars, and in the second half people lived in fear of a nuclear catastrophe. Many blamed science in general and looked elsewhere, such as to the irrational, for a solution.

In the last quarter of the century it became clear to many that not only military conflicts, but also the well-meaning but unbridled striving for a higher standard of living, could lead to disaster. For material well-being, one needs energy, food, and other consumer goods in ever-increasing quantities. The undesirable consequences are air pollution, climate change, and the endangerment of many plant and animal species.

The wounds caused by science and technology can be healed only by science and technology, say the scientists. However, restraint and political will are also necessary factors for any solution.



▲ **Figure 5.91** The Hall effect, discovered in 1879 by Edwin Herbert Hall: If a conductor with a current flowing through it, as shown in the figure, is brought into a homogeneous magnetic field, then a voltage arises perpendicular to the field direction and to the direction of the current. The size of this Hall voltage is $U_H = C_H BI(1/a)$, where C_H is the Hall constant. It depends on the charge-carrier density n in the simplest case as follows: $C_H = 1/nq$, where q is the charge of the current-carrier particles.

The Hall voltage arises because the charge carriers in the conductor are deflected by the Lorentz force $q\mathbf{v} \times \mathbf{B}$ and accumulate on the boundary surface along the side until the Lorentz force is equal to the electric force field being built up in this way. The Hall voltage can be used, among other things, for investigating the conduction mechanism of metals and semiconductors and for the measurement of magnetic fields.

continued on next page

only one hundred years earlier! In those hundred years, the concept of an immutable chemical element whose individual atoms each possess identical properties was the firm and incontrovertible foundation of the rapidly developing science of chemistry. And now it turns out that the confusing tangle of physical and chemical phenomena around radioactivity would become clear and understandable only if we gave up immutability: in other words, if in nature an element might spontaneously transmute into another, just as the alchemists had imagined. At the same time, it also appeared that atoms of an element that is homogeneous by every chemical criterion could have different masses, and consequently other differences in their radioactive behavior. *Here we meet for the first time the phenomenon of transmutation of elements, and also the concept of isotopes.*

And there was yet a third revolutionary phenomenon: At this time it was also realized that these phenomena involve very large releases of energy that are many orders of magnitude greater than what is usual in chemistry.

Toward the end of the first decade of the twentieth century, radiation from natural radioactivity, primarily alpha rays, was already used as a tool for research. As we saw in the previous section, this was how Rutherford discovered the atomic nucleus.

During the next 20 years, not much happened in nuclear physics except for a single outstanding—indeed sensational—event: In 1919, Rutherford succeeded in carrying out the first artificial transmutation of a nucleus. The crucial turning point, as we have already mentioned, was the discovery of the neutron in 1932. In addition, a revolution in experimental methodology also took place. This was the time of the introduction of particle accelerators, whose projectiles, more powerful than those resulting from natural radioactivity, could be used to "smash" atoms to study their inner structure. Theoretical and experimental results relating to the nucleus steadily accumulated, leading to the first models of the nucleus, and then in 1938, a theoretically unexpected nuclear reaction was discovered marking the beginning of the age of practical nuclear energy: the *fission of uranium*. From this point in time, the focus of research in nuclear physics shifted to the problems of practical applications, above all to questions of energy creation. At the same, the interest of the vanguard of physicists turned more and more to a new direction; for them the physics of elementary particles represented the push toward the unknown.

5.4.3 *Becquerel: Why Do Uranium Salts Fluoresce?*

Mankind's first encounter with a phenomenon originating in the atomic nucleus was due to an error, and then finding the right path to its explanation was thanks to an accident.

As we have already mentioned, few discoveries have aroused as much interest, both in the scientific world as well as the general public, as the discovery of x-rays. At the meeting of the French Academy of Sciences on January 10, 1896, two researchers presented x-ray photographs that they had made. Becquerel was present at this meeting, and Poincaré had brought along Roentgen's article. In the course of a discussion following the lecture, Poincaré, in reply to a question raised by Becquerel, informed him that the x-rays are emitted from the glass wall of the discharge tube at a point of fluorescence. Becquerel, who at the time, following in his father's footsteps, was investigating fluorescent materials, remarked that there might be a connection between the two phenomena, fluorescence and the emission of x-rays, in the sense that they might have the same cause. This opinion was written about by Poincaré and was widely disseminated. Becquerel,

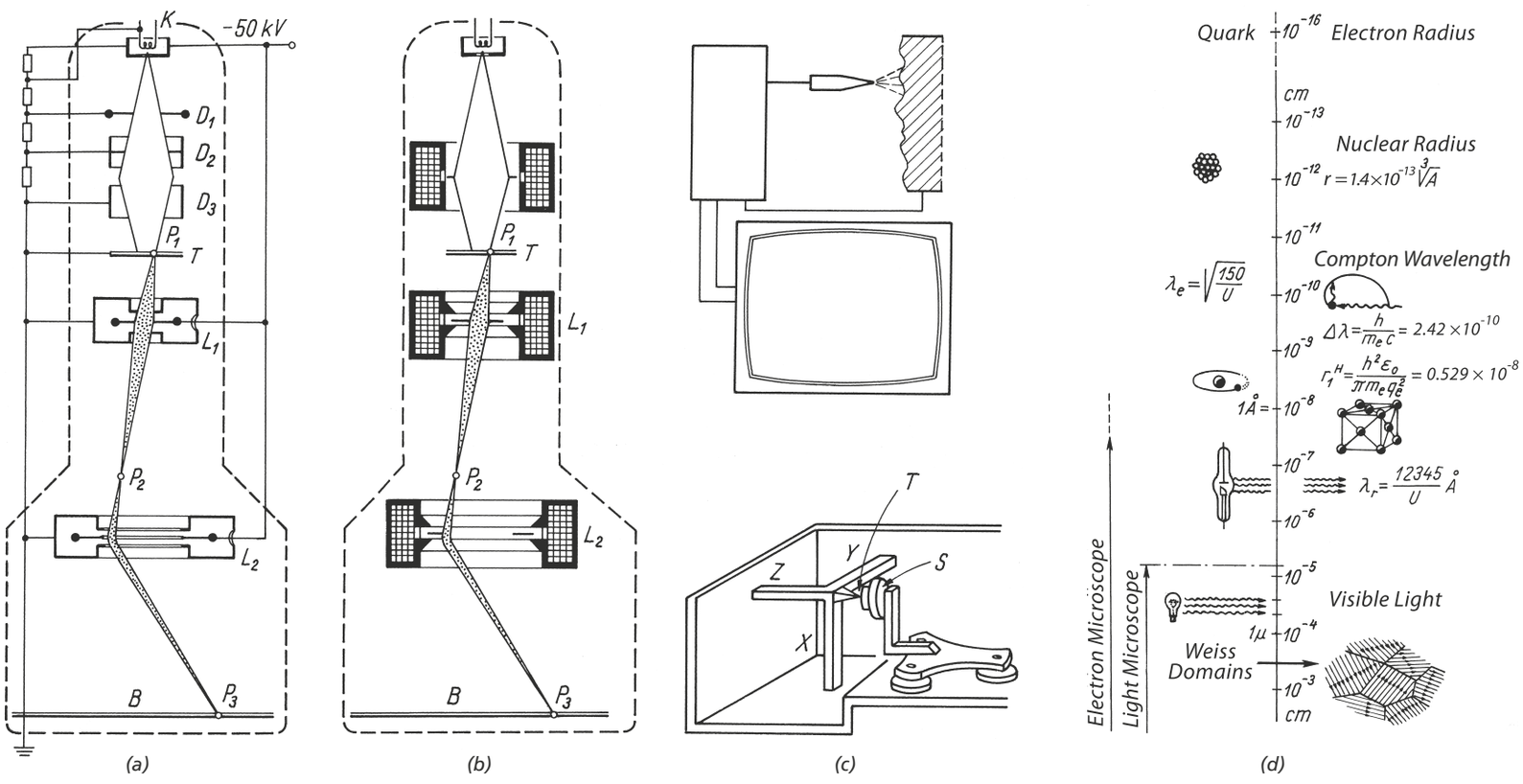
hoping to prove the hypothesis, carefully wrapped a photographic plate in layers of black paper and made sure that the plates did not show any blackening even when exposed to sunlight for several hours. He placed a uranium salt crystal that he had earlier experimented with as a fluorescing material on the wrapped plate. He then placed the whole assembly in sunlight in order to make the uranium salt fluoresce.

After the plate was developed, Becquerel could clearly see a blackening in the shape of the uranium salt crystal, which he took as an indication that fluorescence was accompanied by an emission of x-rays. It is surprising that in his article, Becquerel cites several other researchers who claimed to have observed a similar effect under conditions that, as we know today, should not have caused any effect because their fluorescent crystals did not contain any radioactive material. The association of the two phenomena, fluorescence and x-rays, is, of course, not correct. Today, we know that the x-rays were created by the electrons that struck the glass wall and were thereby slowed down. The fluorescence was only an incidental side effect. In modern

Figure 5.91 *continued*

The idea arose early (Landau, 1930) that under extreme experimental conditions (near absolute zero, in strong magnetic fields, in two-dimensional conductors), quantum effects should occur, with the Hall resistance R_H defined by the equation $U_H = R_H I$ changing by discrete steps with increasing magnetic field strength.

A two-dimensional system can be realized with the surface layer between a semiconductor and a metal (J. R. Schrieffer, 1957). Japanese physicists (T. Ando and colleagues, 1976–1981) have suggested that the quantum jumps in conductivity are closely related to the quantity e^2/h. Klaus von Klitzing and colleagues demonstrated these quantum jumps experimentally in 1980. In doing so, they showed that the jumps are very precisely constant (to one part in 10^7); this made possible a number of theoretical and practical applications: precise determination of fundamental constants, realization of resistance standards, etc.



▲ **Figure 5.92** An electron microscope—whether equipped with electrostatic **(a)** or magnetic **(b)** lenses—has a certain structural similarity to the optical microscope. An STM (scanning tunneling microscope) has a completely different structure **(c)**.

Electron microscopy in the narrow sense begins with the work of H. Busch (1926), who found that certain rotationally symmetric electrostatic and magnetic fields possess quasioptic imaging properties.

E. Ruska and M. Knoll built the first electron microscope with magnetic lenses (1931–1932), and the first with electrostatic lenses was constructed by E. Brüche and H. Johansson (1931); in 1934, the resolving power of the light microscope was exceeded. By the end of the 1930s in Germany as well as in other countries, practical electron microscopes useful in research and industry began to appear on the market.

The STM, developed by G. Binnig and H. Rohrer (IBM Research Laboratory, Zurich) is based on the tunnel effect (Figure 5.63): The surface to be investigated is scanned by a needle with a very sharp point (with radius of about 1 nm = 10^{-9} m), where the point of the needle is held also about 1 nm from the surface. The electron stream generated via the tunnel effect depends on the structural inhomogeneity of the surface. In this way the computer can build a three-dimensional image of the surface.

Notation in the diagrams **(a)**, **(b)**, and **(c)**: P_1: object point; P_2: intermediate image point; P_3: image point on a photographic plate or luminescent screen; L_1: objective lens; L_2: ocular lens; T: object; D_1, D_2, D_3: condenser lens; N: metallic needle; T: surface; X, Y, Z: piezoelectric actuators to move the needle.

continued on next page

Figure 5.92 *continued*

Diagram **(d)** compares the resolutions of light and electron microscopes. It is a roadmap of the sizes of atomic structures in terms of orders of magnitude.

The de Broglie wavelength of the electron, $\lambda_e = \sqrt{150 / U}$ (Å) , is determined from the equation $\lambda_e = h/mv$; v can be calculated from the nonrelativistic energy equation $eU = ½\ mv^2$. The x-ray wavelength $\lambda_r = 12345/U$ (Å) is derived from the equations $hv = eU$ and $\lambda_r v = c$. Here and in diagram **(d)**, U is the voltage, A is the atomic mass number, and Å (Ångstrom) is 10^{-8} cm.

The Weiss domains are the spontaneously magnetized regions of a ferromagnetic material, named for the French physicist Pierre Weiss (1865–1940).

Quotation 5.40

[Madame Curie wrote] "We had no money, no laboratory and no help in the conduct of this important and difficult task ... It was like creating something out of nothing, and ... I may say without exaggeration that this period was, for my husband and myself, the heroic period of our common existence."

"... And yet it was in this miserable old shed that the best and happiest years of our lives were spent, entirely consecrated to work. I sometimes passed the whole day stirring a mass in ebullition, with an iron rod nearly as big as myself. In the evening I was broken with fatigue. ..."

"In agreement with me Pierre Curie decided to take no material profit from our discovery: in consequence we took out no patent and we have published the results of our research without reserve, as well as the processes of preparation of radium. Moreover, we gave interested persons all the information they requested. This was a great benefit to the radium industry, which was enabled to develop in full liberty, first in France and then abroad, furnishing to scientists and doctors the products they needed. As a matter of fact this industry is still using today, almost without modification, the processes which we pointed out. ..."

Humanity certainly needs practical men, who get the most out of their work, and, without forgetting the general good, safeguard their own interests. But humanity also needs dreamers, for whom the disinterested development of an enterprise is so captivating that it becomes impossible for them to devote their care to their own material profit.

Without the slightest doubt, these dreamers do not deserve wealth, because they do not desire it. Even so, a well-organized society should assure to such workers the efficient means of accomplishing their task, in a life freed from material care and freely consecrated to research.

—Ève Curie, *Madame Curie*, [pp. 169, 204–205, 336]

x-ray tubes, the electrons strike a metal target, typically made of tungsten, instead of a glass wall, and there is no accompanying fluorescence whatsoever.

To be noted here is the enormous development of the experimental apparatus of nuclear physics starting from the photographic plate wrapped in black paper up to the large accelerators, which are extremely complex engineering achievements (Figure 5.93(a,b,c,d)).

A chance event led to the realization that the blackening of the photographic plate actually had nothing to do with the fluorescence of the uranium salt.

We have already observed that it is considered a fortunate circumstance in the history of science when from the work of a scientist one can glean something about the birth of an idea and the false starts made on the way to insight. A case in point would be the fact that Becquerel did not wait until he completed his investigations, but published the half-finished results posthaste. Thus, we learn from Becquerel himself—and not from his memoirs but from his scientific publication—of the crucial moment in his discovery:

> The experiments which I shall report were done with the rays emitted by crystalline crusts of the double sulfate of uranyl and potassium [SO_4(UO) K+H_2O], a substance whose phosphorescence is very vivid and persists for less than 1/100th of a second. The characteristics of the luminous rays emitted by this material have been studied previously by my father, and in the meantime I have had occasion to point out some interesting peculiarities which these luminous rays manifest.
>
> One can confirm very simply that the rays emitted by this substance, when it is exposed to sunlight or to diffuse daylight, pass through not only sheets of black paper but also various metals, for example a plate of aluminum and a thin sheet of copper. ...
>
> I will insist particularly upon the following fact, which seems to me quite important and beyond the phenomena which one could expect to observe: The same crystalline crusts, arranged the same way with respect to the photographic plates, in the same conditions and through the same screens, but sheltered from the excitation of incident rays and kept in darkness, still produce the same photographic images. Here is how I was led to make this observation: among the preceding experiments, some had been prepared on Wednesday the 26th and Thursday the 27th of February, and since the sun was out only intermittently on these days, I kept the apparatuses prepared and returned the cases to the darkness of a bureau drawer, leaving in place the crusts of the uranium salt. Since the sun did not come out in the following days, I developed the photographic plates on the 1st of March, expecting to find the images very weak. Instead the silhouettes appeared with great intensity. I immediately thought that the action had to continue in darkness, and I arranged the following experiment. ...
>
> —Becquerel, "Sur les radiations invisibles émises par les corps phosphorescents" [translated by Carmen Giunta]

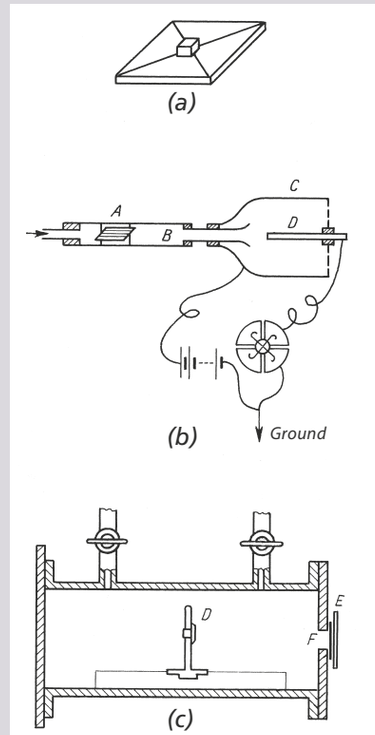
5.4.4 *The Protagonists of the Heroic Age: The Curies and Rutherford*

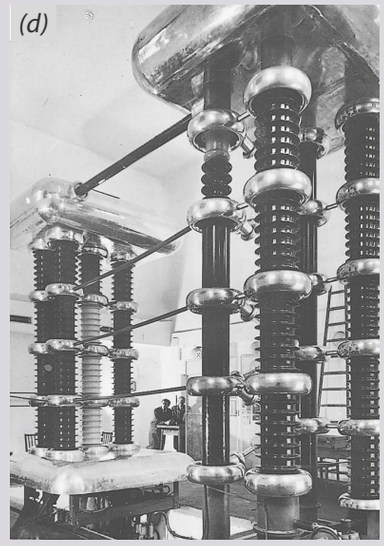
Error and chance played no role in Becquerel's further experiments, which were carried out systematically and with extreme care. In this work he determined that the phenomenon observed did not, in fact, have anything to do with fluorescence and was due entirely to the presence of uranium. Becquerel then observed that the effect of the uranium salt did not depend on its physical or chemical state and that, just like x-rays, the rays, which for a while were called *Becquerel rays*, were capable of ionizing air. However, Becquerel's interest in the rays diminished in the following year, and he turned his attention to the more promising *Zeeman effect*. He entrusted further research to his assistant, Marie Skłodowska-Curie (Figure 5.94). A decisive step was the working out of a method for measuring the intensity of emitted radiation. Marie's husband, Pierre Curie, and his brother, Jacques Curie, built an instrument for her based on the recently discovered piezoelectric effect that made it possible to measure the small currents resulting from the ionization (Figure 5.95, Quotation 5.40).

After it was established that the phenomenon of radioactivity (the word "radioactivity" was first used in 1898 by the Curies) was connected with the element uranium, the question arose whether other elements might exhibit similar properties, in other words, whether there were other naturally occurring radioactive elements. Marie Curie soon discovered that thorium was also radioactive. (She shared the credit for this discovery with G. K. Schmidt.) At this point, Pierre Curie also became involved in research on radioactivity. Only the heroes of classical epics had the persistence and industriousness to triumph over all trials that would be a worthy comparison with the work of the Curies that in 1898 resulted in the discovery of polonium and then radium.

In one of their articles published in 1902, they express their conception of radioactivity as follows:

> In what follows, we would like to present the ideas that served as the foundations of our experiments. Since the beginning of our research, we have always assumed that radioactivity is a *property of the atoms* of matter; this assumption also sufficed to determine the methodology of our research.
>
> Every individual atom of radioactive material represents a permanent source of energy. On the basis of this hypothesis we may draw a number of conclusions without the necessity of describing where the radioactive material obtains its energy.
>
> Experiments over many years have shown that the radiation activity of uranium, thorium, radium, and perhaps actinium as well remains completely unchanged when the radioactive material is returned into the same chemical or physical state and that this activity does not change over time.
>
> If we attempt to determine the source of the energy of radioactivity, we may make a number of assumptions, all of which, however, can be grouped around two general hypotheses: 1. Every radioactive atom has at its disposal the energy that it radiates in the form of potential energy. 2. The radioactive atom is a mechanism that condenses within itself from without at each moment the energy that it radiates. ... According to the hypotheses of the second group, radioactive materials are essentially energy transformers. This energy can come from the heat of the environment, which would have to lead to the material's cooling and represent a violation of the Carnot principle. It can also stem from unknown sources, for example, from rays that we





▲ **Figure 5.93** The development of research techniques in nuclear physics.

(a) A photographic plate wrapped in black paper with a uranium salt crystal lying on top (Becquerel, 1896).

(b) The modern form of the radioactive decay law was inferred from measurements taken with this apparatus (Rutherford, 1900).

(c) Rutherford's setup for "atom smashing" (1919).

(d) In 1932, Cockcroft and Walton were the first to produce a nuclear reaction using an accelerator similar to the one shown in the figure, freeing research from the limitations of having to use naturally radioactive substances as particle sources.

don't know about. It is indeed imaginable that we know very little about the medium that surrounds us, since our knowledge is limited to those phenomena that act immediately on our senses.

—P. Curie and Mme. S. Curie, "Sur les corps radio-actifs"

This opinion is logical, sober, well thought out, and not without imagination. We are quoting it to show that many scientists would rather abandon the second law of thermodynamics than to give up on the immutability of elements. It is possible that the primary reason for this is that Mme. Curie was a chemist. It remained for the physicist Rutherford to reject what was for the chemists a sacred dogma.

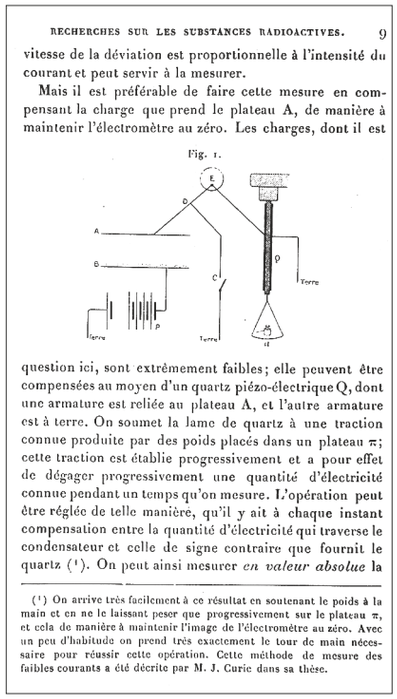
However, let us not get ahead of ourselves. In the meantime, the Curies were not the only ones intensely involved in research on radioactivity, as such work was also going on at many universities in Europe and even in America. Rutherford entered the field with a strong background. In 1896, J. J. Thomson had started to investigate the ionization of gases due to the effect of x-rays. The young Rutherford took part in this research and began studying the ionizing effect of Becquerel



▶ **Figure 5.94** The best-known member of the Curie family, Marie Curie-Skłodowska (1867–1934): born in Warsaw, went to Paris in 1891. Studied physics and mathematics at the Sorbonne. 1895, married Pierre Curie. Assistant to Becquerel, then to her husband. 1906, after the death of her husband, professor.

In 1898, she discovered polonium and radium jointly with her husband. 1910: production of metallic radium; 1903: Nobel Prize in physics jointly with her husband and Becquerel; 1911: sole recipient of the Nobel Prize in chemistry.

Pierre Curie (1858–1906): In addition to his work in radioactivity, made significant contributions to the studies of piezoelectricity (1880) and of the temperature-dependence of magnetic permeability (Curie law, Curie point). With regard to their scientific activity and personal courage, both Curies can serve as models for young scientists. Their daughter Irène Joliot-Curie (1897–1956) studied at the Sorbonne and then worked at the Radium Institute. 1926, married F. Joliot; from this point on, the couple did almost all their research jointly. From 1937, professor at the Sorbonne and from 1948 till her death, director of the Radium Institute. Her husband, Frédéric Joliot-Curie (1900–1958), studied physics and chemistry, receiving a degree in engineering. From 1925, work at the Radium Institute; from 1935, professor, first at the Sorbonne, then at the Collège de France. In 1932, he and his wife came close to discovering the neutron. In 1933, they studied pair generation and annihilation radiation and in 1934 discovered artificial radioactivity. 1939, intensive work on nuclear fission, where they recognized the possibility of the release of nuclear energy in large quantities. Active participation in public life: Joliot-Curie was active in the resistance movement and later became president of the World Council of Peace.



RECHERCHES SUR LES SUBSTANCES RADIOACTIVES. 9

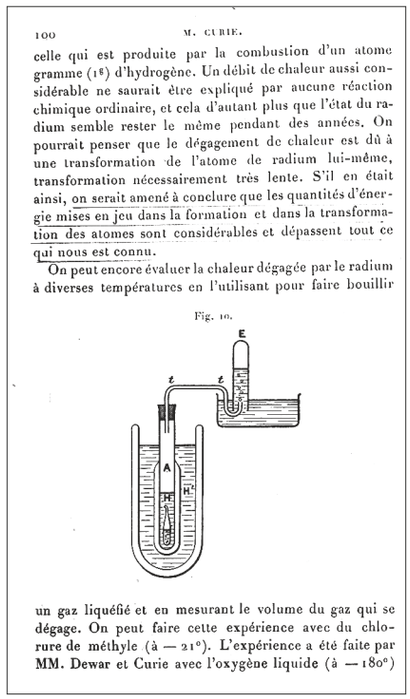
vitesse de la déviation est proportionnelle à l'intensité du courant et peut servir à la mesurer.

Mais il est préférable de faire cette mesure en compensant la charge que prend le plateau A, de manière à maintenir l'électromètre au zéro. Les charges, dont il est

Fig. 1.

question ici, sont extrêmement faibles; elle peuvent être compensées au moyen d'un quartz piézo-électrique Q, dont une armature est reliée au plateau A, et l'autre armature est à terre. On soumet la lame de quartz à une traction connue produite par des poids placés dans un plateau π; cette traction est établie progressivement et a pour effet de dégager progressivement une quantité d'électricité connue pendant un temps qu'on mesure. L'opération peut être réglée de telle manière, qu'il y ait à chaque instant compensation entre la quantité d'électricité qui traverse le condensateur et celle de signe contraire que fournit le quartz (1). On peut ainsi mesurer *en valeur absolue* la

(1) On arrive très facilement à ce résultat en soutenant le poids à la main et en ne le laissant peser que progressivement sur le plateau π, et cela de manière à maintenir l'image de l'électromètre au zéro. Avec un peu d'habitude on prend très exactement le tour de main nécessaire pour réussir cette opération. Cette méthode de mesure des faibles courants a été décrite par M. J. Curie dans sa thèse.



100 M. CURIE.

celle qui est produite par la combustion d'un atome gramme (1g) d'hydrogène. Un débit de chaleur aussi considérable ne saurait être expliqué par aucune réaction chimique ordinaire, et cela d'autant plus que l'état du radium semble rester le même pendant des années. On pourrait penser que le dégagement de chaleur est dû à une transformation de l'atome de radium lui-même, transformation nécessairement très lente. S'il en était ainsi, on serait amené à conclure que les quantités d'énergie mises en jeu dans la formation et dans la transformation des atomes sont considérables et dépassent tout ce qui nous est connu.

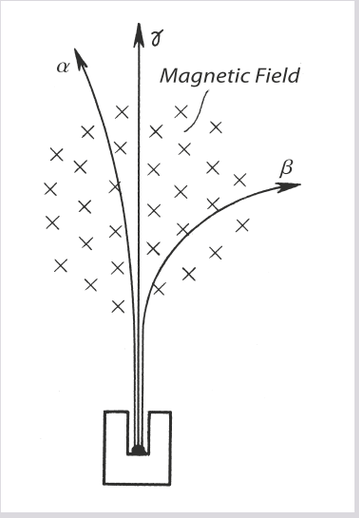
On peut encore évaluer la chaleur dégagée par le radium à diverses températures en l'utilisant pour faire bouillir

Fig. 10.

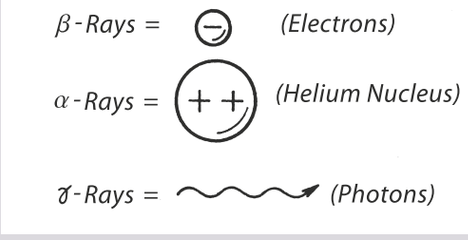
un gaz liquéfié et en mesurant le volume du gaz qui se dégage. On peut faire cette expérience avec du chlorure de méthyle (à — 21°). L'expérience a été faite par MM. Dewar et Curie avec l'oxygène liquide (à — 180°)

◀ **Figure 5.95** Two facsimile pages from M. Curie's book *Recherches sur les substances radioactives*, which appeared in 1905 and contains her habilitation degree thesis. The first page shows an apparatus for measuring the intensity of ionization, built by Pierre Curie and his brother Jacques. It relies on the piezoelectric effect. The second page shows the measurement of heat released in radioactive transformations. The underlined words state that this heat is significantly greater than all previously known heats of reaction.

rays in 1898. He concluded that the radiation is not homogeneous. It has a strongly ionizing component that can nevertheless be absorbed by a thin layer of paper, and even in air its effect extends for only a few centimeters. There is also another component that is responsible for only a very small fraction of the ionization, but whose ability to penetrate is considerably greater. Without any particular justification, Rutherford named these rays for the first letters of the Greek alphabet: alpha and beta rays. We should immediately add that in 1900, Villard discovered a third form of radiation with an even greater power of penetration, the gamma rays (Figure 5.96). Research now turned to the investigation of the nature of these rays. Becquerel, who had returned to this subject, showed that the beta rays are of a corpuscular nature, and carry a negative charge; furthermore, after the measurement of the ratio of their charge and mass, they could be identified with the recently discovered electrons. The identification of alpha rays was not so simple. In 1902, Rutherford already suspected that alpha rays consist of doubly ionized helium atoms, but he was not able to prove this until 1909 (Figures 5.97, 5.98).



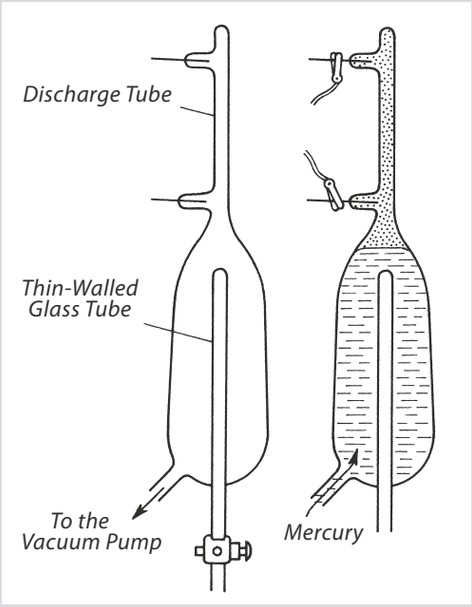
▲ **Figure 5.96** Rutherford and Villard determined that radioactive substances emit three kinds of radiation with differing penetration abilities.



▲ **Figure 5.97** The nature of the rays emitted by radioactive substances was clarified in the first decade of the twentieth century.

The year 1903 saw a great deal of work on the question of the release of energy. Rutherford and Soddy estimated this quantity theoretically, while Pierre Curie and A. Laborde measured it (Quotation 5.41) and found that a gram of radium supplies a constant power of 880 kcal/year from a completely unknown source and, as it seemed then, without any diminution over time.

To recognize that radioactivity was linked to the transmutation of elements, an experimental phenomenon had to be found that would contradict the idea of the Curies quoted above that the strength of radioactivity does not depend on time. Such a suspicion must have already come to them: we read in a footnote of their article that the element polonium that they had just discovered did not seem to exhibit quite the required constancy. Rutherford, too, had reported on a similar observation earlier in 1899. With his colleague Owens, he investigated the radioactivity of thorium applying the same method that had been used in the study of uranium. Owens observed that the measured intensity of the radiation



▲ **Figure 5.98** Rutherford's experiment, which provided a convincing proof that alpha particles are actually helium nuclei. The alpha particles penetrate the thin glass wall of the inner container, which contains the radioactive substance; they are then collected in the outer container as helium atoms and finally compressed in the discharge tube. From the gas-discharge spectrum, one can conclude the presence of helium [Harvard 1970].

Quotation 5.41

We discovered that radium salts continually develop heat.

Namely, a iron-constantan thermocouple one of whose junctions is surrounded by radium-containing barium chloride, the other by pure barium chloride, shows a temperature difference between the two materials.

We used two different receptacles in our experiments; one of them contained 1 g radium-containing barium chloride, of which—by weight—about one-sixth was radium chloride, the rest barium chloride. The junctions of the thermocouple were placed in the middle of each receptacle, where they were completely surrounded by the material. The small receptacles were located—isolated in air—within two identical larger containers, which in turn were located in a third, thermally isolated container. The temperature within this third container was essentially uniform.

Under such conditions, any temperature variations in the environment would have affected the two junctions equally and would not influence the readings of the thermocouple.

Measured in this way, we found a temperature difference of 1.5° between the radium-containing

continued on next page

from thorium changed just *when the door of the laboratory was swung open*. It was quickly realized that air currents played a role. This gave Rutherford the idea to study the radioactivity of the air flowing over the thorium preparation whereupon he discovered that this radioactivity disappears very quickly. Then Rutherford plotted the dependence of the radiation intensity on time. Here for the first time we come across the exponential law of radioactive decay (Figure 5.99). At first, Rutherford called the radioactive gas emitted by thorium into the air an "emanation," and later, to distinguish it from the radium emanation, "thorium emanation." Searching for the origin of this emanation by using chemical methods, Rutherford and Soddy eventually realized that radioactivity was accompanied by a transmutation of the element into a chemically different element. We can see from Quotation 5.42 how carefully the two men formulated their results.

In the first years of the twentieth century, then, researchers had, by and large, arrived at a correct picture of radioactivity. The nature of the radiation was known, and it was also known that alpha rays consist of positively charged ions, with the suspicion being that they are helium ions. The beta rays were identified as electron rays, and finally, it was suspected that the gamma rays were related to x-rays. It was also known that in a given situation, an element emitted either alpha or beta rays, but not both, and in the process, the atoms change chemically as a result of having emitted these rays. The radioactive decay law had been documented and the notion of half-life had been introduced. However, we must not forget that these observations were not yet regarded as phenomena of the atomic nucleus simply because it was not yet known that atoms consist of a central core nucleus and an outer shell. For almost the entire decade that followed, the great problem became how to incorporate into the periodic table the materials that resulted from the complicated transmutation processes. Should they indeed be considered uniform chemically when they behave uniformly in all chemical tests yet they have different radioactive behaviors? The ultimate answer to this question is again due to Soddy, who introduced the notion of isotope of an element and assigned these isotopic variants the identical place in the periodic table as the element itself. Isotopes have the same chemical properties but may differ in other aspects, for example, in their radioactivity. Soddy and Fajans determined the radioactive displacement law that bears their names, according to which in alpha decay, the decay product is to be found two columns to the left in the periodic table from the original element with the atomic weight reduced by four units. In the case of beta decay, the decay product lands one column to the right in the periodic table, and the atomic weight does not change (Figure 5.100).

Thus, it turned out that the principal distinctive characteristic of isotopes, which are chemically identical elements, is that they have different atomic weights, or, as we put it today, different mass numbers. The displacement law of Soddy and Fajans fits perfectly into the framework of ideas about the structure of the atom that meanwhile arose from the work of Rutherford, Moseley, and Bohr (Figure 5.101).

In addition to the scientists in the principal roles, there were many other lesser-known physicists and chemists who participated in this field of research. Among them, we count two who, 30 years later, would achieve renown: Otto Hahn and Lise Meitner. In addition to his work in chemistry, Hahn, together with Schmidt and Meitner, carried out experiments in 1907 on beta radiation and observed the fact, theoretically difficult to explain, that it is emitted from the nucleus not with a fixed energy but with a continuous energy distribution. An explanation was finally given in 1932, by Fermi.

5.4.5 *The Rutherford–Bohr Model Begins to Take Shape*

In Section 4.6.5, we described the scattering experiment that led Rutherford in 1911 to the creation of his atomic model in which a central nucleus is surrounded by orbiting electrons. Figure 4.95 shows Rutherford's experimental setup. The angular distribution of the scattered alpha particles was measured by counting the number of such particles at various angles at which the scintillating screen was placed. As we have seen earlier, in 1903, Crookes found that when alpha particles strike a zinc sulfide screen, they trigger a flash of light, one flash for each alpha particle. One can observe these light flashes under a microscope once the eye has adapted to darkness. They give the impression of stars that light up in a dark sky in a completely irregular pattern. Anyone who witnesses these scintillations for the first time is deeply impressed by this picture of an atomic process in action. It is not surprising—as was noted earlier—that having seen this phenomenon, Mach, toward the end of his life, began to believe in the existence of atoms. Incidentally, Ostwald was also eventually "converted," but by the experiments of Perrin on Brownian motion.

Observing the light flashes is indeed an uplifting experience, but just imagine how tiring it must be to spend a long period of time, perhaps months or even years, conducting experiments in a darkened room, straining one's eyes to count faint flashes of light. The dreary monotony of such work is made more pleasant if one could do it in the company of Rutherford and if the otherwise idle time spent adapting to the dark can be spent in fruitful scientific exchange. Still, it is not surprising that both Rutherford and his colleague Geiger also worked on the problem of mechanizing the counting process. As a result, in 1913 the *Geiger counter* was born, and its perfected variant of 1928, the *Geiger–Müller counter*, would soon become the most important basic tool in all kinds of particle research (Figures 5.102, 5.103).

Quotation 5.41, ***continued***

barium chloride and the pure barium chloride, with the radium-containing salt at the higher temperature.

We also attempted to determine quantitatively the heat quantity developed within a certain period of time by the radium.

A gram of radium develops hourly a heat quantity of roughly 100 small calories.

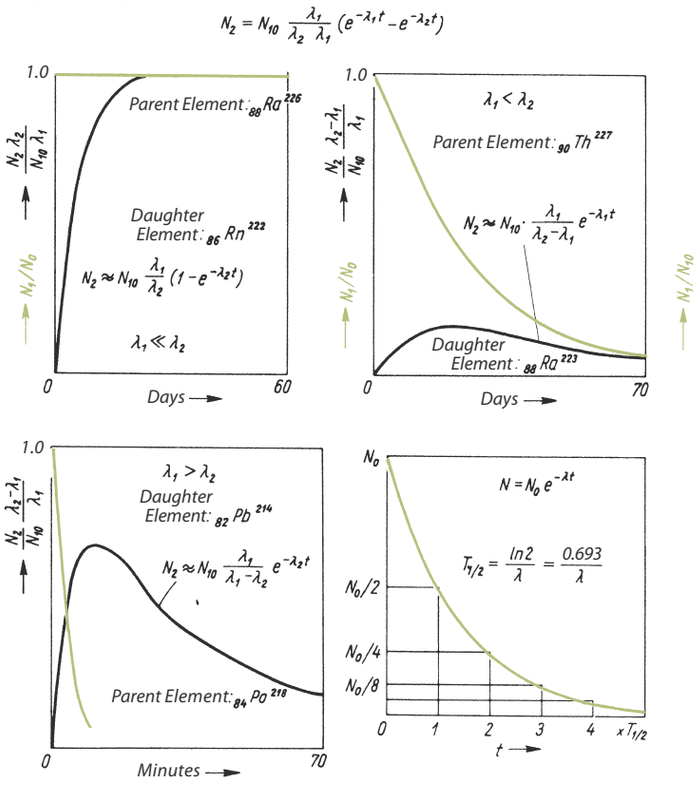
A gram-atom of radium (225 g) would develop 22,500 calories hourly, that it, on the order of the amount yielded in the combustion of a gram-atom of hydrogen in oxygen.

A continual development of such quantities of heat cannot be explained by known chemical reactions. If we seek the source of the heat development in an internal transformation, this must be of a deeply rooted nature and attributable to a transformation of the radium atom. However, such a transformation, if it takes place at all, does so extraordinarily slowly. Indeed, over years, no significant change in the properties of the radium have been observed, and Demarçuay has found no differences in the color spectrum of the same radium chloride preparation, having made observations in intervals of five months.

If the above assumptions turn out to be correct, the amount of energy released in the transformation of the atom is extraordinarily large.

The assumption of a continual transformation of the atom is not the only possible explanation for the heat developed by radium. This heat development can also be explained by saying that radium makes use of an external energy of an unknown nature.

—P. Curie and A. Laborde, "Sur la chaleur dégagée spontanément par les sels de radium," 1903



◀ **Figure 5.99** The exponential law of radioactive decay appeared first in an article written by Rutherford in 1900, in the form $dn/dt = -\lambda n$, $n = Ne^{-\lambda t}$. From the decay constant λ, one can calculate the half-life using the relation $T = (\ln 2)/\lambda = 0.693/\lambda$.

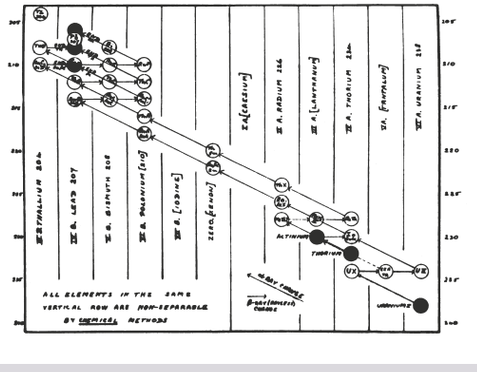
Quotation 5.42

The position is thus reached that radioactivity is at once an atomic phenomenon and the accompaniment of a chemical change in which new kinds of matter are produced. The two considerations force us to the conclusion that radioactivity is a manifestation of subatomic chemical change.

There is not the least evidence for assuming that uranium and thorium are not as homogeneous as any other chemical element, in the ordinary sense of the word, so far as the action of *known* forces is concerned. The idea of the chemical atom in certain cases spontaneously breaking up with evolution of energy is not of itself contrary to anything that is known of the properties of atoms, for the causes that bring about the disruption are not among those that are yet under our control, whereas the universally accepted idea of the stability of the chemical atom is based solely on the knowledge we possess of the forces at our disposal. ...

It is a significant fact that the radioactive elements are all at the end of the periodic table. If we suppose that radium is the missing second higher homologue of barium, then the known examples—uranium, thorium, radium, polonium (bismuth), and lead are the five elements of heaviest atomic weight. Nothing can yet be stated of the mechanism of the changes involved, but whatever view is ultimately adopted it seems not unreasonable to hope that radioactivity affords the means of obtaining information of processes occurring within the chemical atom.

—E. Rutherford and F. Soddy, "The Radioactivity of Thorium Compounds," 1902



▲ **Figure 5.100** A diagram from the 1913 article by Soddy illustrating the displacement law.

With the discovery of isotopes, an old idea was revived: Is it possible that the nucleus of every element is made up of hydrogen nuclei—essentially as it was envisioned almost one hundred years earlier by William Prout (1785–1850)? As we know, the argument against this hypothesis was that the atomic weights of some elements differ considerably from the predicted near-integer values. However, if an element is a mixture of isotopes, then the weight can naturally assume any value between two integers. The existence of radioactive isotopes was experimentally proven; but could there also be stable isotopes? In 1913, J. J. Thomson used his parabola method to investigate the gas neon, whose atomic weight, at 20.2, deviated considerably from an integer value. He in fact found a parabola that corresponded to the atomic weight 22; however, he was not convinced that this indicated the element neon itself, but suspected rather the presence of a neon compound. His colleague Aston, however, took the isotope idea seriously and tried to enrich this isotope using the diffusion method, a process that would later in history become extremely important. The attempt succeeded, and after the enrichment, a significantly more intense line appeared in the mass spectrometer without any difference in the optical spectrum of the gas from the know lines of neon.

In the following five years, Aston undertook a close examination of the periodic table and was able to establish the *Aston whole-number rule*, according to which all atomic weights are very close to integers.

To review: In 1913, the physicists' concept of the atom was as follows: The atom consists of a central part, the atomic nucleus, and a number of electrons orbiting the nucleus, to which the term *atomic shell* or *electron shell* is usually applied. In the neutral atom, the number of electrons agrees with the number of the atom in the periodic table, the atomic number. The atomic number is usually denoted by Z, and so the positive charge of the atom's nucleus is $+Ze$, where e is the magnitude of the charge of an electron. If we call the nearest integer to the atomic weight the mass number (whose usual symbol is A), then the atomic nucleus of mass A must contain precisely A hydrogen nuclei if we insist on thinking of the nuclei of all the elements as being composed of hydrogen nuclei. The total charge on this nucleus would then be not $+Ze$, but $+Ae$. However, we may imagine joining an electron to each of the $A - Z$ extra protons in order to compensate for the excess positive charge. In this case, the nucleus must consist of Z protons and an additional $A - Z$ electron–proton pairs. (We note here that in 1920, Rutherford proposed the name *proton* for the nucleus of a hydrogen atom.)

From this point on, nuclear physics and the physics of the electron shell split into distinct branches of physics. Both chemical processes and spectra are determined by the number and arrangement of the outer electrons, and radioactive phenomena are associated with a change in the structure of the nucleus.

The idea that there must be proton–electron pairs in the nucleus was very attractive: after all, in beta decay electrons are emitted from the nucleus. Today we know that, as compelling as it seemed at the time, this argument is false. At the same time, we can understand how revolutionary it must have been when, two decades later, Fermi dared to assert that the atomic nucleus *contains no electrons*, but that these electrons *come into being* only during the decay process, just as we do not think that photons are components of the atomic shell but that they are created during the transition from one energy state to another.

Rutherford had already suspected, and he expressed the opinion in 1920, that the connection between a proton and an electron in the nucleus must be of quite a different character than that of a hydrogen atom: much tighter, so that the

proton–electron pair forms a single neutral particle. Rutherford therefore suspected the existence of the neutron, even though experiments aimed at direct detection of such a particle were all unsuccessful. As we shall see presently, it was not until 1932 that Chadwick was able to discover the free neutron through the interpretation of a series of experiments in quite a different direction.

5.4.6 *The First Artificial Nuclear Transformation*

In the two decades following the year 1913, *Bohr's quantum theory* and then quantum mechanics were in the limelight. But Rutherford was not idle either, and in 1919, he succeeded in performing the first transmutation of a chemical element. At this time he was still using very simple experimental setups. Making use of alpha rays emitted by a naturally radioactive substance as projectiles, he found that helium nuclei of the alpha rays sometimes collide with a nitrogen nucleus on their paths through air, knocking a hydrogen nucleus, that is, a proton, out of it (Figure 5.93(c)). It was only in 1923 that Blackett succeeded in capturing this event photographically in a *cloud chamber* (Figure 5.104), which proved definitively that the helium nucleus not only knocks a proton out of the nitrogen atom but is absorbed by it in the process. Rutherford thereby brought about the following nuclear reaction:

$$_2\mathrm{He}^4 + {}_7\mathrm{N}^{14} = {}_8\mathrm{O}^{17} + {}_1\mathrm{H}^1.$$

Furthermore, he pointed out the important fact that the energy of the emerging proton is greater than the energy of the helium nucleus that triggers the reaction and that therefore there is a release of energy during the nuclear transformation.

This discovery unleashed a worldwide search for nuclear reactions in a variety of substances. Methods were developed to reduce the energy of the radiation from the naturally radioactive sources using various moderators in a well-defined way in order to plot the "resonance functions," from which information could be gleaned about the internal energy states of the nucleus. These experiments were not in and of themselves particularly spectacular; however, a large quantity of data related to the structure of the atomic nucleus was diligently gathered by scientists, who compensated for the simplicity of their equipment with experimental inventiveness.

5.4.7 *Quantum Mechanics Can Be Applied to Nuclear Phenomena*

As we have seen, quantum mechanics was created primarily as an interpretation of phenomena relating to the atomic shell, and it is there that it enjoyed its greatest success. However, it was not clear whether it is at all applicable to nuclear phenomena as well. It is for this reason that Gamow's theory of alpha decay was of such importance. Gamow envisaged the alpha particles as being located inside the nucleus at specific energy levels as though in a potential well (Figure 5.105). According to the rules of classical mechanics, a particle with such energy cannot escape the nucleus, whereas the *tunnel effect* of quantum mechanics does allow for an escape from the nucleus with a certain probability. The probability of "tunneling out" depends on the energy of the particle as well as on the height and width of the potential barrier. Thus, Gamow was able to interpret to a certain degree of approximation the *Geiger–Nuttall rule*, which gives a relationship between the half-life of a radioactive substance and the energy of the emitted particles (Figure 5.106).

The uncertainty principle of quantum mechanics, however, confounded nuclear physicists. For an electron to remain inside of the nucleus, its location must, of course,

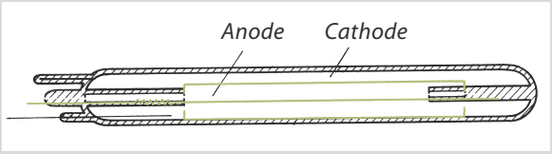
1920

Shell :	*Z Electrons*
Nucleus :	*A Protons*
	A - Z Electrons

1932

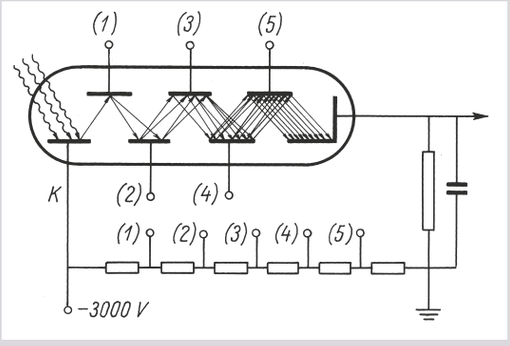
Shell :	*Z Electrons*
Nucleus :	*Z Protons*
	N = A – Z Neutrons

▲ **Figure 5.101** The structure of an atom of atomic number *Z* and mass number *A* as understood by physicists in 1920 and in 1932 (as well as today).



▲ **Figure 5.102** The Geiger counter. A voltage of about 1000 V is maintained between the anode and the cathode. The ionization that occurs from the passage of a particle suffices to trigger a discharge in the gas filling the tube, making possible the recording of the event.

Hans Geiger (1882–1945) made another important contribution relating the half-life of radioactive materials to the energies of the alpha particles emitted (Geiger–Nuttall rule).



▲ **Figure 5.103** Particle counters of another type (scintillation counter) register light flashes triggered by charged particles when they strike certain solid or liquid materials, for example, a sodium iodide crystal that has been activated with thallium (NaI[Th]).

Light emission from very fast particles in transparent media (see Figures 5.150, 5.151, 5.152) can be detected with the Cherenkov counter.

The light quanta from either type of counter are sensed by a photoemissive cathode, which is an electrode alloyed, for example, with cesium and antimony, from which electrons are released. The number of these photoelectrons is multiplied in an electrode system using secondary electron emission. The idea and early (1938) application of electron multipliers for particle counting are due to Zoltán Bay (1900–1992). Bay also did pioneering work in radar astronomy (demonstration of reflected radar signals from the Moon's surface, 1946).

Quotation 5.43

A few months after his Bakerian Lecture of June 1920, in which he first mentioned what had been in his mind for some time, the possible existence of a neutral particle formed by the close combination of a proton and an electron, Rutherford invited me to join him in following up the experiments on the artificial disintegration of nitrogen which he had made in Manchester.

There were a number of reasons for this invitation, so welcome to me. Among them was that I had made some improvements in the technique of counting scintillations, better optical arrangements and a strict discipline; but also he wanted someone to talk to, to while away the tedium of working in darkness.

continued on next page

be limited to the region defined by the dimensions of the nucleus. The relationship

$$\Delta p \cdot \Delta q \approx h$$

then yields such a great fluctuation in the momentum and therefore in the energy that at those energy levels the electrons would certainly escape the nucleus. Therefore, by the uncertainty principle, inside of the nucleus there can be no electrons. As we have already mentioned, this contradiction was resolved with the discovery of the neutron and Fermi*'s theory of beta decay*.

5.4.8 *Predicted by Rutherford, Found by Chadwick: The Neutron*

The year 1932 was an outstanding year for nuclear physics and an outstanding year for the Cavendish laboratory. It is this laboratory, under the direction of Rutherford, that we have to thank for the two discoveries of this year that were to determine the subsequent direction of nuclear physics: the discovery of the neutron by Chadwick and the successful transmutation of nuclei using artificially accelerated particles by Cockcroft and Walton. As we shall soon see, the year 1932 also witnessed a significant advance in the physics of elementary particles: The first antiparticle, the positron, was discovered by Carl David Anderson, and at about the same time, Fermi created his theory of beta decay, introducing the notion of weak interaction, and predicting the existence of a new elementary particle, the neutrino.

If we consider the immediate antecedents of the discovery of the neutron, we might get the impression that Chadwick was merely the fortunate harvester of the fruits of the arduous work of German and French researchers. That, however, was not the case. We have noted that Rutherford had suspected as early as 1920—notably in discussions with Chadwick—that there must exist a neutral particle similar in mass to the proton, namely, the neutron (Quotation 5.43). From this time onward, both Rutherford and Chadwick were continually thinking about possible experiments to demonstrate the existence of the neutron. Chadwick himself lavished considerable efforts on this work. Of special interest is one of their ideas, according to which protons and electrons would somehow unite spontaneously, making a neutron, with a release of energy in the form of extremely penetrating gamma radiation. The analogy to the atom itself was rather plausible: If an electron in a hydrogen atom jumps from an outer orbit to an inner one, closer to the nucleus, it emits a light quantum. To explain the neutron, they assumed that the electron moves to an orbit with a very small radius, roughly the size of the nuclear radius. Rutherford and Chadwick then conjectured that in the surroundings of a material containing a large number of hydrogen atoms, the radiation emitted by this transition must be detectable. As they saw it, the neutrons arising in this way could be split again by cosmic radiation back into protons and electrons. However, Chadwick searched in vain for the hypothetical gamma radiation, and today we know that, by energy considerations alone, the conjectured phenomenon is impossible. In fact, the opposite is the case: It is the free neutron that splits spontaneously into a proton and an electron.

Let us now return to the events leading up to the discovery. As we have seen, in a number of laboratories around the world, alpha particles emitted by naturally radioactive substances were used as projectiles to trigger nuclear reactions. In 1930, Bothe and Geiger, as well as Webster, noted a peculiarity when they bombarded beryllium with alpha particles. They discovered that as a reaction to the bombardment, radiation of such great penetrating power was emitted that it was able to pass through even thick lead plates. It was not ionizing, meaning that it carried

no charge, so they naturally thought that they were dealing with extremely hard gamma radiation. Several laboratories now began to investigate these mysterious gamma rays. In 1932, Irène and Frédéric Joliot-Curie found that these rays would cause high-energy protons to be emitted from hydrogen-containing matter (Figure 5.107). From the energy of the photons, they calculated the energy that a gamma quantum would have to have in order to be able to transmit so much energy to a proton through the *Compton effect*. They arrived at a value of 50 MeV. But this value was much too great to be believed relative to the energies that had been seen before in nuclear physics. Even Rutherford, who always took his experimental results and those of other researchers seriously, on reading this article, said simply, "I don't believe it." Chadwick, who at the back of his mind was always searching for a proof of the existence of the neutron and in whose research beryllium had also played a role, soon realized that the contradictory results of Bothe and colleagues and those of the Joliot-Curies made sense if one assumed that neutrons are emitted from the beryllium under the influence of alpha rays according to the reaction

$$_2\mathrm{He}^4 + {}_4\mathrm{Be}^9 = {}_6\mathrm{C}^{12} + {}_0\mathrm{n}^1.$$

Figure 5.108 shows the entire article in which Chadwick announces the discovery of the neutron.

Since the beginning of the 1930s, the use of particles emitted from naturally radioactive elements for bombardment was gradually replaced by particles accelerated by "atom smashers." The different kinds of particles that can be accelerated, the ability to regulate their energies, and the great intensity of the particle beams contributed greatly to the rapid increase in our knowledge of the nuclear reactions and the structure of the nucleus. After the first experiments by Cockcroft and Walton, Lawrence (Figures 5.109, 5.110) brought his cyclotron into service. An invention of Van de Graaff (1901–1967), the Van de Graaff generator, became a precision instrument of the nuclear physicist because with its help one could produce a highly stable beam of particles that was homogeneous with respect to energy of those particles.

5.4.9 *Nuclear Structure and Nuclear Models*

The discovery of the neutron also aroused the interest of theoretical physicists in the problems of nuclear physics. By 1932, Heisenberg and Tamm had already worked out a theory of the atomic nucleus consisting of protons and neutrons (Figure 5.111). The most important conclusion not requiring difficult mathematics concerns the energy of the nucleus. Let us assume that the nucleus is composed of Z protons and $N = A - Z$ neutrons. Then the total mass of all the particles in the *free* state forming the nucleus is given by

$$Zm_\mathrm{p} + (A - Z)m_\mathrm{n},$$

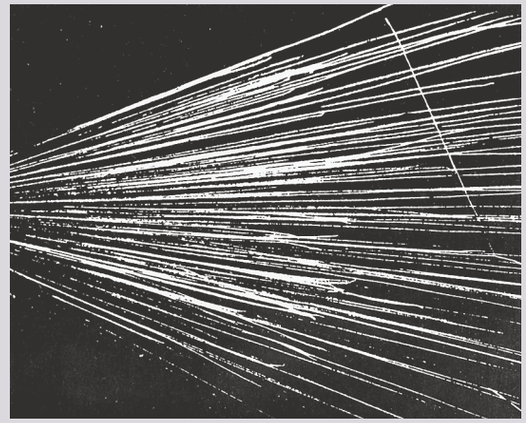
where m_p is the mass of a proton, and m_n is the mass of a neutron. If one measures the precise mass of the nucleus—and at this time, the methods of mass spectroscopy were so refined that the mass of the atomic nucleus could be determined with great precision—then we find that in every case, the total mass of the free particles is greater than the mass of a nucleus comprising these particles. Denoting the mass

Quotation 5.43, ***continued***

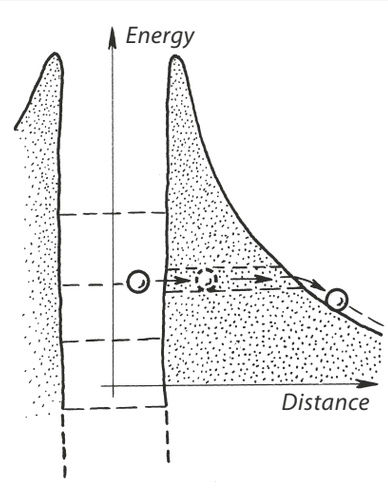
It was during the periods of waiting to begin counting that he expounded to me at length his views on the problems of nuclear structure, and in particular on the difficulty in seeing how complex nuclei could possibly build up if the only elementary particles available were the proton and the electron, and the need therefore to invoke the aid of the neutron. He freely admitted that much of this was pure speculation, and, being averse to speculation without some basis of experiment, he seldom mentioned these matters except in private discussion. ...

He had not abandoned the idea, and he had completely converted me. From time to time in the course of the following years, sometimes together, sometimes myself alone, we made experiments to find evidence of the neutron, both its formation and its emission from atomic nuclei. [Some were] so desperate, so far-fetched as to belong to the days of alchemy.

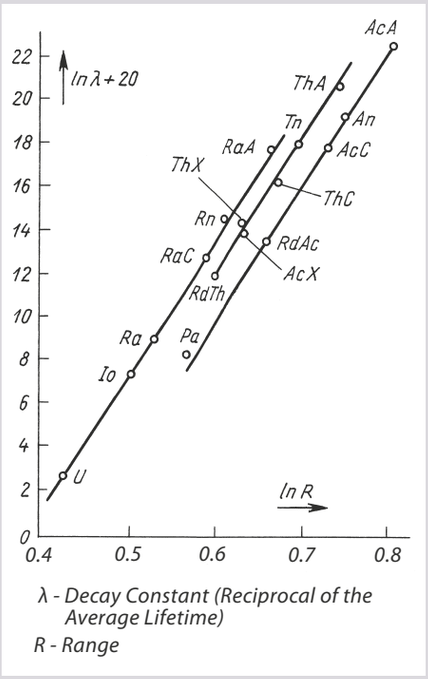
—Sir James Chadwick, "Some Personal Notes on the Search for the Neutron" [p. 159]



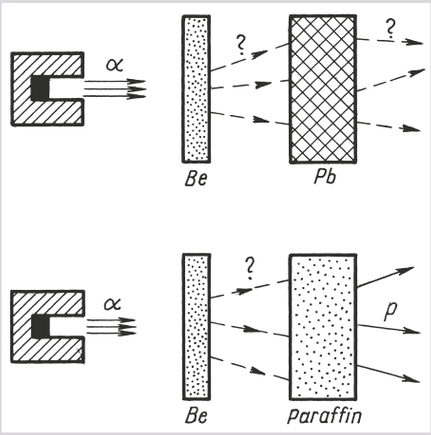
▲ **Figure 5.104** Cloud chamber photograph of the first artificial nuclear transformation.



◀ **Figure 5.105** Gamow's theory of radioactive alpha decay.



▲ **Figure 5.106** The Geiger–Nuttall rule.



▲ **Figure 5.107** Schematic representation of the experiments of Bothe and of the Joliot-Curies [Harvard 1970].

of the nucleus by M, we obtain the difference

$$\Delta M = Zm_p + (A - Z)m_n - M,$$

which is called the *mass defect*. The mass defect represents the binding energy of the nucleus (Figure 5.112). We should imagine that a stable nucleus can form only when the energy of the protons and neutrons in the nucleus is less than their energy when they are not bound together. In the formation of the atomic nucleus, the excess energy leaves in some form, for example, as radiation or as the kinetic energy of a particle ejected from the nucleus. However, according to the equivalence theorem of mass and energy, the excess energy is proportional to the mass defect; that is, $\Delta E = \Delta Mc^2$. If we wish to decompose the nucleus back into its components, then this energy has to be reintroduced. This energy, given per nucleon, is plotted in Figures 5.113 and 5.114. We can see that its value is roughly constant over a broad portion of the periodic table, with a value of about 8 MeV, while at the beginning and end of the table, the binding energy per nucleon is smaller. If we know the masses of the particles that take part in a nuclear reaction, we obtain very useful information about the energy released in the reaction or that must be supplied to it. If we note further that the conservation of momentum law—although here appearing in a relativistic form—must also hold for nuclear reactions, then we can say even more about the characteristics of a variety of nuclear reactions. In Figure 5.115(a,b,c,d), we find information about some nuclear reactions of particular interest.

It is considerably more difficult to determine the probability that a particular nuclear reaction will occur. In the case of the neutron, the situation is relatively simple because it has no charge, so the repulsive electric forces from the nucleus will have no effect on it. We can say approximately that the longer a neutron stays in the proximity of a nucleus, the larger is the probability of a reaction caused by it. Put a bit more concretely, the probability of a nuclear reaction is roughly inversely proportional to the velocity of the neutron.

For charged particles, the situation is more complex. Confining ourselves to a very simple description, we may follow Gamow's train of thought in reverse and assume that a positively charged particle can penetrate the nucleus even if it does not have enough energy to overcome the *Coulomb barrier*. The probability of this "tunneling through" the wall will, of course, decrease if low-energy particles are used to bombard the nucleus. To express these relationships, experimental physicists introduced the concept of a reaction cross section (Figure 5.116), an abstract surface indicating the size of the target that a given nucleus offers a given projectile for a given reaction type. Experimental physicists measure the reaction cross section as a function of the energy of the excited particles. The task of the theory is to interpret the functional dependence under the assumption of various models of the nucleus.

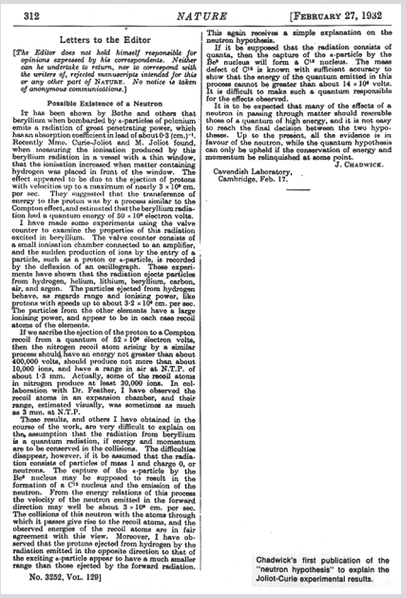
As we have already seen, energy considerations play a major role in the interpretation of phenomena in the field of nuclear physics, with great reliance being placed on mass–energy equivalence. Today this connection is universally regarded as a natural consequence of the special theory of relativity. But if we look back at how this equivalence first appeared, or how we prove this equivalence even today in the context of special relativity, we see that we are investigating how elastic collisions look from different inertial coordinate systems, and we naturally insist that the description of the phenomena should be the same in each. Although Einstein surmised that the equivalence of mass and energy would play a role in the explanation of the phenomena of radioactivity, it was still a qualitatively new insight that

the difference in mass of bodies measured at rest, and also macroscopically shown to be different, may be due to the fact that some energy has left the system. As a concrete example, when a radium atom emits an alpha particle, the combined masses of the resulting radon atom and the emitted helium at rest will be less than the mass of the original radium nucleus. Or, to take a more modern example: If two deuterium nuclei join to form a helium atom, the mass of the helium nucleus is less than the sum of the original masses of the two deuterium nuclei. It was a long time before physicists became fully cognizant that there is a connection between the exact values of the masses of the atomic nuclei and the binding energies.

It is said that this idea was first expressed by LANGEVIN. In any case, in 1927, ASTON noted, but only as an interesting fact, that whenever he found a nucleus that can be regarded as the union of two other nuclei—that is, its atomic number is the sum of the atomic numbers, and its mass number is the sum of the mass numbers, of the two constituents—then the mass of the purportedly compound nucleus is less than the sum of the individual masses (Figure 5.117). Of course, it was impossible to clarify the actual numerical values as long as the composition of the nucleus was unknown, that is, prior to the discovery of the neutron. The main point here is just that the connection between energy–mass equivalence and nuclear physics was by no means self-evident to physicists at the time.

The discovery of the neutron immediately provided the answer to the question, What does the nucleus consist of? The experimentally derived curves for the binding energies and for the reaction cross sections also clarified the changes in the nucleus that can happen spontaneously or that can be caused to happen by bombardment. The task now fell to the theoretical physicists to interpret these important curves. However, it is not enough to know the components of the nucleus, but also how these components are arranged; in other words, the question of the internal structure of the nucleus now arose. Quantum mechanics had already been able to answer very precisely the question of the structure of the atomic shell, or, simply stated, the arrangement of the electrons in an atom. Could the same principles be used for the investigation of the structure of the nucleus? There were many signs at the beginning of the 1930s that the general laws of quantum mechanics should be applicable to the nucleus as well—a belief that to this day has not been refuted. The difficulties as compared to investigating the shell are twofold: First, one would have to understand the forces between the nucleons, the components of the nucleus. Next, even with this knowledge of the nuclear forces, one would not get very far. Namely, in the atomic shell, the electrons move in the dominating force field of the positively charged nucleus, and their mutual interaction can be seen as a small correction factor. In contrast, in the nucleus, there is no distinguished central force field, and the individual particles are therefore subject to strong mutual interactions. Despite these difficulties, attempts were made to apply the schema taken from the physics of the atomic shell to the nucleus.

In 1934, ELSASSER created the first *independent-particle model*, the name referring to the major assumption of the model, that inside the nucleus an average potential is created to which all nucleons contribute and within which they can move independently of one another. The Schrödinger equation is solvable for such a model, and the solutions can be characterized in the well-known way by four quantum numbers. We can then assign various energy levels to the nucleons by way of the *Pauli principle*. In this way, we arrive at the simplest *nuclear shell model*, in which the nucleons are arranged in shells in which the energies of the nucleons within a given shell are more or less equal (Figure 5.118).



312 NATURE [FEBRUARY 27, 1932

Letters to the Editor

[The Editor does not hold himself responsible for opinions expressed by his correspondents. Neither can he undertake to return, nor to correspond with the writers of, rejected manuscripts intended for this or any other part of NATURE*. No notice is taken of anonymous communications.]*

Possible Existence of a Neutron

It has been shown by Bothe and others that beryllium when bombarded by α-particles of polonium emits a radiation of great penetrating power, which has an absorption coefficient in lead of about 0·3 (cm.)⁻¹. Recently Mme. Curie-Joliot and M. Joliot found, when measuring the ionization produced by this beryllium radiation in a vessel with a thin window, that the ionization increased when matter containing hydrogen was placed in front of the window. The effect appeared to be due to the ejection of protons with velocities up to a maximum of nearly 3×10^9 cm. per sec. They suggested that the transference of energy to the proton was by a process similar to the Compton effect, and estimated that the beryllium radiation had a quantum energy of 50×10^6 electron volts.

I have made some experiments using the valve counter to examine the properties of this radiation excited in beryllium. The valve counter consists of a small ionization chamber connected to an amplifier, and the sudden production of ions by the entry of a particle, such as a proton or α-particle, is recorded by the deflexion of an oscillograph. These experiments have shown that the radiation ejects particles from hydrogen, helium, lithium, beryllium, carbon, air, and argon. The particles ejected from hydrogen behave, as regards range and ionizing power, like protons with speeds up to about $3{\cdot}2 \times 10^9$ cm. per sec. The particles from the other elements have a large ionizing power, and appear to be in each case recoil atoms of the elements.

If we ascribe the ejection of the proton to a Compton recoil from a quantum of 52×10^6 electron volts, then the nitrogen recoil atom arising by a similar process should have an energy not greater than about 400,000 volts, should produce not more than about 10,000 ions, and have a range in air at N.T.P. of about 1·3 mm. Actually, some of the recoil atoms in nitrogen produce at least 30,000 ions. In collaboration with Dr. Feather, I have observed the recoil atoms in an expansion chamber, and their range, estimated visually, was sometimes as much as 2 mm. at N.T.P.

Those results, and others I have obtained in the course of the work, are very difficult to explain on the assumption that the radiation from beryllium is a quantum radiation, if energy and momentum are to be conserved in the collisions. The difficulties disappear, however, if it be assumed that the radiation consists of particles of mass 1 and charge 0, or neutrons. The capture of the α-particle by the Be^9 nucleus may be supposed to result in the formation of a C^{12} nucleus and the emission of the neutron. From the energy relations of this process the velocity of the neutron emitted in the forward direction may well be about 3×10^9 cm. per sec. The collisions of this neutron with the atoms through which it passes give rise to the recoil atoms, and the observed energies of the recoil atoms are in fair agreement with this view. Moreover, I have observed that the protons ejected from hydrogen by the radiation emitted in the opposite direction to that of the exciting α-particle appear to have a much smaller range than those ejected by the forward radiation. This again receives a simple explanation on the neutron hypothesis.

If it be supposed that the radiation consists of quanta, then the capture of the α-particle by the Be^9 nucleus will form a C^{13} nucleus. The mass defect of C^{13} is known with sufficient accuracy to show that the energy of the quantum emitted in this process cannot be greater than about 14×10^6 volts. It is difficult to make such a quantum responsible for the effects observed.

It is to be expected that many of the effects of a neutron in passing through matter should resemble those of a quantum of high energy, and it is not easy to reach the final decision between the two hypotheses. Up to the present, all the evidence is in favour of the neutron, while the quantum hypothesis can only be upheld if the conservation of energy and momentum be relinquished at some point.

J. CHADWICK.

Cavendish Laboratory,
Cambridge, Feb. 17.

No. 3252, Vol. 129]

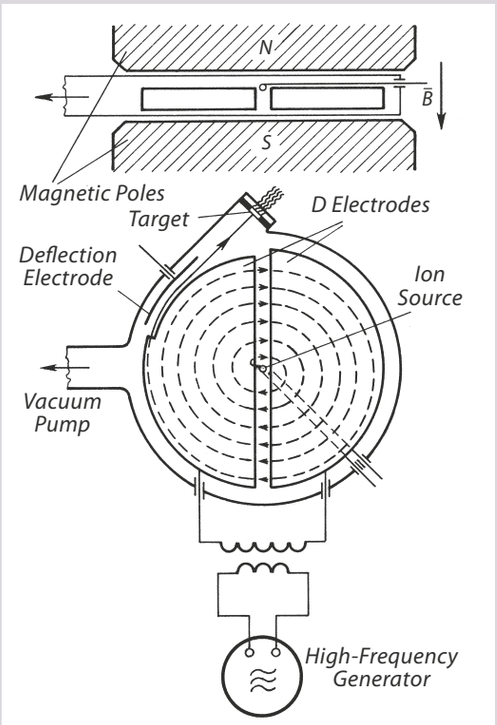
Chadwick's first publication of the "neutron hypothesis" to explain the Joliot-Curie experimental results.

▲ **Figure 5.108** CHADWICK's article on the discovery of the neutron.



▲ **Figure 5.109** ERNEST ORLANDO LAWRENCE (1901–1958): Inventor of the cyclotron. Studies at Yale University, later, from 1930, research and teaching at the University of California, Berkeley. From 1936, director of the radiation laboratory. During the Second World War, worked out the method of electromagnetic separation for the U-235 isotope. The element number 103, lawrencium, is named in his honor.

The idea of the cyclotron in 1929. By 1930, experimental model, an important tool for modern physics research by 1932. In 1937, it was fully understood that particle energies achievable with cyclotrons working with a constant frequency are greatly limited. Moving beyond the limit (about 100 MeV) required a new idea, phase stabilization, which was developed independently by V. I. VEKSLER (USSR, 1944) and E. M. MCMILLAN (USA, 1945).



▲ **Figure 5.110** Diagram of a cyclotron. A high-frequency (≈ 15 MHz) electrical voltage is placed on the hollow (D-shaped) electrodes.

Particles (e.g., protons, deuterons) are emitted by an ion source and accelerated in the electric field between the two D electrodes; they then follow a semicircular path within a D electrode in the strong magnetic field created by the electromagnet. Within the electrode the electric field is excluded and has no effect. After traveling the half circle, the particles again reach the gap between the two D electrodes, but by that time the direction of the electric field has reversed and the particles are further accelerated, reaching a semicircle of greater radius and in this way continue to gain energy. This process repeats until the radius of the path becomes large enough and the particles escape the magnetic field. Here they are deflected from an auxiliary electrode in the direction of the target, where the appropriate reaction can take place.

The cyclotron principle is based on the fact that the orbital period of an electrically charged particle in a homogeneous magnetic field is independent of its velocity, and hence its energy, provided that its mass m may be considered as constant, that is, $v \ll c$. Given this period $T = 2\pi m/qB$, the frequency of the field generator can be adjusted to be $f = 1/T = qB/2\pi m$. Therefore, it is possible to adjust the frequency of the acceleration voltage so that it matches the orbital period of the particles.

Even though this model provides a distinct argument for the existence of discrete energy levels within the nucleus, and at the same time hints at numbers for closed shells that characterize stable atoms, this is merely a hint; in reality, the outstandingly stable atoms have completely different numbers of nucleons from those of the model and correspondingly different closed shells. These so-called *magic numbers* of nucleons are

2, 8, 20, 50, 82, 126.

These magic numbers or rather the associated shell structures were first explained only by a modified shell model. Maria Goeppert-Mayer, in 1949, and Jensen, in 1950, created such models by assuming that there exists a very strong interaction between the orbital momentum and the spin of the particles moving independently of one another in the average potential. In this way, the shell model was revived in a somewhat modified form close to 20 years after its first appearance.

Meanwhile, however, another nuclear model, the *liquid drop model*, had greater success in the theoretical interpretation of the experimental observations. This model was based on the recognition that, according to experimental data, the binding energy per particle is more or less constant, so that the total binding energy of a nucleus grows roughly linearly with the number of nucleons it contains. In addition, the radius of the nucleus grows in proportion to the cube root of the number of nuclei, which means that all atomic nuclei have roughly the same density. A water droplet has very similar properties. Its density is constant, independent of its size, and every particle of the droplet is bound to it with approximately the same energy.

To illustrate the usefulness of this model as well as its limitations, let us examine the path by which Weizsäcker derived the curve for the mass defect (Figures 5.112, 5.112, 5.114).

Setting out from the liquid drop model, the following semiempirical formula for the precise mass of the nucleus can be given:

$$M = Zm_{\mathrm{p}} + (A - Z)m_{\mathrm{n}} - \left[a_1 A - a_2 A^{2/3} - a_3 \frac{Z^2}{A^{1/3}} - a_4 \frac{(A/2 - Z)^2}{A} \pm \Omega \right].$$

The expression in brackets is the mass defect, which gives the binding energy of the nucleus. The individual terms in this expression can be interpreted thus: The term $a_1 A$ gives the cohesion energy (volume energy), which grows in proportion to the number of nucleons.

The nucleons at the nuclear surface do not participate in the binding with the same intensity as other nucleons because they have no interaction partners outward. This effect, which weakens the cohesion energy, is proportional to the surface area (surface energy). As we have seen, the radius of the nucleus is proportional to $A^{1/3}$, so that the surface area and surface energy become proportional to $A^{2/3}$.

The *Coulomb energy* also appears with a negative sign; that is, it, too, weakens the total binding energy. It is proportional to the square of the charge and inversely proportional to the nuclear radius, that is, to $A^{1/3}$.

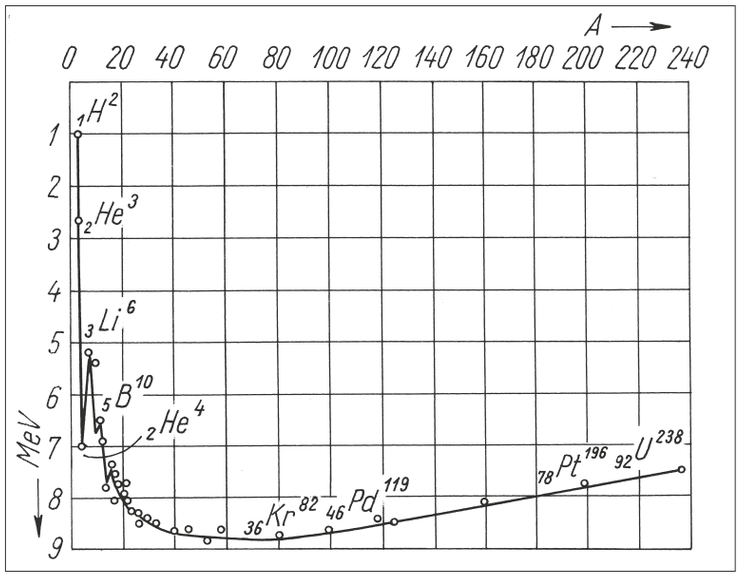
Finally, the fourth term takes into account the empirical fact that the stability curve of the nucleus lies more or less near the line $A/2 = Z$, meaning that for large values of the difference $A/2 - Z$, the nucleus becomes unstable. However, this instability must diminish for large values of A, since the empirical stability curve for heavy atoms deviates from the line $A/2 = Z$, which is why we have to divide by A.

Finally, the term Ω is a correction for an even or odd number of nucleons.

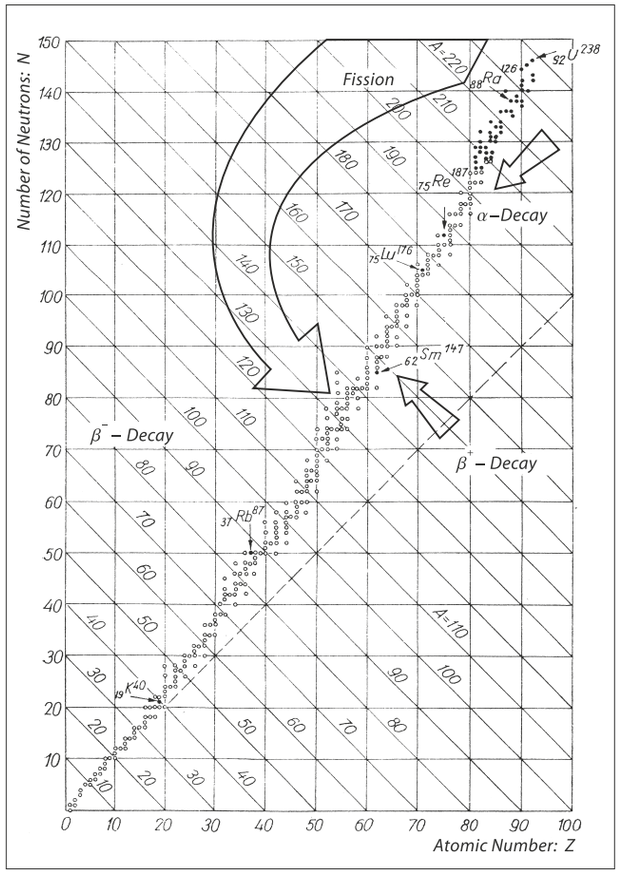
In sum, we can establish that for larger nuclei, the relative contribution of the surface energy decreases, since the ratio of the surface area to volume becomes smaller; but the Coulomb energy increases. The reason for the instability of heavy atoms is exactly this increase. In contrast, the smaller binding energy of light nuclei is due to their relatively larger surface energy.

Bohr essentially relied on the liquid drop model in creating his theory of compound nuclei for nuclear reactions around 1936—a theory that was to be of enormous importance in theoretical nuclear physics. This theory offers the following

description of the course of a nuclear reaction (Figures 5.119, 5.120): The projectile, which can be a charged particle or a neutron, strikes the nucleus and transfers its energy via interaction processes to all the nucleons therein, giving rise to an intermediate highly excited new nucleus. In the second phase of the nuclear reaction, this common energy is eventually transmitted to a single particle, which then exits the nucleus. This second step can take place in many ways, but in the excitation of a given nucleus by a particle of given energy, the probabilities for specific branches or channels out of all possible outcomes can be well determined.



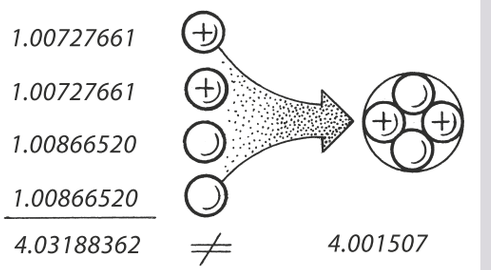
◀ **Figure 5.113** The mass defect, or binding energy, per nucleon as a function of the mass number. This quantity is the measure of the forces holding the nucleons (protons and neutrons, the building blocks of the nucleus) together.



◀ **Figure 5.114** The naturally occurring nuclei. The various nuclei for the same chemical element, meaning the isotopes of that element, lie on a vertical line (Z = constant). The nuclei lying on a horizontal line (N = constant) are called isotones. The nuclei with the same mass number (A = constant) lie on 45° lines (drawn in the figure). The solid circles indicate nuclei that are naturally radioactive.

	Proton	Neutron	Electron
	⊕	◯	⊖
m : Mass (u)	1.00727661	1.00866520	5.485930×10^{-4}
m : Mass (kg)	1.672614×10^{-27}	1.674920×10^{-27}	9.109558×10^{-31}
e : Charge (Coulombs)	$+1.602191 \times 10^{-19}$	0	$-1.6021917 \times 10^{-19}$
s : Spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Magnetic Moment	2.79255 Nuclear Magneton	−1.91315 Nuclear Magneton	1.0011596 Bohr Magneton
	1.41052×10^{-26} Joule/Tesla	-0.96624×10^{-28} Joule/Tesla	9.28485×10^{-24} Joule/Tesla

▲ **Figure 5.111** The building blocks of the atom: proton, neutron (the two nucleons), and electron.



$$\Delta m = 0.0303766 = 28.298 \text{ MeV}$$

$$\frac{\Delta m}{A} = \frac{28.298}{4} = 7.074 \text{ MeV}$$

▲ **Figure 5.112** The mass defect as a measure of the binding energy.

(a) $_1H^1 + {}_3Li^7 = {}_2He^4 + {}_2He^4 + 17.5\ \text{MeV}$

(b) $N^{14} + n = C^{14} + p$

$C^{14} \rightarrow N^{14} + e^- + 0.156\ \text{MeV}$

(c) $_2He^4 + {}_{13}Al^{27} = {}_{15}P^{30} + {}_0n^1$

$_{15}P^{30} \rightarrow {}_{14}Si^{30} + e^+$

(d) $_{10}Ne^{22} + {}_{94}Pu^{242} = {}_{103}Lr^{260} + {}_1H^1 + 3\,{}_0n^1$

▲ **Figure 5.115** **(a)** The first nuclear reaction that was accomplished with artificially accelerated particles (Cockcroft and Walton, 1932).

(b) Due to cosmic radiation, some nitrogen atoms in the air are transformed into radioactive carbon, which decays with a half-life of 5730 ± 40 years. The activity of radioactive carbon incorporated into plant life, and thereafter into animals, decays after the death of the organisms according to the law of exponential decay. By comparing the amount of radioactive carbon left with what is in the atmosphere, one can determine the approximate date of the organism's death (Willard Frank Libby [1908–1980], radiocarbon dating, 1948).

(c) A current physics problem: Up to what atomic number can transuranic elements be created? In laboratories under the directorship of G. Seaborg and A. Ghiorso, and of G. N. Flyorov, the number 106 was achieved at the beginning of the 1970s.

Some transuranic elements can be found in nature. Plutonium-244, the one with the longest half-life ($T_{1/2} = 8 \times 10^7$ years), occurs in trace amounts in minerals. Neptunium-237 ($T_{1/2} = 2.14 \times 10^6$ years) and plutonium-239 ($T_{1/2} = 2.4 \times 10^4$ years) can be found in uranium ore.

Through neutron-capture reactions one can get as far as fermium ($Z = 100$); by bombarding heavy elements with light or medium-heavy ions, it has been possible to create super-heavy atoms well above atomic number $Z = 115$, but these are very short lived due to spontaneous alpha decay and spontaneous fission. Theoretically, there is still a possibility that an island of increased stability will be found in the neighborhood of certain higher "magic" numbers.

▶ **Figure 5.116** A historic document: The fission cross section of the U-235 nucleus as a function of the energy of the bombarding neutrons. The figure plots the values that were measured independently in three countries: USSR, Great Britain (labeled Har in the figure), and the USA (labeled Col and KAPL). These results were top secret until 1955 because knowledge of this information is essential for designing atomic bombs as well as for nuclear power plants [United Nations 1955, p. 285].

The cross section of a reaction is defined in a way that represents the measurement method: From the incident beam, N particles per second fall on a unit area of the target material of density n atoms per cm^3. While passing through the distance dx, the beam encounters $1 \cdot n\ dx$ target atoms. Due to nuclear (or other) reactions, the intensity of the beam will decrease. The value of this decrease is $-dN =$

continued on next page

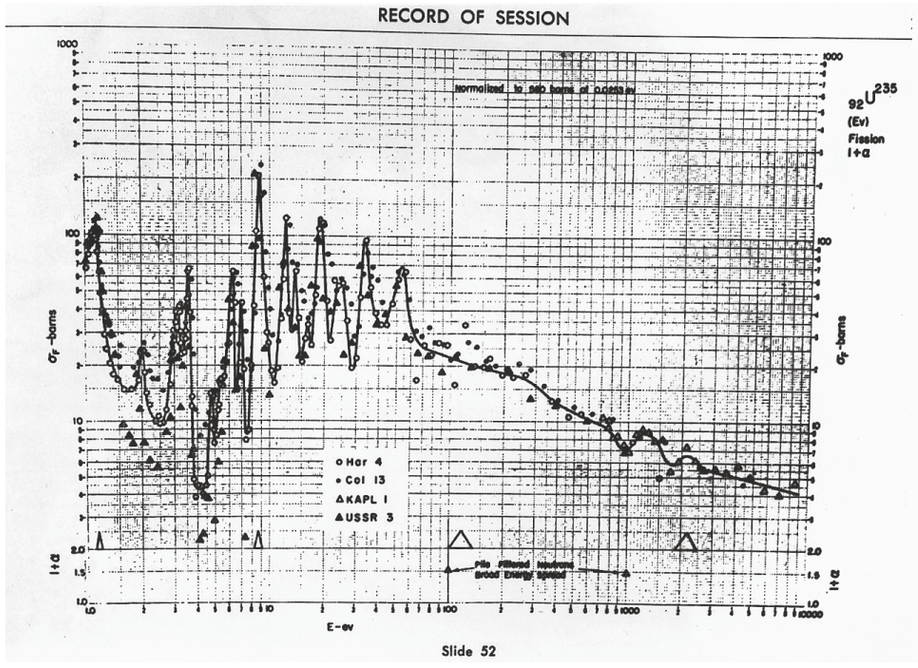
The liquid drop model had its greatest success in the interpretation of nuclear fission. But more about that later.

Because both the independent-particle model—the shell model—and the liquid drop model were able to correctly describe some of the observed phenomena but completely failed in other cases, there was understandably a drive to create a unified model combining the advantages of both. This problem was solved in 1952 by Bohr and Mottelson with their collective model. We must add here that it is Aage Bohr, Niels Bohr's son, to whom we are referring. In the history of physics, five families have produced more than one Nobel Laureate: Marie and Pierre Curie, as well as their daughter and son-in-law; father and son Bragg; the two Bohrs, father and son; the two Thomsons, also father and son; and finally, the Siegbahns, father and son.

Among nuclear models, we must also mention the optical model, offered in 1954 by Feshbach, Porter, and Weisskopf, which is designed primarily for quantitative descriptions of nuclear reactions. Also called the crystal ball model, it likens the nucleus to a sphere composed of a material with a determined index of refraction, but this index also has an imaginary part, so the sphere not only scatters the incident waves, but absorbs them as well. This absorption corresponds to the capture of particles.

5.4.10 *Nuclear Fission: Experimental Evidence, Theoretical Doubt*

The most significant discovery of the 1930s—nuclear fission—merits a separate discussion. Here we are again witnessing a curious turn of events occasioned by the splitting of a heavy atomic nucleus. This was such an unexpected phenomenon that after its discovery, it became clear only in hindsight that it had already been observed five years earlier, but had been given a completely different interpretation. We saw something similar in the discovery of the neutron, but while Chadwick, a member of Rutherford's school, was chomping at the bit, so to speak, to meet the theoretically predicted particle in experimental reality, in the case of fission researchers practically had to be forced into recognizing the reality of the phenomenon by the persuasive power of experiments that excluded the possibility of any other inter-



pretation. Even then they yielded to the evidence only reluctantly. However, Bohr senior, upon hearing of nuclear fission, slapped his forehead and exclaimed, "Oh, what idiots we all have been! Oh, but this is wonderful! This is just as it must be!"

After the discovery of the neutron, a worldwide search for nuclear reactions triggered by neutrons began. Fermi's laboratory in Rome quickly acquired a strong international reputation in the field (Figure 5.121). The other major center of neutron physics research during these years was in Paris, where Joliot-Curie's group was working. By bombarding heavy atomic nuclei, especially uranium, the heaviest nucleus occurring in nature, they expected to get interesting results, specifically the creation of new elements that do not appear in nature, the so-called transuranium elements. Adding a neutron to a uranium nucleus would produce a nucleus with an extra neutron, and one could expect the nucleus to free itself of the redundant neutron via beta decay. But beta decay creates an additional proton, which entails the shifting of the element one place to the right in the periodic table, and so from the element with atomic number 92 (uranium), one would end up with an element with atomic number 93. In 1934, Fermi reported that he had been able to produce an element with atomic number greater than 92. Although he formulated his claim very cautiously, the Italian press jumped in to report that Fermi had presented the queen with a bottle full of this new material created by man. The news was coolly received by the scientific community. In the very same year, 1934, a German chemist, Ida Noddack, sharply attacked Fermi's paper with the argument that Fermi had failed to investigate the associated radioactivity thoroughly and had ignored a whole host of other factors as well. For us, however, this is not the important part. The interesting lines in Noddack's article are the following—and let us bear in mind that we are in the year 1934:

> One may just as well assume that in this new form of nuclear fission by neutrons, considerably more additional "nuclear reactions" take place than have thus far been observed from the action of proton and α rays on nuclei. With the latter type of rays one finds only nuclear transformations giving off electrons, protons, and helium nuclei, which in the case of heavy elements changes the mass of the bombarded nucleus only a little, since the elements that arise are close neighbors. It is conceivable that in bombarding heavy nuclei with neutrons, these nuclei would break up into several *larger* pieces, which would be isotopes of known elements but not neighbors of the bombarded elements.
>
> —Ida Noddack, "Über das Element 93" (On the Element 93) [p. 654]

Here for the first time we encounter the idea of nuclear fission, and it is perhaps not an accident that this idea was broached by a chemist and not by a physicist. We met an analogous phenomenon in our discussion of naturally occurring radioactivity, except that there the possibility of the transmutation of an element was considered first by a physicist rather than by a chemist.

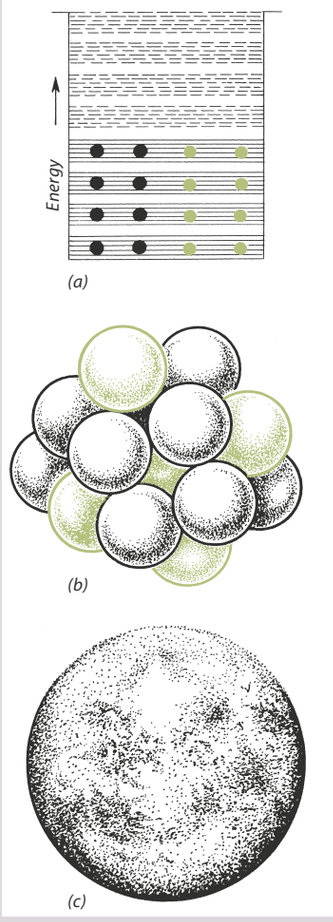
A correct interpretation of new observations requires both genius and scientific knowledge, but scientific knowledge can also be a shackle that someone working in a neighboring discipline might be able to discard. In the case of nuclear fission, it was the chemists who were in such a favored position. The physicists already had very definite ideas—perhaps not in detail, but in a qualitative form—regarding the processes of nuclear reactions. To a physicist it seemed completely impossible that a particle with a charge greater than that of a helium nucleus could be ejected from an atom via the tunnel effect. Therefore, it is understandable that Noddack's ideas had absolutely no influence among physicists.

Figure 5.116 *continued*

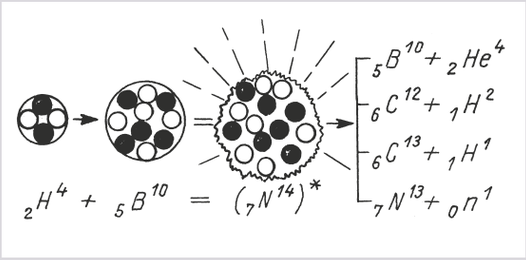
$N\sigma n\,dx$, where the constant of proportionality σ is the cross section of a single target atom with respect to the given reaction (or the net total of other interactions). We can interpret this equation as meaning that each target atom appears as an area σ for each incident particle. Every 1 cm^2 of the incident beam, sees the "effective" target area $(n\,dx)\sigma$, and this is also the probability that one beam particle will react with one target atom in the volume $(1 \cdot dx)$. Therefore, the total number of reactions will be N $(n \cdot 1 \cdot dx)\sigma$. The unit of the cross section: 1 barn = 10^{-24} cm^2.

$$
\begin{aligned}
&{}_1H^1 + {}_3Li^7 \\
&= 1.007825 + 7.016005 \\
&= {}_2He^4 + {}_2He^4 + 17.5\ \text{MeV} \\
&= 4.002604 + 4.002604 + 0.018622
\end{aligned}
$$

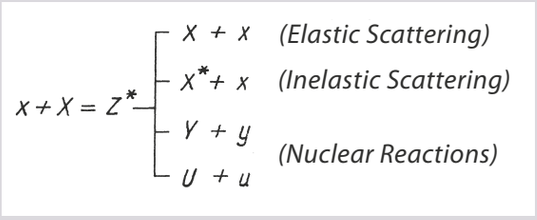
▲ **Figure 5.117** Energy balance of a nuclear transformation.



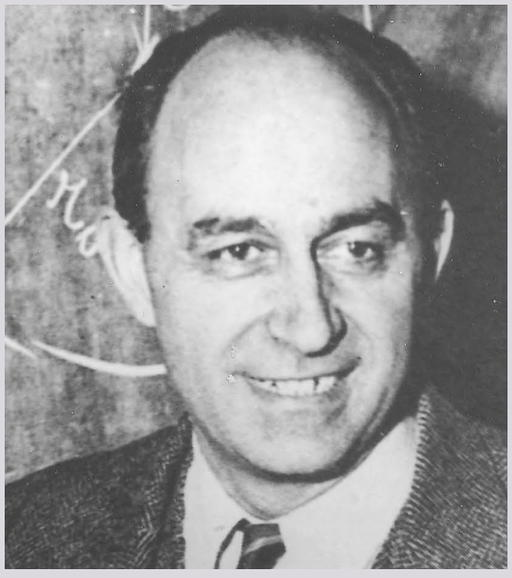
◀ **Figure 5.118** Models of the nucleus. **(a)** Simple shell model. **(b)** Liquid drop model. **(c)** Optical model (after Peierls).



▲ **Figure 5.119** Theory of the intermediate nucleus.



▲ **Figure 5.120** Schematic representation of various types of nuclear reactions.



▲ **Figure 5.121** Enrico Fermi (1901–1954): Studies in physics in Rome; after a brief period of study in Germany, appointed professor of theoretical physics in Rome. There he developed the theory of what is now called the Fermi–Dirac statistics. 1934, complete theory of beta decay. Also in this year, radioactivity artificially caused by neutron bombardment. 1938, after receiving the Nobel Prize, emigrated to United States, where he worked until his death. His best-known achievement: On December 2, 1942, the first sustained chain reaction took place under his direction. Intensive participation in the atomic bomb program. He gave a possible theoretical interpretation for the origin of cosmic radiation (1949), and in particle physics, his research on resonances is of significance.

Fermi is the only physicist of the twentieth century who made outstanding contributions both as a theoretician and as an experimentalist.

continued on next page

The next dramatic episode on the road to the discovery of nuclear fission is linked with the names Joliot-Curie and Savić. In order to find the reaction products of bombardment of uranium with neutrons, it was necessary to separate the material that was responsible for the radioactivity resulting from the bombardment by chemical means. The usual problems with such separation were that one was dealing with truly miniscule quantities of material and also that elements from the same column of the periodic table can be separated from each other only with great difficulty because it is precisely their chemical similarity that caused their ordering into the same column in the first place. In 1938 Joliot-Curie and Savić found that the resulting material was very similar to lanthanum. Lanthanum is one of the rare earth elements, and its atomic number is 57. In the periodic table, it is located in the same column as actinium and is therefore related to that element chemically. Joliot-Curie and Savić were convinced that actinium was responsible for the radioactivity, assuming that as the result of irradiation, an element that is close to uranium in atomic weight would be created. Actinium, whose atomic number is 89, seemed to fit the bill perfectly. In order to separate the actinium from the lanthanum, they used the well-known technique of fractional crystallization, with the lanthanum crystallizing out first because of its lesser atomic weight. For Joliot-Curie and Savić, it was surprising that the radioactivity seemed to follow the lanthanum. In the end it looked to them as if they had managed to separate the radioactivity from the lanthanum; but today we know that the principal portion of the radioactivity they measured must have been in fact due to the lanthanum, and what they had therefore observed was a fission product. Because they were not ready even to consider the possibility of nuclear fission, they gave a different interpretation to their observations in spite of the evidence staring them in the face.

It is perhaps worth mentioning at this point that this was the second time that the Joliot-Curies were observing phenomena of the kind whose eventual correct interpretation would turn out to be of groundbreaking importance. Let us recall that they also laid the foundations for the discovery of the neutron by detecting the high-energy protons resulting from neutron bombardment, though they attributed the effect to unrealistically high energy gamma radiation. In the end, they serve as a telling example of the fact that, in comparison to genius and hard work, the whim of chance plays only a subordinate role. In 1935, the couple deservedly were awarded the Nobel Prize for their work on radioactive isotopes.

From 1935 on, intensive work on the decay products of uranium bombarded with neutrons was also underway in Berlin, under the direction of Otto Hahn (Figure 5.122). We have previously encountered his name and that of his colleague Lise Meitner. Later, Strassmann joined this group, which carried out their research primarily with techniques of chemistry. In the end it was this group that first determined, excluding all other possibilities, that when a uranium nucleus is bombarded with neutrons, a completely new type of reaction occurs, in the course of which moderately heavy nuclei result. The publication of this crucial finding occurred in January 1939 in the journal *Naturwissenschaften* under the title "Über den Nachweis und das Verhalten der bei der Bestrahlung des Urans mittels Neutronen entstehenden Erdalkalimetalle" (On the detection and characteristics of the alkaline earth metals formed by irradiation of uranium with neutrons; Figure 5.123). How greatly the experimenters themselves were surprised by the result of their work is shown by the following quotation from their article:

However, we now come to several recent investigations that we publish with some hesitation on account of the strange results. To remove all doubt about the chemical nature of the initial term of the series precipitated with barium and called a "radium isotope," we have performed fractionated crystallization and fractionated precipitation with the active barium salts in the well-known manner for enrichment (or depletion) of radium in barium salts.

Barium bromide strongly enriches radium in fractionated crystallization, and barium chromate even more under a not too rapid appearance of small crystals. Barium chloride enriches less strongly than bromide, while barium carbonate depletes somewhat. Corresponding experiments that we have made with our disintegration product of purified active barium preparations *are uniformly negative: The activity remained uniformly distributed among all barium fractions*, at least to the extent that we could measure within the not very small possibilities of experimental error. Then a couple of fractionation experiments were carried out with the radium isotope ThX and with the radium isotope MsTh. They proceeded exactly as one would have expected from all previous experience with radium. Then the "indicator method" was applied to a mixture of purified long-lived "RaIV" with pure radium-free $MsTh_1$: The mixture with barium bromide as substrate was fractionally crystallized. *The $MsTh_1$ was enriched, the "RaIV" was not*, its activity remaining the same for the same barium content of the fractions. We arrive at the following conclusion: Our "radium isotopes" have the properties of barium; as chemists, we must say that the new material is not radium, but barium; for any elements other than radium or barium are out of the question.

In an article that appeared one month later (February 1939), this hesitation had already gone by the wayside, and the existence of fission products was now maintained with certainty. The moment of discovery of nuclear fission was marked by the publication of the articles by Hahn and Strassmann. They then had no further role, because research into fission moved away from chemists to physicists, theoreticians, as well as experimentalists. The significance of these articles was that they freed the physicists from their inhibitions. Scientists all over the world now took up this area of research, and by no later than the first half of 1939, they were able to clarify all of the phenomena related to uranium fission and provide a theoretical explanation as well. But before we move on to this topic, we would like to say a few more words about the apparatus, or more accurately, the experimental setup that was used to achieve these results.

To bombard uranium nuclei with neutrons, of course, requires neutrons. The discovery of the neutron was made possible by the fact that in bombarding beryllium with alpha particles, neutrons are emitted according to the reaction $_4Be^9 + _2He^4 = _6C^{12} + _0n^1$. Both Fermi and Hahn obtained their alpha particles from the alpha radiation of naturally radioactive substances. In Figure 5.124, we can see the neutron source used by Fermi. It consists only of a glass test tube containing beryllium (in powdered form) and the gas radon, used as a source of alpha rays. Fermi obtained the radon gas by "tapping" into a container of radium being kept in a hospital for therapeutic purposes. Radon is the decay product of radium, and was at first called "radium emanation."

Figure 5.12 shows an astonishingly simple experimental table used by Hahn and Strassmann, which is today displayed in a museum. Here as well, a radon–beryllium mixture, held in a test tube, served as a neutron source. We also see simple laboratory glassware used for fractional distillation and crystallization.

Figure 5.121 *continued*

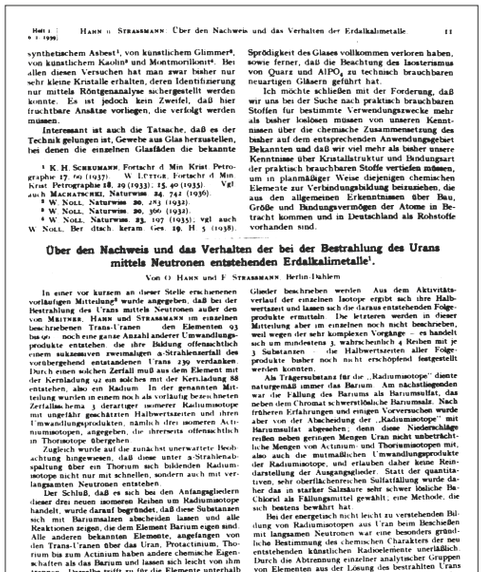
Fermi's importance can be appreciated from the number of different areas of physics where his name appears; we have already mentioned the Fermi–Dirac statistics. In nuclear physics, the fermi is a unit of length, equal to 10^{-13} cm. There is also the Fermi energy: for a system of particles obeying the Fermi–Dirac statistics, the energy level up to which all states are occupied at temperature $T = 0$ K. Fermi surface: the surface in the quasi-momentum space that separates at $T = 0$ K the occupied states from the unoccupied states. Fermium (Fm): the element with atomic number 100 and mass number 244 to 258, discovered in 1953.

Fermion (Fermi particle): Particles with half-integer spin; these include all baryons (protons, neutrons, hyperons), leptons (electrons, muons, neutrinos), as well as particles that are compounds of an odd number of fermions (nuclei, atoms, molecules). The Pauli principle and Fermi–Dirac statistics hold for fermions.

Spin and magnetic moment are considered characteristic quantities of elementary particles just like charge and mass. Measured in units of $h/2\pi$, spin can be integral or half-integral. The former are the bosons; they obey the Bose–Einstein statistics, and their state function is symmetric. The latter are fermions; for them, the Fermi–Dirac statistics hold, and their Ψ function is antisymmetric.



▲ **Figure 5.122** Otto Hahn (1879–1968) and Fritz Strassmann (1902–1980), discoverers of uranium fission, at their workbench.



Heft 1. 6. 1. 1939. HAHN u. STRASSMANN: Über den Nachweis und das Verhalten der Erdalkalimetalle. 11

synthetischem Asbest[1], von künstlichem Glimmer[2], von künstlichem Kaolin[3] und Montmorillonit[4]. Bei allen diesen Versuchen hat man zwar bisher nur sehr kleine Kristalle erhalten, deren Identifizierung nur mittels Röntgenanalyse sichergestellt werden konnte. Es ist jedoch kein Zweifel, daß hier fruchtbare Ansätze vorliegen, die verfolgt werden müssen.

Interessant ist auch die Tatsache, daß es der Technik gelungen ist, Gewebe aus Glas herzustellen, bei denen die einzelnen Glasfäden die bekannte

[1] K. H. SCHRAMMEN, Fortschr. d. Min. Krist. Petrographie 17, 69 (1933). W. LETTOW, Fortschr. d. Min. Krist. Petrographie 18, 89 (1934); 19, 15 (1935). Vgl. auch MACHATSCHKI, Naturwiss. 24, 741 (1936).
[2] W. NOLL, Naturwiss. 20, 283 (1932).
[3] W. NOLL, Naturwiss. 20, 360 (1932).
[4] W. NOLL, Naturwiss. 23, 197 (1935); vgl. auch W. NOLL, Ber. dtsch. keram. Ges. 18, H. 5 (1938).

Über den Nachweis und das Verhalten der bei der Bestrahlung des Urans mittels Neutronen entstehenden Erdalkalimetalle[1].

Von O. HAHN und F. STRASSMANN, Berlin-Dahlem.

In einer vor kurzem an dieser Stelle erschienenen vorläufigen Mitteilung[2] wurde angegeben, daß bei der Bestrahlung des Urans durch Neutronen außer den von MEITNER, HAHN und STRASSMANN in einzelnen beschriebenen Trans-Uranen — den Elementen 93 bis 96 — noch eine ganze Anzahl anderer Umwandlungsprodukte entstehen, die ihre Bildung offensichtlich einem sukzessiven zweimaligen α-Strahlenzerfall des vorübergehend entstandenen Trans 239 verdanken. Durch einen solchen Zerfall muß aus dem Element mit der Kernladung 92 ein solches mit der Kernladung 88 entstehen, also ein Radium. In der genannten Mitteilung wurden in einem noch als vorläufig bezeichneten Zerfallsschema 3 derartiger isomerer Radiumisotope mit ungefähr geschätzten Halbwertszeiten und ihren Umwandlungsprodukten, nämlich drei isomeren Actiniumisotopen, angegeben, die ihrerseits offensichtlich in Thorisotope übergehen.

Zugleich wurde auf die zunächst unerwartete Beobachtung hingewiesen, daß diese unter α-Strahlenabspaltung über ein Thorium auch bislang bekannte Radiumisotope nicht nur mit schnellen, sondern auch mit verlangsamten Neutronen entstehen.

Der Schluß, daß es sich bei den Anfangsgliedern dieser drei neuen isomeren Reihen um Radiumisotope handelt, wurde darauf begründet, daß diese Substanzen sich mit Bariumsalzen abscheiden lassen und alle Reaktionen zeigen, die dem Element Barium eigen sind. Alle anderen bekannten Elemente, angefangen von den Trans-Uranen über das Uran, Protactinium, Thorium bis zum Actinium haben andere chemische Eigenschaften als das Barium und lassen sich leicht von ihm trennen. Dasselbe trifft zu für die Elemente unterhalb Radium, also etwa Wismut, Blei, Polonium, Eka-Iridium.

Es bleibt also, wenn man das Barium selbst außer Betracht läßt, nur das Radium übrig.

Im folgenden soll kurz die Abscheidung der Isotopengemisches und die Gewinnung der einzelnen Glieder beschrieben werden. Aus dem Aktivitätsverlauf der einzelnen Isotope ergibt sich ihre Halbwertszeit und lassen sich die daraus entstehenden Folgeprodukte ermitteln. Die letzteren werden in dieser Mitteilung aber im einzelnen noch nicht beschrieben, weil wegen der sehr komplexen Vorgänge — es handelt sich um mindestens 3, wahrscheinlich 4 Reihen mit je 3 Substanzen — die Halbwertszeiten aller Folgeprodukte bisher noch nicht einschränkend festgestellt werden konnten.

Als Trägersubstanz für die „Radiumisotope" diente naturgemäß immer das Barium. Am nächstliegenden war die Fällung des Bariums als Bariumsulfat, das neben dem Chromat schwerstlösliche Bariumsalz. Nach früheren Erfahrungen und entgegen Versuchen wurde aber von der Abscheidung der „Radiumisotope" mit Bariumsulfat abgesehen, denn diese Niederschläge reißen neben geringen Mengen Uran nicht unbeträchtliche Mengen von Actinium- und Thorimsotopen mit, also auch die mutmaßlichen Umwandlungsprodukte der Radiumisotope, und erhalten daher keine Reindarstellung der Ausgangsglieder. Statt der quantitativen, sehr oberflächentreuen Sulfatfällung wurde daher das in starker Salzsäure sehr schwer lösliche Barium-Chlorid als Fällungsmittel gewählt; eine Methode, die sich bestens bewährt hat.

Bei der energetisch nicht leicht zu verstehenden Bildung von Radiumisotopen aus Uran beim Beschießen mit langsamen Neutronen war eine besonders gründliche Bestimmung des chemischen Charakters der neu entstehenden, künstlichen Radioelemente unerläßlich. Durch die Abtrennung einzelner analytischer Gruppen von Elementen aus der Lösung des bestrahlten Urans wurde außer der großen Gruppe der Transurane eine Aktivität stets bei den Erdalkalien (Trägersubstanz Ba), den seltenen Erden (Trägersubstanz La) und bei Elementen rechts der Bariumfällungen, die offensichtlich der Gruppe der vierten Gruppe des Periodischen Systems (Trägersubstanz Zr) gefunden. Eingehender untersucht wurden zunächst die Bariumfällungen, die offensichtlich die Anfangsglieder der beobachteten isomeren Reihen enthalten. Es soll gezeigt werden, daß Transurane, Uran, Protactinium, Thorium und Actinium sich stets leicht und vollständig von der mit Barium ausfallenden Aktivität trennen lassen.

1. Zu diesem Zweck wurden aus einem bestrahlten Uran mittels Schwefelwasserstoff die Transurane mit Platinsulfid zusammen abgeschieden und in Königswasser gelöst. Aus dieser Lösung wurde Bariumchlorid mit Salzsäure gefällt. Aus dem Filtrat des Bariumniederschlages wurde das Platin nochmals mit Schwefelwasserstoff gefällt. Das Bariumchlorid war inaktiv, das Platinsulfid hatte noch eine Aktivität von ∾ 500 Teilchen/Minute. Entsprechende Versuche mit den langlebigeren Transuranen hatten das gleiche Ergebnis.

2. Eine Fällung von Bariumchlorid aus 10 g nichtbestrahltem Uranylnitrat, das im Gleichgewicht mit $UX_1 + UX_2$ (Thor- und Protactiniumisotope) war und eine Aktivität von ∾ 400000 Teilchen/Minute hat, zeigte eine Aktivität von ∾ 14 Teilchen/Minute, war also ebenfalls praktisch inaktiv; d. h. weder Uran, noch Protactinium, noch Thorium fallen mit dem kristallisierenden Bariumchlorid aus.

3. Schließlich wurde noch aus der Lösung eines Actiniumpräparates ($MsTh_2$) von ∾ 2500 Teilchen pro Minute ein Bariumchloridniederschlag gefällt, der ∾ 3 Teilchen/Minute aufwies, also praktisch ebenfalls inaktiv war.

In ähnlicher Weise wurden die aus bestrahltem Uran gefällten stark aktiven Bariumchloridniederschläge sorgfältig geprüft, doch waren Sulfidniederschläge aus neutraler, schwach essigsaurer oder schwach mineralsaurer Lösung des aktiven Bariums praktisch inaktiv, während die Lanthan- und Zirkonfällungen nur Aktivitäten hatten, deren Entstehung aus der Aktivität der Bariumfällungen leicht nachgewiesen werden konnte.

Die einfache Fällung mit $BaCl_2$ aus stark saurer Lösung gestattet natürlich keine Unterscheidung zwischen Barium und Radium. Nach diesen hier nur sehr summarisch aufgeführten Reaktionen kann die mit den Bariumsalzen abgeschiedene Aktivität nur von Radium herrühren, wenn man das Barium selbst als allzu unwahrscheinlich vorerst außer Betracht läßt.

Wir gehen jetzt kurz auf die mit Bariumchlorid erhaltenen Aktivitätskurven ein, die einerseits zu Aussagen über die Anzahl der „Radiumisotope" führen und außerdem deren Halbwertszeiten zu bestimmen erlauben.

Die Fig. 1 zeigt den Aktivitätsverlauf des aktiven Bariumchlorids nach vierfältiger Bestrahlung des Urans. Die Kurve a gibt die Messungen über die ersten 3 Stunden; die Kurve b die Meßwerte für das gleiche Präparat über 800 Stunden fortgeführt. Der Maßstab der unteren Kurve ist zehnmal kleiner als der für die obere.

Die anfänglich schnelle Abnahme wird allmählich langsamer und geht nach etwa 12 Stunden in eine langsame Zunahme über. Nach ungefähr 120 Stunden beginnt dann wieder eine sehr allmähliche Aktivitätsabnahme, sie erfolgt exponentiell mit einer Halbwertszeit von rund 13 Tagen.

Der Verlauf der Kurven zeigt deutlich, daß hier mehrere Substanzen vorliegen müssen. Man kann aber nicht ohne weiteres sagen, welchos die Körper sind; ob mehrere „Radiumisotope" oder ein „Radiumisotop" mit einer Reihe von Folgeprodukten den Aktivitätsverlauf bestimmen.

Es sei hier gleich vorweggenommen, daß die schon in der ersten Mitteilung angegebenen drei isomeren „Radiumisotope" bestätigt wurden. Sie seien vorerst als Ra I1 und Ra II und Ra IV bezeichnet (wegen eines mutmaßlichen Ra I s. weiter unten).

Ihr Nachweis und die Ermittlung ihrer Halbwertszeiten wird an Hand der folgenden Figuren kurz dargestellt.

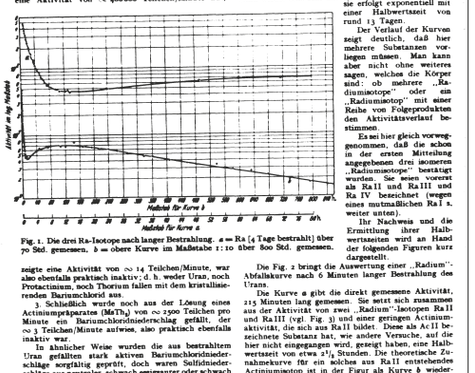


Fig. 1. Die drei Ra-Isotope nach langer Bestrahlung. a = Ra [4 Tage bestrahlt] über 70 Std. gemessen, b = obere Kurve im Maßstabe 1:10 über 800 Std. gemessen.

Die Fig. 2 bringt die Auswertung einer „Radium"-Abfallskurve nach 6 Minuten langer Bestrahlung des Urans.

Die Kurve a gibt die direkt gemessene Aktivität, 215 Minuten lang gemessen. Sie setzt sich zusammen aus der Aktivität von zwei „Radium"-Isotopen Ra II und Ra III (vgl. Fig. 3) und einer geringen Actiniumaktivität, die sich aus Ra I bildet. Diese als Ac II bezeichnete Substanz hat, wie andere Versuche, auf die hier nicht eingegangen wird, gezeigt haben, eine Halbwertszeit von etwa 2½ Stunden. Die theoretische Zunahmekurve für ein solches aus Ra I entstehendes Actiniumisotop ist in der Figur als Kurve b wiedergegeben. Als Halbwertszeit für das Ra II ist dabei schon der Wert von 14 Minuten vorweggenommen. Zieht man die Werte der Kurve b von denen der Kurve a ab, dann ergibt sich die Kurve c der Figur. Diese Aktivität rührt nun praktisch nur noch von Radiumisotopen her, und zwar in der Hauptsache von dem kurzlebigen Ra II und in untergeordnetem Maße dem langerlebigen Ra III. Letzteres hat, wie sich weiter unten aus Fig. 3 ergibt, eine Halbwertszeit von ungefähr 86 Minuten. Den Aktivitätsverlauf von Ra III zeigt die Kurve d der

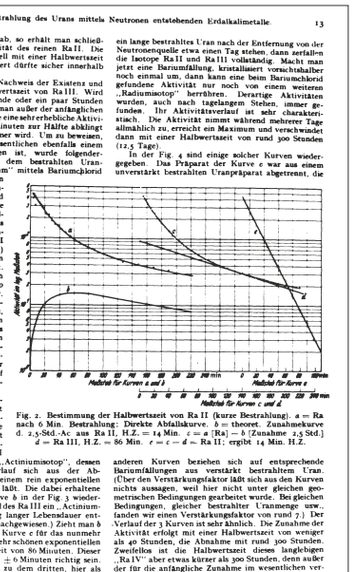


Fig. 2. Bestimmung der Halbwertszeit von Ra II (kurze Bestrahlung). a = Ra nach 6 Min. Bestrahlung; Direkte Abfallskurve. b = theoret. Zunahmekurve d. 2,5-Std.-Ac aus Ra II, H.Z. = 14 Min. c = a [Ra] − b [Zunahme 2,5 Std.] d = Ra III, H.Z. = 86 Min. e = c − d = Ra II; ergibt 14 Min. H.Z.

Fig. 2. Zieht man d von e ab, so erhält man schließlich in Kurve c die Aktivität des reinen Ra II. Die Abnahme erfolgt exponentiell mit einer Halbwertszeit von 14 Minuten. Dieser Wert dürfte sicher innerhalb ± 2 Minuten richtig sein.

Wir kommen nun zum Nachweis der Existenz und zur Bestimmung der Halbwertszeit von Ra III. Wird ein Ra-Präparat eine Stunde oder ein paar Stunden lang bestrahlt, dann findet man außer der anfänglichen schnellen Aktivitätsabnahme eine sehr erhebliche Aktivität, die in etwa 100—110 Minuten auf Hälfte abklingt und schließlich noch langsamer wird. Um zu beweisen, daß diese Aktivität im wesentlichen ebenfalls einem Radiumisotop zuzuschreiben ist, wurde folgendermaßen vorgegangen. Aus dem bestrahlten Uranpräparat wurde das „Radium" mittels Bariumchlorid abgetrennt; 2½ Stunden später wurde das Bariumchlorid wieder aufgelöst und erneut gefällt. Das kurzlebige Ra II ist in dieser Zeit vollständig zerfallen, und das aus dem Ra II in dem Bariumchlorid entstandene Ac II (2½ Stunden Halbwertszeit) wird bei der Umkristallisation des Bariumchlorids entfernt. Das Bariumchlorid ist noch erheblich aktiv; es liegt also noch ein „Radiumisotop" vor. Es wird hier also so vorgegangen, wie dies vorher Meitner, Strassmann und Hahn schon bei der Aufklärung der aus dem Thorium entstehenden künstlichen Umwandlungsprodukte durchgeführt worden war[1]. Der nunmehr erhaltene Aktivitätsverlauf ist in Fig. 3, Kurve a wiedergegeben.

Fig. 3. Bestimmung der Halbwertszeit von Ra III nach 2,5-stündiger Bestrahlung. a = Ra III [2,5 Std. bestrahlt]. 3 Std. n. Bestr. wurde Ac abgetrennt. b = ∾ Zunahmekurve v. langem Ac aus Ra III v. 86 Min. H.Z. c = a − b = Ra III. H.Z. = ∾ 86 Min.

Die anfängliche Abnahme erfolgt exponentiell mit einer Halbwertszeit von rund 13 Minuten. Sie ist also praktisch gleich der des Ra II. Die Kurve b gibt die Messungen über die gleiche Präparat über 800 Stunden fortgeführt. Der Maßstab der unteren Kurve ist zehnmal kleiner als der für die obere.

Die anfänglich schnelle Abnahme wird allmählich langsamer und geht nach etwa 12 Stunden in eine langsame Zunahme über. Nach ungefähr 120 Stunden beginnt dann wieder eine sehr allmähliche Aktivitätsabnahme, sie erfolgt exponentiell mit einer Halbwertszeit von rund 13 Tagen.

Der Verlauf der Kurven zeigt deutlich, daß hier mehrere Substanzen vorliegen müssen. Man kann aber nicht ohne weiteres sagen, welchos die Körper sind; ob mehrere „Radiumisotope" oder ein „Radiumisotop" mit einer Reihe von Folgeprodukten den Aktivitätsverlauf bestimmen.

Es sei hier gleich vorweggenommen, daß die schon in der ersten Mitteilung angegebenen drei isomeren „Radiumisotope" bestätigt wurden. Sie seien vorerst als Ra I1 und Ra II und Ra IV bezeichnet (wegen eines mutmaßlichen Ra I s. weiter unten).

Ihr Nachweis und die Ermittlung ihrer Halbwertszeiten wird an Hand der folgenden Figuren kurz dargestellt.

Fig. 1. Die drei Ra-Isotope nach langer Bestrahlung. a = Ra [4 Tage bestrahlt] über 70 Std. gemessen, b = obere Kurve im Maßstabe 1:10 über 800 Std. gemessen.

Die Fig. 2 bringt die Auswertung einer „Radium"-Abfallskurve nach 6 Minuten langer Bestrahlung des Urans.

Die Kurve a gibt die direkt gemessene Aktivität, 215 Minuten lang gemessen. Sie setzt sich zusammen aus der Aktivität von zwei „Radium"-Isotopen Ra II und Ra III (vgl. Fig. 3) und einer geringen Actiniumaktivität, die sich aus Ra I bildet. Diese als Ac II bezeichnete Substanz hat, wie andere Versuche, auf die hier nicht eingegangen wird, gezeigt haben, eine Halbwertszeit von etwa 2½ Stunden. Die theoretische Zunahmekurve für ein solches aus Ra II entstehendes Actiniumisotop ist in der Figur als Kurve b wiedergegeben. Als Halbwertszeit für das Ra II ist dabei schon der Wert von 14 Minuten vorweggenommen. Zieht man die Werte der Kurve b von denen der Kurve a ab, dann ergibt sich die Kurve c der Figur. Diese Aktivität rührt nun praktisch nur noch von Radiumisotopen her, und zwar in der Hauptsache von dem kurzlebigen Ra II und in untergeordnetem Maße dem langerlebigen Ra III. Letzteres hat, wie sich weiter unten aus Fig. 3 ergibt, eine Halbwertszeit von ungefähr 86 Minuten. Den Aktivitätsverlauf von Ra III zeigt die Kurve d der

Fig. 2. Bestimmung der Halbwertszeit von Ra II (kurze Bestrahlung). a = Ra nach 6 Min. Bestrahlung; Direkte Abfallskurve. b = theoret. Zunahmekurve d. 2,5-Std.-Ac aus Ra II, H.Z. = 14 Min. c = a [Ra] − b [Zunahme 2,5 Std.] d = Ra III, H.Z. = 86 Min. e = c − d = Ra II; ergibt 14 Min. H.Z.

Die im vorhergehenden gebrachten Ergebnisse zusammenfassend haben wir also drei als Ra I1, Ra III und Ra IV bezeichnete isomere Erdalkalimetalle festgestellt. Ihre Halbwertszeiten sind 14 ± 2 Minuten, 80 ± 6 Minuten, 250 ... 300 Stunden. Es wird aufgefallen sein, daß der 14-Minuten-Körper nicht als Ra I, die weiteren Isomeren nicht als Ra II und Ra III bezeichnet worden sind. Der Grund liegt darin, daß wir an ein noch instabileres „Ra" glauben, das bisher nicht nachgewiesen wurde. In unserer ersten Mitteilung über die neuen Umwandlungsprodukte haben wir ein Actinium von etwa 40 Minuten Halbwertszeit angegeben und

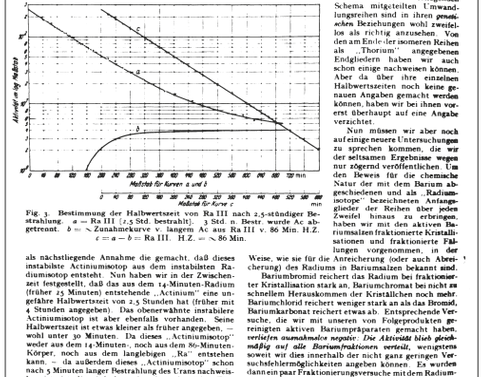


Fig. 3. Bestimmung der Halbwertszeit von Ra III nach 2,5-stündiger Bestrahlung. a = Ra III [2,5 Std. bestrahlt]. 3 Std. n. Bestr. wurde Ac abgetrennt. b = ∾ Zunahmekurve v. langem Ac aus Ra III v. 86 Min. H.Z. c = a − b = Ra III. H.Z. = ∾ 86 Min.

als nächstliegende Annahme die gemacht, daß dieses instabilste Actiniumisotop aus dem instabilsten Radiumisotop entsteht. Nun haben wir in der Zwischenzeit festgestellt, daß das aus dem 14-Minuten-Radium (früher 25 Minuten) entstehende „Actinium" eine ungefähre Halbwertszeit von 2,5 Stunden hat (früher mit 4 Stunden angegeben). Das obenerwähnte instabilere Actiniumisotop ist aber ebenfalls vorhanden. Seine Halbwertszeit ist etwas kleiner als früher angegeben, — wohl unter 30 Minuten. Da dieses „Actiniumisotop" weder aus dem 14-Minuten-, noch aus dem 86-Minuten-Körper, noch aus dem langlebigen „Ra" entstehen kann, — da außerdem dieses „Actiniumisotop" schon nach 5 Minuten langer Bestrahlung des Urans nachweisbar ist, ist die einfachste Annahme für seine Entstehung ein „Radiumisotop", dessen Halbwertszeit kürzer als 1 Minute sein muß. Mit einer größeren Halbwertszeit als eine Minute hätten wir es nämlich nachweisen müssen; wir haben sehr danach gesucht. Wir bezeichnen deshalb diese bisher unbekannte, mit einer stärkeren Strahlenquelle wohl zweifellos nachweisbare Muttersubstanz des instabilsten „Actiniumisotops" als „Ra I".

Das in unserer ersten Mitteilung gebrachte Schema muß dadurch eine gewisse Korrektur erfahren. Das folgende Schema trägt dieser Änderung Rechnung und gibt für die Anfangsglieder der Reihen die nunmehr genauer bestimmten Halbwertszeiten:

$$\text{„Ra I"?} \xrightarrow[<1\text{ Min.}]{\beta} \text{Ac I} \xrightarrow[<10\text{ Min.}]{\beta} \text{Th?}$$

$$\text{„Ra II"} \xrightarrow[14 \pm 2\text{ Min.}]{\beta} \text{Ac II} \xrightarrow[\sim 2{,}5\text{ Std.}]{\beta} \text{Th?}$$

$$\text{„Ra III"} \xrightarrow[86 \pm 6\text{ Min.}]{\beta} \text{Ac III} \xrightarrow[\text{mehrere Tage}]{\beta} \text{Th?}$$

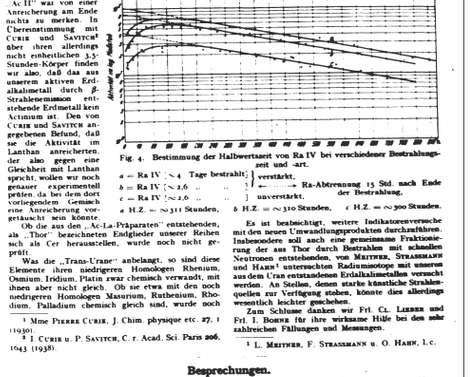
$$\text{„Ra IV"} \xrightarrow[250\text{–}300\text{ Std.}]{\beta} \text{Ac IV} \xrightarrow[<40\text{ Std.}]{\beta} \text{Th?}$$

Die große Gruppe der „Transurane" steht bisher in keinem erkennbaren Zusammenhang mit diesen Reihen.

Die in dem vorliegenden Schema mitgeteilten Umwandlungsreihen sind in ihren genetischen Beziehungen wohl zweifellos als richtig anzusehen. Von den am Ende der isomeren Reihen als „Thorium" angegebenen Endgliedern haben wir auch schon einige nachweisen können. Aber da über ihre einzelnen Halbwertszeiten noch keine genauen Angaben gemacht werden können, haben wir bei ihnen vorerst überhaupt auf eine Angabe verzichtet.

Nun müssen wir aber noch auf einige neuere Untersuchungen zu sprechen kommen, die wir der seltsamen Ergebnisse wegen nur zögernd veröffentlichen. Um den Beweis für die chemische Natur der mit dem Barium abgeschiedenen und als „Radiumisotope" bezeichneten Anfangsglieder der Reihen über jeden Zweifel hinaus zu erbringen, haben wir mit den aktivsten, bariumsalzen fraktionierte Kristallisationen und fraktionierte Fällungen vorgenommen, in der Anreicherung (oder auch Abreicherung) des Radiums in Bariumsalzen bekannt sind.

Bariumbromid reichert das Radium bei fraktionierter Kristallisation stark an. Bariumchromat bei nicht zu schnellem Herauskommen der Kristallchen noch mehr. Bariumchlorid reichert weniger stark an als das Bromid, Bariumkarbonat reichert etwas ab. Entsprechende Versuche, die wir mit unseren von Folgeprodukten gereinigten aktiven Bariumpräparaten gemacht haben, *verliefen enumahmslos negativ. Die Aktivität blieb gleichmäßig auf alle Bariumfraktionen verteilt*, wenigstens soweit wir dies innerhalb der nicht ganz geringen Versuchsfehlermöglichkeiten angeben können. Es wurden dann ein paar Fraktionierungsversuche mit dem Radiumisotop ThX und mit dem Radiumisotop $MsTh_1$ gemacht. Sie verliefen genau so, wie man aus allen früheren Erfahrungen mit dem Radium erwarten sollte. Es wurde dann die „Indikatorenmethode" auf ein Gemisch des gereinigten langlebigen „Ra IV" mit reinem Bariumbromid als Trägersubstanz wurde fraktioniert kristallisiert. *Das $MsTh_1$ wurde angereichert, das „Ra IV" nicht*, sondern seine Aktivität blieb bei gleichem Bariumgehalt der Fraktionen wieder gleich. Wir kommen zu dem Schluß: Unsere „Radiumisotope" haben die Eigenschaften des Bariums, als Chemiker müßten wir eigentlich sagen, bei den neuen Körpern handelt es sich nicht um Radium, sondern um Barium.



denn andere Elemente als Radium oder Barium kommen nicht in Frage.

Schließlich haben wir auch einen Indikatorversuch mit unserem rein abgeschiedenen „Ac II" (H. Z. rund 2,5 Stunden) und dem reinen Actiniumisotop $MsTh_2$ gemacht. Wenn unsere „Ra-Isotope" kein Radium sind, dann sind die „Ac-Isotope" auch kein Actinium, sondern sollten Lanthan sein. Nach dem Vorgehen von Mme CURIE haben wir eine Fraktionierung von Lanthanoxalat, das die beiden aktiven Substanzen enthielt, aus salpetersaurer Lösung vorgenommen. Das $MsTh_2$ fand sich, wie von Mme CURIE angegeben, in den Endfraktionen stark angereichert. Bei unserem „Ac II" war von einer Anreicherung am Ende nichts zu merken. In Übereinstimmung mit CURIE und SAVITCH[1] über ihren allerdings nicht einheitlichen 3,5-Stunden-Körper finden wir also, daß das aus unserem aktiven Erdalkalimetall durch β-Strahlenemission entstehende Erdmetall kein Actinium ist. Den von CURIE und SAVITCH angegebenen Befund, daß sie die Aktivität im Lanthan anreicherten, der also gegen eine Gleichheit mit Lanthan spricht, wollen wir noch genauer experimentell prüfen, da bei dem dort vorliegenden Gemisch eine Anreicherung vorgetäuscht sein könnte.

Ob die aus den „Ac-La-Präparaten" entstehenden, als „Thor" bezeichneten Endglieder unserer Reihen auch als Cer herausstellen, wurde noch nicht geprüft.

Was die „Trans-Urane" anbelangt, so sind diese Elemente ihren niedrigeren Homologen Rhenium, Osmium, Iridium, Platin zwar chemisch verwandt, mit ihnen aber nicht gleich. Ob sie etwa mit den noch niedrigeren Homologen Masurium, Ruthenium, Rhodium, Palladium chemisch gleich sind, wurde noch nicht geprüft. Daran konnte man früher ja nicht denken. Die Summe der Massenzahlen Ba + Ma, also z. B. 138 + 101, ergibt 239!

Als Chemiker müssen wir aus den kurz dargelegten Versuchen das oben gebrachte Schema eigentlich umbenennen und statt Ra, Ac, Th die Symbole Ra, La, Ce einsetzen. Als der Physik in gewisser Weise nahestehende „Kernchemiker" können wir uns zu diesem, allen bisherigen Erfahrungen der Kernphysik widersprechenden, Sprung noch nicht entschließen. Es könnten doch noch vielleicht eine Reihe seltsamer Zufälle unsere Ergebnisse vorgetäuscht haben.

Fig. 4. Bestimmung der Halbwertszeit von Ra IV bei verschiedener Bestrahlungszeit und -art.

a = Ra IV [∾ 4 Tage bestrahlt] } verstärkt,
b = Ra IV [∾ 2,6 ..] } = } = Ra-Abtrennung 15 Std. nach Ende der Bestrahlung,
c = Ra IV [∾ 2,6 ..] } unverstärkt,
a H.Z. = ∾ 311 Stunden, b H.Z. = ∾ 310 Stunden, c H.Z. = ∾ 300 Stunden.

Es ist beabsichtigt, weitere Indikatorenversuche mit den neuen Umwandlungsprodukten durchzuführen. Insbesondere soll auch eine genauere Fraktionierung der aus Thor durch Bestrahlen mit schnellen Neutronen entstehenden, von MEITNER, STRASSMANN und HAHN[1] untersuchten Radiumisotope mit unseren aus dem Uran entstandenen Erdalkalimetallen versucht werden. An Stellen, denen starke künstliche Strahlenquellen zur Verfügung stehen, könnte dies allerdings wesentlich leichter geschehen.

Zum Schlusse danken wir Frl. Cl. LIEBER und Frl. I. BOEHE für ihre wirksame Hilfe bei den sehr zahlreichen Fällungen und Messungen.

[1] L. MEITNER, F. STRASSMANN u. O. HAHN, Z. Physik 109, 538 (1938).

Besprechungen.

STUBBE, H.: **Spontane und strahleninduzierte Mutabilität.** (Probleme der theoretischen und angewandten Genetik und deren Grenzgebiete, redigiert von W. F. REINIG.) Leipzig: Georg Thieme 1937. 190 S. und 12 Abbild. 13 cm × 21 cm. Preis kart. RM 6.80.

Nach 10 Jahren emsiger Arbeit hat die experimentelle Mutationsforschung heute ein Stadium erreicht, das dazu berechtigt, die Ergebnisse dieser Forschungen zusammenfassend darzustellen und einem größeren Leserkreis nahezubringen. Nachdem erst kürzlich von TIMOFÉEFF-RESSOVSKY eine Darstellung der experimentellen Mutationsforschung gegeben wurde, behandelt STUBBE diesen Gebiet in den „Problemen der Genetik". Er versucht auch auf die spontane und die strahleninduzierte Mutabilität zu beschränken, die strahleninduzierte Mutabilität zu beschränken, doch ergeben sich verschiedene Schwierigkeiten, wenn man die spontane Mutabilität behandeln will, ohne auf die Frage der Abhängigkeit der Mutationsrate von verschiedenen physiologischen Bedingungen einzugehen. Wenn auch erst wenig brauchbare Ergebnisse auf diesem Gebiet vorliegen, so kann doch an den hier aufgeworfenen Problemen nicht vorbeigegangen werden. Sie werden denn auch vom Verf. angeschnitten bei der Behandlung der Abhängigkeit der „Mutationsrate"

▲ Figure 5.123
Five pages from HAHN and STRASSMANN's historic article.

The events related to nuclear fission followed each other with dramatic speed. In December 1938, HAHN reported the results of his research to his colleague LISE MEITNER, who had by that time emigrated to Stockholm. This is how it was possible for an article by FRISCH and MEITNER to appear in the January 1939 issue of *Nature*. The article gives a clear picture of the processes that take place during fission based on the liquid drop model of the nucleus. At a physics conference held in the United States in 1939, also in January, this topic, promulgated primarily by BOHR, was the focus of interest. While the conference was still taking place, several researchers managed to confirm the fact of fission, this time with the methodology of physics. Also during 1939, BOHR was able to show, using an improved version of the liquid drop model, that of the two uranium isotopes, only uranium-235 undergoes fission.

5.4.11 *The Chain Reaction: The Large-Scale Liberation of Nuclear Energy*

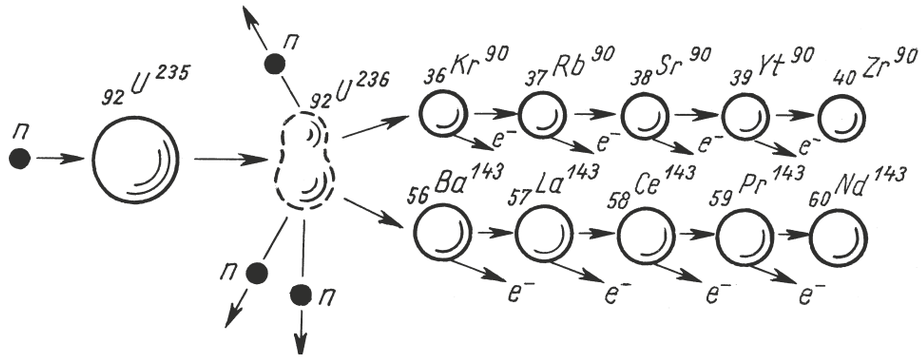
The question was immediately raised whether during fission, in addition to the two medium heavy fission products, free neutrons might also be emitted because this could be of crucial significance for the creation of a chain reaction. In March 1939, experiments in three different locations showed that several neutrons are indeed released in each fission reaction. Because of the importance of this phenomenon, it is worth mentioning the names of the experimenters: HALBAN, JOLIOT, and KOWARSKI (Paris); ANDERSON, FERMI, and HANSTEIN (Columbia University); and SZILÁRD and ZINN (New York University). Their measurements were, of course, not very precise,

but it was already possible to estimate that the average number of neutrons released in each fission reaction was about three. For a chain reaction to be created, this number is of such crucial importance that the precise value was kept secret until 1955, when it was published simultaneously in the United States and the Soviet Union that, on average, 2.47 neutrons are released per fission reaction.

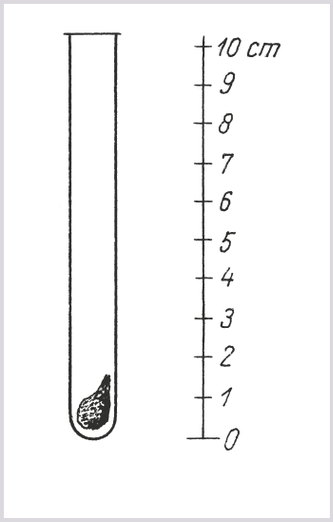
Thus, in March 1939, physicists the world over were aware of the possibility of achieving a chain reaction and along with it the release of nuclear energy for peaceful or military purposes (Quotations 5.44, 5.45). Therefore, the crucial work on uranium fission was carried out in secret everywhere. Thus, it is very difficult to fix the dates, names, and countries for particular discoveries. The date December 2, 1942 is of special significance, for it was on this day that under Fermi's direction, a sustained chain reaction was achieved, producing, for the first time, continuous power from nuclear energy. The power level of this experiment was 200 watts (Figure 5.125).

It is worthwhile to examine in greater detail how the energy stored in the nucleus of the uranium atom is transformed into another form of energy, primarily heat energy in an atomic bomb or nuclear power plant. In general, we think of the release of chemical energy, for example, burning fuels, as something simple. But if we delve more deeply into the physics of the phenomenon, it turns out that our belief in its simplicity is nothing more than the result of sheer habit. We feel that the release of atomic energy is somehow alien, even though, from the standpoint of physics, it is much easier to track the energy balance for the fission of uranium than, for example, for the chemical process whereby one carbon atom and two oxygen atoms combine during ordinary combustion. Figure 5.126 shows the course of the fission reaction that is triggered by a neutron. As the neutron enters the nucleus, it transmits its energy to the nucleus as a whole, and as a result, the nucleus begins to vibrate. In the course of this vibration, the nucleus can assume the shape of a dumbbell. At this point, the electrostatic repulsive forces, which had been compensated for by the nuclear forces when the nucleus was spherical in shape, become dominant and push the two partial nuclei at opposite ends of the dumbbell apart with great force. From this picture, the energy can be estimated, as was done by Frisch and Meitner, who got the value 200 MeV, which is a good approximation. The precise energy balance is shown in Figure 5.127. The fission products flying apart at great velocity collide with the surrounding atoms, giving up their energies until they slow down. The kinetic energy of these atoms increases, and this is what we perceive as an increase in temperature.

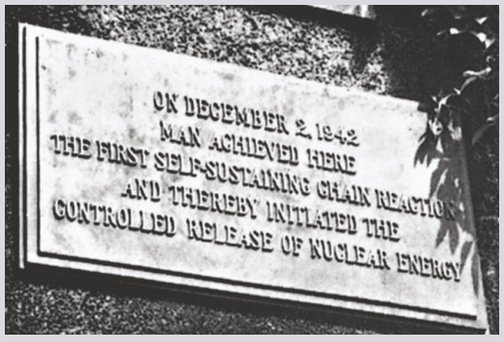
We have already seen that nuclei have more excess of neutrons with increasing atomic weight. This is necessary so that the nuclear binding forces can overcome the repulsive forces of the positively charged protons. If a heavy nucleus splits into



▲ **Figure 5.126** Diagram of a fission reaction.



▲ **Figure 5.124** The neutron source used by Fermi. Radon gas and beryllium powder are contained in a sealed test tube. The alpha particles emitted by the radon liberate neutrons from the beryllium via the reaction ${}_4Be^9 + {}_2He^4 = {}_6C^{12} + {}_0n^1$.



▲ **Figure 5.125** Memorial plaque at the site of the first chain reaction [Harvard 1970].

This first chain reaction took place in a nuclear pile at the University of Chicago. The pile consisted of six tons of pure uranium metal and 50 tons of uranium oxide with natural isotope composition and 400 tons of graphite serving to slow down the neutrons; the number of neutrons was measured with boron trifluoride counters. For control, cadmium rods were used.

After the successful experiment, the following telephone conversation took place between H. A. Compton (discoverer of the Compton effect, head of the Uranium Committee) and James B. Conant (prominent chemist, president of Harvard University, head of National Defense Research Committee):

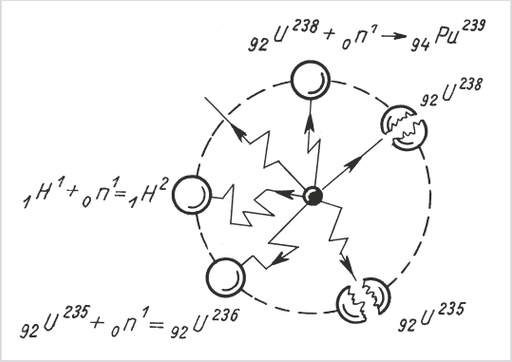
> "Jim," Compton said. "You'll be interested to know that the Italian navigator has just landed in the new world." Conant asked, "Were the natives friendly?" "Everyone landed safe and happy" was Compton's reply.

Leo Szilárd (discoverer of the Szilárd–Chalmers effect, along with E. Wigner the most important assistant in the construction of the reactor), recalls the following:

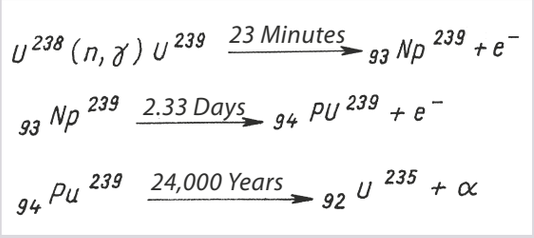
> There was a crowd there and when it dispersed, Fermi and I stayed there alone. Enrico Fermi and I remained. I shook hands with Fermi and I said that I thought this day would go down as a black day in the history of mankind.

Kinetic Energy of the Fission Products	*167 MeV*
Energy of the Beta Decay	*5 MeV*
Energy of the Gamma Rays Following Beta Decay	*5 MeV*
Energy Removed by Neutrinos	*11 MeV*
Energy of the Neutrons Arising from Fission	*5 MeV*
Energy of the Prompt Gamma Rays After Fission	*5 MeV*
Total	*198 MeV*

▲ **Figure 5.127** Energy balance in a fission reaction.

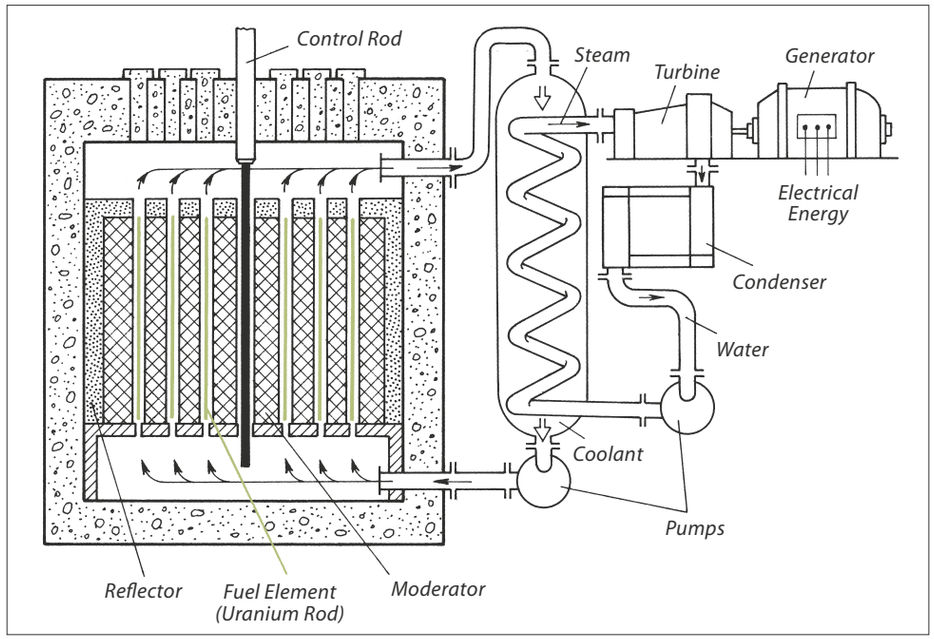


▲ **Figure 5.128** Possible fates of a neutron in a reactor containing naturally occurring uranium.



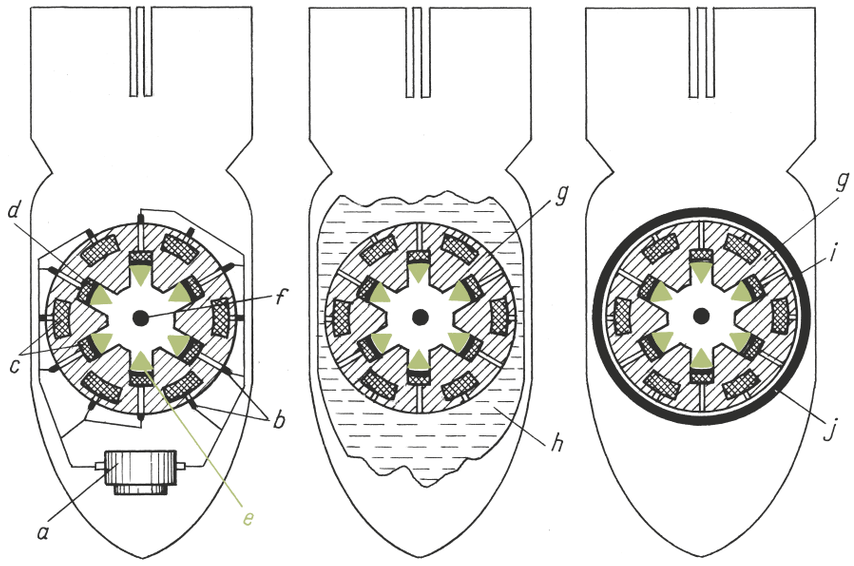
▲ **Figure 5.129** The synthesis of plutonium. Like U-235, plutonium is also fissionable.

two parts, then each part will now have an excess of neutrons too great for their sizes. It was therefore to be expected that during nuclear fission, some of these excess neutrons would be immediately released. Other excess neutrons in the fission products would be eliminated through beta decay; this explains the strong beta activity of the fission products. But why do the excess neutrons have such a special significance? As late as 1933, RUTHERFORD was still able to say with confidence that it was all "nonsense" to speak of a practical use for the enormous quantities of energy released in nuclear reactions. He pointed out correctly that the probability of the occurrence of a nuclear reaction is very small and that millions and millions of particles would have to be accelerated to a high energy before one of them would strike an atom and trigger a reaction with large energy. However, if during the reaction the same particle that initiated the reaction was to be created, and especially in multiple numbers, then there is the possibility that these new particles will set off more and more fissions with more and more excess neutrons, so that finally a "divergent chain reaction" occurs with an ever increasing liberation of energy. One problem is, naturally, whether the released neutrons can actually start new fissions, or whether they will be captured by the nucleus of some other element and therefore become lost to the chain reaction. And indeed, a number of dangers lurk for neutrons (Figure 5.128). For example, they are absorbed by vari-



▲ **Figure 5.130** Diagram of a nuclear reactor. The fuel of the reactor is natural uranium or uranium enriched with the isotope U-235; plutonium can also be used (1 gram of U-235 delivers about 24 MWh of heat energy; that is, 1000 kW of thermal power over 24 hours). If the neutron-absorbing control rods (containing cadmium or boron) are removed from the reactor, the number of neutrons in the reactor space will increase exponentially (although with a large time constant). At a specified neutron population—which means a specific power level—the reactor state can be stabilized. The moderator will slow down the neutrons produced by the fission processes as some of their kinetic energy will be lost via collisions with the nuclei of the moderator material. The slower neutrons can then cause the U-235 nuclei to fission with greater probability.

Some facts about the use of nuclear energy, in chronological order: Discovery of fission (1938); after the first successful sustained chain reaction (1942), nuclear energy was used for military purposes. Usable electrical power (250 kW) was generated (1951; Arco, Idaho), which was used to illuminate the halls of EBR-1 (Experimental Breeder Reactor). The first civilian nuclear power plant was connected to the electrical power grid in 1954, in the Soviet Union, with a power of 5 megawatts. The first large-scale nuclear power plant with the respectable power of 60 megawatts was placed in service in 1956, in Calder Hall, UK.



Legend:
a: fuse; *b*: detonator; *c*: conventional explosives; *d*: neutron reflector; *e*: fissile material; *f*: initiating neutron source; *g*: fission trigger, identical to arrangement in left-hand figure; *h*: deuterium–tritium mixture; i: deuterium–lithium mixture; *j*: natural uranium U-238.

▲ **Figure 5.131** The sword of Damocles hanging over mankind: the atomic bomb. A conventional high explosive charge causes the separate sections consisting of U-235 or plutonium to be compressed into a sphere whose mass is above the critical mass for a divergent chain reaction to begin, explosively releasing large amounts of energy. If the fission bomb is enveloped in an additional deuterium–tritium mixture, then one has a fusion, or hydrogen, bomb. Chronology: July 16, 1945: first test (Alamogordo, New Mexico); August 6, 1945: Hiroshima; August 9, 1945: Nagasaki. Scientific director of the American atomic program: J. R. Oppenheimer. The first test explosion in the Soviet Union took place in 1949. The "father" of the hydrogen bomb is the Hungarian-born Edward Teller, who suggested a method for the realization of such a bomb in 1945. The first H-bomb tests were carried out in 1951 in the USA and in 1953 in the USSR. Great Britain has had atomic bombs since 1952 (uranium bombs) and 1957 (H-bomb). Then came France in 1960, China in 1964, and then, a decade later, India and others. In addition to its use in bombs, atomic energy plays an important role in military applications in providing power for submarines (*Nautilus*, 1955) and aircraft carriers.

A key problem in the construction of atomic bombs was initially—but also later for the realization of an atomic energy program in general—the separation of the U-235 isotope from natural uranium that contains only 0.7% U-235. The most successful method was the diffusion method: U-235 diffuses somewhat more rapidly through a porous barrier, and therefore the material passing through will be enriched to a very slight extent. With a huge cascade of such diffusion elements, we can, in principle, obtain an arbitrary degree of enrichment. At the very beginning of the atomic bomb project, a different separation method was used: the *Calutron,* developed by Lawrence on the principle of the mass spectrograph. With an ion stream of one ampere, three grams of U-235 could be produced per day. The simplest path to pure fissile material, however, goes by way of plutonium: In high-powered reactors, fissile Pu-239 can be synthesized from natural U-238 (see Figure 5.129), and it can be separated with a "simple" chemical method adapted to a man-made element. The first atomic bomb, tested in New Mexico, contained plutonium. The one dropped on Hiroshima was a uranium bomb, which had not been tested.

ous contaminants with a large cross section, so that one of the most serious problems in the early years of reactor technology was to create materials of the requisite purity. Because of the limited physical size of the fissionable fuel elements, many neutrons leave the active zone of the reactor without triggering a fission event. From this, it follows that there is a critical dimension for both energy-generating reactors and for fission bombs below which a chain reaction cannot occur.

We will mention here another useful form of neutron capture: After the capture of a neutron, the uranium isotope U-238 is transformed via two beta decays into plutonium, a transuranium element with the atomic number 94, which when hit by a slow neutron will fission, just like U-235 (Figure 5.129).

Quotation 5.44

Newton: ... It's nothing more or less than a question of the freedom of scientific knowledge. It doesn't matter who guarantees my freedom. I give my services to any system, providing that system leaves me alone. I know there's a lot of talk nowadays about physicists' moral responsibilities. We suddenly find ourselves confronted with our own fears and we have a fit of morality. This is nonsense. We have far-reaching, pioneering work to do and that's all that should concern us. Whether or not humanity has the wit to follow the new trails we are blazing is its own look-out, not ours.

Einstein: Admittedly we have pioneer work to do. I believe that too. But all the same we cannot escape our responsibilities. We are providing humanity with colossal sources of power. That gives us the right to impose conditions. If we are physicists, then we must become power politicians. ...

Möbius: ... Our knowledge has become a frightening burden. Our researches are perilous, our discoveries are lethal. For us physicists there is nothing left but to surrender to reality. It has not kept up with us. It disintegrates on touching us. We have to take back our knowledge and I have taken it back. There is no other way out, and that goes for you as well. ... Only in the madhouse can we be free. Only in the madhouse can we think our own thoughts. Outside they would be dynamite. ...

Newton: Let us be mad, but wise.

Einstein: Prisoners but free.

Möbius: Physicists but innocent.

—Dürrenmatt, *The Physicists* [pp. 76, 81–82, 84]

Quotation 5.45

We are natural scientists from various countries, various races, various beliefs, various political convictions. We are bound externally only by the Nobel Prize that we have all received.

We have happily placed our lives in the service of science. It is, we believe, one route to happier life for mankind. We see with shock that this very science also provides mankind with the means to destroy itself.

An all-out use in war of the weapons possible today can contaminate the earth with such radioactivity that entire nations would be annihilated. Neutral countries would share in this death just as the belligerents.

continued on next page

Quotation 5.45, ***continued***

If a war were to break out between the great powers, who could guarantee that it would not develop into such a death struggle? Thus a nation that opens the door to total war brings about its own destruction and endangers the entire world.

We do not deny that perhaps peace today is preserved precisely because of fear of these deadly weapons. Nevertheless, we regard it to be self-delusion if governments believe they can avoid a war for a long period of time only by fear of these weapons. Fear and tension have so often produced war. Likewise, it appears to us a self-delusion to believe that small conflicts will continue to be decided by traditional weapons. In extreme danger no nation will deny any weapon that scientific technology can produce. All nations must decide voluntarily to refrain from violence as the last means of politics. If they are not prepared to do so, they will cease to exist.

Mainau, 15 July 1955

[partial list of signatories:]
Kurt Adler, Cologne
Max Born, Bad Pyrmont
Adolf Butenandt, Tübingen
Arthur H. Compton, St. Louis
Gerhard Domagk, Wuppertal
Hans von Euler-Chelpin, Stockholm
Otto Hahn, Göttingen
Werner Heisenberg, Göttingen
Georg V. Hevesy, Stockholm
Richard Kuhn, Heidelberg
Fritz Lipmann, Boston
Hermann Joseph Muller, Bloomington
Paul Hermann Müller, Basel
Leopold Ruzicka, Zurich
Frederick Soddy, Brighton
Wendell M. Stanley, Berkeley
Hermann Staudinger, Freiburg
Hideki Yukawa, Kyoto

—Mainau Declaration [Shea 1983, p. 17]

$$ {}_1H^2 + {}_1H^2 = {}_2He^3 + {}_0n^1 + 3.25\ MeV $$

$$ {}_1H^2 + {}_1H^2 = {}_1H^3 + {}_1H^1 + 4\ MeV $$

$$ {}_1H^2 + {}_1H^3 = {}_2He^4 + {}_0n^1 + 17.6\ MeV $$

$$ {}_1H^2 + {}_2He^3 = {}_2He^4 + {}_1H^1 + 18.3\ MeV $$

$$ {}_2He^3 + {}_2He^3 = {}_2He^4 + 2\cdot{}_1H^1 + 12.8\ MeV $$

$$ {}_3Li^6 + {}_1H^2 = 2\cdot{}_2He^4 + 22.4\ MeV $$

$$ {}_3Li^7 + {}_1H^1 = 2\cdot{}_2He^4 + 17.3\ MeV $$

◀ **Figure 5.132** Some possible usable fusion reactions for the creation of energy.

Figure 5.130 shows how nuclear energy is used today in power plants for peaceful purposes; the destructive version—the bomb—is shown in Figure 5.131. See also Plate XXVIII.

5.4.12 *Energy through Nuclear Fusion: The Fuel of the Stars in the Hands of Mankind*

The utilization of the energy released during the fission of heavy atomic nuclei is not the only source of usable nuclear energy. The greatest challenge at present with respect to the use of nuclear energy is the controlled release of thermonuclear, or fusion, energy. Nuclear fusion is the direct opposite of nuclear fission and is much more similar than fission to the classical processes of combustion. We can see from the binding-energy curve displayed in Figures 5.113 and 5.114 that both the decay of heavy nuclei into lighter nuclei and the fusion of light nuclei into heavier nuclei are linked to a release of energy. Figure 5.132 shows processes in which fusion energy is released. These processes take place when the ingredients in a reaction collide with sufficiently high kinetic energy. In an accelerator, one can bring the particles up to a high enough velocity with no difficulty, but the efficiency is so poor that a gain in energy is out of the question. However, if we increase the velocity of the particles by raising the temperature of the entire system, where we are talking in terms of several hundred million degrees, then nuclear reactions will take place with a high enough frequency both to maintain the temperature of the system and to allow for removal

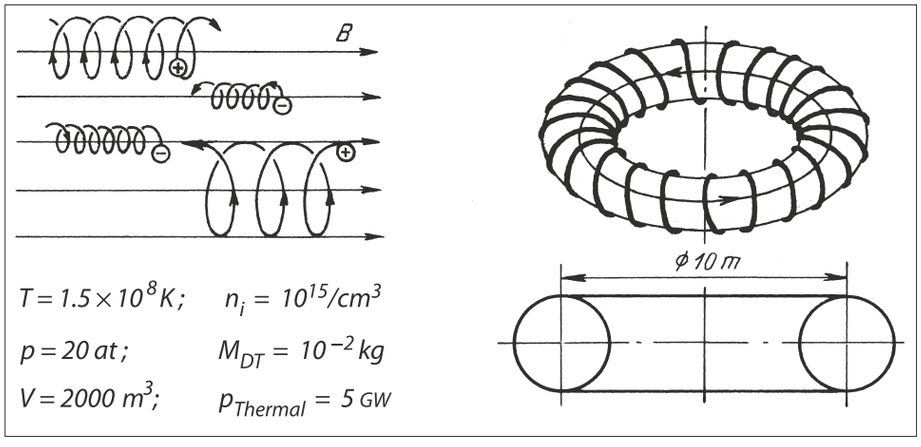


Figure 5.133 The problem of realizing controlled nuclear fusion for creating energy: Heating the deuterium–tritium gas mixture to a sufficiently high temperature (≈ 10^8 K), realization of an electromagnetic wall that can hold the plasma together at these temperatures for a sufficiently long time, the transformation of the energy released as heat in order to generate, for example, steam for the generation of electricity.

In a homogeneous magnetic field, a charged particle describes a spiral path around a field line with radius $r = mv_\perp / eB$. The particle is enclosed in a "tube," so to speak, but this tube is open at both ends. The simplest solution for closing the tube is to close up the field lines as a circle, employing a doughnut-shaped coil. However, this torus is "porous" as an electromagnetic wall because of the collision of the particles and also because of the inhomogeneity of the field. To minimize the "porosity," various auxiliary fields can be used. In the most promising type, called a *tokamak*, the auxiliary field is created by a very high current induced in the plasma. The current also serves to heat the plasma (ohmic heating). However, other methods of heating are also available: through high-frequency fields, through bombardment with energy-rich neutral particles, through magnetic compression (rapid change in the intensity of the toroidal magnetic field), etc. Great hopes have also been placed in installations in which extraordinarily intensive laser beams are used to create temperatures of several hundred million degrees in *solid* DT mixtures to ignite the reaction.

The removal of released energy can be accomplished with liquid lithium metal. One advantage of lithium is that, when activated by neutrons from the reactions, it "breeds" the tritium fuel that does not otherwise exist in nature.

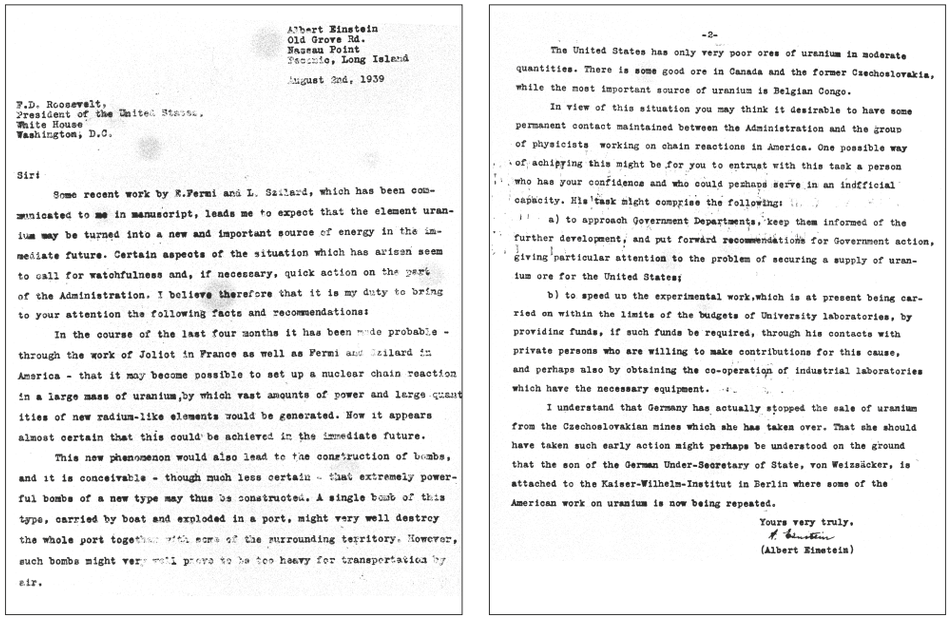
of energy from the system. That it is indeed possible to release energy in huge quantities is evidenced by the existence of the hydrogen bomb, in which fusion reactions take place when the bomb is set off. The controlled release of fusion energy has yet to be realized, but if success were ever achieved in this area, mankind would be liberated from all the oppressive shortages of energy now being faced (Figure 5.133).

5.4.13 *The Responsibility of Physicists*

The fact that from a practical viewpoint most significant discoveries of the twentieth century have been primarily put in the service of mass destruction raises the question of the responsibilities of science, especially physics, or more to the point, of physicists. Countless articles, studies, books, novels, poems, plays, and films have been devoted to this topic (Quotations 5.44, 5.45). We would merely like to observe that a physicist is a member of human society just like any other human being, with neither greater nor lesser responsibilities. Most people do their work with dedication and joy, especially in case of creative work, and frequently give no thought at all to future consequences. It is well worth reading the recollections of Weisskopf about the years in which the atomic bomb was being built (Quotation 5.46).

We can see that physicists were willing and will most likely continue to be willing to make themselves useful to governments for the fabrication of ever more destructive weapons (Figure 5.134, Figure 2.69), while in peacetime participating in peace conferences.

Anyone interested in studying the attitude of physicists regarding the societal consequences of their scientific discoveries should be aware of the fact that the first



Albert Einstein
Old Grove Rd.
Nassau Point
Peconic, Long Island

August 2nd, 1939

F.D. Roosevelt,
President of the United States,
White House
Washington, D.C.

Sir:

Some recent work by E.Fermi and L. Szilard, which has been communicated to me in manuscript, leads me to expect that the element uranium may be turned into a new and important source of energy in the immediate future. Certain aspects of the situation which has arisen seem to call for watchfulness and, if necessary, quick action on the part of the Administration. I believe therefore that it is my duty to bring to your attention the following facts and recommendations:

In the course of the last four months it has been made probable - through the work of Joliot in France as well as Fermi and Szilard in America - that it may become possible to set up a nuclear chain reaction in a large mass of uranium,by which vast amounts of power and large quantities of new radium-like elements would be generated. Now it appears almost certain that this could be achieved in the immediate future.

This new phenomenon would also lead to the construction of bombs, and it is conceivable - though much less certain - that extremely powerful bombs of a new type may thus be constructed. A single bomb of this type, carried by boat and exploded in a port, might very well destroy the whole port together with some of the surrounding territory. However, such bombs might very well prove to be too heavy for transportation by air.

-2-

The United States has only very poor ores of uranium in moderate quantities. There is some good ore in Canada and the former Czechoslovakia, while the most important source of uranium is Belgian Congo.

In view of this situation you may think it desirable to have some permanent contact maintained between the Administration and the group of physicists working on chain reactions in America. One possible way of achieving this might be for you to entrust with this task a person who has your confidence and who could perhaps serve in an unofficial capacity. His task might comprise the following:

a) to approach Government Departments, keep them informed of the further development, and put forward recommendations for Government action, giving particular attention to the problem of securing a supply of uranium ore for the United States;

b) to speed up the experimental work,which is at present being carried on within the limits of the budgets of University laboratories, by providing funds, if such funds be required, through his contacts with private persons who are willing to make contributions for this cause, and perhaps also by obtaining the co-operation of industrial laboratories which have the necessary equipment.

I understand that Germany has actually stopped the sale of uranium from the Czechoslovakian mines which she has taken over. That she should have taken such early action might perhaps be understood on the ground that the son of the German Under-Secretary of State, von Weizsäcker, is attached to the Kaiser-Wilhelm-Institut in Berlin where some of the American work on uranium is now being repeated.

Yours very truly,

(Albert Einstein)

Figure 5.134 Einstein's letter to F. D. Roosevelt, president of the United States, in which he called the president's attention to the possibility of the creation of an atomic bomb. (See also Figure 2.69.)

My action concerning the atomic bomb and Roosevelt consisted merely in the fact that, because of the danger that Hitler might be the first to have the bomb, I signed a letter to the President which had been drafted by Szilard. Had I known that that fear was not justified, I, no more than Szilard, would not have participated in opening this Pandora's box. For my distrust of governments was not limited to Germany. Unfortunately, I had no share in the warning made against using the bomb against Japan. Credit for this must go to James Franck. If they had only listened to him!

—Einstein, Letter to Laue, March 19, 1955 [Schweber 2008, p.61]

Quotation 5.46

It is very difficult to describe—let alone to judge—one's feelings and one's decisions after 30 years. There are many now who say that a physicist should devote his time to the search for truth and should not devote his abilities to produce weapons of destruction. It is easy to say, and it may even be right. There were very few who actually followed it. Pauli had to leave Europe at that time and was in Princeton and he never even thought of joining any of these war projects. But for a man in my position, and for so many others, not to join would have been a very difficult path to follow. I cannot tell you why I went to Los Alamos. The fear that Hitler would make the bomb had much to do with it. If someone suggests I went just because everybody else did, he may be right. The significance of physics is a very difficult thing to describe. Physics is not only the search for truth, it is also potential power over nature; the two aspects cannot be separated. One may say that a physicist should search only for truth and should leave the power over nature to somebody else; but this attitude actually skirts the subject and does not look at reality. The important point about physics—and most of science—is not only that it represents natural philosophy but that it also is deeply involved in action—in life, in death, in tragedy, in abuse, in the human predicament. Whether this is good or bad, who is to judge? Both sides can be argued. Some terrible things have happened since man made use of fission. However, the first bomb may have ended the war. At least, that is what we thought; we thought we had saved a million live; maybe we were right the second bomb was much harder to defend. We did participate in all that and I don't want to condemn or to justify it.

I cannot deny that those four years in Los Alamos were a great experience, from a human as well as from a scientific point of view, in spite of the fact that they were devoted to the development of the most murderous device ever created by man. Such are the contradictions of life. From the human point of view, to live together, intellectually and otherwise, with the best physicists from the whole world (such as Niels Bohr, Enrico Fermi, J. Chadwick, R. Peierls, E. Segrè, and many more) was quite an experience. The discussions we had on philosophy, art, politics, physics, and on the future world under the shadow of a super-weapon remain unforgettable. But also from the purely professional level, we had to face tasks that never were faced before. It was a remarkable experience to deal with matter under unusual conditions. We attempted to predict the behavior of matter and make experiments about it under conditions that were by factors of thousands

continued on next page

Quotation 5.46, ***continued***

outside of the ordinary. To hold in your hand a piece of metal that man made all new, to see events and processes actually develop before your eyes which never had been seen before, that was something very remarkable.

Then came the great test, the first nuclear explosion.

—Victor F. Weisskopf, "My Life as a Physicist" [pp. 14–16]

Quotation 5.47

But after the thirty fat years in the beginning of the present century, we are now dragging through the lean and infertile years, and looking for better luck in the years to come. In spite of all the efforts of the old-timers like Pauli, Heisenberg, and others, and those of the younger generation like Feynman, Schwinger, Gell-Mann, and others, theoretical physics has made very little progress during the last three decades, as compared with the three previous decades. ... [H]undreds of articles have been published on the problem of elementary particles, and still we are in darkness and uncertainty on the subject. Let us hope that in a decade or two, or, at least, just before the beginning of the twenty-first century, the present meager years of theoretical physics will come to an end in a burst of entirely new revolutionary ideas similar to those which heralded the beginning of the twentieth century.

—Gamow, *Thirty Years That Shook Physics* [pp. 161, 163]

The state of research in physics seems to me to be very closely related to the situation in the years around 1492, the time when Columbus discovered America. The most important personalities of that time were the navigators and the master of shipbuilding. I would like to compare them with the physicists and engineers our time who are planning the construction of accelerators. The experimental physicists resemble the soldiers and sailors who debarked from the ships, went ashore, and encountered on the new coasts new and rare forms of—not particles!—plants and people.

Finally, the theoreticians are to be compared with those who sat comfortably in Madrid and informed Columbus that he had reached India.

—V. Weisskopf, Lecture given at Eötvös Loránd University, 1976

reaction of the physicists in Los Alamos on hearing the news of Hiroshima's total destruction was one of relief: It worked! This was, of course, the hitherto-untried uranium bomb. In his memoirs, Otto Frisch writes:

> Somebody opened my door and shouted "Hiroshima has been destroyed!"... I still remember the feeling of unease, indeed nausea, when I saw how many of my friends were rushing to the telephone to book tables at the La Fonda hotel in Santa Fe, in order to celebrate. Of course they were exalted by the success of their work, but it seemed rather ghoulish to celebrate the sudden death of a hundred thousand people, even if they were "enemies."
>
> —Otto Frisch, *What Little I Remember* [p. 176]

5.5 Law and Symmetry

5.5.1 *The Historian's Role in the Description of Modern Physics*

In 1955, at the first Geneva conference on the peaceful uses of nuclear energy, both the Soviet Union and the United States presented some research results in nuclear physics that had until then been kept secret. With the publication and wide dissemination of data related primarily to nuclear fission and the neutron reaction cross sections that determine how chain reactions form, a major branch of nuclear physics became an engineering science. The design of nuclear plants for energy production is now no different from other engineering practice: Calculations begin with the classical equations for diffusion and neutron deceleration, while the phenomena relating to nuclear physics can be dealt with by using tables of fission cross sections.

As a consequence, nuclear physics was dethroned from its exalted position as the study of nature's most fundamental processes, and its place was taken over by the physics of elementary particles. The passing of the baton occurred under circumstances similar to the dramatic shift in the focus of research in 1932 from the physics of the outer atomic shell to the physics of the nucleus. In 1955, the vanguard of physicists were excited by the so-called *theta–tau puzzle*, whose solution, namely, *parity violation*, followed quickly and became the great sensation of 1956–1957. At this time accelerators were already in service that were capable of producing high-energy particles with energies comparable to those of cosmic rays. But instead of leading to the solution of the mysteries surrounding the particles already discovered, the number of open questions kept proliferating. The implication of all this is that the shift in the focus of research must have begun quite a bit earlier, as early as immediately following the Second World War.

With this, we begin to approach the physics of our own day, and the problematic nature of bringing the history of physics up to the present appears in earnest. What should we call history in this instance? What is the role of the historian? A summary of open questions and the most recent discoveries is the task of survey papers and status reports in the various scientific journals. Writing them is a job for the leading practitioners of the field. The author feels himself incapable of assuming this task. At the same time he would also recoil from the task of seeking historical analogies for the present situation. Could it be like the state of the art in 1902 on natural radioactivity or in 1910 on that of atomic spectra? Of all the data collected back then, and of the associated ad-hoc theories that appeared, only the tiniest portion remains as part of the standard storehouse of physics or even of the history of science. There is also the nagging feeling that trying to understand the phenomena of particle physics

with today's notions might be as pointless as were the attempts toward the end of the nineteenth century to interpret electromagnetic phenomena with the help of cleverly constructed mechanical models (Quotation 5.47).

Of course, experimental facts remain as experimental facts for the future, and their enumeration in chronological order can, if one restricts attention to what is *essential*, be useful in a presentation of the history of physics that will be written a half-century from now. The same holds for the experimental equipment, the accelerators and detectors, as well as their development. By emphasizing the adjective *essential*, we are also expressing the conviction that today there already exists a basis for distinguishing the essential from the incidental. Naturally, we also need to discuss with particular emphasis the experimental observations and associated theoretical ideas that have overturned old physical beliefs and also the new view on natural laws in which symmetries and related conservation laws play a special role.

5.5.2 *The Elementary Particles, in Order of Appearance*

What are the basic constituents of the world and what are the interactions between them? These questions were already raised in classical Greece. The physics of elementary particles now seeks answers at the subatomic level. Physicists at first arrived at the idea of the atoms of chemical elements as the basic building blocks and then discovered that the "indivisible" atoms have a structure in which the atomic nucleus can be distinguished from the electrons in the shell. The next step was to decompose the nucleus into its components, the protons and neutrons. Do these particles also have an internal structure, and can they be seen as systems with internal degrees of freedom? Or in everyday terms, can one find constituent parts for these particles that are even more elementary? Is it appropriate even at this level to ask, "What is something composed of?"

Figure 5.135 presents a timeline of information on the discovery of the elementary particles, the development of experimental equipment that made the discoveries possible, and the development of theoretical ideas about the nature of the elementary particles. From this figure, we can summarize the principal steps as follows:

Until well into the 1930s, our ideas about the structure of the universe were disquietingly simple: It seemed that the chemical elements in all their variety were composed of only two basic units, protons and electrons. Even if photons are included as the carriers of electromagnetic energy, there would be just three elementary particles (see Figure 5.101).

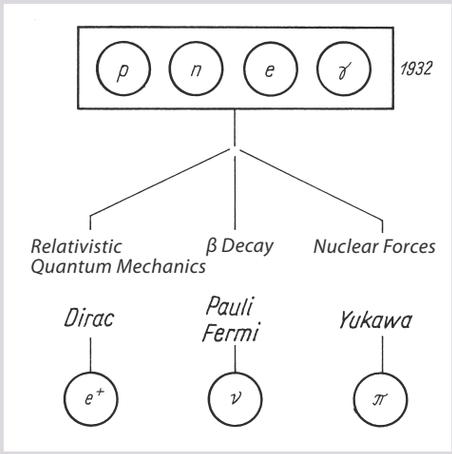
We have already mentioned the reason for the problem with this simplicity: According to the uncertainty relation, it would not be possible for electrons to remain within the nucleus, so they could *not* have been among the nuclear building blocks. With the discovery of the neutron, in the 1930s, it looked as though the problem would be resolved. True enough, the number of elementary particles had now risen to four, but with the photon (γ), proton (p), neutron (n), and electron (e), all physical phenomena of the everyday world, so to speak—including the processes in the nuclear power plants to be developed later—could be not only interpreted but understood with concepts accessible to a good high-school student. However, those physicists who dug deeper found it necessary to introduce even more particles (Figure 5.136): Dirac's relativistic quantum mechanics led us to a positive counterpart for the electron (Quotation 5.48); Pauli was forced to introduce a new particle, the *neutrino*, to explain beta decay, while the theory of nuclear forces demanded a new particle of nonzero rest mass. Let us examine the arguments in order.

Quotation 5.48

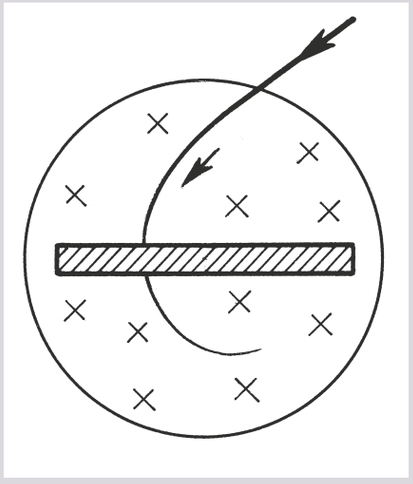
It was a very exciting time, because of the invention of the Dirac equation. Everybody was completely flabbergasted that Dirac, out of sheer intuition, could write down an equation that completely explained the properties of the relativistic electron. In fact, the equation had more in it than even Dirac thought: It contained the positron also. But that came later. I was in Göttingen 1928 to 1931; the Dirac equation was published in 1927, and everybody discussed the Dirac equation and the fact that the gyromagnetic factor of the electron turns out to be two. How wonderful!

A good example of the arrogance of theorists, which has not changed in 40 years, is the following story: There was a seminar held by the theoretical group in Göttingen, and Otto Stern came down to Göttingen from Hamburg and gave a talk on the measurements he was about to finish of the magnetic moment of the proton. He explained his apparatus, but he did not tell us the result. He took a piece of paper and went to each of us saying, "What is your prediction of the magnetic moment of the proton?" Every theoretical physicist from Max Born down to Victor Weisskopf said, "Well, of course, the great thing about the Dirac equation is that it predicts a magnetic moment of one Bohr magneton for a particle of spin one-half." Then he asked us to write down the predictions; everybody wrote "one magneton." Then, two months later, he came to give again a talk about the finished experiment, which showed that the value was 2.8. He projected the paper with our predictions on the screen. It was a sobering experience.

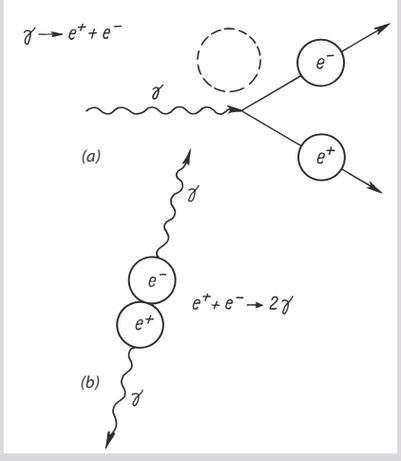
—V. Weisskopf, "My Life as a Physicist" [pp. 5–6]



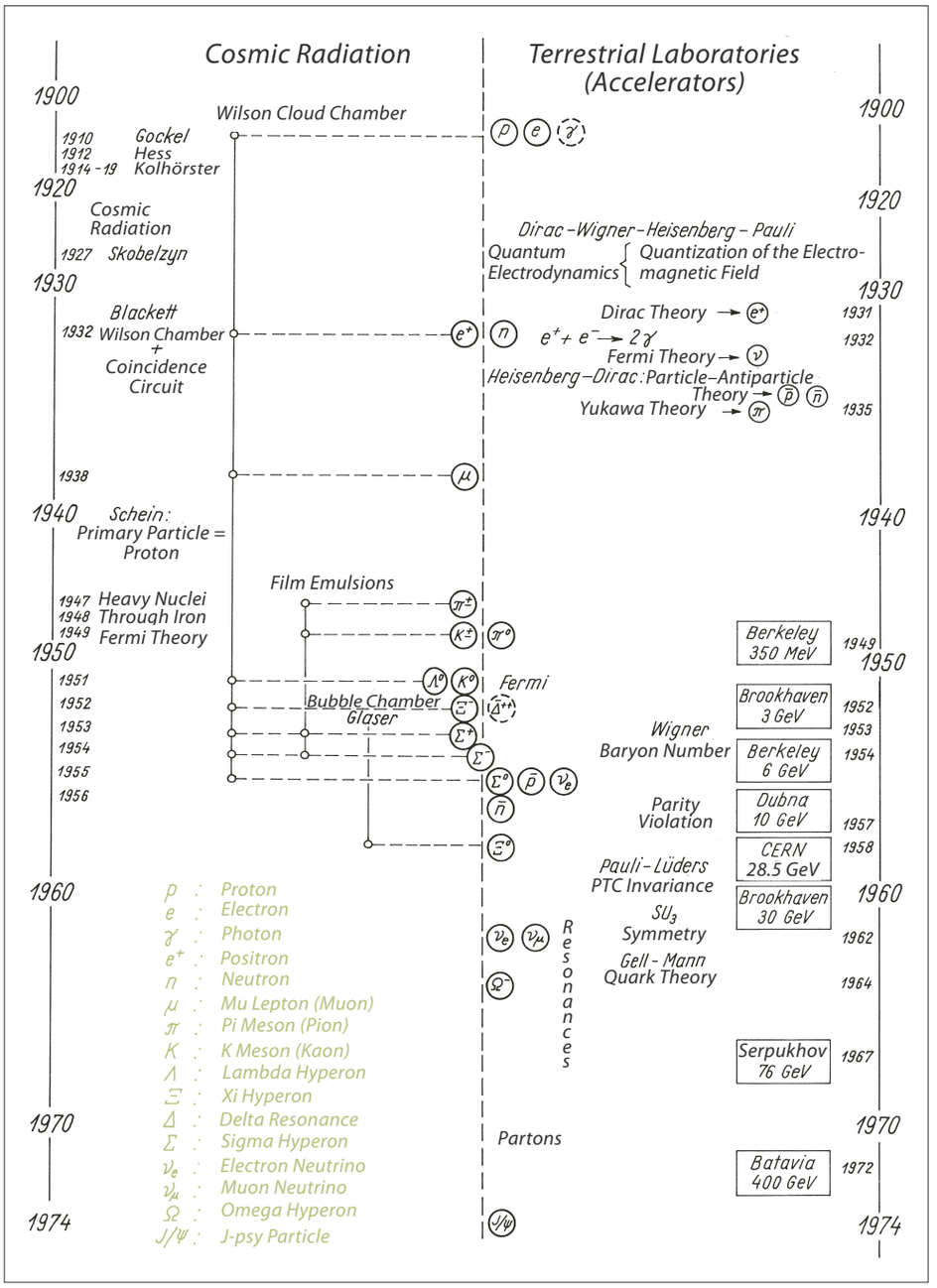
▲ **Figure 5.136** In addition to the postulated four elementary particles as of 1932, relativistic quantum theory required the positron (e^+); the theory of β decay required the neutrino (ν); and the theory of nuclear forces required the existence of a particle with a mass significantly greater than that of the electron (the π meson).



◀ **Figure 5.137** Carl Anderson proved the existence of the positron with a cloud chamber image like this one: the particle loses energy in the lead plate in the middle; therefore, its trace is more strongly bent in the given magnetic field and it follows with no ambiguity that the particle must have come from above. From the direction of the magnetic field and the direction of the curve, we can conclude that the charge of the particle is positive.



◀ **Figure 5.138** **(a)** Pair creation can take place only in the presence of matter; in vacuum, the energy and momentum laws cannot be simultaneously satisfied.
(b) Annihilation can proceed in vacuum with the emission of two photons.



▲ **Figure 5.135** Chronology of progress in particle physics (continued in Figure 5.173).

We have already spoken about the Dirac equation. However, from this theory, one cannot unambiguously deduce the existence of the positron, and it was only in 1931 that the full picture was finally put together. Complete certainty came only with the experiments in which the first antiparticle was found (Anderson, 1932; Figure 5.137). The following year, 1933, brought confirmation of more of Dirac's hypotheses, according to which the gamma quantum can turn into an electron–positron pair (Joliot-Curie, Anderson, Blackett, and Occhiallini), or an electron and a positron can annihilate to form gamma quanta (Joliot-Curie and Thibaut, Figure 5.138). In 1934, Heisenberg and Dirac generalized this theory to one of particles and antiparticles. This, in itself, did not seem to overly complicate our view of the basic building blocks of the universe. Indeed, it favored our

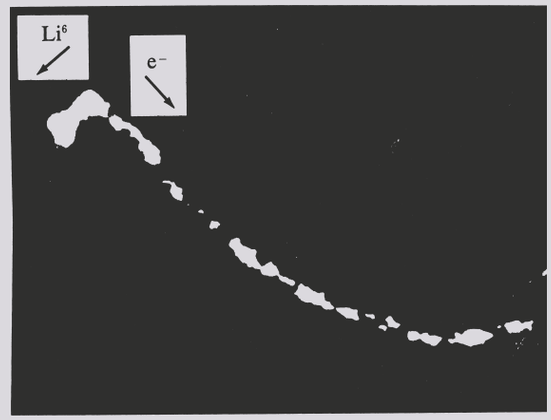
sense of symmetry by giving for every particle an antiparticle such that when the two are united, their rest masses disappear and their energy, for example, in the form of gamma quanta, is radiated away, or conversely, the particle may be created as a pair, given sufficient energy.

The other hypothetical particle, which had no role in any previous schema, was the neutrino. FERMI formulated his theory of beta decay at the beginning of the 1930s, and, as we have seen in the previous sections, a peculiarity of beta decay is that the energy of the emitted electron has continuous values, not the discrete values that we would expect in view of the well-defined energy level of the nucleus. As early as 1927, PAULI had expressed the idea that a new, as yet unknown, particle emitted along with the electron was responsible for the differences in energy. PAULI at first thought that this particle was a constituent of the nucleus. But FERMI proposed the idea that this particle comes into being during the decay process, exactly as does the electron. From pictures taken in the Wilson cloud chamber (Figure 5.139), it is clear that with the neutrino hypothesis, the conservation of momentum can also be satisfied. Understandably, the majority of physicists doubted the actual existence of such a particle because its existence was only supported by indirect evidence. They were hardly satisfied with the explanation that the neutrino has no observable positive effects because its interaction with matter is extremely weak. Direct experimental proof of the neutrino, by REINES and COWAN (Figure 5.153, discussed later), did not come about until 1955.

Incidentally, the inactivity of the neutrino does not mean that it is emitted with little energy: In a nuclear reactor with a power of 100 MW, about 5 MW, more or less, radiates away in the form of neutrinos. Because they do not interact much with matter, neutrinos tend to accumulate in the universe and add to its average mass density.

With this, a large number of problems were solved, but there still remained a need for a theory to account for the forces acting between the proton and the neutron. It is clear that the protons in the nucleus repulse each other. The stability of the nucleus requires that this electrical Coulomb repulsion must be offset by a very intense short-range force. The decisive step in the explanation of this force was the theory put forward by YUKAWA in 1935. To determine the forces between the nucleons, YUKAWA (Figure 5.140) took as his basis the interpretation of the Coulomb forces in quantum electrodynamics. He thought that a mediating particle that plays a role similar to that of the photon in the electromagnetic interaction could be the cause of the force. Building on empirical knowledge of the intensity and range of the nuclear forces, YUKAWA determined that the rest mass of the mediating particle should be—in contrast to the photon—not zero, but about 200 times the mass of an electron. After ANDERSON and NEDDERMEYER discovered in 1936 a particle in cosmic rays of just such a mass, to be known as the mu (μ) meson, theory and experiment seemed to be in perfect accord.

Thus, by the beginning of the 1940s, the picture again looked reassuring: Only the neutrino was still unable to find its place in the family of the elementary particles. The first half of the 1940s, understandably, brought no new surprises in the field of particle physics because the top physicists in the leading industrialized nations were at work on military projects. After the war, it became apparent that the mu meson discovered in 1936 could not be the particle predicted by YUKAWA because neither its mass nor its lifetime corresponded to the values required by the theory. But the main reason that it would not be able to play the role of mediator of forces is that it interacts with matter only very weakly. Today, it is no longer counted among the mesons and its new name is *muon*. Then finally, in 1947, the missing π mesons (pions) were

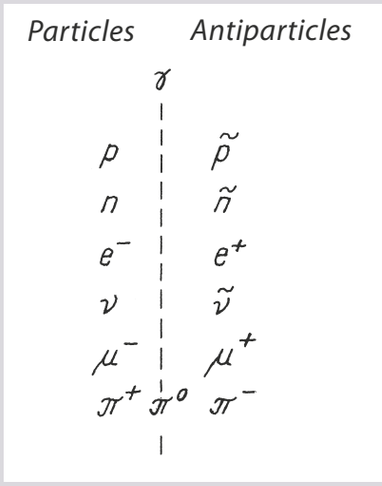


▲ **Figure 5.139** The missing momentum testifies to the existence of the neutrino. This photograph, unique in its clarity, was made by CSIKAI and SZALAY in their investigation of the decay $_2He^6$ ə $_3Li^6 + e^- + \nu_e$. The decaying nucleus has a small mass number, and the energy of the electron is considerable (3.5 MeV). Therefore, one expects a markedly visible effect. Measurements are made difficult because the isotope $_2He^6$ has a short half-life ($T_{1/2}$ = 0.8 s), and for this reason it has to be produced immediately before the measurement [Csikai 1959].



◀ **Figure 5.140** HIDEKI YUKAWA (1907–1981): Completed his studies in 1929, in Kyoto; then taught first in Osaka, and from 1939 again in Kyoto. 1942–1952: visiting professor at Columbia University.

In 1934, YUKAWA created the meson theory of nuclear forces: In Coulomb's law, e_1e_2/r^2, the quantities e_1 and e_2, according to the quantum-electrodynamic description of the interaction, are a measure of the capacity of the two particles to emit or absorb the photons that mediate the force; the factor $1/r^2$ comes purely from the geometry. If we assume with YUKAWA that particles with a finite rest mass serve to mediate the nuclear force, then instead of e_1 and e_2 we have to substitute two analogous quantities g_1 and g_2, as the measures of the strength of the emission and absorption of these particles, and the geometric factor must be multiplied by the exponential $e^{-r/r'}$ ($r' = \hbar/mc$). We see at once that the finite mass of the mediator particles is responsible for the short range of the nuclear forces. If we consider that this range is about 10^{-13} cm, then from $\lambda\!\!\!^{-}/mc \approx 10^{-13}$ cm, we find that the mass of the mediator particle should be a few hundred electron masses.



▲ **Figure 5.141** A temporary system of elementary particles of satisfying simplicity as of 1947. The photon (γ) is responsible for the electromagnetic interaction. The nucleus is composed of protons and neutrons, and with the addition of electrons, one has an atom. The neutrino (ν) plays a role in β decay, and the π mesons account for the binding of the nucleons. The μ meson is still without a role. Of these particles, γ, p, e, and ν are stable; the rest decay according to the following schema:

$$n \xrightarrow{918\ \mathrm{s}} p + e^- + \bar{\nu}_e \quad [100\%]$$

$$\pi^+ \xrightarrow{2.2\times10^{-6}\ \mathrm{s}} \begin{cases} \mu^+ + \nu_\mu & [99.99\%] \\ e^+ + \nu_e & [0.01\%] \end{cases}$$

$$\pi^0 \xrightarrow{0.83\times10^{-16}\ \mathrm{s}} \begin{cases} \gamma + \gamma & [98.85\%] \\ \gamma + e^- + e^+ & [1.15\%] \end{cases}$$

$$\pi^- \xrightarrow{2.6\times10^{-8}\ \mathrm{s}} \begin{cases} \mu^- + \bar{\nu}_\mu & [99.99\%] \\ e^- + \bar{\nu}_e & [0.01\%] \end{cases}$$

$$\mu^- \xrightarrow{2.2\times10^{-6}\ \mathrm{s}} e^- + \bar{\nu}_e + \nu_\mu \quad [100\%]$$

$$\mu^+ \xrightarrow{2.2\times10^{-6}\ \mathrm{s}} e^+ + \nu_e + \bar{\nu}_\mu \quad [100\%]$$

discovered, first the charged π^+ mesons in cosmic radiation, by Lattes, Occhialini, and Powell, and then after the Berkeley synchrocyclotron began "production" of pions in 1948, a detailed investigation of their properties could be undertaken. Thus, for example, in 1950, the transformation $\pi^\circ \rightarrow 2\gamma$ was discovered simultaneously in cosmic radiation (Carlson and colleagues) and in the cyclotron at Berkeley (Bjorklund and colleagues). With the discovery of π mesons, it seemed as though theory and experiment were essentially in harmony, that the big questions had been clarified, and that the theoreticians could now take a breather to work out the finer details (Figure 5.141). But already in the year of the hoped-for closure, in 1947, new heavy particles were found in cosmic radiation (Rochester and Butler). By 1953, it had become clear that there not only exist particles—called hyperons—that are heavier than nucleons, but that there also exist K particles, or kaons, whose masses are between the mass of nucleons and that of mesons. If we look again at Figure 5.135, we can see the unprecedented ferment of the 1950s in particle physics. One after another, new particles were discovered, including the antiproton and antineutron. Nor were the theoreticians idle: They were busy introducing new types of quantum numbers to characterize the given elementary particles and setting up their associated conservation laws.

But the sensation of the 1950s was undoubtedly the proposal by Lee and Yang that parity symmetry in weak interactions may be violated, a prediction that was then experimentally proved in 1957 by Wu and colleagues. The significance of this discovery was reinforced by the fact that, contrary to custom, Lee and Yang were awarded the Nobel Prize immediately after the experimental proof of parity violation, in 1957.

In the 1960s, the task physicists faced was to bring order to the plethora of already accumulated and constantly accumulating experimental results. It may be counted as a success of these theoretical works that the existence of a new elementary particle, the Ω^- hyperon, was predicted, and it was found in 1964. Also, it was during these years that the quark theory was born, according to which the multiplicity of elementary particles could be reduced to three fundamental particles.

Looking at Figure 5.142, we see that all the elementary particles mentioned thus far can be arranged according to well-known "classical" characteristics—mass, spin, and charge—into more or less organized groups. There are light particles (leptons), heavy particles (baryons), and finally, particles with intermediate mass (mesons). In this division according to mass, the spin of the particles gives us an indication of where we should draw the line between the groups: The leptons and baryons have half-integer spin, and the mesons have integer spin (to be precise, spin 0). The former are fermions, and the latter are bosons, just like the photon, whose spin is equal to 1. The charge of the particles can be $+e$ or $-e$, but it can also be neutral. The assignment of baryons and mesons into the group of hadrons is because of their participation in the strong interaction, whereas for the leptons, the weak interaction is characteristic. But more about this later.

We can truly no longer speak of a "disquietingly simple" basic structure of the universe. But things become even more complicated. By the beginning of the 1950s, Fermi was already investigating the interaction between π mesons and protons and found that the reaction cross section as a function of energy at a particular energy E_0 shows a bell-curve-like local maximum with a particular width, implying a resonance phenomenon that is known in nuclear physics and can be described by the Breit–Wigner formula:

$$\sigma(E) = \sigma_0 \frac{(\Gamma/2)^2}{(E_0 - E)^2 + (\Gamma/2)^2}.$$

From the location E_0 of the maximum and the characteristic of the width Γ ($\Gamma\tau \approx \hbar$), we obtain the mass and lifetime τ of the "excited" particle: τ is of order of magnitude 10^{-23} seconds. In the early days, these

resonances were categorized as the excited states of the relatively stable "real" elementary particles, whose lifetimes of roughly 10^{-10} seconds loom infinitely long in comparison with 10^{-23} seconds. Great hopes were placed in the investigation of resonances: In the study of atoms and nuclei and in the clarification of their structure, excited states have played an important role. But detailed investigations also showed that the "real" or "not real" distinction has no theoretical meaning: The usual way of saying this is that the distinction contradicts "particle democracy." Indeed, the only difference is that resonances decay via the strong interaction, whereas the weak interaction is responsible for the decay of the others.

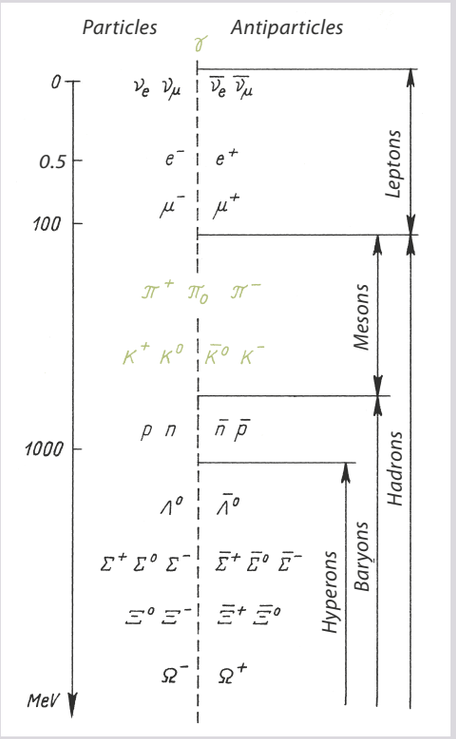
In 1964, there was once more a glimmer of hope: the possibility of a synthesis that could be, if not final, at least enduring. But what happened instead was something like 1947 all over again: Both theoretical arguments and experimental results cropped up that required more work. Nonetheless, the basic idea turned out to be correct, and so it is helpful to divide the discussion into the period before 1964 and the period thereafter.

5.5.3 *A Few Words about Cosmic Rays*

The crucial role of cosmic radiation in extending our knowledge of the elementary particles becomes clear when we look at Figure 5.135. We can see that until 1953 new particles and interactions were discovered primarily from the investigations of cosmic radiation. The explanation for this is very simple. In the interaction processes of the elementary particles, we are dealing in most cases with very large energies, several GeV (giga electron volts). It was only in the second half of the twentieth century that it became possible to produce such energies in the laboratory. Indeed, in the figure we see that after the mid 1950s the focus of experimental particle physics shifted to investigations of those phenomena that could be created by using large accelerators. The study of cosmic rays nevertheless remains important even today for two reasons: First, one can find particles in cosmic rays with energies that we have no hope of attaining with any accelerators in the near future (cosmic particles with energies of 10^{18} eV = 0.16 Watt seconds have been found; this is already a macroscopically significant value!), so that the observation of the effects of this kind of radiation might yield important new information. Second, the importance of cosmic rays as messengers from outer space remains, because they are our witnesses to events in distant regions of the cosmos.

For the sake of completeness, we have indicated some of the stages in the investigation of cosmic radiation. It is customary to ascribe the discovery of cosmic radiation to Hess (Quotation 5.49), who in 1912 conducted balloon experiments in the atmosphere up to an altitude of 5000 meters. He found that the ionization of air increases with altitude. The background ionization of air was measured by the time it took for a well-insulated electroscope to lose its charge. The reason for ionization was assumed to be the radioactivity in Earth's mantle. The external origin of the ionizing radiation was accepted only from the mid 1920s, and that is also the time when the term *cosmic rays* came into use. The most exciting questions—namely, what are the primary components of cosmic rays? or in other words, what are the characteristics of the particles that come from the cosmos, and what kinds of particles arise from the interaction of these primary rays with the atmosphere?—were basically solved by the beginning of the 1940s (Figure 5.143).

Today, Earth-orbiting satellites make it possible to measure the primary components of cosmic radiation *before* they enter the atmosphere, so now we have a more precise answer to the above question. It has been determined that the composition of this radiation is similar to the composition of our galaxy: It consists primarily of protons (92%), helium nuclei (6%), electrons (1%), and gamma quanta (0.1%),



▲ **Figure 5.142** The classification of stable and relatively stable elementary particles (half-life 10^{-6} to 10^{-20} seconds) according to simple "classical" principles. The values given on the vertical axis in mega electron volts are meant to give only a rough indication of the masses.

The spin of the particles shown in color is either 1 (photon) or 0 (mesons). Leptons are distinguished from hadrons by their interactions: For the leptons, the weak interaction is characteristic, whereas the hadrons interact strongly, although they also participate in weak interactions. The hyperons are hadrons whose masses are greater than those of the nucleons (protons and neutrons).

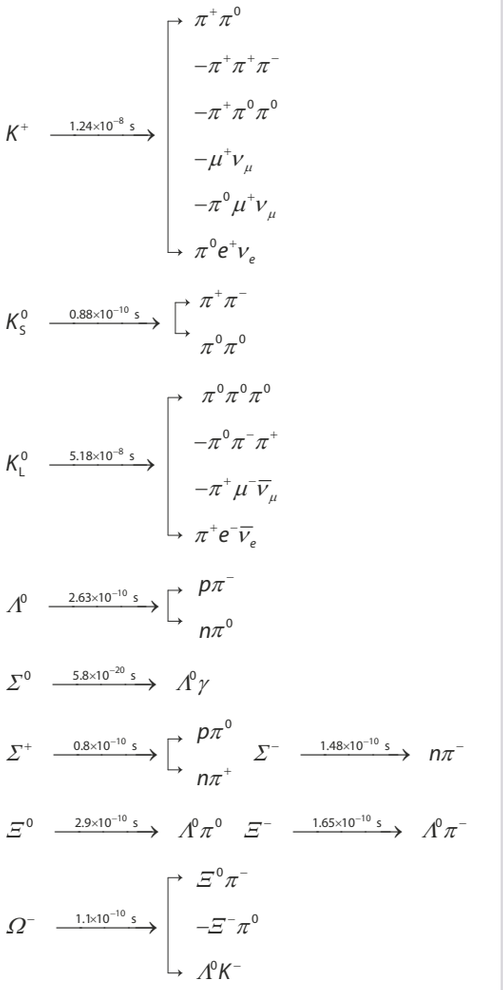
The dashed line separates the particles from the antiparticles. If a particle lies on the line (π^0, γ), then it is identical to its antiparticle. The masses and spins of the antiparticles are the same as the masses and spins of the corresponding particles; however, the electrical charges, the baryon charges, the lepton numbers, and the strangeness appear with opposite signs.

The K meson has some special properties: The strangeness of K^+ and K^0 is +1, whereas that of K^- and $\bar{K}^0$ is −1. Since K^0 and $\bar{K}^0$ are distinguished in their strangeness, they behave like different particles in strong interaction processes. However, under weak interaction, the transformation of the particles into each other is possible (the transformation of other particles into their antiparticles is forbidden by a strict conservation law), so that mixed states of the two particles can occur. The state $K_L^0 = 1/\sqrt{2}(K^0 - \bar{K}^0)$ has a longer half-life (5.2×10^{-8} s) than the state $K_S^0 = 1/\sqrt{2}(K^0 + \bar{K}^0)$, for which the half-life is 0.88×10^{-10} s); below, we show a characteristic decay process. The peculiar behavior of K mesons plays an important role in symmetry violation (see Section 5.5.9).

continued on next page

Figure 5.142 *continued*

The decay schemas for the particles not already listed in Figure 5.141:



Quotation 5.49

The best explanation of these observations may be to assume that highly penetrative radiation enters our atmosphere from above and causes ionization in detectors placed even at the lower layers. The intensity of the radiation changes—it seems—from one hour to the next. Because no diminution of the radiation was observed during balloon flights undertaken by night or during a solar eclipse, one can scarcely view the Sun as the cause of this hypothetical radiation ...

—Viktor F. Hess, "Observations of Penetrating Radiation in Seven Free Balloon Flights," 1912

as well as atomic nuclei of heavier elements up to iron and beyond. There are also detectable but negligible amounts of positrons and antiprotons.

The composition of the primary components rules out certain theories about the origin of cosmic rays: It cannot be interpreted as some sort of leftover from the big bang because of the presence of heavy nuclei. Nor can it be the consequence of some sort of high-energy explosions of old stars because, in that case, the number of heavy nuclei would have to be greater.

Here we can mention Fermi's theory about the origin of cosmic radiation that was popular in the 1950s: The ions of midrange energies emanating from stellar atmospheres might have gradually attained their enormous energies during their wanderings through space by "collisions" with the magnetic fields of moving macroscopic objects (e.g., plasma clouds, shock waves of supernova explosions).

More recent thinking assumes that the particles acquire their final energy in highly localized regions of the universe (in quasars, pulsars, etc.) by a mechanism that is not well understood (e.g., in very strong, rapidly changing magnetic fields).

5.5.4 *Particle Accelerators and Detectors*

Accelerators have enormous advantages over cosmic radiation as a source of particles for use in experiments. With accelerators the experimenter has the ability to choose and regulate the energy of the particles as well as the number of particles—more precisely, the particle current—which is greater, and thus the occurrences of the expected phenomena will be more frequent. Finally, it is also advantageous to have control over the entire setup, including the detectors. The great disadvantage of accelerators is, of course, their enormous size, with the additional requirement of the most demanding technical precision. Their construction and operation entails huge costs, which understandably brings political issues into the picture.

This raises the question of why physicists continue to strive for higher and higher particle energies. We already know that the energy binding the electron to the atom is on the order of several electron volts. To remove an electron from the atom, particles accelerated to energies between 10 and 100 electron volts would be sufficient. In nuclear physics, the corresponding energy is roughly 1 to 10 million electron volts. In contrast, to obtain an antiproton, the accelerated particle would have to have at least the rest energy of an antiproton (about a billion electron volts, or in other words, 1000 mega electron volts = 1 giga electron volt = 1 GeV). This minimal value, however, is insufficient for a number of reasons. First of all, most elementary particles—because of the various conservation laws of particle physics—can come into existence only in pairs or in larger groupings. There is also a "classical" reason from everyday physics. When an accelerated particle collides with a stationary one, according to the law of the conservation of momentum, any newly created particles will also have a velocity; therefore, a part of the kinetic energy of the incoming particle will show up as the kinetic energy of the created particle. This portion of the original energy must be viewed as lost for the purposes of particle creation. The situation at higher energies is even worse because of the relativistic increase in mass. The accelerated particle, with its larger mass colliding with the stationary particle with smaller mass, has a smaller and smaller fraction of its energy available for the reaction. This problem is solved by having the particles collide head-on, in which case the resulting momentum is zero, leaving the total energy available for the desired purpose.

We have already spoken about the "classical" accelerators of nuclear physics, namely, the linear accelerators (cascade generators, Van de Graaff generators) and

the basic form of the cyclic-orbit accelerators, the cyclotron. We know that the working principle of the cyclotron rests on the condition that the orbital period of a charged particle in a magnetic field is constant ($T = 2\pi m/qB$), that is, independent of its velocity. However, if we are working in the relativistic domain, the mass m does not stay constant but grows according to the formula $m = m_0\sqrt{1 - v^2/c^2}$. Consequently, the particle will not arrive at the correct time in the constant-frequency electric field of the electrodes and will not be able to accelerate further.

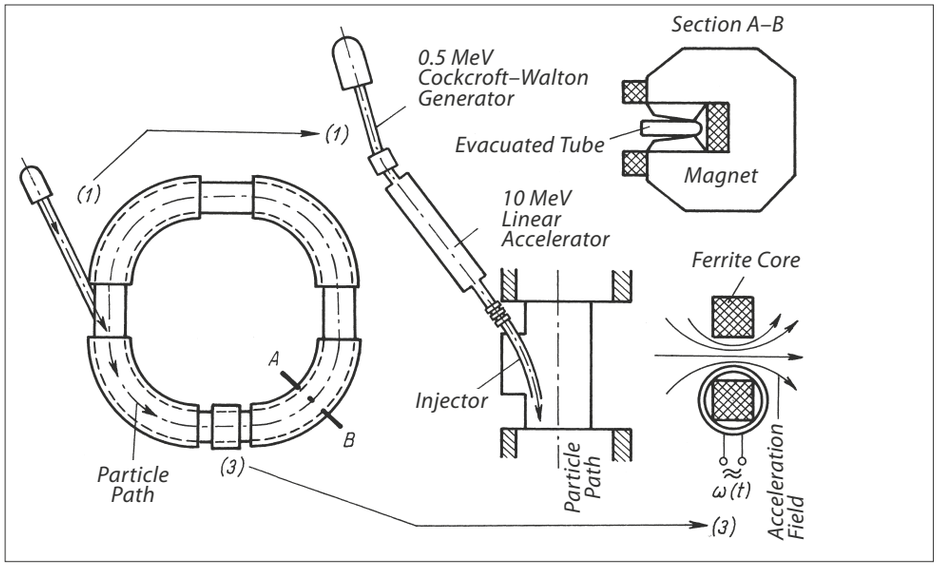
Vladimir Veksler (USSR, 1944) and Edwin McMillan (USA, 1945) independently came up with the principle of phase stability for increasing the attainable energy. The resulting synchrocyclotrons, which are, in effect, frequency-modulated cyclotrons, raised the energy levels to about 500 MeV.

At the beginning of the 1950s, developments took on a new direction. The Norwegian physicist Rolf Wideröe had specified the working principle of the betatron (electron accelerator) in 1928, in which the idea was to accelerate electrons in an alternating magnetic field in circular orbits of constant radius—but it was not until 1941 that Kerst succeeded in its implementation.

The combination of a variable magnetic field, variable frequency, and new methods of strong focusing made it possible to break through into the realm of giga electron volts beginning in the 1950s. Some dates: Dubna 1957: 10 GeV; CERN 1959: 28 GeV; Serpuchow 1967: 76 GeV; Batavia, USA 1973: 300 GeV; CERN 1976: 400 GeV (Plate XXVII), to name only the most significant.

Figure 5.144 shows a diagram of the cosmotron (Bevatron).

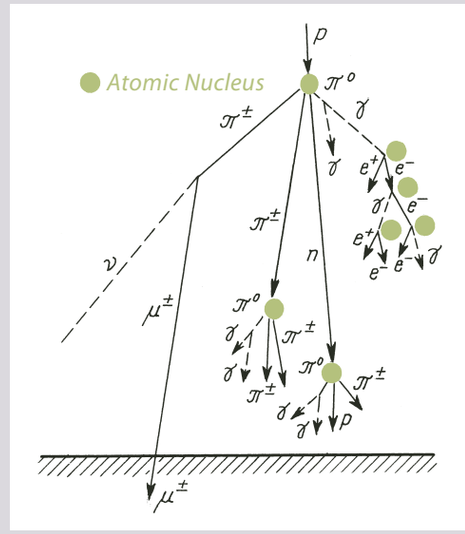
Some of the later accelerators are *SPS*, CERN: 900 GeV, 1986; *Tevatron*, Fermilab: 2000 GeV = 2 TeV, 1987; *SLAC* (e^+, e^-), Stanford: 100 GeV, 1987; *LEP.I.* (e^+, e^-), CERN: 100 GeV, 1989; and *HERA*, Hamburg: (e, p) 320 GeV, 1991.



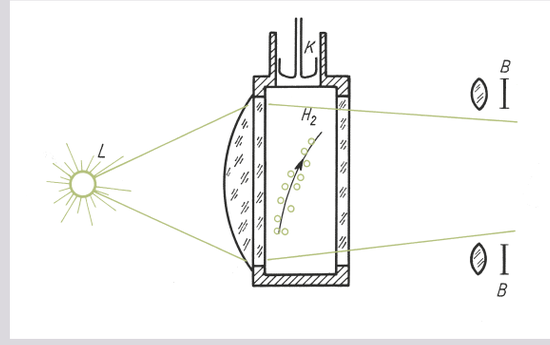
▲ **Figure 5.144** Principal parts of a Bevatron used for accelerating particles to energies of several giga electron volts (see also Plate XXVII).

The dimensions, costs, and feasibility of an installation are determined by the greatest achievable electric field strength. Today, this is about 0.1 GeV/m. This may well increase by a factor of 10^3 in the future if a whole new generation of accelerators (laser beat-wave plasma accelerators) can be constructed.

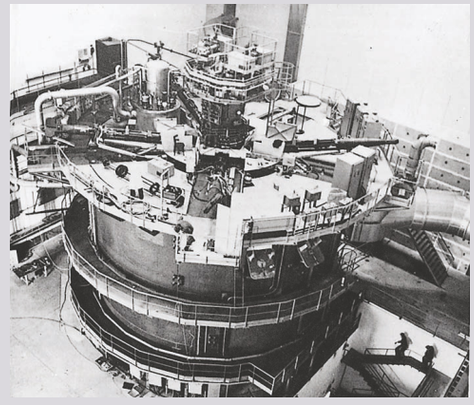
A few words about detectors of higher-energy particles. In Figure 5.135, the detection methods used for the discoveries of the individual elementary particles are



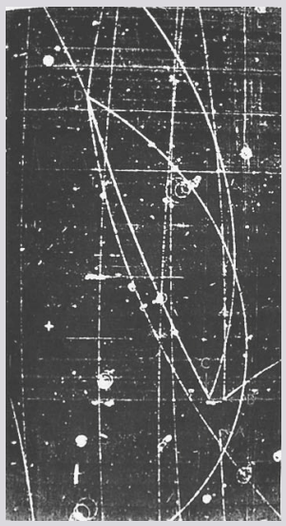
▲ **Figure 5.143** A cosmic ray air shower produces a cascade of transformations of elementary particles into one another.



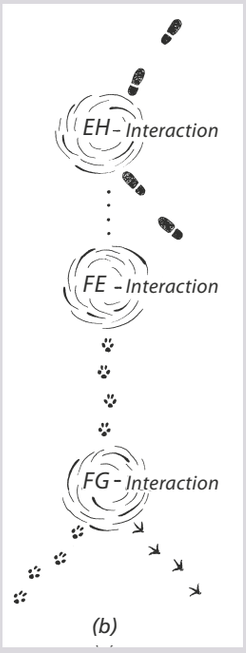
▲ **Figure 5.145** Schematic diagram of a bubble chamber.



▲ **Figure 5.146** Photograph of one of the largest bubble chambers in the world.



(a)



(b)

▲ **Figure 5.147** **(a)** Unusually clean and beautiful photograph from a bubble chamber, although, even here, there are some traces of background noise that do not belong to the process under investigation. The solution for the events seen here is given in Figure 5.157. The thickness of the traces and the curvature of the paths in the magnetic field are carriers of important information.

(b) One may compare the methods that are used in particle physics for identifying elementary particles and interaction processes with those for tracking animals in the wild. From the traces on the figure, one can identify the interacting particles (fox and goose) and establish the interaction laws between the two (the number of geese decreases; $\Delta G = -1$; the number of foxes remains constant; $\Delta F = 0$). The strengthening of the F trace after the FG interaction hints at the mass of G ($\Delta m_{fox} = m_{goose}$). As a result of the interaction, the fox moves into an excited state (its mass has increased), but it will return to its ground state after a certain time constant. From a break in the trace, we conclude that a new interaction process FE is involved; the new particle, which leaves no trace, we may call an "eagle." After interaction EH with a hunter who shoots the eagle and takes it with him, the mass of the eagle can be determined ($m_{eagle} = \Delta m_{hunter} - m_{fox}$).

We recognize the mirror symmetry of the FG interaction and can also establish the law for the irreversibility of the direction of time. However, the discovery of some rare occurrences in which this law is violated (geese that have been captured and carried off freed themselves and ran away) caused a great sensation.

also shown. It is clear that until the middle of the twentieth century, the most important tool was the Wilson cloud chamber. The first to use the cloud chamber for the investigation of cosmic rays was Skobelzyn in 1927. An improved version was the continuously operating diffusion chamber, and also the bubble chamber (Figures 5.145, 5.146) constructed by Glaser in 1952, which can be seen as a variant of the cloud chamber, but where the trace of a particle is visible as a trail of bubbles along its path (Figures 5.147, 5.148). One advantage of the bubble chamber over the gas-filled cloud chamber is the shortening of the very-high-energy particle's path in the denser medium, and also that liquid hydrogen in the bubble chamber already contains one of the most interesting target particles, namely, the protons. Figure 5.149 shows a spark counter (Keuffel, 1948), which led to the spark chamber. This type of detector, in a number of variations, has found more widespread application (Figure 5.149).

The first method chronologically for detecting nuclear radiation, the blackening of light-sensitive plates, had a rebirth in 1947. Traces of particles in the form of grains of silver became visible in photographic emulsions of suitable composition and thickness when they were developed. The most complex transformations could be followed with these traces (Powell, 1946). Along with this, even if only for a brief period of time, came the laborious, monotonous, eye-straining visual evaluation of the results, performed mostly by lab technicians, just as in the early days of the experiments with the counting of scintillations. Finally, we must not overlook about the extensive application of the Geiger counter and ionization chambers, and the more recent emphasis on two new types of detectors. One is the scintillation counter, which uses the well-established scintillation phenomenon but combines it with a photomultiplier (Kallmann, 1947), and the other is the Cherenkov counter, which is based on the phenomenon that when an electrically charged particle moves in a transparent medium with a velocity that is faster than the speed of light in that material, the particle will emit light, the so-called Cherenkov radiation. While this was discovered by Cherenkov in 1934 (Figures 5.150, 5.151, 5.152), and interpreted theoretically shortly thereafter by Frank and Tamm, it was only used for particle detection after the Second World War. Cherenkov's work was eventually recognized with a Nobel Prize. These detectors are extremely significant because, with their help, the speed of a particle can be determined independently of its other properties from the distribution of the directions of the Cherenkov radiation.

Auxiliary electronic circuits have had a significant role in the use of such detectors and for expanding their range of application. One pathbreaking idea was to use coincidence circuits for selecting out the events of interest. This first occurred in 1932. After the Second World War, various amplitude analyzers for the evaluation of optical and electrical signals by particle detectors proved to be indispensable, and later, information from the detectors had begun to be fed directly into computers, whose software carries out the required calculations.

Figure 5.153 shows one of the crowning achievements of the post-war detection technology in particle physics, the experiment of Reines and Cowan in 1956 for a direct proof of the existence of the neutrino.

5.5.5 *Fundamental Interactions*

At this point we offer a summary of what kind of picture can be drawn from the information obtained from the experiments about the elementary particles, their classification, and their interactions. We start with the last of these.

Let us once again recall the classical model of the interaction between two particles, as in Rutherford's scattering experiment. For our current considerations, it

is of no significance that we are dealing with the motion of an alpha particle in the electric field of the atomic nucleus. What is essential is only that a force satisfying certain laws exists between the particles, in this particular case, the electrical Coulomb force, which acts in inverse proportion to the square of the distance between the particles. If we substitute this force into Newton's equations of motion, we can determine the course of motion—the path and the position of the particle at every instance—by integrating the equations of motion. The quantitative treatment of the interaction is thereby reduced to the establishment of a force law.

It is not possible to come up with such a clear picture of the interaction between elementary particles, nor can we follow the interaction processes in all of their details. This is also the reason that we prefer the more general term "interaction" rather than "force." As is now understood, there are the following four basic interactions: *the gravitational interaction, the weak interaction, the electromagnetic interaction, and the strong interaction* (Figures 5.154, 5.155). To get to know these better, let us take the well-known interactions of gravitation and electromagnetism, about which we have already spoken extensively, and examine them from the point of view of particle physics. We first write down the force laws:

$$F_{ik} = G\frac{m_i m_k}{r^2}, \qquad F_{ik} = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_k}{r^2},$$

$$G = 6.67\times10^{-11}\ \mathrm{Nm^2kg^{-2}}, \qquad \varepsilon_0 = 8.854\times10^{-12}\ \mathrm{As/Vm}.$$

We can immediately see that both interactions depend on distance in exactly the same way and that, in principle, one should be able to detect these forces at any distance. We say that the effective range of this type of force is infinite. Thus, we have introduced a characteristic of forces and, more generally, of interactions: their range. Another characteristic of interactions is their strength or intensity. At first glance, it would seem to be very difficult to compare the intensities of two forces, for how should the masses and charges be chosen for the purposes of comparison? But for calculating the effect of the electrical force, there is a natural unit: the charge of the electron. For the effect of gravitation, we can choose the masses of the proton and of the electron. We can then consider the case of a hydrogen atom:

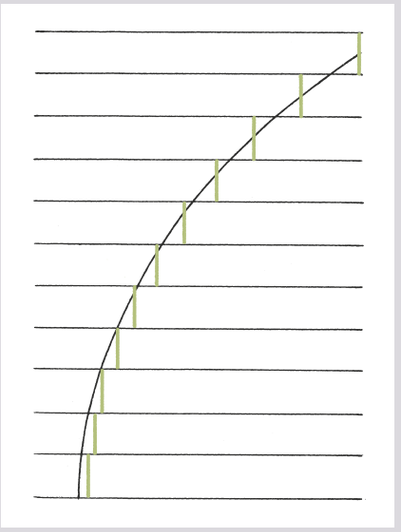
$$F_{\text{gravitation}} = 6.67\times10^{-11}\frac{1.67\times10^{-27}\times9.1\times10^{-31}}{r^2} = \frac{1.01\times10^{-67}}{r^2},$$

$$F_{\text{electric}} = \frac{1}{4\pi\times8.85\times10^{-12}}\frac{\left(1.6\times10^{-10}\right)^2}{r^2} = \frac{2.3\times10^{-28}}{r^2},$$

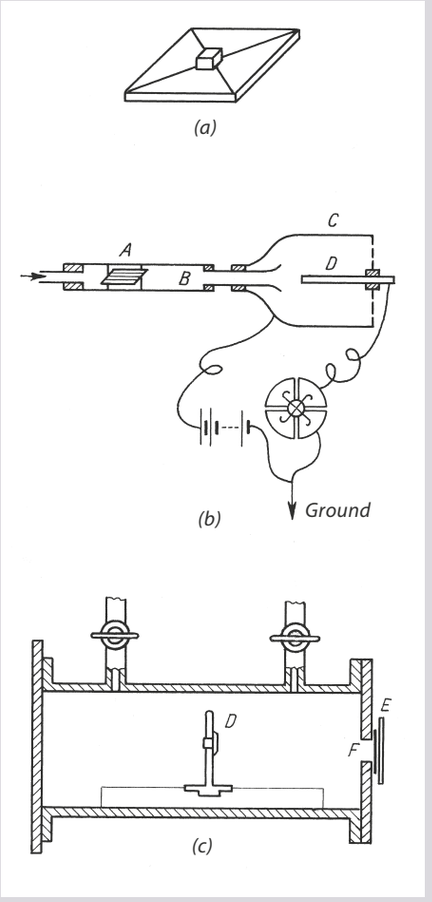
giving us the ratio

$$\frac{F_{\text{gravitation}}}{F_{\text{electric}}} = \frac{1.01\times10^{-67}}{2.3\times10^{-28}} = 4.4\times10^{-40}.$$

It is immediately obvious that the electric interaction is far more intense than the gravitational interaction by many orders of magnitude and that, consequently, gravitation must be viewed as an extremely weak interaction. This seems to go against our everyday experience, but there is a very simple explanation: Though there is only one kind of gravitational force, namely, attractive, the electrostatic force, due to the existence of positive and negative charges, can be both attractive and repulsive. This means that the electric interaction may be shielded, so to speak.



▲ **Figure 5.148** Diagram of a spark counter.

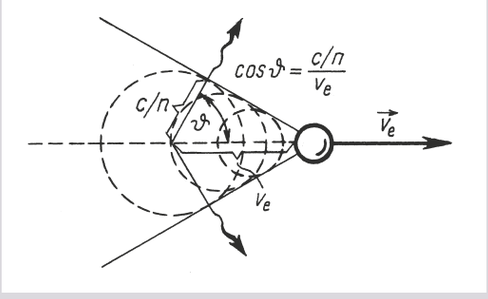


▲ **Figure 5.149** The multiwire proportional chamber; one of the particle detectors developed by Georges Charpak (Nobel Prize, 1992).



▲ **Figure 5.150** Pavel Alekseevich Cherenkov (1904–1990): Completed studies at the State University of Voronezh; from 1930 at the Physical Institute of the Academy of Sciences of the USSR; professor at the Moscow Engineering Physics Institute.

In 1934, on a suggestion of S. I. Vavilov, Cherenkov investigated γ luminescence in various solutions and noted a weak blue light with properties different from those of the luminescence. Vavilov suggested that this radiation is the effect of fast Compton electrons that are the results of collisions with the γ rays in the medium. A theoretical explanation based on classical electrodynamics was given in 1937 by I. E. Tamm and I. M. Frank, and in 1940, V. L. Ginzburg obtained a similar result using quantum theory.



▲ **Figure 5.151** The condition for the appearance of Cherenkov radiation is that the velocity of the particle be greater than $u = c/n$, where u is the speed of light in the given medium and n is the index of refraction. The effect is analogous to the shock waves around supersonic aircraft. From the figure we see that in this approximation the angle ϑ of the emitted light ray and the speed of the particle are related by the formula $\cos\vartheta = (c/n)/v_e$.

The macroscopic bodies in our environment are in most cases electrically neutral and therefore exert no outward electrical force. The gravitational force, however, cannot be shielded, and therefore we can rightly say that its range is infinite. In contrast, the range of the electrical forces is normally on the order of atomic distances, or roughly 10^{-10} meters. We may therefore conclude that, despite its weakness, the gravitational interaction is preponderant on the cosmic scale.

The most characteristic property of the gravitational interaction is its universality; it is always present alongside the other interactions. In general, a characteristic of interactions is in fact their selectivity—the total lack of which we can call universality—where selectivity means that certain interactions are linked to certain properties. The fundamental building blocks of the universe can be classified based on these properties. For example, the electrical interaction makes it possible to classify particles as positively or negatively charged, or as neutral.

As we have seen, the gravitational interaction plays no role in the investigation of elementary particles. This does not mean, of course, that the particularly strong gravitational fields in the large cosmic laboratories of nature do not influence processes among elementary particles. We see in Figure 5.154 a question mark for the gravitational interaction, which indicates that we have no detailed knowledge about the underlying processes.

A closer look at the electromagnetic Coulomb interaction can aid us in our understanding of the other interactions. The theories of its nature developed along a remarkably spiral path. According to the earliest ideas, an interaction is possible only between bodies in immediate contact. In contrast, when the Coulomb law was established, the role of the transmission medium seemed unimportant. The Coulomb law was at first understood as a law of action at a distance. Faraday introduced the idea of force lines, which describe the force in the older spirit of immediate contact. The field lines, like stretched flexible threads, connect the interacting particles. This is interpreted physically by saying that a charged particle creates an electric field around itself that exists independently of any other particle. So it can be viewed as a "ready state," and the interaction will be mediated by this state. The charge may be viewed as the measure of the ability to create the field. The existence of fields that tear themselves away from accelerating charges (electromagnetic waves), then the experimentally proven particle nature of these waves, and finally quantum electrodynamics changed the picture of the interaction again: One charge emits photons while moving; the other absorbs them. The acts of emission and absorption constitute the interaction. Analogously, the electrostatic interaction and even the electrostatic field can be reduced in this picture to a continuous emission and absorption of photons by the electrical charges. To indicate that this is a metaphor, one speaks of virtual photons. The electric charge can now be viewed as a measure of the intensity with which a particle emits and absorbs quanta of the electromagnetic field. Here we have to pay attention to the fact that the charge fulfills a hidden double function, which is particularly interesting because it is not necessarily the case that both roles in an interaction are played by the same characteristic. On the one hand, the charge determines the intensity of the interaction, while on the other hand, it also satisfies the law of conservation of charge, whereby the algebraic sum of the charges of the particles taking part in a reaction remains constant. In the other interactions, we can find one quantity that is characteristic of the intensity of the interaction, and a second quantity that remains constant throughout the transformations.

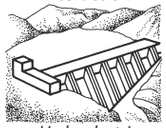
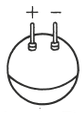
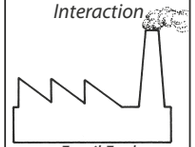
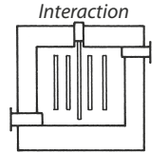
At this point, we view an interaction as follows: An elementary particle emits or absorbs another particle, characteristic of the interaction, and it is this particle that

mediates the interaction. Thus, an interaction, even a static one, can be reduced to the elementary acts of creation and destruction.

The interaction between nucleons, the strong interaction, is mediated by the π mesons, as shown in Figure 5.155. That it deserves the "strong" designation is evident from the fact that it is responsible for the cohesive structure of the nucleus, and we know that to transform this structure we need orders of magnitude more energy than for bringing about changes in the atomic shell.

As we have already mentioned, it is not easy to fit the weak interaction into this picture. Even the nature of the particles that mediate this interaction remained unexplained until quite recently. It seemed possible that the weak interaction might be local and that a mediating particle might not be required. However, interest soon shifted to a theory according to which vector bosons play the role of mediator particles (Klein, Lee, Wheeler, et al. 1949; Gell-Mann, Feynman, Marshak, Sudarsham, Sakurai 1958).

In Figures 5.154 and 5.155, we have also shown the extent to which the various interactions make a classification of the elementary particles possible. We see that the hadrons can take part in all interactions. The proton, for example, can be a partner in an electromagnetic interaction process, yet it can also, for example in nuclear transformations, interact strongly as well with another proton or with a neutron. The reaction $\bar{\nu} + p \rightarrow n + e^+$, inverse beta decay, is an example of the weak interaction

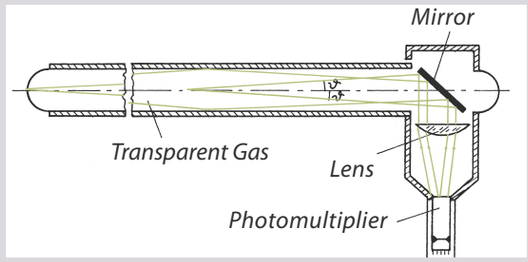
Engineering Form	*Gravitational Interaction*  *Hydroelectric Power Plant*	*Weak Interaction*  *Radioisotope Thermo-electric Generator*	*Electromagnetic Interaction*  *Fossil Fuel Power Plant*	*Strong Interaction*  *Atomic Reactor*
Phenomenological Description	$F_{ik} = G\frac{M_i M_k}{r_{ik}^2}$	$_{38}Sr^{90} \rightarrow {}_{39}V^{90} + e^-$	$C + O_2 = CO_2$	$_{92}U^{235} + n \rightarrow {}_{36}Kr + {}_{56}Ba + 3n$
Basic Phenomenon	*exchange of gravitons (?)*	*exchange of vector bosons*	*exchange of photons*	*exchange of pions*
Range	∞	$< 10^{-16}$ *cm*	10^{-7} *cm* (∞)	$< 10^{-13}$ *cm*
Intensity	10^{-38}	10^{-13}	10^{-2}	1
Characteristic Time	–	10^{-10}–10^{+8} *s*	10^{-20}–10^{-7} *s*	10^{-23} *s*
Subdivisions		*right-handed, left-handed*	*electrically charged, electrically neutral*	*hadrons, leptons*
Conservation Laws	?	*energy, momentum, angular momentum, electrical charge, baryon number, lepton number, CTP*		
			strangeness, parity, CP	*strangeness, parity, CP, isospin*

▲ **Figure 5.154** Summary of the characteristics of the four basic interactions.

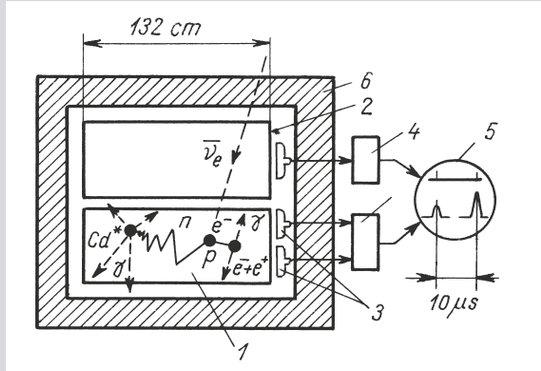
The intensity of an interaction is characterized by the coupling constant, which is closely related to the interaction energy. To obtain a dimensionless number, we divide by ɔc. Thus, for example, we obtain $e^2/\hbar c$ ($\approx$ 1/137, the fine-structure constant) for the intensity of the electromagnetic interaction, and for the gravitational interaction, we obtain $Gm_N^2/\hbar c$ ($\approx 10^{-39}$, where m_N is the nucleon mass).

This figure and Figure 5.155 reflect ideas from the mid 1960s. Since then the mediators of the weak interaction have been established (namely, the vector bosons $W^{\pm}$ and Z^0 with masses 79.5 GeV and 90 GeV, respectively). Also, the strong interaction is now seen as a residual of the color forces acting between the quarks that make up the nucleons.

The electromagnetic interaction also has an effect, of course, on macroscopically neutral objects when these are composed of electrically charged particles. In this sense, the neutrons are also only "macroscopically" neutral because they are composed of charged quarks. In fact, the neutron does possess a magnetic moment and can be affected by an electromagnetic field. In the end, only the neutrinos remain "insensitive" to the electromagnetic interaction.



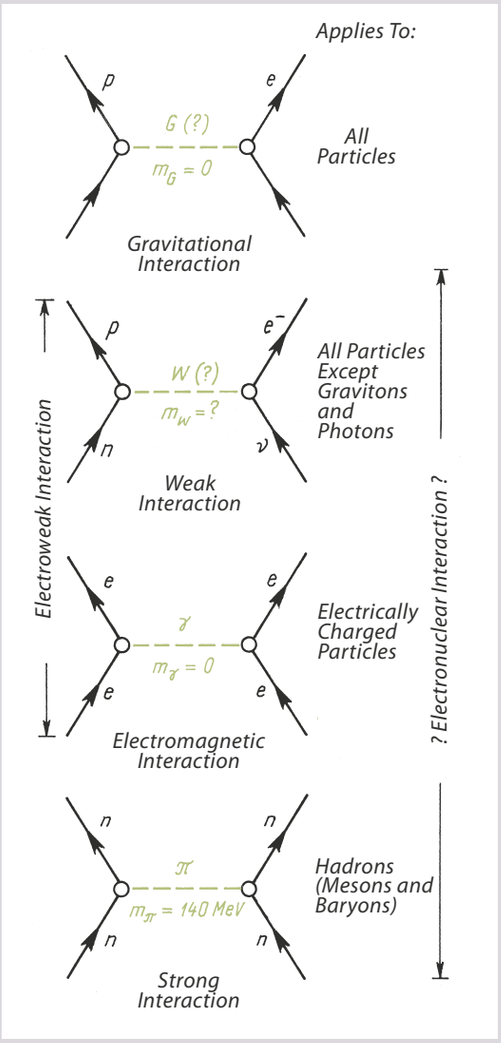
▲ **Figure 5.152** The figure shows how Cherenkov radiation can be used for particle detection. In the 70-GeV accelerator in Serpukhov, a particle with velocity close to the speed of light traverses a very long (about 10 m) gas-filled tube. The emitted radiation is reflected from the mirror-coated tube wall and then arrives via an optical system consisting of a mirror and a lens to a photomultiplier tube that amplifies the signal for the registration of the passage of a particle.



▲ **Figure 5.153** Principle of the experiment by Reines and Cowan (improved 1958 version):

1. 1400-liter liquid scintillation detector for detecting the antineutrino;
2. Scintillation detector to detect the cosmic background radiation; connected for anticoincidence with detector (1);
3. two photomultipliers connected for delayed coincidence;
4. electronics;
5. dual-trace oscilloscope;
6. protective shield made of lead and paraffin against the reactor's radiation.

An antineutrino exiting the nuclear reactor triggers the reaction $p + \bar{\nu}_e \rightarrow n + e^+$ with a hydrogen nucleus contained in the scintillator liquid in detector (1). The resulting positron is annihilated upon collision with an electron, resulting in two γ quanta; this event is detected by one photomultiplier. The resulting neutron is absorbed after a brief random walk lasting on average 5–10 microseconds by one of the nuclei in cadmium salts dissolved into the detector liquid. This event is detected by the other photomultiplier. The anticoincidence switching eliminates the influence of cosmic radiation because the lower detector is automatically switched off when the upper one records an event. The delayed-coincidence switching, in turn, selects the events that occur in the exact sequence described above. Measurement of the cross section gave the value of $\sigma = 11 \pm 2.6 \times 10^{-44}$ cm^2, which agrees well with the theoretically predicted value $\sigma = 10$–14×10^{-44} cm^2 (from the *Great Soviet Encyclopedia*).



▲ **Figure 5.155** Depiction of the fundamental interactions using Feynman diagrams (as of the 1960s).

Working out the details of the strong interaction (Quantum chromo-dynamics, QCD) and of the electroweak interaction was a task for the 1970s. These two together comprise the Standard Model.

The "grand unification," or the theory of the electronuclear interaction, is a project going on now. But the question of whether all four interactions can be subsumed into a Theory of Everything is for the twenty-first century to resolve.

involving a proton. We mention only in passing the fact, obvious from the remarks made above, that protons are also subject to the universal gravitational interaction.

When they carry a charge, leptons participate in the electromagnetic interaction; however, the typical interaction for this group of particles is the weak interaction.

Finally, in addition to the gravitational interaction, the gamma quanta participate only in the electromagnetic interaction, and the neutrinos only in the weak interaction.

5.5.6 *The Conservation Laws*

As ever more particles and their interactions were being observed following the end of the 1940s, the question naturally arose, given the theory, what kinds of new particles might be theoretically expected to be seen, and given these new particles, what would the new interactions be like? Simply from the conservation laws of physics, which are also considered valid in quantum mechanics—the conservation of energy, momentum, and angular momentum—we can already make a number of predictions as to which transformations are *not* possible. For example, the law of the conservation of energy forbids the spontaneous decay

$$p \rightarrow n + e^+$$

because the total mass of the reaction products would be greater than the masses of the initial particle. The conservation of momentum law excludes "single-particle" decays such as $K^+ \rightarrow \pi^+$ because for this to happen, the decay product would have to carry off the mass difference in the form of kinetic energy. The kinetic energy, however, is necessarily linked with momentum, and with only one decay product, there is no counter-partner that could compensate to make the total momentum change zero. This would remain true even if the decaying particle had a large velocity at the outset.

Let us add to the above the conservation law for electric charge, which states that the total charge of the particles participating in a reaction must remain the same before and after the reaction.

Detailed investigation of the transformations has shown that there are many reactions that do not occur even though we could expect them to occur, because they would satisfy all the conservation laws listed above. An example of such a transformation is

$$\pi^- + p \rightarrow K^- + \Sigma^+,$$

but we can also offer a simpler example, one that is much more significant for the existence of our world: In nature, the decay of a proton

$$p \rightarrow \pi^+ + \gamma$$

does not occur, although this decay would be permitted by all four of the conservation laws mentioned above. If the decay were actually possible, it would mean that the proton is not a stable particle. However, it is stable to the most extreme degree imaginable, having a life span certainly greater than 2×10^{30} years, greater, in fact, by many orders of magnitude than the entire age of the universe. The stability of the proton is a very strongly verified experimental fact, and this stability—along with the stability of the electron—underlies the stable existence of our material world and thus ultimately the existence of life.

Perhaps the greatest insight in particle physics was the realization that such reactions cannot occur because they would violate some other, new conservation law. We

may imagine that the particles carry some new kind of "charge," where this term is only a suggestion of a characteristic that is similar to the electric charge insofar as it also obeys a conservation law. For example, the proton decay discussed above is prohibited by the conservation law for *baryonic charge* (Wigner 1949, Stueckelberg 1938). The baryonic charge of the proton is $B = 1$ (for the neutron we also have $B = 1$, and for the corresponding antiparticle we have $B = -1$), but the π^+ meson has $B = 0$. We can now understand why this decay does not occur: The sums of the baryonic charges of the particles on the left- and right-hand sides do not agree. It is easy to verify that the reaction leading to the creation of an antiproton,

$$p + p \rightarrow p + p + \bar{p} + p,$$

satisfies the conservation laws if we assume that the energy of the bombarding proton is greater than 5.6 GeV.

Here, the question as to what further similar conservation laws could be stated arose as a matter of course. In addition to the electric and baryonic charge, there is also the *leptonic charge*, with an associated conservation law (discovered independently by György Marx, J. B. Zeldovich, and E. F. Konopinski in 1953 and confirmed experimentally in 1955 by Davis). But there were further complications. The theory could not explain why a muon decaying into an electron and a photon according to the formula

$$\bar{\mu} \rightarrow e^- + \gamma$$

is *not* possible. Until 1962, this question was called the μ–$e\gamma$ riddle, and it was only in that year that it was discovered that there are two types of neutrino, the electron neutrino and the muon neutrino, and consequently, two variants of the leptonic charge. The conservation law holds separately for each variant (Lederman, Schwartz, Steinberger).

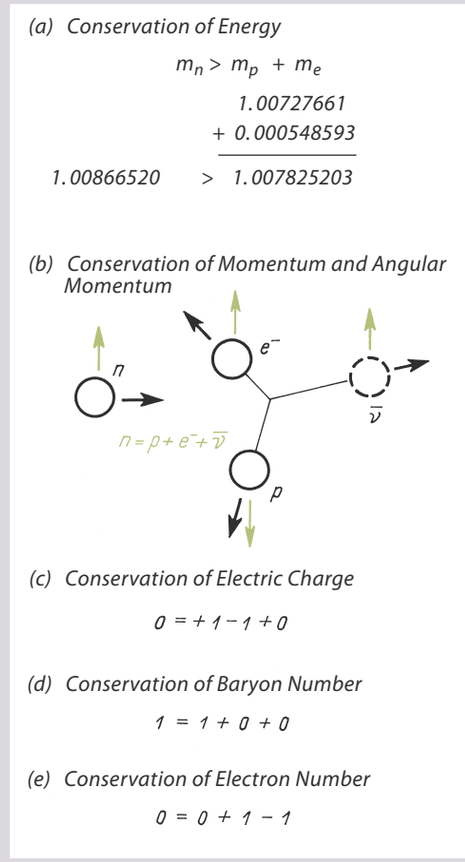
As part of the work at the beginning of the 1950s, another type of charge, strangeness, was introduced in order to explain why certain "strange" particles can be created only in pairs: Now we had a reason, namely that it is the only way they can satisfy the conservation law for strangeness (M. Gell-Mann and K. Nishijima).

Figure 5.154 shows the domains of applicability of the various conservation laws. We see that the conservation laws for energy, momentum, and angular momentum as well as those for electric, baryonic, and leptonic charges are rigorously satisfied for all interactions. It is also evident that the stronger the interaction, the greater the number of conservation laws that can be written for the resulting transformations.

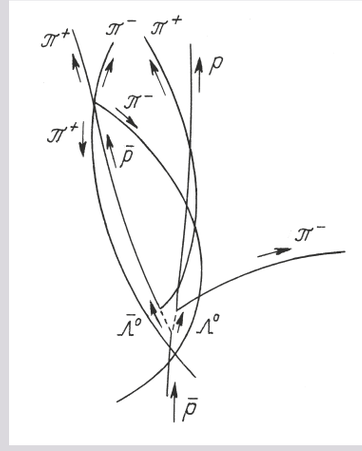
In Figure 5.156, we demonstrate how the conservation laws are satisfied in a very simple process, the decay of the neutron. (The beta decay of the neutron was observed in 1948 by Snell and Miller; in 1951, Robson measured its half-life.) In Figure 5.157, we can see the result of a bubble-chamber experiment that shows a complicated sequence of particle transformations. The tracks suggest that momentum is conserved. By identifying the particles with their individual path segments, we can check them against the conservation laws.

5.5.7 *Symmetry, Invariance, Conservation*

The introduction above of the new quantities that are actually quantum numbers—baryonic charge, leptonic charge, and so forth—may seem forced or ad hoc. Let us review how we arrived at them. Experiments acquainted us with a great



▲ **Figure 5.156** An illustration of the conservation laws as they apply to neutron decay (half-life is about 15 minutes).



▲ **Figure 5.157** Analysis of the bubble chamber photograph shown in Figure 5.147. The antiproton entering from below ($\bar{p}$) collides with a hydrogen atom (p) of the bubble chamber medium, where, from the reaction

$$p + \bar{p} = \Lambda^0 + \bar{\Lambda}^0,$$

a hyperon pair is created consisting of a lambda null particle (Λ^0) and an anti-lambda-null particle ($\bar{\Lambda}^0$). These particles are electrically neutral and therefore leave no trace. However, the

continued on next page

Figure 5.157 *continued*

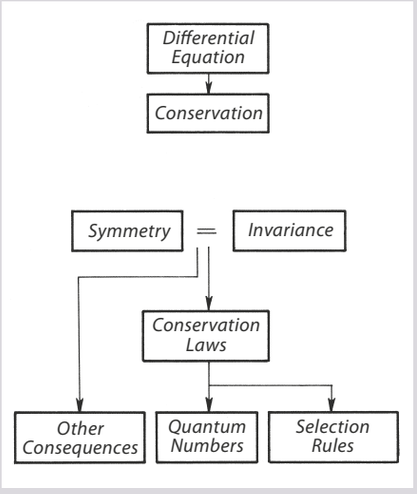
result of their spontaneous decay,

$$\Lambda^0 = \pi^- + p,$$

$$\overline{\Lambda}^0 = \pi^+ + \overline{p},$$

can be seen. The antiproton that arises again collides with a proton of the bubble chamber, resulting in the reaction

$$\overline{p} + p = \pi^+ + \pi^+ + \pi^- + \pi^-.$$



◀ **Figure 5.158** Conservation laws in classical physics and in particle physics.



▲ **Figure 5.159** The discoverers of the fundamental symmetry principles and symmetry violation.

Eugene Wigner (1902–1995): Born in Budapest, studied in Berlin. 1928–1930: Teaching position at the Berlin Technical University, then emigrated to the United States, to work at Princeton. In 1939, Einstein was persuaded by him, Leó Szilárd,

continued on next page

number of elementary particles, and we then made a list of the *conceivable* reactions that could take place between them. Then we eliminated those that contradicted the well-established conservation laws of physics—conservation of energy, conservation of charge, and so on. Looking at the remaining reactions, we see that some can be observed in nature, and some cannot. In order to be able to delete the latter from the list as being impossible, we came up with various conservation laws and said that such reactions do not occur precisely because they would violate one of these new laws.

If we take a closer look at the ideas behind the introduction of these new conservation laws, we have to conclude that these new results of twentieth-century physics are just as significant as the theory of relativity and quantum mechanics.

Notice that even the form of the question is novel: Instead of asking of a phenomenon *observed* in nature *how* it proceeds, we ask about a phenomenon *not observed* why it *does not* take place, in other words, why from among all possible reactions, some occur and others do not.

Let us now consider more closely, with the help of Figure 5.158, the route we followed to arrive at the conservation laws of classical physics. We began by formulating a differential equation describing a phenomenon, for example, Newton's equation of motion, and then concluded that certain quantities, such as energy and momentum, remain unchanged during the process. In other words, armed with the knowledge of how certain things change, we can infer the constancy of something else. In the physics of elementary particles, we gave up trying to follow the interaction processes in detail from the start. In the previous discussion, as well, we tried to establish relationships only between the characteristics of the initial particles and those of the particles produced in each transformation. But here the classical schema is completely useless. The question then becomes the following: Can the conservation phenomena be given a more general theoretical treatment? In other words, can we derive conservation laws without knowing every detail of the course of the events?

As was already known earlier but only emphasized with the advent of particle physics, the conservation laws of classical mechanics can be derived from simple assumptions about the structure of space. For instance, the law of conservation of momentum is a direct consequence of the homogeneity of space. This simply means that because no point in space is singled out, phenomena at one point will unfold exactly as at any other point. In the language of mathematics, we would say that the equations of classical mechanics are invariant with respect to the parallel translation of the coordinate system. Or in everyday language: If a physical process is measured in one laboratory, then the results should be the same in any other laboratory. But when we put it this way, it sounds so trivial—and recall that the objectivity of physics as a science depends on its credibility—that it becomes a challenge to our belief that a concrete law, such as the law of the conservation of momentum, can be derived from such premises. It can also be shown that the conservation of angular momentum is connected to the isotropy of space or, in technical terms, that the laws of nature must be invariant with respect to the rotation of the coordinate system. In everyday language: there is no singled-out direction in space to which reference must be made when we describe physical phenomena. The law of the conservation of energy, in turn, follows from the homogeneity of time; that is, the laws of physics must be invariant with respect to a shift in time. In simple terms, this means that a physical law recognized as correct today could be checked by anyone tomorrow and again found to be valid. This statement, too, seems to be completely trivial.

Let us pause and examine the above statements a bit more carefully. As an example, let us take the law of the conservation of momentum and, for the sake of simplicity, restrict our attention to two interacting particles. The interaction energy depends on the position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ of the two particles. The requirement that the interaction energy be invariant with respect to a translation of the coordinate system means that it is allowed to depend only on the difference of the two position vectors, since a translation would map the position vector $\mathbf{r}_1$ to the new position vector $\mathbf{r}_1 + \mathbf{r}_0$, and $\mathbf{r}_2$ into $\mathbf{r}_2 + \mathbf{r}_0$. The interaction energy U is thus a function of the distance between the two particles: $U = U(\mathbf{r}_1 - \mathbf{r}_2)$. The forces $\mathbf{F}_1$ and $\mathbf{F}_2$ acting on the two particles are then given by

$$\mathbf{F}_1 = -\text{grad}\,U(\mathbf{r}_1 - \mathbf{r}_2), \quad \mathbf{F}_2 = -\text{grad}\,U(\mathbf{r}_1 - \mathbf{r}_2) = \mathbf{F}_1.$$

But from this it follows that the sum of the forces that can effect a change in the total momentum vanishes: $\mathbf{F}_1 + \mathbf{F}_2 = \mathbf{0}$, and therefore the momentum is conserved

The conservation of energy law can be proved similarly.

The usual way to sum this up is by saying that the well-known conservation laws—originally derived with so much effort from the equations of motion—simply follow from the symmetry properties of space. Emmy Noether (1882–1935) as early as 1918 with the theorems now named in her honor demonstrated that every symmetry property has a consequent conservation law. In physics, we must add that symmetry has a somewhat different meaning than in everyday life. For example, we say that the human face is symmetric because the right half of the face is the mirror image of the left half. We regard a building as symmetric if we can locate a vertical line from which, whether going left or right we find the same arrangement of features. Of course, there are other mirror symmetries besides those relative to a vertical line. In our daily lives, we also use the notion of rotational symmetry. A figure is rotationally symmetric if it transforms into itself when rotated through a certain angle. Thus, for example, a square transforms into itself by a rotation through 90°. The circle was already regarded in ancient Greece as the most perfect of figures because it exhibits infinitely many symmetries, among them rotation about the center through an arbitrary angle.

More generally, we shall take the symmetry of an object to mean that some of its properties remain the same when some particular operation is performed on it.

In physics, the first detailed investigation of symmetries was understandably connected with the investigations of crystals. But this was a geometric, or one might say a descriptive, investigation of symmetry. The theory of relativity already recognized the characteristics of symmetry properties in providing a limiting frame in physics: From the isotropy of four-dimensional space, which also incorporates time, it derived the Lorentz invariance of the natural laws. Poincaré had already shown, that the symmetry operations carried out in four dimensions form a *group*, in the mathematical sense. This may well have been the first time that the mathematical tool known as *group theory* was used outside of abstract geometry to uncover general foundations for the laws describing physical processes.

The additional conservation laws that are valid for the transformations of elementary particles can similarly be reduced to symmetry properties, but these symmetries are not so easy to visualize. Let us take a simple example: the law of the conservation of isospin. The notion of isospin was introduced by Heisenberg to describe the charge-independence of the forces acting between nucleons. This can be theoretically explained if we view the proton and neutron as two different states of the same particle, the nucleon. If we now give the nucleon a new quantum number by analogy to its spin number—let us call it isospin—whose magnitude for nucleons will be ½, then we can view the proton and the neutron as the states

Figure 5.159 *continued*

and Edward Teller to write the historic letter (Figure 5.134) that started the research on a fission bomb. During the war, he worked with Fermi on the creation of a chain reaction. Wigner had already studied the structure of spectra with group-theoretical methods in the first rapid development period of quantum mechanics (*Über nichtkombinierende Terme in der neueren Quantenmechanik*, 1927) and symmetry relations. 1936: In joint work with Breit, created the dispersion theory of nuclear reactions (Breit–Wigner formula); 1952: conservation of parity.

Chien-Shiung Wu (1912–1997): Born in China, went to the United States in 1936 to work with Lawrence. Doctorate in 1940; from 1957, professor at Columbia University. Her famous experiment proved the violation of parity.

Chen Ning Yang (born 1922 in China): Studied in China, scholarship to the United States in 1946; from 1949 worked at the Institute for Advanced Study in Princeton; from 1955, professor; and from 1965, Albert Einstein professor of physics at SUNY Stony Brook. His main work was in the physics of elementary particles. Later, he has worked on the theory of reactions taking place at very high energies.

Tsung-Dao Lee (born 1926 in Shanghai): 1946 with Yang at the University of Chicago; after a short hiatus, again worked with Yang at Princeton. 1953: Columbia University; 1960: Princeton; 1964: again Columbia. In addition to parity violation, investigated the limits of validity of the conservation of time-reversal symmetry.

▼ **Table 5.4** Summary of the quantities appearing in the conservation laws:

P	parity
Q	electrical charge
B	baryon charge number
L	electron number
L_e	muon number
S^u	strangeness
I	isospin
L_a	isospin "projection"
$Y = B + S$	hypercharge

The Gell-Mann–Nishijima relationship connects the quantum numbers:

$$Q = I_3 + \frac{1}{2}(B + S) = I_3 + \frac{1}{2}Y.$$

▼ **Table 5.6** Quantum numbers for some of the more important particles.

	I_3	B	Q	S	Y	L_e	L_μ
ν_e	0	0	0	0	0	1	0
ν_μ	0	0	0	0	0	0	1
e^-	0	0	–1	0	0	1	0
π^+	1	0	1	0	0	0	0
$\pi 0$	0	0	0	0	0	0	0
K^+	½	0	1	1	1	0	0
p	½	1	1	0	1	0	0
n	–½	1	0	0	1	0	0
Σ^+	1	1	1	–1	0	0	0
Ξ^-	–½	1	–1	–2	–1	0	0
Ω^-	0	1	–1	–3	–2	0	0

Quotation 5.50

Since there are some who say that there is a right and a left in the heaven, with those who are known as Pythagoreans—to whom indeed the view really belongs—we must consider whether, if we are to apply these principles to the body of the universe, we should follow their statement of the matter or find a better way. At the start we may say that, if right and left are applicable, there are prior principles which must first be applied.

—Aristotle, *On The Heavens* [1984, pp. 470–471]

Quotation 5.51

"...by no single concept, but only by pointing to our left hand and right hand, and thus depending directly on intuition [Anschauung] can we make comprehensible the difference between similar yet incongruent objects (such as oppositely wound snails)."

[This quote of] Kant has been interpreted as follows: If the first creative act of God had been the forming of a left hand, then this hand, at the time when it could be compared to nothing else, would already have possessed that definite character of the left one (in contrast to the right one) which can only intuitively but never conceptually be apprehended. This is incorrect, as Leibniz points out, if we intend this to mean that something else would have happened had God created a 'right' hand first, rather than a 'left' hand. One must follow the process of the world's genesis further in order to uncover a difference: Had God, rather than making first a left and then a right hand, begun by making a right one and proceeded to form another right one, then He would have changed the plan of the universe not in the first but in the second act, by bringing forth a hand which was equally rather than oppositely oriented to the first-created one.

—Hermann Weyl, *Philosophy of Mathematics and Natural Science* [p. 80, footnote p. 97]

Unobservable Phenomenon or Quantity	Symmetry Transformation	Conservation Law or Selection Rule
exchange of identical particles	permutation	Bose–Einstein and Fermi–Dirac statistics
absolute location in space	$r \to r + r_0$	momentum
absolute time	$t \to t + t_0$	energy
absolute direction in space	rotation	angular momentum
absolute velocity	Lorentz transformation	form of the equations
absolute right (left)	$r \to -r$	parity

▲ **Table 5.5** "Classical" symmetry transformations and associated conservation laws.

Just as the theory of optics follows from the principle of least time, and the theory of mechanics is epitomized by the principle of least action, so does the whole kit and caboodle of electromagnetic theory follow from the principle of gauge invariance. All of electromagnetic theory, including the very existence of the photon, and especially the fact that it is a massless particle with a unit spin, may be logically deduced from the principle of gauge invariance alone. It is strange and surprising that a requirement for symmetry has as a consequence the existence of a force. How wise is this principle, so worthy of a Supreme Being that it sayeth for him, "Let there be light!" And how tempting it is to extend this principle so as to generate all the forces of nature. How wonderful it would be if all of the fundamental properties of matter could be understood as deductive consequences of an overriding symmetry principle.

—Sheldon Glashow, *Interactions: A Journey through the Mind of a Particle Physicist and the Matter of This World* [p. 125]

of the nucleon characterized by isospin projection +½ and –½, respectively. Just as ordinary spin can assume states relative to a particular direction differing by a whole number, we can now view this other phenomenon, isospin, as assuming two different set directions in the abstract isospin space.

Similarly, the mesons π^-, π^0, π^+ can be seen as three different states of a single particle. The isospin of this particle is equal to 1 and has three possible projections: +1, 0, and –1. In the former case of nucleons, again making use of the same analogy, we speak of a charge doublet, and for the latter case of π mesons, of a charge triplet. The conservation law for isospin can be reduced to the symmetry properties of the abstract isospin space (Gell-Mann, Nishijima, 1953–1954). It should be noted here that the isospin conservation law is satisfied only for the strong interactions. Table 5.4 presents a summary of the quantities appearing in the conservation laws.

In Table 5.5, we present a simplified summary of the connections between symmetry transformations and conservation laws, based on a lecture given by T. D. Lee. The table shows emphatically that the starting point for these considerations is always a physical principle: We assume that certain quantities are not measurable, which leads to the invariance of the laws under some symmetry transformations, and from this, the conservation laws follow.

In Table 5.6, we assemble the most important characteristics of some elementary particles.

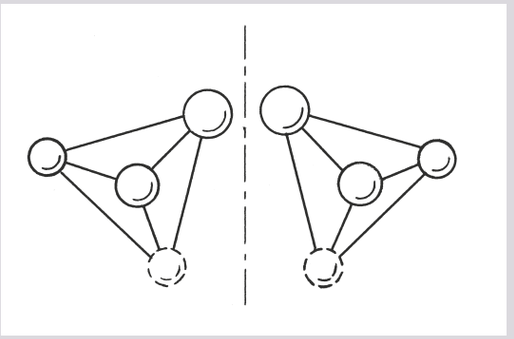
5.5.8 *Mirror Symmetry?*

We shall now look a bit more closely at the law of conservation of parity. The symmetry leading to the conservation law in this case can be simply stated in the language of quantum mechanics, and at the same time it is also easily visualized. In 1927, Wigner (Figure 5.159) pointed out that an empirical rule found in 1924 by Laporte, according to which the energy levels of complex atoms can be divided into two groups of different types, is simply the consequence of the mirror

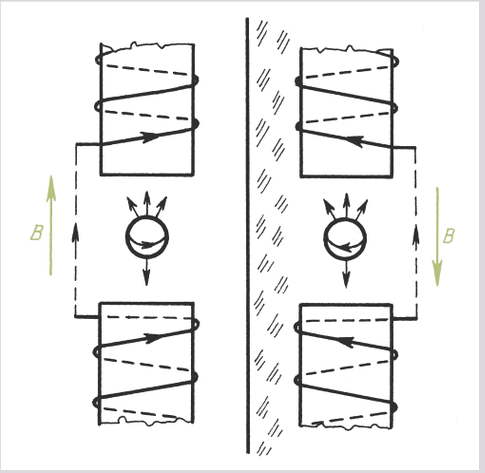
symmetry of the electromagnetic forces responsible for the atomic structure. From this point on, the notion of parity became established in quantum mechanics. Particle states have an even parity if the wave function remains unchanged under the transformation $x \to -x$, $y \to -y$, $z \to -z$; if the sign of the wave function changes in this transformation, then the parity is odd. In the former case, the particle's parity is +1, and in the latter case, it is –1. It had become a self-evident truth in quantum mechanics that parity is conserved in the different interaction processes.

Translated to the language of everyday life, the conservation of parity expresses the following: The laws of physics are invariant with respect to reflection in a mirror. Even more simply: If a phenomenon is possible, then its mirror image is also possible. Mirror symmetry and the problem of right and left did, of course, already appear earlier in history, not only as a geometric but also as a philosophical question (Quotations 5.50, 5.51). Our daily experience seems to contradict mirror symmetry. For starters, the human body is not completely symmetric, the most obvious asymmetry being that the heart is on the left, and so the right and left sides can definitively be distinguished. Even in simple molecules, we can see violations of mirror symmetry. Figure 5.160 shows two molecules built from the same component atoms, but one molecule is the mirror image of the other. This means that there are no realizable geometric operations that could turn one arrangement into the other: here, we must not forget that reflection is an abstract operation and does not have a actual counterpart in nature. Yet, it turns out that bioorganic nature prefers one or the other arrangement. Consequently, living organisms evolved to use just one of the variants. That nature is nonetheless fundamentally mirror-symmetric can be demonstrated in the following way: In the second half of the nineteenth century, it was known that the molecules of tartaric acid can theoretically exist in two mirror-symmetric variants, which can be identified because one rotates the polarization plane of light to the right, the other to the left. In its natural form, only one variant actually occurs, the one that rotates light to the right. When it became possible to synthesize tartaric acid, it was observed that the synthetic tartaric acid did not change the polarization plane of light at all, presumably because each of the two symmetric variants was present in equal proportions. The picture was made complete by Pasteur's imaginative experiment: We know that certain microbes feed on tartaric acid and that they can utilize only the natural molecule, the one rotating to the right. If microbes are fed synthetic tartaric acid, then the percentage of "inedible" molecules should increase, and the remaining tartaric acid should increasingly rotate the polarization plane of light *to the left*. Experiments have borne out Pasteur's expectations, proving the hypothesis that inorganic nature is fundamentally mirror-symmetric. How, then, can we explain the origin of asymmetry in nature in general and in bioorganic nature in particular? These asymmetries can be reduced to random asymmetries (Earth's rotation, fluid flow, etc.) that were present at the places and times at which the molecules were formed. (See Plate XXVI.)

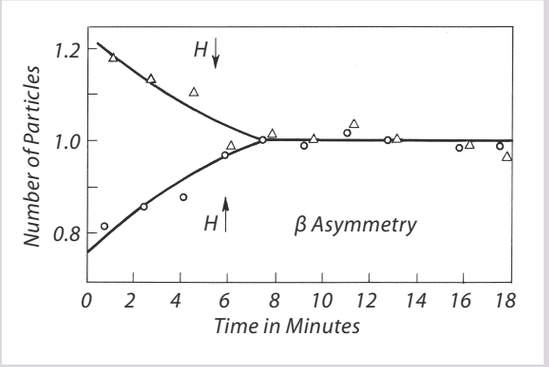
Let us recapitulate our progress so far. The interactions of the elementary particles, their transformation into each other—hence, in essence, the most fundamental processes in our material world—are governed by conservation laws. This means that all those transformations, that is, elementary processes, that these conservation laws allow will take place in the world. Or, to put in another way, whatever is possible—naturally within the bounds of the conservation laws—will actually happen. At the same time, the conservation laws are the consequences of the symmetry principles. Thus:



▲ **Figure 5.160** Two mirror-symmetrical molecules composed of the same combination of atoms.



▲ **Figure 5.161** Mirror symmetry requires that the same number of particles be emitted in the direction of the magnetic field as in the opposite direction.



▲ **Figure 5.162** Results of the measurements made by Wu and colleagues.

Quotation 5.52

In April, 1956, during a conference on nuclear physics at the University of Rochester, in New York, there was a spirited discussion of the theta-tau puzzle. Richard Phillips Feynman, a physicist at the California Institute of Technology, raised the question: Is the law of parity sometimes violated? In corresponding with Feynman, he has given me some of the details behind this historic question. They are worth putting on record.

The question had been suggested to Feynman the night before by Martin Block, an experimental physicist with whom Feynman was sharing a hotel room. The answer to the theta-tau puzzle, said Block, might be very simple. Perhaps the lovely law of parity does not always hold. Feynman responded by pointing out that if this were true, there would be a way to distinguish left from right. It would be surprising, Feynman said, but he could think of no way such a notion conflicted with known experimental results. He promised Block he would raise the question at next day's meeting to see if anyone could find anything wrong with the idea. This he did, prefacing his remarks with, "I am asking this question for Martin Block." He regarded the notion as such an interesting one that, if it turned out to be true, he wanted Block to get credit for it.

Chen Nin Yang and his friend Tsung Dao Lee, two young and brilliant Chinese-born physicists, were present at the meeting. One of them gave a lengthy reply to Feynman's question.

"What did he say?" Block asked Feynman later.

"I didn't know," replied Feynman, "I couldn't understand it."

"People teased me later," writes Feynman, "and said my prefacing remark about Martin Block was made because I was afraid to be associated with such a wild idea. I thought the idea unlikely, but possible, and a very exciting possibility. Some months later an experimenter, Norman Ramsey, asked me if I believed it worth while for him to do an experiment to test whether parity is violated in beta decay. I said definitely yes, for although I felt sure that parity would *not* be violated, there was a possibility it would be, and it was important to find out. 'Would you bet a hundred dollars against a dollar that parity is not violated?' he asked. 'No. But fifty dollars I will.' 'That's good enough for me. I'll take your bet and do the experiment.' Unfortunately, Ramsey didn't find time to do it then, but my fifty dollar check may have compensated him slightly for a lost opportunity." ...

"Now after the first shock is over," Pauli wrote to Weisskopf on January 27, after the staggering news had reached him, "I begin to collect myself. Yes, it

continued on next page

Symmetry principles determine the fundamental processes.

Some go as far as to claim that the conservation laws, or rather the symmetry properties, also determine the intensity—the frequency—with which various events occur.

In this way we formed a simple—and we may now add, disquietingly simple—picture of our world and of the laws that govern it. Does this indicate a return to Platonic ideals? Is it really sufficient to study the symmetries of things in order to gain knowledge of the real world? Isn't this too similar to the idea that heavenly bodies must move in circles because the circle is the most symmetric of the curves? And even if it were so, imagine two long-distance runners apparently running neck and neck on a track; their movements are identical, but one of them is about to set a new world record because she has in fact lapped the other.

Nature has made sure that the physicists did not have to endure for too long such disquieting simplicity. We have spent this much time discussing the details of mirror symmetry so that we would be better able to appreciate the general consternation among both experimental and theoretical physicists when it became known that during weak interactions, mirror symmetry is not satisfied; in other words, parity is not conserved, and the law of conservation of parity is violated.

The preliminary work for the discovery of the violation of parity goes back to 1953, when Dalitz and Fabry suggested that the parity of the newly discovered theta (θ) and tau (τ) mesons could be determined from the decay equations

$$\theta \to \pi^+ + \pi^0, \quad \tau \to \pi^+ + \pi^0 + \pi^0.$$

Since the parity of the π meson had already been determined from earlier experiments to be $P = 1$, the above decay processes suggest that the parity of the θ meson is +1, while that of the τ meson is −1. (The total parity on the right side is the product of the parities of the individual particles.) More precise investigations strengthened this picture, and at the same time it also became clear that the two mesons have otherwise completely *identical* properties. This phenomenon was hard to understand, and so it was called the *tau-theta puzzle*. In spring 1956, Lee and Yang voiced their suspicion that *one was dealing here with a single particle, and that during the decay of this particle, for which the weak interaction is responsible, the conservation of parity law is not satisfied*. (This single particle is today denoted by K^+.) They looked at the earlier experimental results and found that under the strong interaction, the conservation law for parity had been strictly established, but for the weak interactions, the accumulated experimental records did not really constitute a proof of parity conservation. Continuing in this vein, in the summer of 1956 they suggested a series of experiments that would unequivocally decide whether the law of conservation of parity is valid in the case of weak interactions, in other words, whether the weak interactions are mirror symmetric in nature. The decisive experiment was carried out by C. S. Wu, a Chinese-born physicist working at Columbia University, together with colleagues at the National Bureau of Standards (E. Ambler, R. W. Hayward, D. H. Hoppes, and R. P. Hudson). The essentials of this experiment are these: Wu and her colleagues cooled a sample containing the β-active Co^{60} isotope to a temperature of 0.1 K and placed it in an external magnetic field that was strong enough to align the spins of the cobalt atoms in the direction of the field. The low temperature is necessary so that the movement of the atoms due to heat would not disturb the ordered alignment of the spins. The purpose of the measurement was to find out whether the electrons

from the beta decay are emitted with the same probability in the direction determined by the magnetic field—let us say toward its north pole—as in the opposite direction. As shown in Figure 5.161, mirror symmetry requires that the emissions in one direction occur with the same probability as in the other; otherwise, the mirror image would represent a process that is not realized in nature. The measurements definitively and unambiguously showed a lack of mirror symmetry in this case. The graph of the measurements shown in Figure 5.162 is surprising also in view of the fact that the deviations from the values that were expected assuming mirror symmetry are not just at the limits of detectability, but amount to as much as 40%. Quotation 5.52 gives us an insight into the state of mind of the theoretical physicists before the discovery: confident that symmetry will prevail, but a little bit curious, also. Many laboratories had been, in fact, reluctant to risk wasting their efforts on the costly and complicated experiment believing that the theoretically all-but-obvious negative effect would inevitably follow.

However, experiments have shown that the law of conservation of parity is violated in the weak interaction. It should go without saying that this result was a shock to theoretical physicists most of all. As the theoreticians incredulously exchanged the sensational news, one experimental physicist remained unperturbed. When asked why he was not excited that parity conservation was violated, he answered, "Why should I be excited? I never believed that it was conserved anyway." It is worth reviving this real-life anecdote because a number of lessons can be drawn from it. To be surprised that a law is not valid for all phenomena, we must first believe in its general validity. This means that from the psychological and pedagogical points of view, it is a good idea first to emphasize the validity of the law so that the fundamental nature of its limitations can be better understood. There is also the implication that when laws or paradigms have finally become hopelessly outmoded, their replacements can be taught without reference to them.

But here is an additional possibility. The experimental physicist of the anecdote relied perhaps on "common sense" and might have thought it self-evident that that spin's direction, indicated by an arrow, already establishes an asymmetry, an up and a down direction, so to speak. But the spin only determines the *direction of rotation*; the distinction of the direction of arrow is just a convention and could have been made according to the sense of either a right-handed screw or a left-handed screw. Had he realized that, he would have been just as shocked as the others because the experiment shows that nature can distinguish right-handed and left-handed coordinate systems. That is, a sense of "left" and "right" is inherent in nature.

After the experimental proof of the violation of parity, a number of physicists (Landau, Abdus Salam, Lee, and Yang) expressed the opinion that neutrinos have a certain helicity, with the neutrino having the direction of a left-handed screw and the antineutrino that of a right-handed screw (Figures 5.163, 5.164). Thus, it is the neutrino alone that is responsible for the violation of mirror symmetry, and because the neutrino plays a special role in the weak interaction, it is readily understandable why the parity violation would be seen here.

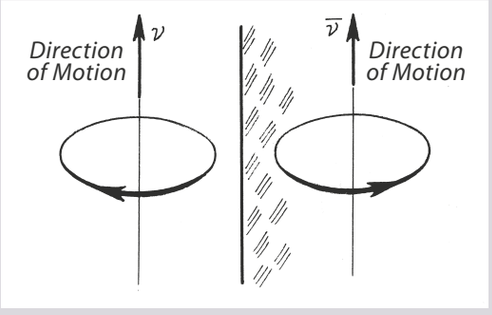
5.5.9 *"A Bit of Asymmetry Improves the Aesthetics"*

Mirror symmetry and its orderly calm can be completely restored, or so it seemed at first: If along with the mirroring we also change all particles to their antiparticles, then we again obtain processes that are realizable and will take place in nature. The transition from particle to antiparticle is called charge conjugation, denoted

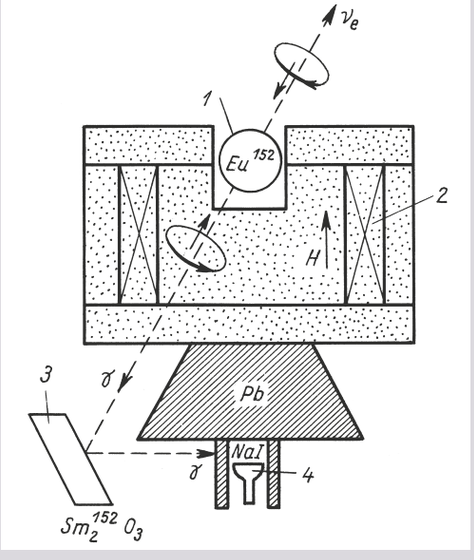
Quotation 5.52, ***continued***

was very dramatic. On Monday, the twenty-first, at 8 pm I was supposed to give a lecture on the neutrino theory. At 5 pm I received three experimental papers [reports on the first three tests of parity].... I am shocked not so much by the fact that the Lord prefers the left hand as by the fact that he still appears to be left-handed symmetric when he expresses himself strongly. In short, the actual problem now seems to be the question: Why are strong interactions right-and-left symmetric?

—Martin Gardner, *The Ambidextrous Universe* [pp. 239–240, 247]



▲ **Figure 5.163** Neutrinos have a particular rotational orientation with respect to their direction of motion and are therefore responsible for the violation of parity.

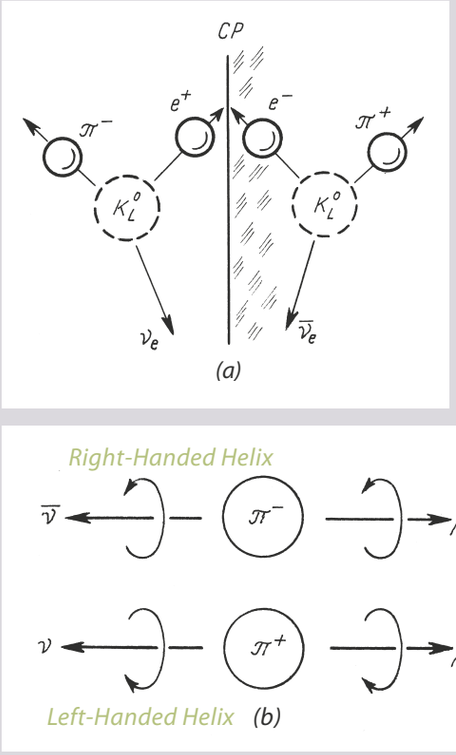


▲ **Figure 5.164** The experiment of Goldhaber and colleagues to prove the helicity of the neutrino: The nucleus of the europium isotope Eu-152 (1) captures an electron from an orbit near the nucleus (*K* capture), resulting in the emission of a neutrino and an excited samarium nucleus, which transitions to the ground state with the emission of a γ quantum. If the neutrino and γ quantum are emitted in opposite directions, they will have the same helicity. The

continued on next page

Figure 5.164 *continued*

γ quantum traverses the analyzing iron magnet (2) whose field can be turned up or down, which allows the helicity of the circular polarization of the γ rays to be determined; the ray then undergoes resonance scattering on a samarium nucleus in the samarium oxide target in its path (3). The resonance condition is that the recoil of the emitting Sm nucleus has to be small, which is best satisfied when the γ quantum and the neutrino are emitted in opposite directions. The sodium iodide (NaI) scintillation crystal (4) and associated photomultiplier counts the numbers N_- and N_+ of the scattered quanta with the analyzer field set up or down. From this the effect $(N_- - N_+)/\frac{1}{2}(N_- + N_+)$ of the magnetized iron on the γ rays can be calculated, which should be according to theory +0.025 for a left-helical polarization, or –0.025 for the right; the experiment gave the value +0.017 ± 0.0003, showing that the neutrino has left helicity once the experimental uncertainties are also taken into account (from the *Great Soviet Encyclopedia*).



▲ **Figure 5.165** CP violation makes it possible **(a)** to identify the plus and minus charges and **(b)** to distinguish between right and left.

by the letter *C*, while the mirror reflection is denoted by *P*, for "parity," so that the twofold transformation is the "*CP* operation." The new invariance theorem can now be simply stated: Natural phenomena are invariant under the *CP* operation.

The absolute validity of this new symmetry law did not have a particularly long life, for soon "*CP*-violating" processes were discovered. We can say, then, that this reflection symmetry also holds only approximately—although it is a very good approximation.

It can be shown using special relativity's Lorentz invariance, and also the causality principle, that we can get from any realizable transformation to another realizable transformation if in addition to spatial reflection and charge conjugation, we also reflect the time, in other words, change the sign of time. It is usual to denote this operation by *T*. We therefore expect every process to be invariant with respect to the *CPT* transformation.

To illustrate the significance of this, let us consider, following Lee, an example:

Let us assume that an exchange of information between the civilizations of our Earth and those of a distant star were possible and we wished to compare the results of physical experiments. If *CP* invariance held, then it would be impossible to determine through a comparison of experimental data whether the two worlds consist of the same particles or are "antiworlds" with respect to each other. Thus, for example, one could not determine whether the hydrogen nucleus of one world was a positively charged proton and that of the other world a negatively charged antiproton. Of course, in such an antiworld, the role of the electron would be taken over by the positron. Nor could we agree unambiguously with respect to experimental measurements what is meant by right and left. Before the civilizations met, it would be absolutely essential for each side to determine whether the other world was made of matter or antimatter because when matter and antimatter meet, they annihilate each other and are transformed into radiation.

Yet the fact that *CP* symmetry is not strictly valid makes it possible for each of the two worlds to determine the composition of the other. One begins with the following two decay processes of the *K* meson:

$$K_L^0 \to e^+ + \pi^- + \nu_e, \quad K_L^0 \to e^- + \pi^+ + \bar{\nu}_e.$$

We see that the decay products in the second process are antiparticles with respect to the first. Measurements show that the two decay processes *do not occur with the same frequency*, and although the deviation is very small, it is significant enough to be measurable. Its magnitude is

$$\frac{W\left(K_L^0 \to e^+\pi^-\nu_e\right)}{W\left(K_L^0 \to e^-\pi^+\bar{\nu}_e\right)} \approx 1.003.$$

Based on this fact, we can inform other civilizations unambiguously what *we* understand by a positive charge. We must request that they measure the probabilities with which the two decays of K_L^0 mesons occur and assign a positive sign to the *e* particles that are created in the somewhat more frequent process.

In this way, we can also distinguish left from right by measuring the helicity of the neutrino arising from the decay $\pi^+ \to \mu^+ + \nu$ (Figure 5.165(a,b)). The result is the definition of the left-handed screw.

For the cultural historian, it is interesting to ponder (Figure 5.166) whether it is by chance that it was Chinese scientists who predicted theoretically and proved

experimentally the violation of mirror symmetry. One could point to the fact that well-known symbols of Western culture such as the cross and five-pointed star are mirror-symmetric, whereas the yin–yang symbol, which is of importance in Chinese culture and, like the European symbols, appears on national flags, does not possess mirror symmetry.

If we assume that *CPT* symmetry is strictly satisfied, then it follows from the violation of *CP* symmetry that there is also a violation of time-reversal symmetry. And indeed, there exist experimental indications that time-reversal symmetry is violated.

The fact that mirror symmetry is not strictly valid does not mean at all that the symmetry principles have lost their significance for particle physics. As was shown in the course of later investigations, using abstract symmetries, which are no longer accessible to intuition and can be described only with the help of group theory (SU_2 and SU_3 symmetries), the elementary particles have been systematized so successfully that with their help, new phenomena have been accurately predicted.

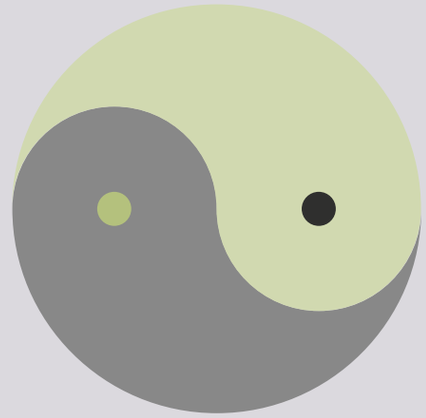
In 1960–1961, Murray Gell-Mann and Yuval Neeman proposed the "eightfold way" model. According to this principle of systemization—the name refers to the teachings of the Buddha—hadrons, mesons, and baryons, whose particles form families with eight members, can be grouped into multiplets: Six particles are located at the vertices of a regular hexagon, and two additional particles are located at the center. Figure 5.167 shows the octet of baryon with half-integer spin: Even the neutron and proton have their places here. The quantum numbers for strangeness (S), charge (Q), hypercharge ($Y = B + S$), and isospin projection (I_3) are identified.

This eightfold way can be enlarged. Figure 5.168 shows an arrangement of 10 hyperons with spin 3/2 and parity +1 (the $3/2^+$ decuplet). Beginning with this arrangement, Gell-Mann and Nishijima predicted the existence of the Ω^- particle. Contrary to the usual depiction, we have pointed the apex of the triangle upward in order to emphasize the similarity of the figure to the "holy" Pythagorean tetractys (Figure 1.35). The subsequent experimental detection of the Ω^- particle strengthened belief in the entire concept.

Following this line of argument, we select the triangle as the basic element of the geometric ordering in our I_3-Y coordinate system and obtain three particles at the vertices (Figure 5.169). The most striking and unexpected—and until the later experimental proof, almost unacceptable—property of these particles is their electric charges: Expressed in units of the elementary charge, they are the fractions ±⅓ and ±⅔.

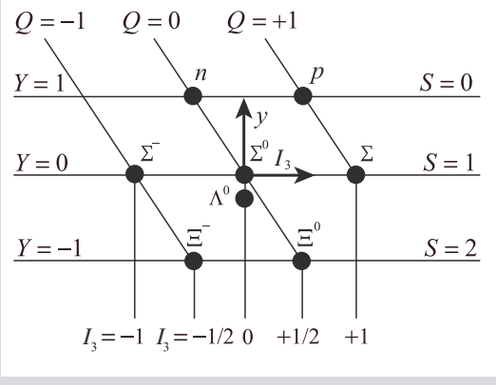
Thus, in 1964 Gell-Mann and George Zweig arrived at—for the time at least—the fundamental building blocks of matter, the quarks. The name was proposed by Gell-Mann and refers—perhaps reflecting slightly his own doubts—to a sentence in James Joyce's novel *Finnegans Wake*. Every now and then, a drunken innkeeper calls to the bartender, "Three quarks for Muster Mark."

Figure 5.170 shows how the most important elementary particles are constructed from quarks based on this hypothesis. At first glance, we are astonished at the classical simplicity and clarity of the model. An engineer could practically call the workshop with a work order: Build a neutron from quarks that are forged, milled, and welded together. Boltzmann and Lord Kelvin would have been pleased to finally understand the underlying principle of all things, because a mechanical model seems to be at hand. But all of that is just appearance. If we wish to compare this figure with anything at all, then we should not compare it with the mechanical

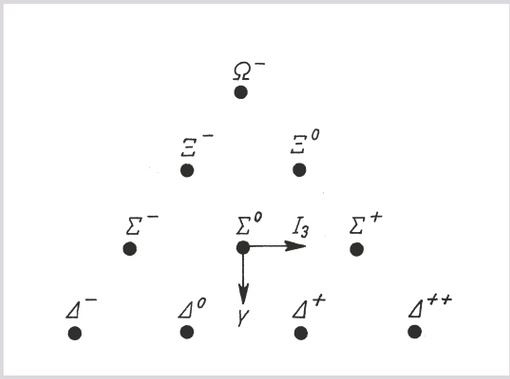


▲ **Figure 5.166** An important symbol in the Orient, the yin–yang symbol, is not mirror-symmetric, and perhaps it is therefore not completely by chance that the violation of mirror symmetry was discovered by Chinese scientists. In the guest book of the Department of Atomic Physics at the University of Budapest, we find Heisenberg quoting Plato, while Tung-Dao Lee, in contrast, wrote the following few lines of a poem by Lao Tzu (sixth century BCE):

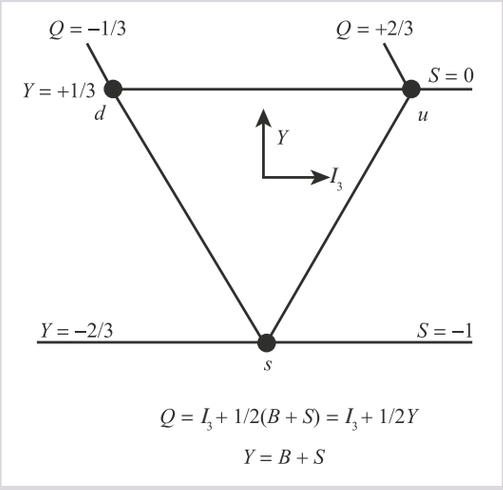
> The Tao that can be told
> is not the universal Tao.
> The name that can be named
> is not the universal name.
>
> In the infancy of the universe
> there were no names.
> Naming fragments the mysteries of life
> into ten thousand things
> and their manifestations.
>
> —Lao Tzu, *The Tao Te Ching*



▲ **Figure 5.167** With the help of the SU-3 symmetry, the elementary particles can be organized into "multiplets." The figure shows the octet of baryons with half-integer spin.



▲ **Figure 5.168** The $(3/2)^+$ decuplet (spin 3/2, parity +1). This arrangement led to the prediction of the Ω^- hyperon. One hundred years earlier they would have said that a hole in the periodic table would have to be filled.



▲ **Figure 5.169** If we place a triangle, the basic element of the hexagon of the octet in Figure 5.167, in our coordinate system (I_3, Y), we represent the basic elements of physical matter: the three quarks, *u* (up), *d* (down), and *s* (strange); their quantum numbers are listed in Figure 5.170. The anti-quarks could also be found in an upward pointing triangle.

Quotation 5.53

I too play with symbols, and have planned a little work, *Geometric Cabala*, which is about the Ideas of natural things in geometry; but I play in such a way that I do not forget that I am playing. For nothing is proved by symbols, nothing hidden is discovered in natural philosophy through geometric symbols; things already known are merely fitted [to them]; unless by sure reasons it can be demonstrated that they are not merely symbolic but are descriptions of the ways in which the two things are connected and of the causes of this connection.

—Kepler, Letter to Joachim Tanckius, 1608 [Vickers 1984, p. 155]

continued on next page

models of Boltzmann and Kelvin or even of Maxwell; instead, we need to go back two thousand years, for this figure is much more reminiscent of the Platonic ideals (Quotation 5.53).

5.5.10 *Back to the Apeiron?*

In the foregoing we have followed the attempts within particle physics to answer questions of the form: "What is something composed of?" The method used in the search had proved itself very successful in the investigation into the structure of the atom and the atomic nucleus: One transfers energy to the entity that is thought to be complex, and with that, it is broken up into its components. To separate the parts of the atomic shell, it sufficed to bombard the atom with electrons or protons having an energy of a few electron volts. In the case of the nucleus, the projectile had to have several million electron volts of energy in order to break off from the nucleus one or more nuclear building blocks, the nucleons. For the investigation of elementary particles, literally gigantic energies are required because only with energies in the giga-electron-volt range can one expect new results. But is this method really the right one? Will it actually answer the question that was posed? Heisenberg's reply was a resounding no:

> If we speak of protons, pions, hyperons, and so on as consisting of smaller, as yet unobserved, particles, of quarks, partons, then we are forgetting that the sentence "they consist of..." has an acceptably clear sense only if we succeed in decomposing these protons and so on into parts with a small expenditure of energy, where the rest masses of the particles thus arising are much greater than the expended energy.

Heisenberg thus arrived at the idea that the elementary particles are to be seen as different manifestations, different quantum states, of one and the same "primordial substance." The elementary particles, it would follow, are the only possible manifestations of matter. Because of its similarity to the primordial substance hypothesized by Anaximander, Born called this substance *apeiron*.

The various elementary particles come into existence under the influence of the primordial substance on itself. It is because of this self-interaction that a fundamental equation of Heisenberg proposed that unified field theory has to be nonlinear. From the solution to this equation, the properties of the elementary particles, including their masses, should follow, and the theory should also provide an answer to the question of why it is precisely these particles with these properties that exist.

> The final equation of motion for matter will probably be some quantized nonlinear wave equation for a wave field of operators that simply represents matter, not any specified kind of waves or particles. This wave equation will probably be equivalent to rather complicated sets of integral equations, which have "eigenvalues" and "eigensolutions," as physicists call it. These eigensolutions will finally represent the elementary particles; they are the mathematical forms which shall replace the regular solids of the Pythagoreans. We might mention here that these "eigensolutions" will follow from the fundamental equation for matter by much the same mathematical process by which the harmonic vibrations of the Pythagorean string follow from the differential equation of the string. But, as has been said, these problems are not yet solved.
>
> If we follow the Pythagorean line of thought we may hope that the fundamental law of motion will turn out as a mathematically simple law, even if

its evaluation with respect to the eigenstates may be very complicated. It is difficult to give any good argument for this hope for simplicity—except the fact that it has hitherto always been possible to write the fundamental equations in physics in simple mathematical forms. This fact fits in with the Pythagorean religion, and many physicists share their belief in this respect, but no convincing argument has yet been given to show that it must be so.

—Heisenberg, *Physics and Philosophy* [pp. 72–73]

To give the reader a taste of what such a universal equation looks like, we write it down explicitly, although without attempting an explanation:

$$\gamma_\nu \frac{\partial \Psi}{\partial x_\nu} \pm l^2 \gamma_\mu \gamma_s \Psi \left(\Psi^* \gamma_\mu \gamma_s \Psi \right) = 0.$$

After very promising initial successes, Heisenberg was unable during the remainder of his life to set this theory on a sufficiently firm foundation so that it could displace the theories that rely on the structural model.

Among the "thought structures" of our mind, the theory of relativity called into question the a priori existence of space and time; quantum mechanics brought new problems to the idea of causality; and finally, particle physics shed a new light on the category of substance. We are not going to explore, even superficially, the philosophical arguments surrounding the category of substance—that is, the idea of matter—as we did with causality. Here also, we find the anthropomorphic beginnings as well as the thesis and antithesis of Hume and Kant (Quotations 5.54, 5.55, 5.56). We note here only that the primordial substance with its various manifestations—the elementary particles representable by symmetry groups—can be compared with the matter of Aristotelian physics. We refer again to Weyl:

Aristotelian philosophy considers matter as the determinable, in contrast to the determining form. Matter is [the] possibility of becoming formed. In a production process of several stages its matter appears "more formed" at each step, and thereby the range of possibilities for further forming becomes more restricted. And at the same time matter, the component of merely potential rather than actual being, shrinks more and more. Substantiality is ascribed to the forms rather than to matter. The forms push matter from potentiality to actuality ....

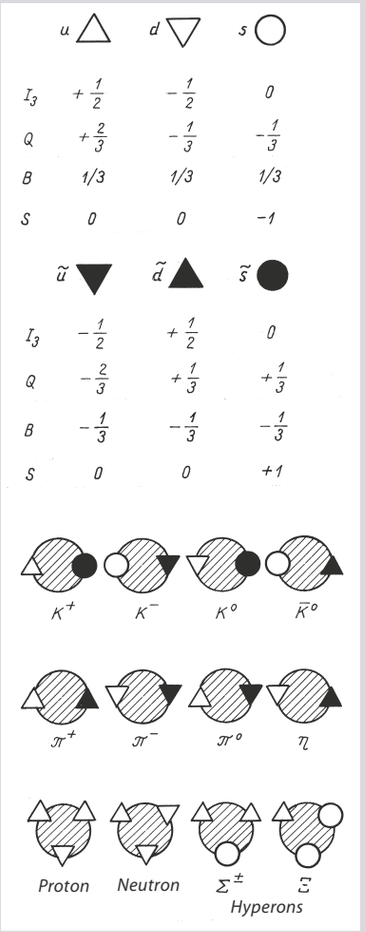
—Hermann Weyl, *Philosophy of Mathematics and Natural Science* [p. 178]

Even if we do not view the primordial substance as a "possibility," that is, "brought to life" only through form, the permanent matter structures that we perceive within the manifold diversity of the real world are secured—or more modestly and precisely put, described—by the form-giving laws of quantum mechanics. So, for instance, in measurement practices of sufficiently high precision, we exploit the fact that all atoms of an element are identical. Moreover, even after they are destroyed, they can be identically reproduced. As we see in Quotation 5.57, Newton thinks that this is possible only because the smallest building blocks of matter are "indestructible." In the end, these building blocks are an embodiment of the classical notion of substances, the invariable carriers of particular properties, the *accidentia*. But what makes it possible that, after the interaction of two systems, the original state can be recovered? Just imagine how different the situation would be if, say, two solar systems collided. Following Weisskopf, we can emphasize

Quotation 5.53, ***continued***

I am convinced that we can discover by means of purely mathematical construction the concepts and the laws connecting them with each other, which furnish the key to understanding of natural phenomena. Experience may suggest the appropriate mathematical concepts, but they most certainly cannot be deduced from it. Experience remains, of course the sole criterion of the physical utility of a mathematical construction. But the creative principle resides in mathematics. In a certain sense, therefore, I hold it true that pure thought can grasp reality, as the ancients dreamed.

—Einstein. *Ideas and Opinions,* 1954



◀ **Figure 5.170** Top: Quantum numbers of the three quarks, *u* (up), *d* (down), and *s* (strange), and the three antiquarks. Bottom: the quark models of some elementary particles (Gell-Mann and George Zweig, 1964).

Quotation 5.54

And since I conceive that other beings can also have the right to say "I", or that it can be said for them, it is through this that I conceive what is called substance in general.

—Leibniz, Letter to Queen Sophie Charlotte [p. 188]

Quotation 5.55

[W]hich view of things being destructive of its primary and more natural motions, obliges the imagination to feign an unknown something, or *original* substance and matter, as a principle of union or cohesion among these qualities, and as what may give the compound object a title to be call'd one thing, notwithstanding its diversity and composition. ...

[O]ur propension to confound identity with relation is so great, that we are apt to imagine something unknown and mysterious, connecting the parts, beside their relation ...

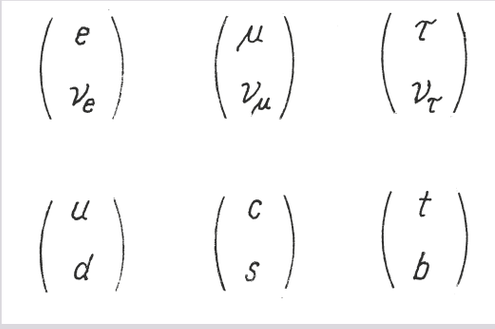
—David Hume, *A Treatise on Human Nature* [pp. 144, 164]

Quotation 5.56

All appearances contain the permanent (substance) as the object itself, and the transitory as its mere determination, that is, as a way in which the object exists. ...

[T]he permanent in the appearances is therefore the substratum of all determination of time, and, as likewise follows, is also the condition of the possibility of all synthetic unity of perceptions, that is, of experience.

—Kant, *Critique of Pure Reason* [footnote p. 212, p. 214]



▲ **Figure 5.171** Lepton–quark symmetry: at first, heuristic principle, then experimentally verified fact.

that it is the quantum-mechanical laws, that is, the rules of quantization, that are responsible for maintaining all these stable characteristics.

5.5.11 *The Quark Theory Is Completed*

The three quarks and the simple, clear, and orderly world built out of them were, for the researchers, just the first way station on a long—indeed, perhaps endless—journey, a station that nevertheless represented a good approximation and the potential for further progress. That improvements were needed in at least two areas was already clear in 1964. Let us first consider the Δ^{++} particle: Its spin is 3/2; its charge is $Q = 2$. The only way it could be composed is out of three u quarks with their spins all pointing in the same direction. Hence, the system is symmetric with respect to the exchange of two particles. However, the particles with half-integer spin obey the Pauli exclusion principle; therefore, their state can only be antisymmetric. We must therefore assume that for a complete determination of the quarks' states, another quantum number is required, with whose help these baryon states, considered as symmetric, can be made antisymmetric. This new quantum number was given the name *color* by Greenberg. We need three colors, say red, green, and blue. Thus, we have nine quark states, as each of the three quark flavors u, d, and s can be red, green, or blue.

The step forward in the second area was the extension of the number of quark flavors. Drawing attention to the lepton–quark symmetry, in 1964, Bjorken and Glashow postulated the existence of a fourth quark. The leptons are easily grouped into two pairs: (e, ν_e) and (μ, ν_μ). Equally natural is the quark pairing (u, d), where now, in contrast, the s quark is left without a partner. The postulated fourth quark based on these considerations is the *charmed* quark c. Its characteristic quantum number is called "charm," with the value +1 for c and −1 for its antiparticle $\bar{c}$. Thus, the number of possible quark states has now grown to 12.

A decade had to pass, until 1974, before two independent groups of researchers under the direction of Ting and Richter discovered a new boson, which was named the J/Ψ particle and whose mass is more than 20 times that of the pions. This particle does not fit into the old quark schema at all and can only be interpreted as a bound state of the charmed quark and the charmed antiquark $(c\bar{c})$. The discovery in 1977 of the Y particle, with 10 times the mass of the proton, required the introduction of yet another quark, the b quark (*beauty* or *bottom*). In the meantime (1975) the discovery of the heavy lepton τ^-, with 20 times the mass of the muon, and its corresponding ν_τ neutrino upset the lepton–quark symmetry. To restore the order, the existence of the so-called t quark (*top* or *truth*) to pair up with the b quark was predicted (Figures 5.171, 5.172).

One of the most important results during these two decades certainly was the promising progress made in the unification of the fundamental interactions. In 1967, Weinberg, Salam, and Glashow, more or less independently of each other, while to some extent building on one another's work, developed a unified theory of the electromagnetic and weak interactions. The essence of this theory is that the weak and electromagnetic interactions are two different manifestations of one and the same more fundamental interaction. The particles that mediate the interaction form a family; its four members are the already-known photon and three vector bosons: the W^+, with charge +1; the W^-, with charge −1; and the neutral Z^0. The theory produces experimentally verifiable predictions for their masses (79 GeV for the pair W^+ and W^-, and 90 GeV for the Z^0; see also Figure 5.173). At the

beginning of the 1980s, several accelerators became available to supply the energies required to produce these particles. We can summarize the theory as follows: At very high energies, when the $W^{\pm}$ and Z^0 particles are created as easily as photons, the two interactions are identical. At lower energies, the two types of interactions separate. Figuratively speaking, this is similar to the effect when, with decreasing temperatures, a molten material freezes in one of two different phases. The timeline of the discoveries is summarized in Figure 5.174.

5.6 Mankind and the Universe

5.6.1 *New Information Channels*

No matter how much the details may change in the course of history, the relationship between the cosmic—or, simply put, the heavenly—phenomena and mankind has always had two defining characteristics: First, the heavens–Earth connection is part of every ideology (Quotation 5.58); and second, the cosmos is the theater—and we may justifiably say laboratory—for such physical phenomena, the understanding of which can advance, perhaps decisively, our understanding of what happens on Earth.

Along with creation myths—and tightly connected with them—calendars made their appearance. They were the first practical results of a descriptive science of motion. The laws of motion of our solar system, seen as part of an infinitely extended universe, inspired the understanding of terrestrial dynamics.

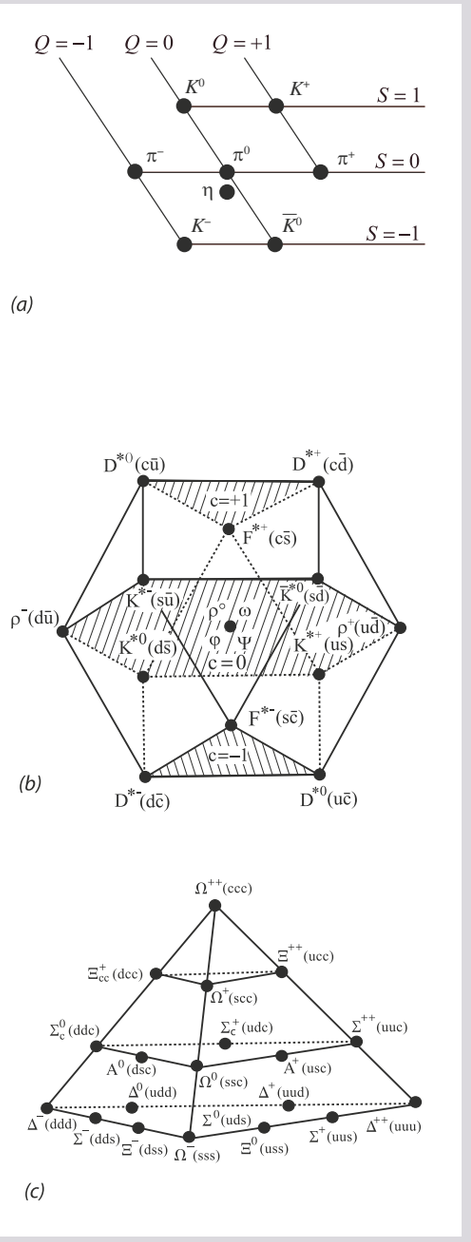
Physicists today are interested above all in the laboratory aspect of the universe: they can interpret the information arriving through a great variety of channels about the behavior of matter under extreme conditions. But no one can avoid the powerful aesthetic impression that the cosmos exerts in forming one's self-knowledge and worldview.

Astronomy is an independent branch of science, with its own specific theoretical and practical problems. We will not attempt to discuss, even in outline form, the entire history of astronomy. As was the case in the earlier text, as well as here, we select the phenomena to be discussed on the basis of the above two key factors, and give the chronology of their discoveries and interpretation. We wish above all to deal with those questions of astrophysics that are pertinent to nuclear and particle physics, or to find phenomena that can serve as experimental proof or disproof for theoretically predicted processes under conditions that are not realizable in terrestrial laboratories.

The study of our world in its entirety, including its geometry and its evolution, is of fundamental importance both for the expansion of our direct knowledge of physical reality and for the formation of our worldview.

In Figure 5.175 we give an overview of the development of astronomical knowledge, taking into account the above criteria. The list calls attention to the fact that the perfection of our knowledge of the universe is the consequence of our ability to collect information through more and more channels. Until the beginning of the seventeenth century, mankind was essentially limited to observations with the naked eye, and therefore, an eternal and unchanging heavenly order could still be believed in. Even Pascal was bewildered by the "eternal silence of these infinite spaces" because he had no way of knowing that there are cataclysms playing out in the cosmos in comparison to which the hydrogen bomb is a mere puffball.

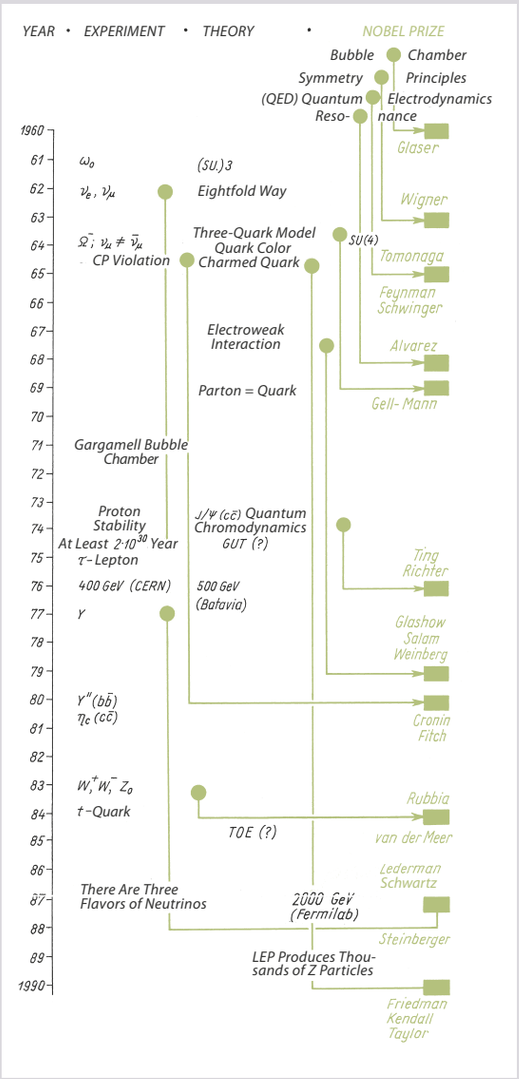
The telescope expanded the dimensions of space, and the spectra obtained with its help proved that the materials in the stars and on Earth are one and the same.



▲ **Figure 5.172** **(a)** Here is another octet model (based on a drawing by Glashow) of the mesons with spin 1, which was the basis for further extensions.

(b) This extension includes mesons with quark flavors $c = -1$, $c = 0$, and $c = +1$. They can be placed at the vertices of a semiregular Archimedean solid (eight equilateral triangles, six squares). For $c = 0$, we get the octet in diagram (a).

(c) The decuplet in Figure 5.168 can also be extended for $c = -1, 0, +1$. The result is a regular tetrahedron [Solomey 1997].



▲ **Figure 5.173** Chronology of progress in particle physics (continued from Figure 5.136).

Quotation 5.57

All these things being consider'd, it seems probable to me, that God in the Beginning form'd Matter in solid, massy, hard, impenetrable, moveable Particles, of such Sizes and Figures, and with such other Properties, and in such Proportion to Space, as most conduced to the End for which he form'd them; and that these primitive Particles being Solids, are incomparably harder than any porous Bodies compounded of them; even so very hard, as never to wear or break in pieces; no ordinary Power being able to divide what God himself made one in the first Creation.... And therefore, that Nature may be lasting, the Changes of corporeal Things are to be placed only in the various Separations and new Associations and Motions of these permanent Particles.

—Newton, *Opticks* [p. 400]

Electromagnetic radiation remains even today the most important carrier of information. But the range of frequencies we can observe has widened to encompass the full spread from extremely long radio waves, through microwaves, to the infrared and visible light, the ultraviolet and x-rays, and up to the hard gamma rays. Because not all frequencies reach the surface of Earth, until recently mankind did not have all the "windows" open for viewing the universe. With space satellites, however, all signals from the cosmos can be received and investigated unimpeded (Figures 5.176, 177).

We have already mentioned the ordinary optical telescope as well as the use of spectral analysis for astronomical investigations, and we now turn our attention to the more important accomplishments of radio astronomy, which has already grown into a widely employed research technique. In addition, we should mention the "neutrino telescope," which certainly has a bright future (Figure 5.178).

Radio astronomy can be seen as having been born in 1931, when Jansky observed a signal at wavelength 15 meters coming from the center of the Milky Way. In 1946, J. S. Hey, J. Phillips, and S. Parsons were able to determine the source of these radio waves more precisely by using radio interferometry. Among these objects, the quasars (quasi-stellar sources of radio emission), discovered by Sandage and Matthews in 1960, deserve special attention. In 1963, it was determined by M. Schmidt that the light from quasars is subject to a particularly large red shift, and this is why they can be considered to be among the most distant objects in the universe. In 1967, pulsars (pulsating sources of radio emission) were discovered, and it turned out that they are rapidly rotating high-density neutron stars.

Our ideas about the early evolution of the universe have been largely formed by an immensely important observation by two physicists working at Bell Laboratories: In 1965, Arno Penzias and Robert Wilson discovered a uniform background radiation corresponding to a temperature of 2.7 K.

We have still to mention two additional information channels: one, the study of cosmic radiation, and the other, neutrino astronomy, still under development. Because of their very weak interaction with matter, neutrinos are capable of giving us immediate information from inside stars, about which we would otherwise be able to learn only very indirectly and with great time delay. At the same time, in extremely dense materials, such as in a neutron star, their interaction may be intensive, and they could have a decisive influence on the energy balance. The feeble interaction of neutrinos with matter is, of course, a disadvantage for their detection. This is the key difficulty in trying to wrest the information they carry from them.

In the following discussion, we first look at the physical processes in the energy production of stars as an application of nuclear physics. Knowing these processes, we can also understand the developmental history of the stars—how they are born and how they die. Here we can meet with extraordinary states of matter whose creation under terrestrial conditions we could not even dream of. Finally, we discuss the theory of the evolution of the universe that is today the most probable, the theory of the big bang, which, as we hinted, is supported above all by radio astronomy's discovery of the background radiation.

5.6.2 *Energy Production in the Stars*

Poets write hymns in praise of sunlight, which nurtures all (Quotation 5.59), and engineers capture sunlight to make it do useful work: The power reaching our Earth per square centimeter when the Sun is at the zenith and the measurement is taken outside the atmosphere is 0.135 W—that is, we can count on roughly 1 kW of

power for every square meter. This means that all of mankind's energy needs could be met by the solar radiation impinging on a surface the size of a small country. We obtain even more impressive numbers if we consider the power emitted per square meter from the Sun's surface. From the total radiated power of the Sun of 3.86×10^{33} erg/s (3.86×10^{23} kW) and a solar diameter of 1.392×10^{26} km we obtain a value of 60 MW/m^2, which can be compared with the output of a small power plant.

Where does the Sun get such immense quantities of energy to waste? This question could arise, of course, only after the conservation of energy law was discovered. In the 1870s, Helmholtz opined that the energy radiated by the Sun is balanced by a decrease in its gravitational energy. It is not difficult to calculate the total gravitational energy of the Sun, and then the length of time that this energy could maintain the observed rate of radiation. Helmholtz obtained 30 million years as the result. However, the biologists of the period already had concrete ideas about the evolution of life based on Darwin's theory of evolution. For them, the time span given by the physicist Helmholtz was too small, even though it is very long in comparison to a human life span. They were correct: Radioactive dating, in agreement with other methods, estimates the age of Earth at about five billion years. From this, it follows that the gravitational energy, regardless of its crucial role in many cosmic phenomena such as the origin of stars and the evolution of galaxies, is insufficient to make up for the Sun's energy loss by radiation.

At the beginning of the twentieth century, after the discovery of radioactivity, the question was whether this phenomenon might be directly related to energy production in the stars. But this energy, too, turned out to be insufficient. Here it is worth noting that, in contrast, our ideas about Earth's energy balance were fundamentally changed when the energy from radioactive decay could be taken into account.

In 1920, Eddington conjectured that, within stars, radiation arising from the annihilation of electrons and protons might play a role. The proton and the electron would cease to exist, and their mass—or we could say energy—in the sense of mass–energy equivalence reappears as photons. Later it was shown that this would lead to a lifetime for the Sun that is much too long, and also, Eddington himself suspected that, and as we know from the previous sections, the mass of the proton and electron—in other words, their rest energy—cannot be transformed completely into radiation; this process is prohibited, among other reasons, by the conservation law of baryonic charge—one of the most precisely verified laws of physics.

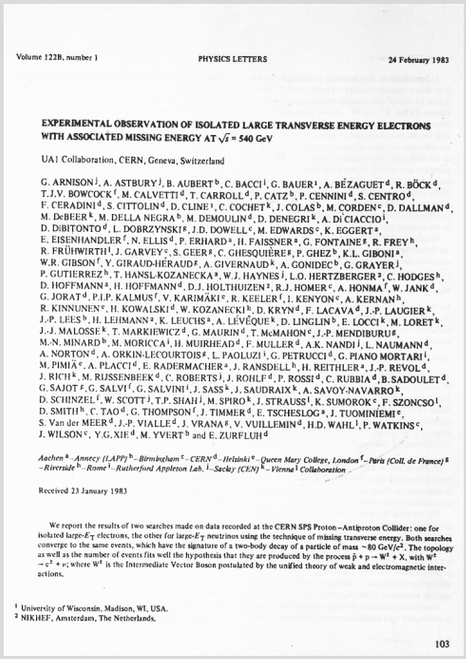
In 1929, Atkinson and Houtermans showed the correct path for a solution to the problem, namely that the energy production of the Sun and all other stars is due to the nuclear reactions within them that take place at high temperatures. At that time, however, knowledge of nuclear physics was far from sufficient to identify what the plausible reactions might be and, based on those, provide numerical estimates that could be compared with the existing observations.

The Sun's composition and its thermodynamic characteristics have been well known from general thermodynamic, energetic, and equilibrium considerations: the temperature at the center of the Sun is 15×10^6 K, the pressure is 2×10^{11} atm, and the density is 100 g/cm^3. Weizsäcker had the idea that at the roughly ten-million-degree temperatures of the Sun's interior, a sequence of fusion reactions could occur that would transform four protons, or hydrogen atoms, into a helium atom. The precise calculations were carried out by Bethe and Critchfield in 1938. Called the proton–proton (or p–p) cycle, this reaction consists of the

Quotation 5.58

There are two things which imbue the mind with a feeling of admiration and reverence, ever renewed, and ever on the increase, the more frequently and the more perseveringly our thoughts are occupied with them: *the star-clad sky* there *above us*, *and the moral law within ourselves*. There is no need that we should search after them, or merely surmise them, as hidden in obscurities, or as having to be placed in a transcendent region beyond our horizon. We see them before us and connect them directly with the consciousness we have of our existence. The starting-point of the former is the place which we occupy in the external world of the senses, and as for the connection in which we find ourselves placed, its dimension is widened to an immeasurable extent, with worlds upon worlds, systems upon systems, and moreover, limitless times, in respect of their periodic motions, their beginning and their continuance. The starting-point of the latter is our invisible self, our personality, and exhibits us as having our place within a world which has true infinity, while, on the other hand, it is only the understanding to which it is discernible. And as for our way of being connected with this world (and along with this, with all the aforementioned visible worlds), we recognize it as a connection which does not, as happened in the former case, exhibit a merely contingent character but one which is universal and necessary. That which engaged our attention first—the sight, that is, of a countless multiplicity of worlds—annihilates us in importance, as it were, inasmuch as we are *animal creatures*, obliged to give back to the planet (a mere speck in the universe) the matter from which it came, these creatures having been provided for a short time—one does not know how—with a vital force. As regards the second characteristic, on the other hand, it infinitely raises our value; we are thought of as having the attribute of being *intelligences*, and that by virtue of our personality containing within it the moral law which reveals to us a life independent of animality and even of the whole world of sense. Such is the truth of the matter, at least as regards anything which may be gathered from what happens when a purposive destiny is assigned to our existence through the instrumentality of this law, a destiny not restricted to the conditions and limits of this life, but which, on the contrary, reaches out to infinity.

—Kant, *Critique of Practical Reason* [p. 201]



Volume 122B, number 1 PHYSICS LETTERS 24 February 1983

EXPERIMENTAL OBSERVATION OF ISOLATED LARGE TRANSVERSE ENERGY ELECTRONS WITH ASSOCIATED MISSING ENERGY AT $\sqrt{s}$ = 540 GeV

UA1 Collaboration, CERN, Geneva, Switzerland

G. ARNISON[j], A. ASTBURY[j], B. AUBERT[b], C. BACCI[i], G. BAUER[1], A. BÉZAGUET[d], R. BÖCK[d], T.J.V. BOWCOCK[f], M. CALVETTI[d], T. CARROLL[d], P. CATZ[b], P. CENNINI[d], S. CENTRO[d], F. CERADINI[d], S. CITTOLIN[d], D. CLINE[1], C. COCHET[k], J. COLAS[b], M. CORDEN[c], D. DALLMAN[d], M. DeBEER[k], M. DELLA NEGRA[b], M. DEMOULIN[d], D. DENEGRI[k], A. Di CIACCIO[i], D. DiBITONTO[d], L. DOBRZYNSKI[g], J.D. DOWELL[c], M. EDWARDS[c], K. EGGERT[a], E. EISENHANDLER[f], N. ELLIS[d], P. ERHARD[a], H. FAISSNER[a], G. FONTAINE[g], R. FREY[h], R. FRÜHWIRTH[1], J. GARVEY[c], S. GEER[g], C. GHESQUIÈRE[g], P. GHEZ[b], K.L. GIBONI[a], W.R. GIBSON[f], Y. GIRAUD-HÉRAUD[g], A. GIVERNAUD[k], A. GONIDEC[b], G. GRAYER[j], P. GUTIERREZ[h], T. HANSL-KOZANECKA[a], W.J. HAYNES[j], L.O. HERTZBERGER[2], C. HODGES[h], D. HOFFMANN[a], H. HOFFMANN[d], D.J. HOLTHUIZEN[2], R.J. HOMER[c], A. HONMA[f], W. JANK[d], G. JORAT[d], P.I.P. KALMUS[f], V. KARIMÄKI[e], R. KEELER[f], I. KENYON[c], A. KERNAN[h], R. KINNUNEN[e], H. KOWALSKI[d], W. KOZANECKI[b], D. KRYN[d], F. LACAVA[g], J.-P. LAUGIER[k], J.-P. LEES[b], H. LEHMANN[a], K. LEUCHS[a], A. LÉVÊQUE[k], D. LINGLIN[b], E. LOCCI[k], M. LORET[k], J.-J. MALOSSE[k], T. MARKIEWICZ[d], G. MAURIN[d], T. McMAHON[c], J.-P. MENDIBURU[g], M.-N. MINARD[b], M. MORICCA[i], H. MUIRHEAD[d], F. MULLER[d], A.K. NANDI[j], L. NAUMANN[d], A. NORTON[g], A. ORKIN-LECOURTOIS[g], L. PAOLUZI[i], G. PETRUCCI[d], G. PIANO MORTARI[1], M. PIMIÄ[e], A. PLACCI[d], E. RADERMACHER[a], J. RANSDELL[b], H. REITHLER[a], J.-P. REVOL[d], J. RICH[k], M. RIJSSENBEEK[d], C. ROBERTS[j], J. ROHLF[g], P. ROSSI[d], C. RUBBIA[d], B. SADOULET[d], G. SAJOT[g], G. SALVI[f], G. SALVINI[i], J. SASS[k], J. SAUDRAIX[k], A. SAVOY-NAVARRO[k], D. SCHINZEL[f], W. SCOTT[j], T.P. SHAH[j], M. SPIRO[k], J. STRAUSS[1], K. SUMOROK[c], F. SZONCSO[1], D. SMITH[b], C. TAO[d], G. THOMPSON[f], J. TIMMER[d], E. TSCHESLOG[a], J. TUOMINIEMI[e], S. Van der MEER[d], J.-P. VIALLE[d], J. VRANA[g], V. VUILLEMIN[d], H.D. WAHL[1], P. WATKINS[c], J. WILSON[c], Y.G. XIE[d], M. YVERT[b] and E. ZURFLUH[d]

Aachen[a] – Annecy (LAPP)[b] – Birmingham[c] – CERN[d] – Helsinki[e] – Queen Mary College, London[f] – Paris (Coll. de France)[g] – Riverside[h] – Rome[i] – Rutherford Appleton Lab.[j] – Saclay (CEN)[k] – Vienna[l] Collaboration

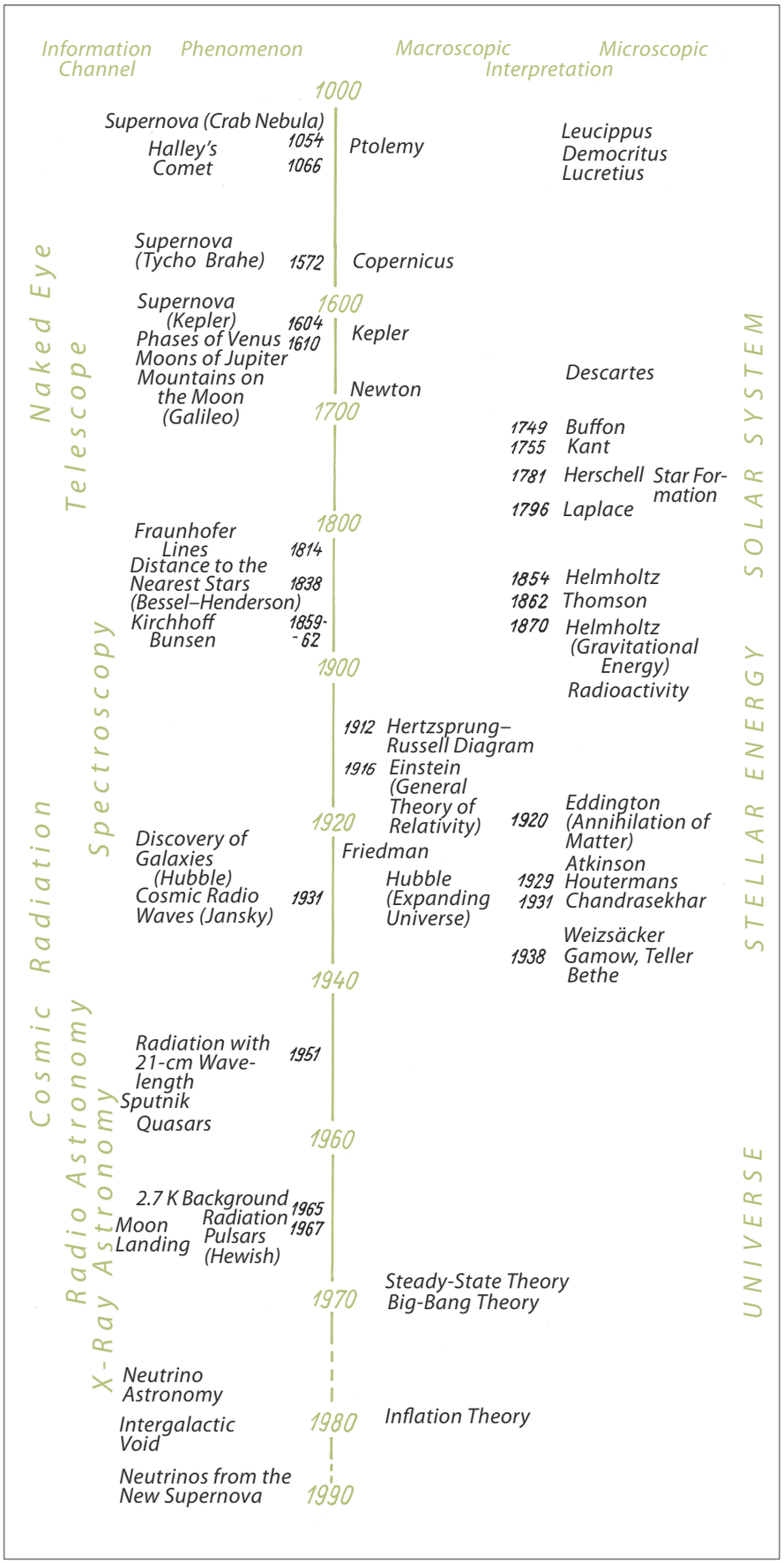
Received 23 January 1983

We report the results of two searches made on data recorded at the CERN SPS Proton–Antiproton Collider: one for isolated large-E_T electrons, the other for large-E_T neutrinos using the technique of missing transverse energy. Both searches converge to the same events, which have the signature of a two-body decay of a particle of mass ~ 80 GeV/c^2. The topology as well as the number of events fits well the hypothesis that they are produced by the process $\bar{p} + p \rightarrow W^{\pm} + X$, with $W^{\pm} \rightarrow e^{\pm} + \nu$; where $W^{\pm}$ is the Intermediate Vector Boson postulated by the unified theory of weak and electromagnetic interactions.

[1] University of Wisconsin, Madison, WI, USA.
[2] NIKHEF, Amsterdam, The Netherlands.

▲ **Figure 5.174** It is not only technical and economic demands that are unimaginably high for the realization of experiments in particle physics, but also the sheer intellectual effort that is involved. Here is the article on the discovery of the $W^{\pm}$ and Z^0 bosons with its more than 100 authors. CHEN-NING YANG correctly notes:

> The inevitable movement against large dimensions is regrettable, since it results in the limitation of free and individual initiative. Research loses intimacy, inspiration, and controllability. Nevertheless, one must accept it as a fact of life; however, we draw courage from the knowledge—despite their overwhelming physical dimensions—that these machines, detectors, and indeed also the experiments themselves are based on ideas that possess that simplicity, intimacy, and controllability that at all times has made research so thrilling and inspiring.



▶ **Figure 5.175** Chronology of our knowledge of the universe.

following steps (Figure 5.179):

$$ {}_1H^1 + {}_1H^1 = {}_1H^2 + e^+ + \nu + 1.44 \text{ MeV}, $$

$$ {}_1H^2 + {}_1H^1 = {}_2He^3 + \gamma + 5.49 \text{ MeV}, $$

$$ {}_2He^3 + {}_2He^3 = {}_2He^4 + 2\,{}_1H^1 + 12.85 \text{ MeV}. $$

The temperature dependence of the energy released in this cycle is illustrated in Figure 5.182, discussed later. In everyday language, we say that this reaction burns the stellar hydrogen into helium.

In 1938, Bethe hit upon the idea of another cycle, the CNO, or carbon–nitrogen–oxygen, cycle. The result of this cycle is again the construction of a helium nucleus from four hydrogen nuclei, with carbon serving merely as a catalyst. Here are the individual steps of the cycle (Figure 5.180):

$$ {}_6C^{12} + {}_1H^1 = {}_7N^{13} + \gamma + 1.95 \text{ MeV}, $$

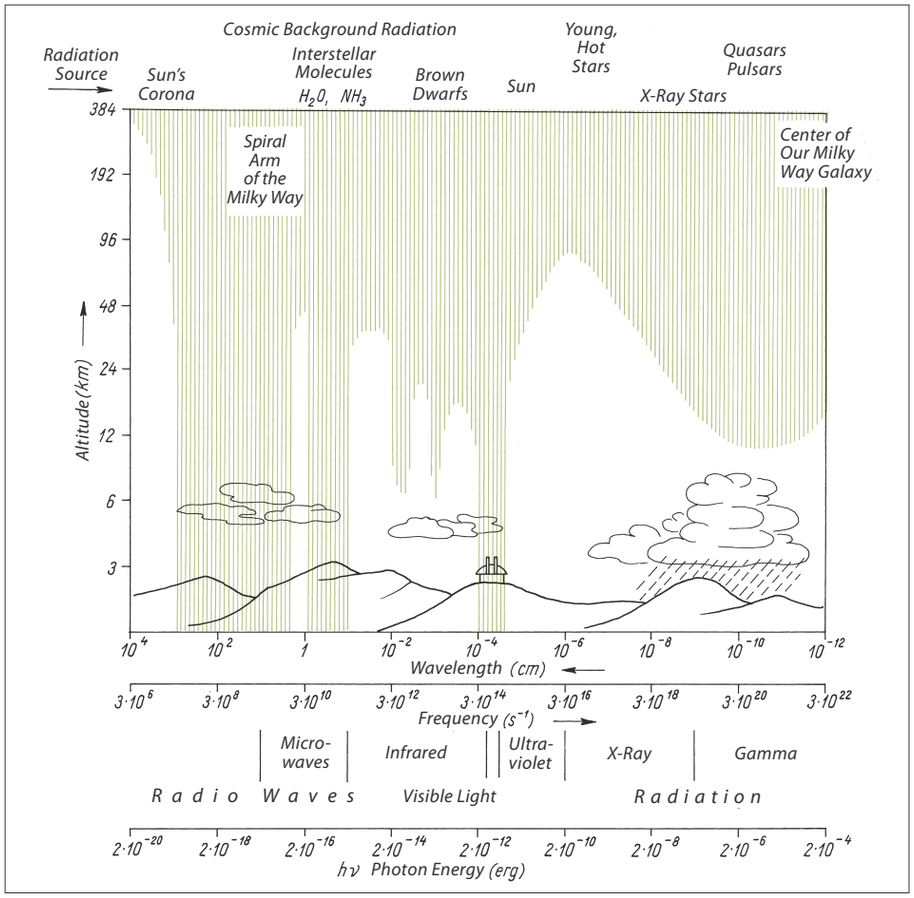
$$ {}_7N^{13} = {}_6C^{13} + e^+ + \nu + 2.22 \text{ MeV}, $$

$$ {}_6C^{13} + {}_1H^1 = {}_7N^{14} + \gamma + 7.54 \text{ MeV}, $$

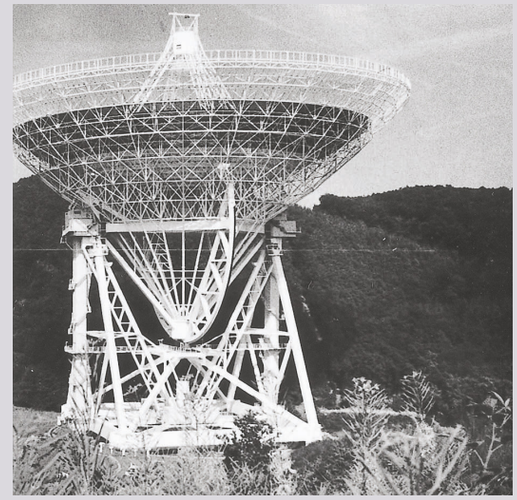
$$ {}_7N^{14} + {}_1H^1 = {}_8O^{15} + \gamma + 7.53 \text{ MeV}, $$

$$ {}_8O^{15} = {}_7N^{15} + e^+ + \nu + 7.21 \text{ MeV}, $$

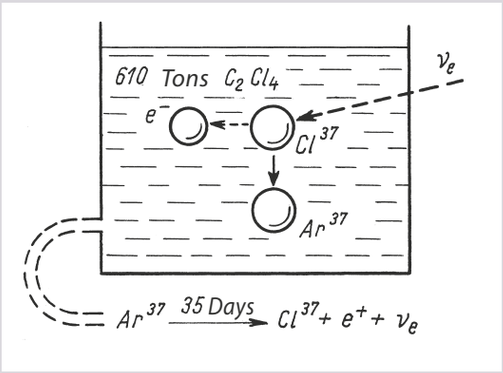
$$ {}_7N^{15} + {}_1H^1 = {}_6C^{12} + {}_2He^4 + 4.96 \text{ MeV}. $$



◀ **Figure 5.176** Some electromagnetic waves are absorbed by the atmosphere, others are able to pass through it, depending on their frequency or, in other words, their wavelength; as messengers from distant parts of the universe, they reach our Earth's surface through a narrow "window." To evaluate the entire spectrum, instruments must also be placed aboard Earth-orbiting satellites.



▲ **Figure 5.177** Alongside telescopes designed to capture light waves, radio-telescopes, which are powerful antennas, have become ever more important in unraveling the mysteries of the universe. (After Roland Gööck.)



▲ **Figure 5.178** Diagram of a neutrino telescope (R. Davis, J. Bahcall, 1967). This "telescope" has nothing in common with an optical telescope: it is buried nearly a mile (1480 m) beneath the surface of Earth to be shielded from disturbing influences. It consists of a container filled with 610 tons of perchloroethylene (C_2Cl_4), which is ordinary dry-cleaning fluid. Neutrinos with an energy greater than 0.814 MeV can create radioactive argon from the chlorine nuclei; every 100 days, this argon is "washed out" and its activity measured. High-energy neutrinos from the interior of the Sun are released through the relatively rare reaction ${}_5B^8 \rightarrow 2\ {}_2He^4 + e^+ + \nu$, which is part of the energy-producing p-p cycle. It is easy to determine that the Sun is the place of origin of the neutrinos because the measured activity changes through the seasons with the changing distance between the Sun and Earth. However, the strength of the signal did not agree with the theory. The solution to this "solar neutrino" problem—neutrino "flavor" oscillation—was not found until the late 1990s. Davis shared in the Nobel Prize in 2002.

The gravitational collapse preceding a supernova explosion can be detected from the sudden increase in neutrino emission. The direction can be determined from a comparison with measurement data from neutrino telescopes placed at various locations around the world.



▲ **Figure 5.179** The pharaoh Akhenaten and his wife Nefertiti receive the beneficent rays of the Sun god Ra with reverent devotion (ca. 1350 bce). We might imagine that those are photons in the royal hands.

Figure 5.181 also shows the temperature-dependence of this reaction. We see that at lower temperatures, the simple p–p cycle produces the greater quantity of power, whereas at higher temperatures it is the CNO cycle that does so.

At very high temperatures, of order 100 million degrees, and with a sufficient density of helium nuclei, the 3α, or Salpeter, process can occur. In this reaction, a carbon nucleus is created as the result of the collision of three helium nuclei, with a release of energy. Although the probability of a collision of three helium nuclei is very small, a favorable condition intervenes: The collision of two helium nuclei creates a beryllium isotope that has a relatively long lifetime, and we can then consider that this intermediate nucleus collides with the third helium nucleus. The reaction proceeds as follows:

$$ {}_2He^4 + {}_2He^4 = {}_4Be^8 + \gamma - 0.095\ \text{MeV}, $$

$$ {}_4Be^8 + {}_2He^4 = {}_6C^{12} + \gamma + 7.4\ \text{MeV}. $$

Figure 5.181 also shows the temperature-dependence of this third reaction. Moreover, the Sun is shown with its parameters lying near the p–p and CNO curves. Of course, there are other reactions taking place as well. The intermediate products of the p–p cycle, the He^3 nuclei, react with other particles that are present, such as He^3, p, and He^4, and form nuclei that ultimately lead to He^4. What is more important for the purposes of elementary particle research is that neutrinos of various energies arise in the process.

5.6.3 *Birth, Life, Death on a Cosmic Scale*

From what we have discussed, it follows that the Sun at this moment is consuming its hydrogen, while in its center, the "ashes" of this burning accumulate, which may then became the raw material for a combustion process that takes place at higher temperatures. The question arises, what happens when the hydrogen in the center runs out? What will be the fate of our Sun? Before we answer this question, it is worth investigating in greater detail how the Sun arrived at its present state. What caused the high temperatures in the interior at which thermonuclear reactions could begin? Of course, there is no way to track the complete developmental history of a star over the course of a human lifetime or even over the entire history of mankind. Not only with distances but also with times we have to use a cosmic scale. Nonetheless, we shall see that some phases of the changes play out explosively on a human time scale, that is, in seconds, days, or months.

The cataloguing of the properties of existing stars and the interpretation of this catalogue enables us to follow with high probability the developmental phases of a star. We can view stars with different properties as stars in different phases of their lives: Some are just being born, others are at the zenith of their development, and others show signs of old age. This is exactly like acquiring an understanding of the stages of human development by studying, at a single point in time, the population of a city: the infants, adults, and the aged.

From the point of view of studying stellar development, the most useful organization of the characteristics of actual stars is the Hertzsprung–Russell diagram (1912; Figure 5.182). This diagram shows the state of each star by a point, where the

horizontal coordinate represents its surface temperature and the vertical coordinate represents absolute brightness. Let us follow the development of a star with one solar mass on the Hertzsprung–Russell diagram: It begins with the gravitational contraction of a cosmic gas cloud. In the first phase, the gravitational energy is transformed into the ordered motion of the particles, until the shock wave created by the congestion of the high-velocity particles heats up the entire gaseous mass. It is the further slow contraction of the heated cloud that will finally lead to the temperatures in the interior of the star where production of energy by fusion can begin.

Some processes in the initial phases of stellar development take place over time spans on a human scale. For the greatest part, nearly 99%, of its life a star is consuming the available hydrogen in its center, and consequently most stars are located on the main sequence of the Hertzsprung–Russell diagram.

We can now predict the further fate of a star such as our Sun, which is of primary interest to us, at the time when the hydrogen in its center will have been consumed. When energy production in the center of the star ceases, the radiation pressure outward can no longer balance the gravitational pressure inward, and there will be a renewed gravitational contraction.

The energy thereby released further heats up the star, so that in the center, helium "combustion processes" begin, and the burning of the hydrogen can spread to the star's outer layers. Accordingly, these layers expand while the surface will cool, and the star becomes a red giant. Reaching this state, our Sun's expanding radius will extend beyond the orbit of Mercury, or even beyond that of Venus; it is even possible that our Earth will end up in the Sun's interior. Even if that would not happen, Earth's surface temperature will be about 2000 K, so that everything will dry out, burn, and melt, and life on Earth will be impossible.

After a star's supply of helium fuel has been exhausted, the star shrinks; its density increases, and it becomes a white dwarf. In some cases the outer layers expand again, more so than in the star's red-giant era, and the star appears in our sky as a bright "nova."

The white dwarf slowly cools down and then continues its existence as a cold, gray, and unimportant object. It contains the end products of the transformation processes: carbon nuclei as well as nitrogen and oxygen nuclei arising from their capture of protons or helium nuclei.

Much more exciting is the fate of stars with a mass greater than that of the Sun when their supply of helium is exhausted. It had been known earlier that the state of a white dwarf for stars of large mass cannot be stable: With such stars, the gravitational contraction continues until a temperature of a billion degrees is reached in the star's interior. At this temperature, elements of higher atomic number up to iron can be created by interaction processes of carbon nuclei, and through capture of the neutrons released at the even higher temperatures, eventually all the elements of the periodic table can be produced.

According to Hoyle and colleagues, this state of high temperature and large mass is unstable. The excess mass is flung into the cosmos through an explosion-like expansion of the star's shell, and the result is the most striking phenomenon to be observed in the heavens—the explosion of a supernova. It appears, as its name suggests, as a new star in the sky. Its brightness may be so great that it may be seen from Earth even in broad daylight. History has recorded three famous supernovas. In 1054, Chinese astronomers recorded a new star: Its remains today form the Crab Nebula. We have already read, in Quotation 3.9, about the second supernova to be observed,

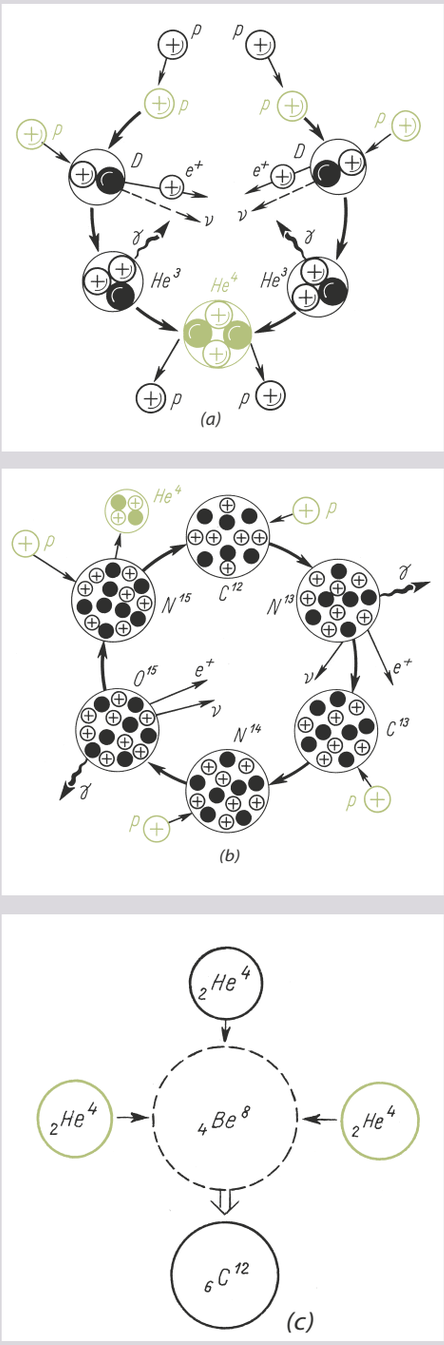
Quotation 5.59

When it is dawn and you rise in the horizon and shine as the sun in the day, you dispel the darkness and shed your beams. The Two Lands keep festival, awake, and stand on their feet, for you have raised them up. They wash their bodies, they take their garments, and their hands praise your arising. The whole land, it does its work.

All beasts are content with their pasture, the trees and herbs are verdant. The birds fly out of their nests and their wings praise you. All wild beasts dance on their feet, all that fly and flutter—they live when you arise for them.

The ships voyage down and up stream likewise, and every way is open, because you arise. The fishes in the river leap up before your face. Your rays are in the sea.

—"Hymn to the Sun," from the time of pharaoh Akhenaten



▲ **Figure 5.180** Nuclear reactions that provide the energy for stars.

(a) p-p cycle.

(b) CNO cycle.

(c) Salpeter reaction.

by Tycho Brahe in 1572. This incident led Tycho to doubt for the first time the infallibility of Aristotle. The third supernova was described by Kepler in 1604.

The appearances of supernovas, these cosmic explosions, are of interest to us for two reasons. First, we would like to know the fate of the remains of the star that survive after the explosion. As we have mentioned, the state of a white dwarf of sufficiently great mass is unstable because the gravitational forces will cause a collapse. The electrons are forced by the gravitational interaction into the nucleus of the atom, so that the density of the star is comparable to the density of an atomic nucleus. One can therefore conceive of such stars as a single huge nucleus: This is the neutron star. Because of the enormous density of the neutron star, the neutrinos come to play an important role. The neutrinos created in the interior of stars interact with the surface layer, bringing their energy to it, and ultimately it is perhaps this energy that is responsible in some measure for the eruption of a substantial portion of the matter of the star, which we observe as a supernova. In Figure 5.183, this development is shown along with characteristic data on neutron stars such as dimensions and density.

As early as 1932, that is, in the year of the discovery of the neutron, Landau suggested the possibility of the existence of neutron stars, and similar ideas were expressed by Oppenheimer. When much later, a pulsar was found in the interior of the Crab Nebula, which is—as we have mentioned—the remains of a supernova, it was at once conjectured that the radio signals received in an exact pulsating rhythm brought news about a neutron star (Figure 5.184).

The explosion of a supernova is also of interest for a second reason: The necessary conditions for the creation of the heavy elements exist in stars in the phase prior to a supernova explosion. During the explosion, these elements are scattered throughout the cosmos, and from them a new star can arise. These elements are all to be found in our solar system; without them, the complex organization of matter that is a condition of life could not have formed. Thus, in all probability, a great portion of the atoms making up our bodies once took part in a supernova explosion.

It seems that the neutron star is still not the last stage in the life of a star of large mass. One can imagine matter being compressed to such an extent that the strength of its gravitation would prevent any information carrier leaving it. Even photons have mass, so under a strong enough gravitational field, just like a material body thrown upward, they will also stop and fall back toward the center. We call these most unusual objects in our universe, understandably, black holes. How can we observe them? Do they exist at all? If next to a visible star there is a black hole, they behave like a double star with one invisible partner; also, the black hole will leach plasma from the companion star and the resulting gamma rays can be observed.

Black holes are of great theoretical importance. The matter within them is in such a state that the otherwise superweak gravitational interaction plays a decisive role. It is clear that to answer the questions about these objects, some blending of the general theory of relativity with quantum mechanics will be necessary.

5.6.4 *The Formation of the Universe*

By bringing the characteristics of the geometry of space, such as its metrics and curvature, into direct relationship with the (average) density of matter, the general theory of relativity sets the stage for the great cosmic play, but with a company of undistinguished actors without roles. The storyline and the casting will be

primarily the responsibility of quantum mechanics, although among the stage, the story, and the characters are connections that are in some cases forced by circumstances but remain for the most part unclear (Quotations 5.60, 5.61, 5.62).

As shown by Friedman at the beginning of the 1920s, the fundamental equations of general relativity can only have an expanding or a contracting universe as solutions (Figure 5.185). This conclusion from the fundamental equations of the general theory of relativity appeared even earlier—Einstein himself, however, considered such a conclusion absurd and added another term to the equation in order to account for a static nature of the universe ("my greatest blunder," as he later called it). Hubble's observations in 1929 unequivocally confirmed the expansion of the universe (Figure 5.186). If we project the expansion back in time, at a specific point in the past, the matter of the universe must have been compressed into a singularity. Lemaître was the first to attempt to give physical content to this picture. As he put it, the matter of the entire universe was at that point in time crammed into a sphere of infinitely small dimensions and at an extraordinarily high temperature, and then began to expand. The theory of the hot and expanding universe was more precisely worked out by Gamow. His goal was to provide an explanation for the origin of the heavy elements found on Earth as well as elsewhere in the cosmos. It had already been recognized that the temperatures within stars such as our Sun are not sufficient to permit the creation of the elements of the second half of the periodic table via nuclear reactions. Therefore, Gamow assumed that in its primordial state, the universe must have been sufficiently "hot" to enable their synthesis. Later, Zeldovich and others showed that while the basic idea of a hot early universe is correct, it was not so for the task for which it had been conceived, namely, for an explanation of the existence of the heavy elements.

In somewhat simplified form, our image today of the origin of the universe is as follows (Figures 5.187 and also 5.196, discussed later): The universe began with a primordial explosion, the big bang. This explosion may have taken place roughly 15 billion years ago. Disputes here are in the range of a few billion years, give or take. It is therefore all the more astonishing that the events immediately after the birth of our universe proceeded with a dramatic swiftness and that the theory can record them with great accuracy.

The first instants following the explosion will be the subject for a later section. By 10^{-11} seconds into the life of the universe, the four known interactions are already working separately. After this, we are in more familiar territory. Around 10^{-6} seconds, the quarks assemble into baryons (protons and neutrons) and into light nuclei. The universe at one minute is a mixture of hydrogen and helium nuclei, electrons, and photons. Its temperature is about a billion degrees. After 300,000 years, the light neutral atoms can form, the energy of the photons is insufficient for ionization, and so the universe has become transparent.

In space, with its essentially homogeneous distribution of matter, only random fluctuations in density occur. As long as the helium and hydrogen gases are in an ionized state, their interaction with photons is strong enough to prevent locations with initially higher density to become even denser due to gravitational attraction. What happens is that the contracting matter "encloses" the photons, which then work against the gravitational attraction from within. But if the size of the denser area is very large, this effect is less significant, especially as the universe cools to about 4000 K where the ionization rate is lower and the interaction of the photons with the neutral gas becomes negligible. So the initial denser regions, whose size is on the order of 100 million light years, can begin their contraction, now under

Quotation 5.60

Having established these distinctions, we may now proceed to the question whether the heaven is ungenerated or generated, indestructible or destructible. Let us start with a review of the theories of other thinkers; for the proofs of a theory are difficulties for the contrary theory. Besides, those who have first heard the pleas of our adversaries will be more likely to credit the assertions which we are going to make. We shall be less open to the charge of procuring judgment by default. To give a satisfactory decision as to the truth it is necessary to be rather an arbitrator than a party to the dispute.

That the world was generated all are agreed, but, generation over, some say that it is eternal, others say that it is destructible like any other natural formation. Others again, with Empedlocles of Acragas and Heraclitus of Ephesus, believe that there is alternation in the destructive process, which takes now this direction, now that, and continues without end.

Now to assert that it was generated and yet is eternal is to assert the impossible; for we cannot reasonably attribute to anything any characteristics but those which observation detects in many or all instances. But in this case the facts point the other way: generated things are seen always to be destroyed. Further, a thing whose present state had no beginning and which could not have been other than it was at any previous moment throughout its entire duration, cannot possibly be changed. For there will have to be some cause of change, and if this had been present earlier it would have made possible another condition of that to which any other condition was impossible. Suppose that the world was formed out of elements which were formerly otherwise conditioned than as they are now. Then if their condition was always so and could not have been otherwise, the world could never have come into being. And if the world did come into being, then, clearly, their condition must have been capable of change and not eternal: after combination therefore they will be dispersed, just as in the past after dispersion they came into combination, and this process either has been, or could have been, indefinitely repeated. But if this is so, the world cannot be indestructible, and it does not matter whether the change of condition has actually occurred or remains a possibility.

Some of those who hold that the world, though indestructible, was yet generated, try to support their case by a parallel which is illusory. They say that in their statements about its generation they are doing what geometricians do when they construct their figures, not implying that the universe really had a beginning, but for didactic

continued on next page

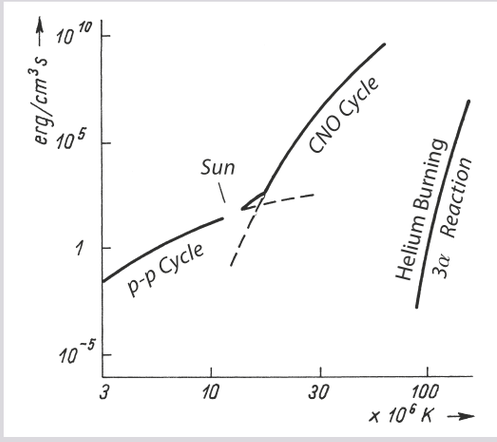
Quotation 5.60, ***continued***

reasons facilitating understanding by exhibiting the object, like the figure, as in course of formation. The two cases, as we said, are not parallel; for, in the construction of the figure, when the various steps are completed the required figure forthwith results; but in these other demonstrations what results is not that which was required. Indeed it cannot be so; for antecedent and consequent, as assumed, are in contradiction. The ordered, it is said, arose out of the unordered; and the same thing cannot be at the same time both ordered and unordered; there must be a process and a lapse of time separating the two states. In the figure, on the other hand, there is no temporal separation. It is clear then that the universe cannot be at once eternal and generated.

To say that the universe alternately combines and dissolves is no more paradoxical than to make it eternal but varying in shape. It is as if one were to think that there was now destruction and now existence when from a child a man is generated, and from a man a child.

If the world is believed to be one, it is impossible to suppose that it should be, as a whole, first generated and then destroyed, never to reappear; since before it came into being there was always present the combination prior to it, and that, we hold, could never change if it was never generated. If, on the other hand, the worlds are infinite in number the view is more plausible.

—Aristotle *On the Heavens,* Book I, Part 10 [2009]

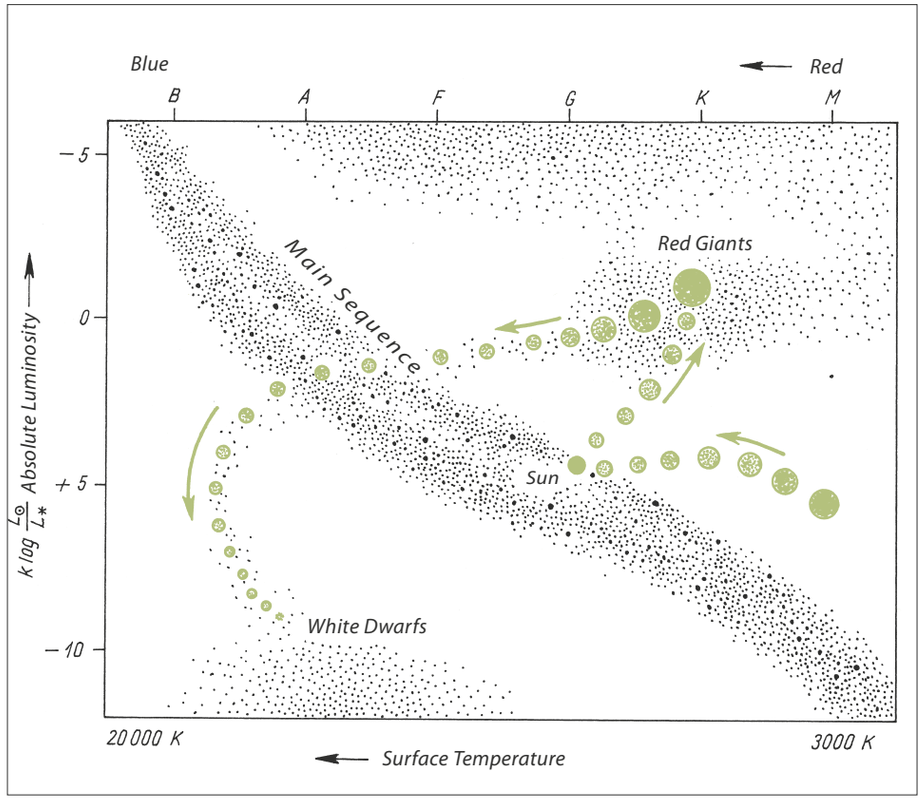


▲ **Figure 5.181** Energy production from the nuclear reactions in the interiors of stars as a function of temperature.

the influence of gravitational force. The first structures in the originally homogeneous universe are formed in this way: the galaxy clusters. One characteristic of the theory of the hot universe is the particular order of the birth of cosmic objects: First come the galaxy clusters, then the galaxies, then the star clusters, and finally the stars arise, that is, the smaller units are formed by a process of differentiation of larger ones. This sequence is not obvious because the opposite order is also conceivable. The primeval cloud could well have contracted first locally into stars, which then, due to their gravitational attraction, would have gradually merged into larger units.

With a more detailed investigation of the phenomena, the form of galaxies, which is similar to a discus, as well as their rotation can be explained. Because the gravitational force field is a potential field, according to the hydrodynamic theorems of Helmholtz and Kelvin, gravitation cannot cause the rotation; however, a satisfactory explanation may be found in the turbulence caused by shock waves through the galaxy disks.

The reader may well be astonished at the boldness of scientists in combining the knowledge obtained from terrestrial laboratories with that gleaned from observations of the cosmos, still registered with terrestrial means, to reach back 15 billion years and reconstruct events that occurred in fractions of a second. Is this story really more believable than the creation myths? I believe that we can all agree that the physicists' "Genesis" is no less rich in imagination than any creation myth. And the results of this "creation" exceed every apocalyptic vision: rapidly rotating neutron stars with densities millions of times greater than that of water; black holes that devour everything that comes into their field of attraction and allow no



▲ **Figure 5.182** Hertzsprung–Russell diagram and the life-cycle of a star with the mass of our Sun.

information to leave them; quasars that bear tidings from the edge of the universe from a distance of 10 billion light years about events that took place 10 billion years ago, in which energies equivalent to billions of Suns were released in a space not larger than one star; an expanding universe whose most distant objects rush away from us at close to the speed of light.

We have already noted that the investigation of a unique phenomenon such as the history of the universe creates theoretical and methodological problems. In order to prove a law, we need a series of repeatable events. But if we recall the schema that was already advanced by Galileo—creating a hypothesis, applying mathematical tools to the hypothesis to compute quantities that are experimentally verifiable, and then verifying them experimentally—that schema can be applied here, also. But that entails a major difficulty. In the case of a concrete, repeatable phenomenon, we can vary the boundary conditions that express the effects of the rest of the world on the system under investigation, and by doing so we can obtain better and better evidence for our hypothesis. In contrast, when the entire universe is under consideration, the boundary conditions—whether in space or time—themselves become hypothetical. In this case, a "proof" simply means that the theory must account for the phenomena observed, but even then the objection can be raised that the theory has simply been designed only to summarize the observations. But if the theory will work for new phenomena, not previously observed, especially if it hints at the kind of phenomena we should be looking for and can expect to find, then its explanatory power could approach even the power of the explanations offered in classical physics.

Let us now take a look at the experimental foundations that support this grandiose vision. As we have already seen, the principal support for the theory is the radiation observed by radio astronomers in 1965 that reaches Earth not from particular stellar objects but equally from all directions. In Figure 5.188, the intensity distribution of this radiation is represented as a function of frequency. It is evident that the measured values match up very well with the curve given by Planck's formula for equilibrium radiation, which for this cosmic background radiation yields a temperature of 2.7 K. The theory of a hot and expanding universe also leads to this value from the consideration that the equilibrium radiation right after the big bang with a temperature of several thousand degrees must have cooled to this same value as a consequence of the expansion of the universe. Moreover, the dimensions of the galaxy clusters resulting from the theory agree with the measured values. Finally, the expansion of the universe is a generally accepted theoretical conclusion that rests on a firm observational foundation.

The energy of the universe must be accepted as a given quantity, and therefore as a sort of initial condition. But where does this energy come from? We find ourselves standing once again before the riddle of origin and eternal being, creation or eternity—the question that was puzzled over by Aristotle, Augustine, and Maimonides. But now the answer must rest on physical laws and must be verifiable by directly or indirectly measurable physical phenomena.

The big bang theory—extended by the corrections based on a grand unification theory (GUT) (inflationary phase, etc.) and strengthened by ever more precise measurements of the background radiation—seems at present to be the only viable theory of a dynamic universe.

Quotation 5.61

Those who ask "What was God doing before he made heaven and earth?" are still steeped in error which they should have discarded. "If he was at rest," they say, "and doing nothing, why did he not continue to do nothing for ever more, just as he had always done in the past? If the will to create something which he had never created before was new in him—if it was some new motion stirring in him—how can we say that his is true eternity, when a new will, which had never been there before, could arise in it?" ...

My answer to those who ask "What was God doing before he made heaven and earth?" is not "He was preparing Hell for people who pry into mysteries." This frivolous report has been made before now, so we are told, in order to evade the point of the question. But it is one thing to make fun of the questioner and another to find the answer. So I shall refrain from giving this reply. For in matters of which I am ignorant I would rather admit the fact than gain credit by giving the wrong answer and making a laughing-stock of a man who asks a serious question. ...

A fickle-minded man, whose thoughts were all astray because of his conception of time past, might wonder why you, who are God almighty, Creator of all, Sustainer of all, and Maker of heaven and earth, should have been idle and allowed countless ages to elapse before you finally undertook the vast work of creation. My advice to such people is to shake off their dreams and think carefully, because their wonder is based on a misconception.

How could those countless ages have elapsed when you, the Creator, in whom all ages have their origin, had not yet created them? What time could there have been that was not created by you? How could time elapse if it never was?

You are the Maker of all time. If, then, there was any time before you made heaven and earth, how can anyone say that you were idle? You must have made that time, for time could not elapse before you had made it.

But if there was no time before heaven and earth were created, how can anyone ask what you were doing "then"? If there was no time, there was no "then."

—Saint Augustine, *The Confessions* [1961, pp. 261, 262–263]

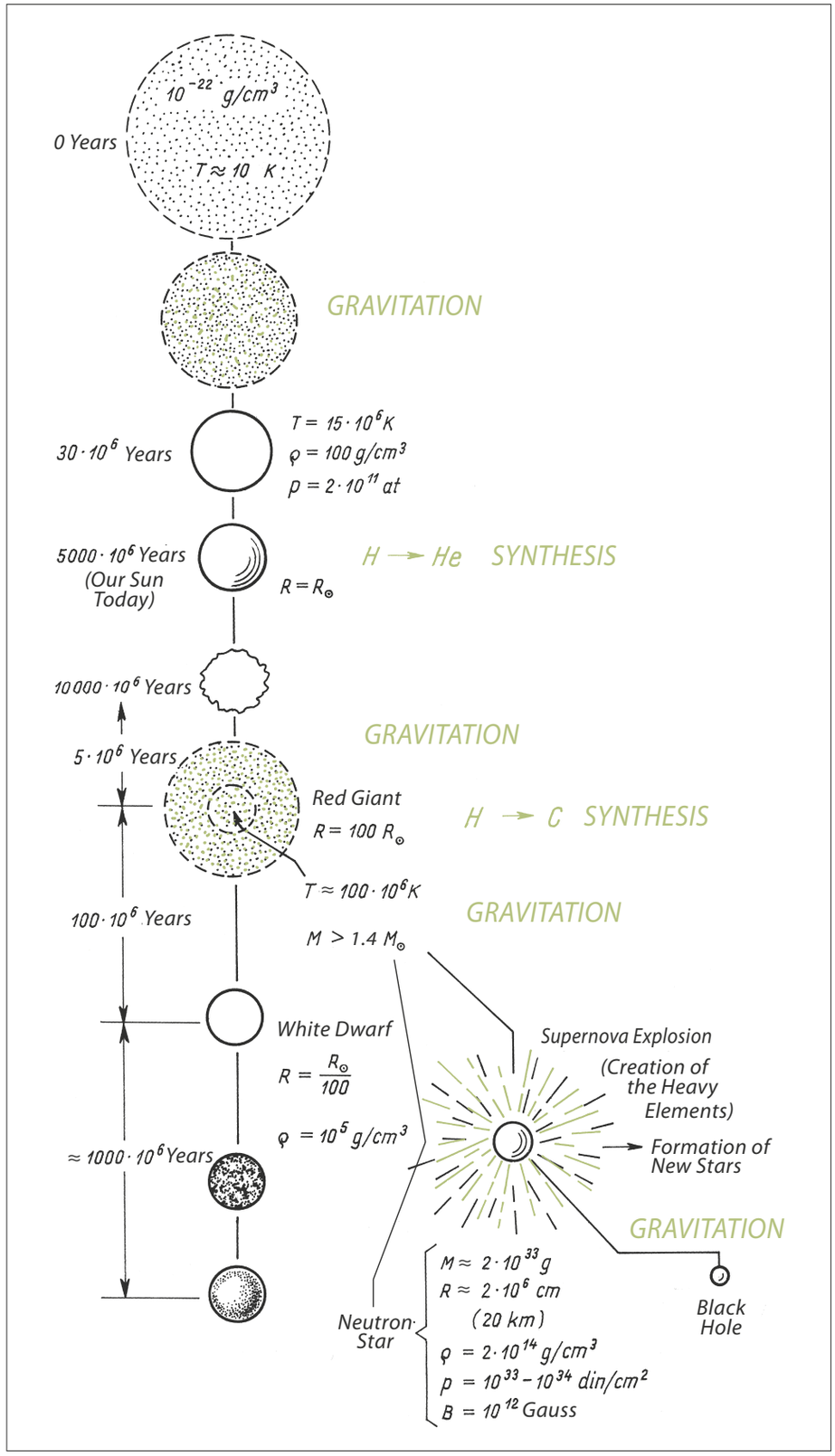
Quotation 5.62

Consequently [to the General Theory of Relativity], very old philosophic problems were taken up that had occupied the mind of man since the earliest phases of philosophy and science. Is space finite or infinite? What was there before the beginning of time? What will happen at the end of time? What will happen at the end of time? Or is there no beginning and no end? These questions had found different answers in different philosophies and religions. In the philosophy of ARISTOTLE, for instance, the total space of the universe was finite (though it was infinitely divisible). Space was due to the extension of bodies, it was connected with the bodies; there was no space where there were no bodies. The universe consisted of the earth and the sun and the stars: a finite number of bodies. Beyond the sphere of the stars there was no space; therefore, the space of the universe was finite.

In the philosophy of KANT this question belonged to what he called "antonomies"—questions that cannot be answered, since two different arguments lead to opposite results. Space cannot be finite, since we cannot imagine that there should be an end to space; to whichever point in space we come we can always imagine that we can go beyond. At the same time space cannot be infinite, because space is something that we can imagine (else the word 'space' would not have been formed) and we cannot imagine an infinite space. For this second thesis the argument of KANT has not been verbally reproduced. The sentence 'space is infinite' means for us something negative; we cannot come to an end of space. For KANT it means that the infinity of space is really given, that it 'exists' in a sense that we can scarcely reproduce. KANT's result is that a rational answer to the question whether space is finite or infinite cannot be given because the whole universe cannot be the object of our experience.

A similar situation is found with respect to the problem of the infinity of time. In the *Confessions* of Saint AUGUSTINE, for instance, this question takes the form: What was God before he created the world? AUGUSTINE is not satisfied with the joke: 'God was busy preparing Hell for those who ask foolish questions.' This, he says, would be too cheap an answer, and he tries to give a rational analysis of the problem. Only for us is time passing by; it is expected by us as future; it passes by as the present moment and is remembered by us as past. But God is not in time; a thousand years are for him as one day, and one day as a thousand years. Time has been created together with the world, it belongs to the world, therefore time did not exist before the universe existed. For God the whole course of

continued on next page



▲ **Figure 5.183** Life course of a star.

SUBRAHMANYAN CHANDRASEKHAR (1910–1995), Indian-born American astrophysicist, proposed that there is a limit (Chandrasekhar limit), of 1.44 solar masses, above which stars suffer a gravitational collapse and become neutron stars or black holes. For this work, Chandrasekhar was much belatedly awarded the Nobel Prize in 1983.



◀ **Figure 5.184** The Crab Nebula is the remnant of a supernova explosion in the year 1054; at its center is a neutron star. A supernova explosion is one of the most interesting spectacles in the universe, occurring in any given galaxy about once every 50 years. In the course of the explosion, energy on the order of magnitude of 10^{51} to 10^{52} ergs is released. Given that our Sun radiates 3.86×10^{33} erg/s, it would require 10^{18} seconds = 32 billion years to radiate such a quantity of energy. We encounter greater releases of energy only in quasars: There, the amount of energy in the radio-wave region alone is around 10^{52} erg/s. In comparison, the energy of an atomic bomb explosion is on the order of 10^{22} to 10^{25} ergs.

The supernova that was observed in 1987 in the Large Magellanic Cloud has special significance. The flare-up could be seen with the naked eye in the Southern Hemisphere. This galaxy is located in our immediate neighborhood, a mere 150,000 light-years away. According to theory, the greatest part (more than 99%) of the energy of the supernova explosion was dissipated in an estimated 10^{58} neutrinos. Some of these neutrinos were, in fact, detected. They arrived three hours earlier than the light from the explosion, since they were generated earlier than the light, further confirming the theory. In two installations built to study proton decay in Japan and in a salt mine in Ohio, 10 and 8 events, respectively, triggered by neutrinos were registered. At the same time, the decay of a proton has yet to be observed.

5.7 Summary and Preview

5.7.1 *Physics, Philosophy, and Society at the Turn of the Millennium*

In the second half of the twentieth century, and even more so in the last two or three decades, the public evaluation of science, especially that of physics, has changed radically, and thus the role of physics in society has also changed. Gone are the times when the man on the street and the official authorities both held science and scientists in almost superstitious awe, and particularly physicists, who made it possible to deter the enemy in the Cold War from attacking through the construction of a previously unimaginable arsenal of offensive and defensive weapons. On the right and left sides of the world, uncountable billions of dollars and rubles have been squandered to this end.

For its part, philosophy at the same time hopes to be able to maintain its prestige through a willingness to help, indeed with a certain homage to physics (Quotation 5.7).

Euphoria, but fright as well, has disappeared, and society now rightfully asks with a returned consciousness for what purpose the money of the community is being spent. Only such tasks are approved that serve to fulfill needs that are recognized as important by society itself.

An initial result of this fundamental change in point of view is this: The taxpaying citizen is intervening forcefully in the development of physics.

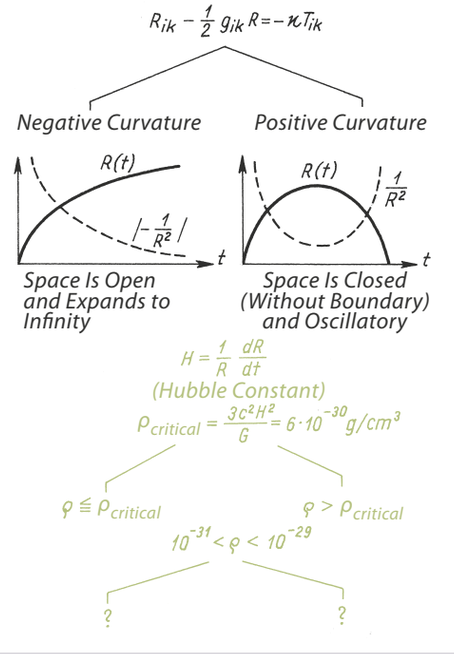
That best-known and most conspicuous event of this sort was the rejection by the United States of the SSC (Superconducting Super Collider), with proton en-

Quotation 5.62, ***continued***

the universe is given at once. There is no such time before He created the world. It is obvious that in such statements the word 'created' at once raises all the essential difficulties. This word as it is usually understood means that something has come into being that has not been before, and in this sense it presupposes the concept of time. Therefore, it is impossible to define in rational terms what could be meant by the phrase 'time has been created'. This fact reminds us again of the often discussed lesson that has been learned from modern physics: that every word or concept, clear as it may seem to be, has only a limited range of applicability.

In the theory of general relativity these questions about the infinity of space and time can be asked and partly answered on an empirical basis. If the connection between the four-dimensional geometry in space and time and the distribution of masses in the universe as a whole. At least one can build 'models' of the universe, cosmological pictures, the consequences of which can be compared with the empirical facts.

—Werner Heisenberg, *Physics and Philosophy* [pp. 109–111]



▲ **Figure 5.185** The fundamental equations of Einstein's general theory of relativity establish a relationship between the density of matter and the geometry of space. Under the assumption of a homogeneous and isotropic distribution of matter, depending on the value of the (average) mass density ρ, both an open cosmological model extending to infinity with a negative curvature and a

continued on next page

Figure 5.185 *continued*

closed but unbounded oscillatory model are possible. The value $\rho_{critical}$ separating the two solutions and corresponding to zero curvature (Euclidean space) depends, according to the formula printed in color, on the Hubble constant H, describing the expansion rate of the universe, and the gravitational constant G. In the diagram, $R(t)$ denotes the distance between two cosmic objects at time t, and $1/R^2$ is proportional to the curvature.

Not only is the expanding universe given by Einstein's equation in accord with Hubble's experimental results, together they also provide an answer to such old questions as Olbers' paradox. Wilhelm Olbers (1758–1840) asked, why is the sky dark at night? In a homogeneous star-filled infinite cosmos, the entire sky should be uniformly brightly lit, for along every line of view there should be a star.

The relationship between the Hubble constant H and the critical density $\rho_{critical}$ can be derived as follows: We consider a sphere of radius R that contains a large number of galaxies. Let ρ denote the average density of matter. The total mass of this sphere is therefore $M = (4\pi R^3/3)\rho$.

A galaxy of mass m on the surface of this sphere possesses potential energy $-GMm/R$ and kinetic energy $\frac{1}{2}mv^2$. Since its velocity, expressed in terms of the Hubble constant, is $v = RH$, its total energy will be

$$E = -\frac{GmM}{R} + \frac{1}{2}mH^2R^2 = mR^2\left(\frac{1}{2}H^2 - \frac{4\pi}{3}\rho G\right).$$

If the total energy E is negative, then this galaxy cannot escape to infinity because there its potential energy would be zero and its kinetic energy cannot in principle be negative. However, if the total energy is positive, then the movement of the galaxy can continue without bound. The critical mass density $\rho_{critical}$ that separates these two possibilities comes from the equation

$$\frac{1}{2}H^2 - \frac{4\pi}{3}\rho_{critical}G = 0,$$

yielding

$$\rho_{critical} = \frac{3H^2}{8\pi G}.$$

The relativistic treatment of the relationship begins with Einstein's equation of general relativity above to arrive at the Friedmann equation

$$H^2 = \left(\frac{1}{R}\frac{dR}{dt}\right)^2 = \frac{8\pi G}{3}\rho + \frac{k}{R^2} + \frac{\lambda}{3}.$$

The first term on the right-hand side is the square of the Hubble constant used above, which in this context is usually denoted by H_0; λ is the cosmological constant. The value k is the curvature constant. Its value can be positive (closed universe), negative (open universe), or zero (flat universe). We see that the simple case $H = H_0$ occurs under the assumption $k = 0$, $\lambda = 0$. The quantities appearing here allow us to pose some basic questions about the structure and future—the development over time—of the universe:

1. What is the density of matter? (determination of ρ)
2. Is the universe curved or flat? (determination of k)
3. Is the expansion of the universe accelerating or decelerating? (determination of λ)

continued on next page

ergy 20 TeV, circumference 83 km, cost of $8 billion; Congress failed to appropriate the necessary funds, even though in the opinion of leading physicists, this installation would have provided a new, deeper view of the fundamental building blocks of the universe.

In view of the enormous sums, this attitude seems just, sober, and objective. Indeed, in this concrete case, the decision made by Congress brought with it—even if only indirectly—a fruitful result: The United States (and Japan) joined the European CERN project for the construction of the LHC (Large Hadron Collider). Thus, Congress assisted in the establishment of the first global research institute of this kind.

However, it is not only the usefulness of physics that was called into question: Physicists felt under even more pressure as doubt about the legitimacy and authenticity of physics was being enunciated by philosophers, sociologists, and anthropologists, thus threatening the conviction, previously considered secure by physicists, that the object of their research was to establish the objective laws of the real world.

Nobel Laureate Steven Weinberg, one of the creators of the Standard Model, very cleverly embodied the misdeeds of physicists in three metaphors: The physicist acts simultaneously as Frankenstein, Mandarin, and Adding Machine.

In the famous novel of Mary Shelley, which appeared in 1818, Frankenstein, with best intentions, creates a living being, a monster, who after committing several atrocities, kills his creator. The atrocities of the monster that the physicists created are these: nuclear war, accidents in nuclear power plants, radioactive waste, degradation of the environment, climate change, the ozone hole, genetic manipulation, extinction of plant and animal species, and so on. Physicists understand their own indirect responsibility to some extent, but in their view, the guilt falls more on technology. The engineers push the responsibility off in a different direction: They have to satisfy society's demands. People need more and more energy, and consumer goods are required for an increasing population with increasing demands (Plate XXIII).

Yet one should not forget the benefits wrought by technology. There are too many to list. Moreover, the harm caused by science can be redeemed only by science itself.

The Mandarin physicist embodies caste morality, the morality of closed social classes: It allows no new ideas, particularly when they come from young people, and efforts at reform are suppressed even more when they come from "outside," that is, from someone not a member of the club. This attitude harms both physics and society.

On the contrary, reply the physicists. They claim that they rejoice over every new idea, regardless of its origin. Almost 80% of the articles submitted to scientific journals are accepted. The reason for a rejection is usually a lack of originality; which is to say that they contain nothing new. Einstein was young and unknown, but his revolutionary new ideas were immediately accepted for publication in one of the leading journals. We might even claim that physics and physicists have never suppressed new ideas and also that no essential idea in physics ever came from an outsider.

The Adding Machine physicist is rebuked for offering mankind a cold, soulless, inhuman world, even if it is an objective one. The physicist counters that there is a certain compensation in the beauty and harmony of the laws of nature.

As we have mentioned, physicists have suffered from arguments about the legitimacy of physics. The entire methodology of physics was created, and repeatedly refined over the course of its development, to ensure the objectivity and universality of our knowledge, to be able to separate truth, that is, actual facts, from myth, superstition, and pseudoscience. It is this established goal that has been called into question.

The attacks against all sorts of claims of legitimacy can be summed up in the word "relativism." Under relativism, concepts lose their truth value, and laws lose their sense of having absolute validity; one must specify the point of view—historical, sociological, biological, and perhaps even individual—before one can assert even a relative truth.

Relativism is a venerable way of thought. PLATO quotes PROTAGORAS thus: Man is the measure of all things. The entire history of physics is a battle to liberate science from such subjectivity. In the modern era, following the path of DESCARTES and NEWTON, KANT also created a metaphysical background: Researchers of the nineteenth century rightly believed that they were, in fact, investigating the objective laws of the universe. This belief (or superstition, as FEYERABEND would correct us) was particularly strong in Germany. GAUSS did not dare to publish his research on non-Euclidean geometry, perhaps because he felt that others would see in it a veiled attack on the core principles of Kantian philosophy.

The open confrontation against the metaphysical foundation of confidence in our knowledge began with philosophical positivism (AUGUST COMTE: *Cours de philosophie positive*, 1830–1842). ERNST MACH formulated this view in physics even more strongly (*The Analysis of Sensations*, 1886).

Positivism does more than prescribe what to a physicist is obvious: that the claims of his theories must be proved by experiment. Its demands extend further: In all its details, in every relation, a theory can contain only those concepts and quantities that are observable and, at least in principle, measurable. MACH had a great influence on the young EINSTEIN, although EINSTEIN later "slandered" him (Un bon mécanicien, mais un philosophe déplorable). It is not by chance that even the coordinate systems in EINSTEIN's working out of his ideas, in his thought experiments, are fully fitted out with clocks, measuring rods, and experimental physicists. EINSTEIN hoped thereby to be able to assimilate the Kantian metaphysical entities into the structure of positivism.

However, physicists soon observed that with positivism, they had taken on, in addition to its purifying effect, the unnecessary ballast of a restrictive obligation because the "superfluous" thought structures contain an enormous heuristic force.

We turn again to an observation of EINSTEIN's on KEPLER's work: Knowledge of the precise information available from experiment in no way suffices to establish Kepler's laws. In addition, the human mind must first construct the concept of an abstract entity, the ellipse. That the thought constructions of mathematics conform so readily with the facts of the real world is astonishing. EUGENE WIGNER expresses this surprise with the expression: the unreasonable effectiveness of mathematics (an aside: WEINBERG offers, somewhat derisively, a parallel slogan: the unreasonable ineffectiveness of philosophy).

Thus, positivism fails as a driving force. How could one, for example, deal with superstring theory, not a single element of which is accessible to measurement or observation?

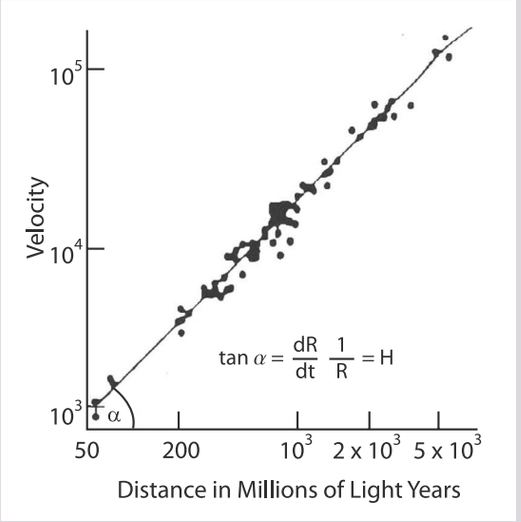
So it was that positivism developed along the line of MACH, POINCARÉ (conventionalism in mathematics, 1902), WHITEHEAD and RUSSELL (axiomatic system of symbolic logic, 1910–1913), the Vienna Circle (MORITZ SCHLICK, LUDWIG WITTGENSTEIN), to RUDOLF CARNAP's synthesis (*The Logical Structure of the World*, 1928; *Introduction to Symbolic Logic*, 1952), retreated into its internal problems, divorced from physics, not helping in its development, but not hindering it either.

In the second half of the twentieth century, attacks increased against the objective reliability of the assertions of physics, under the aegis of relativism. Efforts in

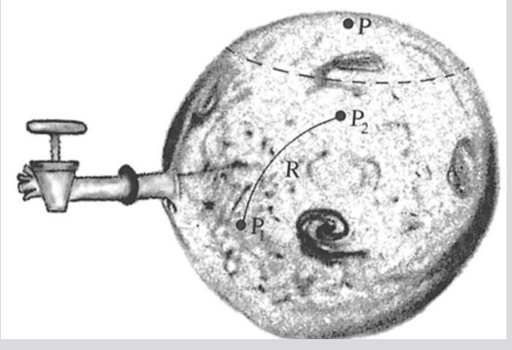
Figure 5.185 *continued*

Based on the critical evaluation of new observational material, the following answers can be considered reasonable at present [Perlmutter et al. 1999]:

1. The density of matter, taking into account dark matter, amounts to about one-third of the critical value ($\rho = (1/3)\rho_{critical}$).
2. The speed of the expansion of the universe increases over time ($\lambda \neq 0$), due to an unknown dark energy.
3. The universe is flat ($k = 0$).



▲ **Figure 5.186** The Hubble constant is equal to the slope of the line in the diagram (after [Barrow 1994, p. 3]). Different evaluations of the measurement data can, of course, lead to different lines and to different Hubble constants. The newest, more precise methods have, in fact, established an acceleration of the expansion.



▲ **Figure 5.187** We would like to draw attention here, emphatically, to the danger of relying on intuition when dealing with questions of cosmology. The common-sense picture of the big bang is that a small sphere of extremely dense matter in some specific point of otherwise empty three-dimensional space suddenly begins to expand while the border between the portion of space filled with matter and the external empty space moves ever farther away from the center of the universe. This is not the case at all. If we model our three-dimensional curved space with the

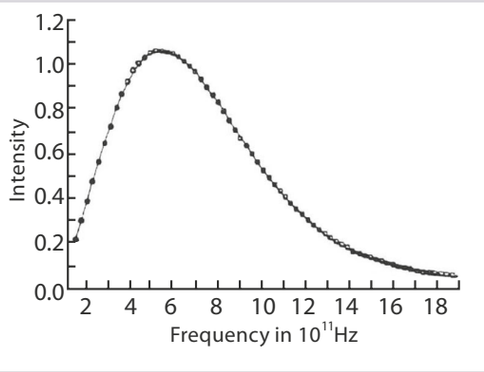
continued on next page

Figure 5.187 *continued*

two-dimensional surface of a sphere, we can think of the expanding universe as the surface of a balloon that is being inflated: Space itself is what is expanding; there is thus no center, or if you will, any point can be chosen as the center. There is no boundary surface, yet the volume is finite.

It is easy to see that the distance between two points on the sphere's surface (that is, in the two-dimensional analogy to our universe) grows at the same rate as R itself; that is, the change with respect to the units of length and time remains constant: $(1/R)(dR/df)$ = constant, and this constant, in fact, represents the Hubble constant.

The dashed circle in the figure is called the event horizon. In the time interval from the big bang until now, light emanating from point P has been able to reach this circle (a hypersurface in the four-dimensional curved universe). It is known that no effect can propagate faster than light, so this circle—the event or effect horizon—represents an impassible boundary for all interactions that originate at P. (See also Figure 5.196.)



▲ **Figure 5.188** The residual radiation from the big bang, an equilibrium radiation corresponding to a temperature of 2.7 K.

This curve, measured in 1990 by the COBE (*Cosmic Background Explorer*) satellite, represents an extremely precise experimental verification of Planck's radiation law. The background radiation is isotropic; that is, it has the same value in every direction to a precision of 1 part in 10^4. Since 1992, however, small deviations in the isotropy of size 1 part in 10^5 have been detected. These can be attributed to two different sources, both of which are exceptionally important. First, our Earth moves with a velocity of 370 km/s relative to the space in which the background radiation is isotropic. This motion is composed of the motion of Earth around the Sun (31 km/s), the motion of the Sun around the galactic center (220 km/s), and the motion of the galaxy in the direction of the Virgo Cluster (600 km/s). The resulting Doppler effect from these combined values leads to a correction of the values of anisotropy. Second, the remaining, irregular, variations come from the time before the formation of galaxies. They bring us information about the deviation from homogeneity of matter when the universe was 3×10^5 years old and the background radiation with a temperature of 6000 K arose. Although these deviations are extremely small (1 part in 10^6), they contain important information on the relationships to earlier times—information about the fluctuation in the energy density in the period of inflation. Combined with the results obtained with TeV accelerators, the boundary of reliable physics can perhaps be pushed back into what was previously only a speculatively accessible past [Hogan 1998].

social philosophy and anthropology were directed toward an attempt to prove not that these assertions are false, but that they represent only one of many possible ways of looking at things, one among many points of view.

"Scientific truth" is ultimately a broadly accepted agreement about what can be considered as "reality." The "scientific" debates that take place as to what is to be accepted as proof or as an appropriate experiment, are no better and no less muddled than those in a political or legal debate. This point of view is often called the strong program of the sociology of science.

Of course, physicists do not deny that the discovery of a physical law or theory depends strongly on historical and social contingencies; however, they are convinced that the content of the assertions themselves are testable objectively and experimentally, predictive, and universal.

Paul Feyerabend (one of his characteristic statements appears as Quotation 0.26) gave his first book (*Against Method*, 1975) the subtitle *Outline of an Anarchistic Theory of Knowledge*. He writes [p. 296]:

> Thus science is much closer to myth than a scientific philosophy is prepared to admit. It is one of the many forms of thought that have been developed by man, and not necessarily the best. It is conspicuous, noisy, and impudent, but it is inherently superior only for those who have already decided in favor of a certain ideology, or who have accepted it without ever having examined its advantages and its limits. And as the accepting and rejecting of ideologies should be left to the individual it follows that the separation of state and *church* must be complemented by the separation of state and *science*, that most recent, most aggressive, and most dogmatic religious institution. Such a separation may be our only chance to achieve a humanity we are capable of, but have never fully realized.

There are researchers who go even further: The physicist does not discover objects of nature and the phenomena of nature, in the way that a seafarer discovers an island, for the island exists independently of whether it is discovered or not. Such objects are formed and constructed through the interactions of the members of the research group, the scientific community. Andrew Pickering expresses this conception in the title of his book, *Constructing Quarks: A Sociological History of Particle Physics* (1984).

Another group recognizes, indeed admires, the accomplishments of physics, its boldness in following the path of the search for truth, not caring where the path leads. It is frightened, on the one hand, because of the dangers that threaten and, on the other, because of the rational, cold world that physics offers us: inhuman, heedless, in which there is no place for the soul, and offering no answer to the essential human questions, indifferent to love, ethics, and beauty.

Physicists can counter that they are, in fact, incapable of answering this question, and just for this reason strive not to present their findings as a coherent worldview. This ethos aims at as true as possible an account of the objects and phenomena of nature; beauty is to be found in the harmony of nature's laws.

And when physicists also intentionally strive to answer ideological questions, then they find themselves nevertheless unintentionally confronted with questions that are also taken up by myths and religions, when, for example, the problems of the big bang are discussed. If our Sun becomes a red giant, human existence and our Earth, with all its beauty and horror, will be swallowed up and turn into ionized gas. Who can deny that this vision, which Russell deals with in his book *Mysticism and Logic* (Quotation 5.63), is not the reflection of an entire worldview?

The soul seeks a way out of this soulless, inhuman world and asks, "Are we certain that the knowledge that science offers us is true knowledge? Is there perhaps a deeper stratum of knowledge that encompasses all that is important for humanity?" (Theodore Roszak). These questions necessarily awaken an association. The problem and its solution come from Plato. "Phenomena must be rescued." That is the task of science, and, in fact, science satisfies this demand. The truth, the shining truth, lies, however, deeply hidden in the world of ideas: The analogy of this idea or conception must be created for modern mankind. Roszak calls this deeper layer of knowledge *gnosis*. He symbolizes it in the form of a spectrum: At one side is science; in the middle, art; and on the other side, religion.

The unconvinced reader, but the physicist also, may ask in this situation whether there might still exist a reliable method for deciding which system of conjectures that can be combined in a unified way and that can be applied to the real world can be seen as an adequate and reasonable description of the world around us. Some of the quotes presented here despair of the existence of such a method. To a certain extent, Thomas Kuhn shares this opinion when he maintains that, in essence, a paradigm shift cannot be viewed as a rational process: One cannot give oneself over to vain hopes that the acceptance of new paradigms always leads closer to the truth. In contrast, Karl Popper asserts that there exists a reliable criterion, that of falsification: The assertions of true science are, at least in principle, falsifiable (Section 0.3).

Physicists may take comfort in the knowledge that the tried and true method of "experiment → theory → experiment" leads to an approximation of the laws of nature. They may recall, for example, the experimental determination of the mass of the vector bosons $W^{\pm}$, Z^0, and ask how such predictions, which were so precisely verified, could have arisen from "muddled debates" between the adherents of an "aggressive type of religious institution." They agree with Weinberg in his point of view and concern:

> These radical critics of science seem to be having little or no effect on the scientists themselves. I do not know of any working scientist who takes them seriously. The danger they present to science comes from their possible influence on those who have not shared in the work of science but on whom we depend, especially on those in charge of funding science and on new generations of potential scientists.
>
> —Steven Weinberg, *Dreams of a Final Theory* [p. 190]

5.7.2 *The Standard Model and Beyond*

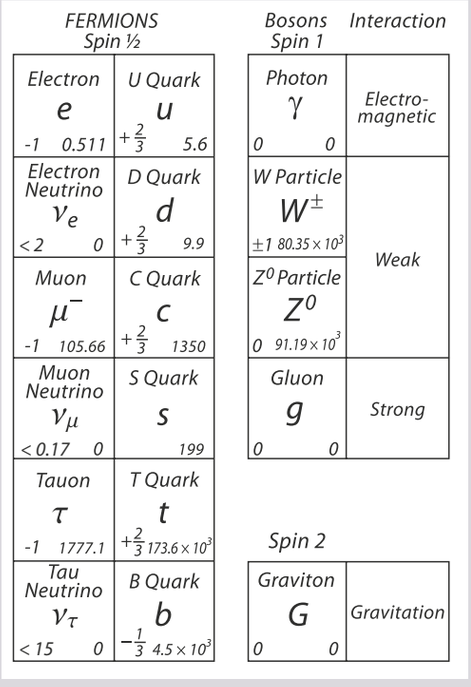
We can summarize the characteristic features of the situation at the turn of the millennium in the theory of elementary particles as follows:

1. There is a theory, called the *Standard Model*, which exhibits great inner coherence and consistency. There is no experimental fact that challenges it, and it has proved its "prophetic" power; yet it leaves a number of questions unanswered.
2. It appears that the path that, with the help of symmetry considerations, has led from quantum electrodynamics to the electroweak theory and quantum chromodynamics can be followed further to the establishment of a grand unified theory, which would be an all-encompassing theory of everything:

Quotation 5.63

That Man is the product of causes which had no prevision of the end they were achieving; that his origin, his growth, his hopes and fears, his loves and his beliefs, are but the outcome of accidental collocations of atoms; that no fire, no heroism, no intensity of thought and feeling, can preserve an individual life beyond the grave; that all the labours of the ages, all the devotion, all the inspiration, all the noonday brightness of human genius, are destined to extinction in the vast death of the solar system, and that the whole temple of Man's achievement must inevitably be buried beneath the debris of a universe in ruins – all these things, if not quite beyond dispute, are yet so nearly certain, that no philosophy which rejects them can hope to stand. Only within the scaffolding of these truths, only on the firm foundation of unyielding despair, can the soul's habitation henceforth be safely built.

—Bertrand Russell, *Mysticism and Logic* [p. 51]



▲ **Figure 5.189** The Standard Model. At the end of the second millennium, the following particles were viewed as fundamental: The particles with half-integer spin, the fermions, were seen as literally the basic building blocks of matter; the bosons, particles with integer spin, are the mediators of the interactions. There are seventeen particles listed here. For the sake of completeness, it must be noted that every type (flavor) of quark appears in three colors. More precisely, each type has a color charge, and furthermore, there are eight distinct gluons. In the hadrons (protons, neutrons, mesons), the quarks are held together by the color forces—mediated by the gluons—between the various colors.

continued on next page

Figure 1.89 *continued*

The Higgs boson, the quantum of the Higgs field, is not listed in the table although it plays a fundamental role in the theory. It has not yet been observed in experiments.

If one adds in the antiparticles, the number of elementary particles again climbs to more than 50. However, the structure of ordinary matter is still simple. It consists of u and d quarks (these are the building blocks of protons and neutrons)—held together by gluons—which, with the addition of electrons, form atoms with nuclei consisting of protons and neutrons, held together by electromagnetic forces mediated by the photons.

In the figure, we find, for each particle, the name, the symbol, the electric charge (in relation to the elementary charge, that is, the charge on the electron, in eV), and the mass (in MeV).

The electron, the three neutrinos, the photon, the gluon, and the graviton are stable, whereas the muon and tauon have half-lives of

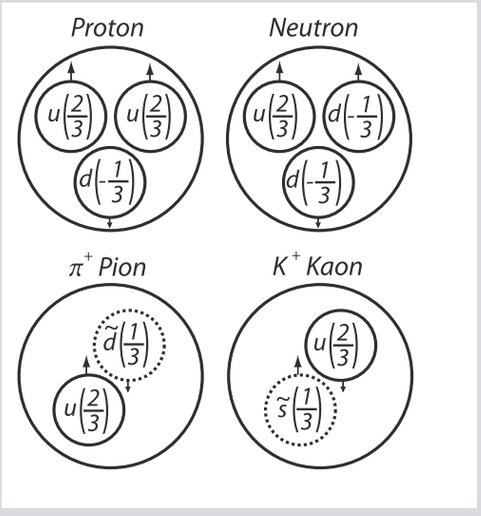
$$\mu^- \rightarrow 2.14 \times 10^{-6} \text{ s} \quad \text{and} \quad \tau^- \rightarrow 0.29 \times 10^{-6} \text{ s}.$$

For the vector bosons W±¬, Z0, one usually gives a quantity that is inversely proportional to the half-life, the energy spread:

$$W^{\pm} \rightarrow 2.08 \text{ MeV} \quad \text{and} \quad Z^0 \rightarrow 2.49 \text{ MeV}.$$

Due to "color confinement," quarks are never observed in a free state, and so one cannot speak of a half-life for them.

The notation for quark flavors comes from the initials of the various names assigned to them, usually by their discoverers, namely, *u*: up; *d*: down; *c*: charmed; *s*: strange; *t*: top or truth; and *b*: bottom or beauty.



▲ **Figure 5.190** The quark structure of some hadrons: the proton and the neutron (the nucleons), the π^+ pion and the K^+ kaon (two mesons). To denote the color charge, the colors red, green, and blue have traditionally been used. The dashed lines indicate anticolor charges.

Once again the vision arises—desirable for some, alarming for others—of an end of basic research in physics.

3. The experimental verification of the assertions of the future theory may be limited by the need for energies so high that their realization in a terrestrial laboratory would lie beyond today's technical and financial capabilities. Consequently, the interplay of theory and experiment breaks down, with the old and well-tried schema for research, "theory → experiment → theory → experiment" and so on, becoming impossible.
4. The problem of the macrocosm, the origin and development of the universe, and the problem of particle physics become essentially identical and can be solved only in tandem. The incorporation of gravitation into a common unified interaction may be the key problem.
5. The big bang, the origin of the universe, raises the problem, more philosophical than physical, of the why and wherefrom of the physical laws.
6. The search for indirect experimental proof demands an ever greater allocation of technological and financial resources, which can strain the budget of even a wealthy country. Thus, a paradoxical situation arises: The more distant and tangential a direction of research lies from the problems of everyday reality, of interest only to theoretical physicists ensconced in their ivory towers, the greater the necessity of relying on the harsh everyday reality of the costs of research.
7. Because of its economic relevance, there is the possibility, perhaps even the necessity, for society, the "taxpaying citizen," to intervene in the development of physics.

In what follows, we shall enlarge upon these remarks, primarily based on the opinions of leading physicists.

In the Standard Model, the weak and electromagnetic interactions are combined into a single interaction, the electroweak interaction (Glashow, Salam, Weinberg 1967/1968). The strong interaction is described by quantum chromodynamics (QCD, Fritzsch–Gell-Mann–Leutwyler, 1973). The standard theory can say nothing new about the gravitational interaction. The most spectacular result of the standard theory is the prediction and subsequent experimental verification of the masses of the mediator particles $W^{\pm}$, Z^0 (Rubbia et al., 1983). The most important assertions of this theory are collected in Figure 5.189.

The fundamental particles can be divided into two large groups. The particles with half-integer spin—the fermions—can be considered the actual building blocks of matter, while those with integer spin—the bosons—are the mediators of their interactions.

The fermions can, in turn, be divided into three well-defined families. Each family consists of two heavy particles—the quarks—and two light particles—the leptons, each consisting of a charged electron-like particle and an associated neutrino.

It is a fundamental fact that the existence of the strong interaction is what makes it possible for the quarks to be "colored," or more precisely, to be equipped with color charge. This color charge plays a role in the strong interaction similar to that played by electrical charge in the electromagnetic interaction. In the electromagnetic interaction, the charges are the sources of the mediators, the photons, whereas in the strong interaction it is the color charges that are the source of the mediators, the gluons. These color forces hold the quarks together in the hadrons (protons, neutrons, mesons), Figure 5.190. The hadrons themselves are "colorless,"

that is, their color charge is zero. In the protons and neutrons, the superposition of the three colors leads to this neutrality; in the mesons, in contrast, it is the combination color–anticolor.

But just as the electrically neutral molecules can be joined together through the electrical residual-force effect, through the van der Waals forces, so the residual-force effect of the color forces is what binds the color-neutral protons and neutrons, what we call nuclear force or residual strong force and what can be shown as equivalent to the earlier description that used π mesons.

The color forces holding the quarks together approach zero as the distance between them approaches zero: there they move almost freely (asymptotic freedom,) 'T HOOFT, 1972; GROSS, WILCZEK, POLITZER, 1973). So, they remain together as in a sack, so to speak, and always in a color-neutral structure. No quark has yet been observed in the free state. If we try to split off a quark by conveying sufficiently high energy to the sack, part of the invested energy will turn into a quark–antiquark pair that emerges as a meson.

One idea that had great heuristic bearing on the formulation of the Standard Model is the fundamental quark–lepton symmetry. At the same time, the theory also demonstrates the striking correctness of GLASHOW's statement that the "act of becoming" consists in the emission or absorption of a particle (Figure 5.191(a,b,c)). All known phenomena can now be newly interpreted: Neutron decay, for example, consists at the quark level in the transformation of a *d* quark into a *u* quark.

We still retain a relatively simple and clear picture of our everyday world (Figure 5.192).

Those are the achievements of the theory. Now for the shortcomings. Here we refer again to Figure 5.1, where the unsolvable problems of the turn of the twentieth century are presented. There we find, in addition to question marks, evidence of contradictions: For some phenomena, classical physics had no answers, and sometimes it offered categorical, but incorrect, answers. In contrast, no contradictions have been found in the Standard Model even though it is bristling with question marks. But it has also made an as-yet-unverified claim: The theory assumes the existence of Higgs bosons, which are the mediators of the Higgs field. They play an important role in the theory, being responsible for the size of the masses of the vector bosons $W^{\pm}$, Z^0, but their existence has not yet been demonstrated.

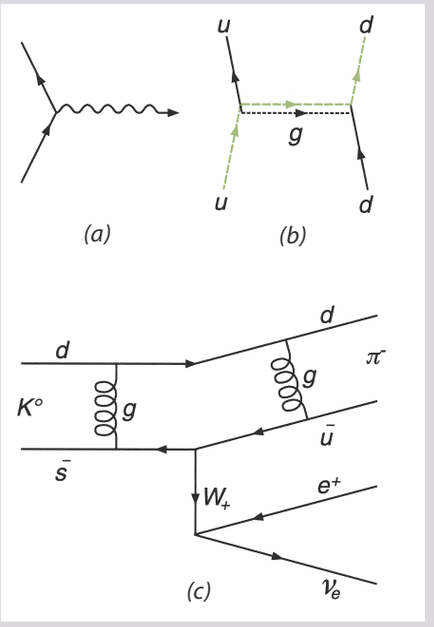
Inner coherence, consistency, prophetic power: All these positive attributes cannot let us forget that the theory fails to answer a host of vital questions.

GLASHOW himself lists 12 such questions, including these: Why are there three fermion families? Why are they so similar? Why are the charges of the proton and the electron identical? Do photons and neutrinos have finite rest mass? Is there a magnetic monopole? What is the life span of the proton? Is it infinite? That is, can it be considered absolutely stable?

Since 1974, GLASHOW and GEORGI have proposed answers to many of these questions: They created the grand unification theory (GUT). They followed the path indicated in point 2 above, that is, the path through symmetry considerations.

5.7.3 *Groups and Symmetries*

The path-breaking success of quantum electrodynamics (QED) in the 1940s was achieved by its ability to remove the appearance of the (indeed, more than one) infinite values that arise in the computation of the electromagnetic interaction energy. This was made possible by a renormalization. This renormalizability follows from the fact that the electromagnetic field is gauge invariant, that is, it possesses gauge symmetry. It can even be maintained—with some exaggeration—that just as all the laws of (geometric) optics can



◀ **Figure 5.191 (a)** According to GLASHOW, the "fundamental act of becoming" consists of the emission or absorption of a particle by another particle.

(b) The role of the gluon in mediating the strong interaction is only superficially similar to how photons mediate the electromagnetic force, because the gluons themselves also have color charges; they can therefore strongly interact with one another. This considerably complicates the gluon's behavior: as a result, for example, the force between quarks within hadrons will diminish, giving the quarks their asymptotic freedom. Each of the eight types of gluon possesses two colors: one of the three colors and an anticolor. (Of the nine possible combinations, the theory allows the realization of only eight of them.)

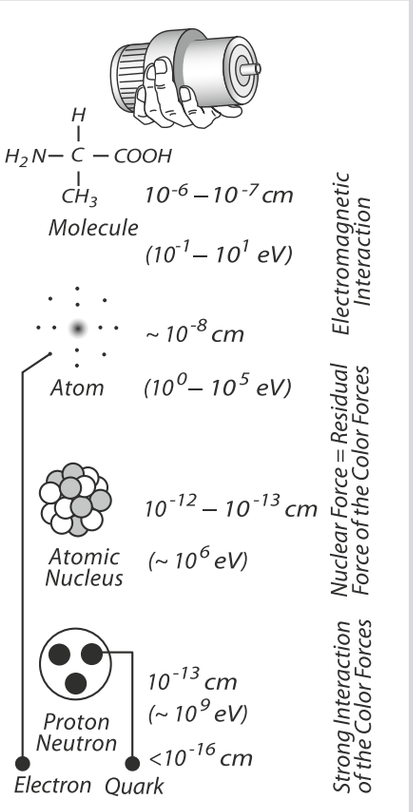
The diagram shows a red *u* quark that emits a red antigreen gluon, transforming itself into a green *u* quark; its interacting partner, a green *d* quark, turns into a red *d* quark.

(c) A real process consists of several successive fundamental events. The decay of a kaon pictured here ($K_0 \rightarrow \pi^- + e^+ + \nu_e$) is the result of six such events.

▼ **Figure 5.192** The structure of the macro world that surrounds us and that we can sense directly—be it inanimate or living, natural or man-made—is relatively simple. As a first step, chemistry divides matter into molecules as constituents. The molecules consist of atoms, the number of which may reach thousands, or even hundreds of thousands, so the dimensions of molecules may be almost macroscopic. Everything considered, the atoms are joined into molecules through the electromagnetic interaction mediated by photons. This same interaction appears in its clearest form in the binding of electrons with positively charged nuclei to create neutral atoms. The stability of the nuclei is ensured by the nuclear forces—the "residuals" of the color forces—acting between the nucleons. Finally, the nucleons are assembled from *u* and *d* quarks by the color forces mediated by gluons.

This means that our own bodies fundamentally consist of *u* and *d* quarks and electrons. In the human body we will find on average 7×10^{28} *u* quarks, 6.5×10^{28} *d* quarks, and 2.5×10^{28} electrons.

Below the sizes of the components, the figure also indicates the energy required to change the state of the given system (in parentheses). The image for a "macroscopic body" here is a human hand holding an electric motor with the same power rating (70 watts) as a human body.



be derived from the principle of shortest time, all of electrodynamics—the existence of the photon with all of its properties (spin 1, rest mass 0)—follows from gauge symmetry. This thrilling idea, so successful in its beauty, represents a great challenge to researchers.

Gauge symmetry is a special case of general symmetry groups. A physical theory possesses gauge symmetry if it is invariant with respect to a gauge transformation, that is, if it possesses gauge invariance. A gauge transformation is an alteration of the field quantities used in the theory by adding in suitable functions of time and location or by multiplication by a phase factor; if no relevant physical quantities (obtained by measurement) are thereby changed, the theory possesses gauge symmetry. The simplest gauge symmetry or gauge invariance is considered to be this one: The effect of an electrical force remains unchanged if we add to the potential an arbitrary (constant) quantity. Or, somewhat more generally, the Maxwell equations remain invariant with respect to a gauge transformation of the charge-current potential.

To describe gauge symmetry, one has at one's disposal the mathematical field of group theory, and matrix notation is well suited for the mathematical treatment of the subject.

Arthur Cayley (1821–1895) introduced matrices in 1858 as independent mathematical objects as a formal way of simplifying the treatment of systems of linear equations.

In what follows, we shall be dealing exclusively with square matrices. We also assume that the determinant of a (square) matrix in nonzero. A matrix with 1's along the diagonal and zero elsewhere is called an identity matrix. These are the most important rules of matrix algebra:

In general, the commutative law does not hold: $\mathbf{BA} \neq \mathbf{AB}$.
The associative law always holds: $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} \; (\equiv \mathbf{ABC})$.
For the identity matrix $\mathbf{1}$, we have $\mathbf{1A} = \mathbf{A1} = \mathbf{A}$.
For a nondegenerate matrix (det $\mathbf{A} \neq \mathbf{0}$), there is always an inverse matrix, denoted by $\mathbf{A}^{-1}$, such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{1}$.

Matrix algebra became of interest to physicists with Heisenberg's formulation of quantum mechanics, that is, with matrix mechanics. The group theory that we are about to sketch has become in recent decades an everyday tool of particle physicists.

The fundamentals of group theory were first written down in essence by Évariste Galois (1811–1832) at the age of 21 on the eve of his death in a duel. He investigated the solvability in radicals of degree-*n* polynomial equations in one unknown with constant coefficients. With the introduction of the notion of group, he was able to solve this ancient problem.

A group is a set of elements with the following properties:

1. It possesses a binary operation (multiplication) such that the product of each pair of elements in the group is again an element of the group.
2. The associative property holds for this operation: $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.
3. There is a distinguished element **E**, the identity, with the property that $\mathbf{EA} = \mathbf{A}$ for every element **A** in the group.
4. For each element **A**, there is an inverse element $\mathbf{A}^{-1}$ in the group such that that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{E}$.

The commutative law is not required to hold. A group for which $\mathbf{BA} = \mathbf{AB}$ for every pair of elements is called an abelian group.

The introduction of the idea of a group makes possible the unified investigation of various sets with a wide variety of multiplication operations. A very simple (infinite) group is that of the integers under the usual addition. Since the multiplication of square matrices is again a square matrix of the same dimension, the nondegenerate square matrices of a given dimension form a group under matrix multiplication. The nondegeneracy condition is required for the existence of inverses. A "model" of a simple finite group is given by the four complex numbers 1, *i*, –1, –*i* with operation the usual complex multiplication.

The symmetry groups play an important role in particle physics. In the simplest and easiest-to-understand cases, the elements of this group are operations that map symmetric (spatial or planar) geometric objects into themselves. A simple example is seen in Figure 5.193. An equilateral triangle can be mapped to itself by six operations:

1. Rotation by 120°.
2. Rotation by 240°.
3. Reflection in the line through point I.
4. Reflection in the line through point II.
5. Reflection in the line through point III.
6. The identity element. The triangle remains unmoved and unchanged.

We denote these operations, that is, the elements of this (finite) symmetry group, by the letters A, B, C, D, F, E. Multiplication is defined by carrying out two of the operations in succession. It is easy to see that multiplication of two elements leads to another element of the group. Moreover, inverse elements exist: For each operation, one can find a group operation that reverses that operation. The most important tool in investigating the structure of a finite group is its multiplication table.

The special groups SU(n) that arise in particle physics are subgroups of the unitary groups U(n). The unitary groups are represented by matrices—unitary matrices—for which the rule $\tilde{\mathbf{A}}^{\mathrm{T}} = \mathbf{A}^{-1}$ holds, where $\tilde{\mathbf{A}}$ is obtained from the matrix $\mathbf{A}$ by replacing each element with its complex conjugate; $\mathbf{A}^{\mathrm{T}}$ is the matrix transpose of $\mathbf{A}$, that is, rows and columns are interchanged; $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$; and n is the dimension of the matrix. The special unitary matrices SU(n) (the letter S stands for "special," U for "unitary") have determinant +1.

The gauge symmetry of the electromagnetic quantum field is given by U(1) symmetry. This describes, in ordinary space, the simplest symmetry, namely rotational symmetry of the circle. As mentioned earlier, this symmetry enables a renormalization and thereby the almost unbelievable precision of the (theoretical) determination of the anomalous magnetic moment of the electron and the Lamb shift.

The symmetry SU(2) leads to the short-lived VA theory of the weak interaction (1957; the letters V and A stand for "vector" and "axial").

The product group U(1) × SU(2) enabled the development of the theory of the renormalizable electroweak interaction (Glashow–Salam–Weinberg, 1967).

The theory of the strong interaction, quantum chromodynamics (QCD), rests on gauge symmetry SU(3) (Fritsch–Gell-Mann, 1973).

The theory of the electroweak interaction and QCD together form the Standard Model, which in the 1980s was generally accepted and applied.

The Standard Model is based on the product SU(3) × SU(2) × U(1). Glashow and Georgi searched for a single group that could be used as the basis of a unified theory of the three interactions. They found the group SU(5) to be suitable for this purpose.

5.7.4 *The Grand Unification*

Thus there arose the grand unification theory (GUT). The theory provides answers to a number of previously unanswered questions: It gives the values of the charges of quarks and electrons, requires a rest mass of zero for the photon, and provides approaches to explain the inflation of the cosmos and asymmetry of matter–antimatter through CP violation. A more recent implication is that there are possibly transitions between the two fermion families and between the leptons and the hadrons. However, this means that the proton is unstable. The theory also gives the life span: 10^{30} years. This value is largely consistent with the values that follow from previous stability studies.

This time is 20 orders of magnitude longer than the age of the cosmos. But if one considers that one kilogram of matter contains approximately 10^{27} nucleons, then in a ton of matter, one nucleon per year should decay; in 1000 tons, one may expect three decays per day. Detection of such a decay clearly involves huge technical and financial difficulties; nonetheless, some attempts are presently underway, thus far without result.

To clarify the difficulties involved, consider the following specifications for an installation built in Japan for measuring the life span of the proton. In a zinc mine, a tank is filled with 45000 tons of highly purified water. In the water are placed 11000 electron multipliers, each with a light-sensitive cathode with a diameter of 50 centimeters that measures the Cherenkov radiation, caused by the rapidly moving electrically charged decay products (Figure 5.194).

Although no proton decay has yet been recorded, the efforts have not been in vain: neutrinos from the 1987 supernova were recorded.

A direct proof of the correctness of the theory would be the experimental observation of the mediator particles, the X bosons, in analogy to how the "*experimentum crucis*" of the theory of the electroweak interaction was the observation of the $W^{\pm}$ and Z^0 bosons. Since the mass of these X bosons would be of an order of magnitude of 10^{15} GeV, a direct proof is far beyond today's technological capabilities.

Regardless of the close to theoretical impossibility of a direct experimental verification, theoreticians continue to pursue this general direction. One of these theories, "SUSY" (supersymmetry, based on the symmetry group SU(8)), allows not only transitions between the two fermion families, as does the grand unification, but also blurs the boundaries between the fermions and bosons. It makes

Quotation 5.64

The requirements of quantum mechanics plus special relativity are so tight that historically constructing string theories was very difficult. The conditions that must be obeyed are highly overdetermined. A vast effort went into the construction of string theories, and by the time the dust cleared in 1984–85 it was found that there were five of them. They differ by very general properties of the strings:

- In two theories (the Type IIA and Type IIB theories, which differ by whether there is invariance under reversal of the orientation of spacetime) the strings are closed and oriented and are electrical insulators.
- In two theories (the heterotic superstrings with gauge groups SO(32) and E8 E8) the strings are closed, oriented, and superconducting.
- In the last case (Type I) the strings are unoriented and insulating and can have boundaries, in which case they carry electric charges on their boundaries. ... It is rather strange to be told that there is a rich new framework for physics which unifies quantum mechanics and gravity and that in this new framework there are five possible theories. If one of those theories describes our universe, who lives in the other four worlds? ...

The five string theories—and a wild card, eleven-dimensional supergravity, that has proved to be important in getting a systematic understanding—are now understand as different limiting cases of a more comprehensive (and little understood) theory known as M-theory. ... *M* stands for magic, mystery, or matrix, according to taste. The magic and mystery are clear enough, while "matrix" refers to a new noncommutativity, roughly analogous to $[p; x] = -ih$ but very different, that seems to enter the theory. Physicists and mathematicians are likely to spend much of the next century trying to come to grips with this theory.

—Edward Witten, "Magic, Mystery, and Matrix" [pp. 1127–1129]

I suppose that there are three basic reasons why string theory has attracted so much interest in the past 20 years. One is that it is there. String theory is the only known generalization of relativistic quantum field theory that makes sense. The framework of special relativity plus quantum mechanics is so rigid that it practically forces quantum field theory upon us. The tightness of the modern framework is one of the main reasons why physicists were able to discover what has become the standard model of elementary particles. A big idea like a consistent generalization of quantum field theory comes along only every now and then. So we are duty-bound to take it seriously. A second reason has to do with what physicists

continued on next page

Quotation 5.64, ***continued***

have learned in developing string theory. String theory forces general relativity upon us, whereas standard quantum field theory apparently makes it impossible to incorporate general relativity. And string theory leads in a remarkably simple way to a reasonable rough draft of particle physics unified with gravity. And finally, string theory has proved to be remarkably rich, more so than even the enthusiasts tend to realize. It has led to penetrating insights on topics from quark confinement to quantum mechanics of black holes, to numerous problems in pure geometry. All this suggests that string theory is on the right track; otherwise, why would it generate so many unexpected ideas? And where critics have had good ideas, they have tended to be absorbed as part of string theory, whether it was blackhole entropy, the holographic principle of quantum gravity, noncommutative geometry, or twistor theory. But what is string theory? It may well be the only way to reconcile gravity and quantum mechanics, but what is the core idea behind it? Einstein understood the central concepts of general relativity years before he developed the detailed equations. By contrast, string theory has been discovered in bits and pieces—over a period that has stretched for nearly four decades—without anyone really understanding what is behind it. As a result, every bit that is unearthed comes as a surprise. We still don't know where all these ideas are coming from—or heading to. One day we may understand what string theory really is. But even if we do, and the theory is on the right track, will we be able to learn how it works in nature? I certainly hope so. Realistically, it all depends on many unknowns, including the nature of the answer, how clever we will be, and the clues we can get from experiment.

—Edward Witten, "Unraveling String Theory"

no distinction between the actual elementary particles and the mediator particles as they all belong to a single superfamily. The characteristic mass, and hence the characteristic energy, for this theory is the Planck mass, 10^{19} GeV.

The seeds of today's quite fashionable superstring theory (TOE, theory of everything) can be found in the investigations of Gabrielle Veneziano (at the beginning of the 1960s). The research of John Schwarz and Michael Green, building on Veneziano's work, was acknowledged only in the latter half of the 1980s. The theory shifts the center of interest to the topology of space: Our space has nine dimensions, but six of them cannot be perceived, since they are somehow "coiled up" in a tiny ball of dimensions on the order of the Planck length. The original and characteristic feature of the theory is the following: The particles have no point-like structures but rather correspond to a looped string through whose various forms of vibration in nine-dimensional space the particular nature of the particle—whether it appears as a photon or quark—is determined.

Edward Witten, the leading researcher in superstring theory, maintains that this theory is capable in its most recent form of uniting quantum field theory and the general theory of relativity (Quotation 5.64). Relying on Stephen Hawking's earlier work on black holes, Witten asserts that according to standard quantum theory, gravitation was impossible, while in superstring theory it is unavoidable.

To be sure, the Standard Model is a huge improvement over the mishmash of approximate symmetries, ill-formulated dynamical assumptions, and mere facts that my generation of physicists had to learn in graduate school. But the Standard Model is clearly not the final answer, and to go beyond it we shall have to grapple with all its failings.

All these problems with the Standard Model touch in one way or another on a phenomenon known as *spontaneous symmetry breaking*. The discovery of this phenomenon has been one of the great liberating developments of twentieth-century science. ...

[T]he symmetry that relates the electron and neutrino and so on is a property of the underlying equations of the Standard Model, equations that dictate the properties of the elementary particles, but this symmetry is not satisfied by the *solutions* of these equations—the properties of the particles themselves. ...

In such cases we say that the symmetry is broken, although a better term would be "hidden," because the symmetry is still there in the equations, and these equations govern the properties of the particles. We call this phenomenon a *spontaneous symmetry breaking* because nothing breaks the symmetry in the equations of the theory; the symmetry breaking appears spontaneously in the various solutions of these equations.

It is principles of symmetry that give our theories much of their beauty. That is why it was so exciting when elementary particle physicists started to think about spontaneous symmetry breaking in the early 1960s. It suddenly came home to us that there is much more symmetry in the laws of nature than one would guess merely by looking at the properties of elementary particles. Broken symmetry is a very Platonic notion: the reality we observe in our laboratories is only an imperfect reflection of a deeper and more beautiful reality, the reality of the equations that display all the symmetries of the theory.

—Steven Weinberg, *Dreams of a Final Theory* [pp. 192, 193–195]

However, does this argumentation suffice to lead one to accept such a bold theory? Glashow rejected it categorically and sarcastically (Quotation 5.65). Feyn-

MAN observed that he knew well that the "wise old ones" frequently assume that new knowledge is nonsense and thus make themselves ridiculous in front of today's young people and posterity; nonetheless, he considers superstring theory to be nonsense.

It is, however, appropriate that in taking leave of this subject we quote the optimistic view of the founders of superstring theory, SCHWARZ and GREEN:

> There is no doubt that string theory is extraordinarily compelling. Theorists speak eloquently of the incredible beauty and richness of the theory. But clearly another incentive for studying the subject derives from the fact that if superstring theory does eventually provide a quantitative explanation of all the particles and forces of nature, it will represent one of the greatest scientific triumphs in the history of mankind. Indeed, one might claim that it would be the culmination of reductionist science, because we would at last have identified the smallest entities from which the world is built and elucidated the fundamental principles on which the universe runs. No wonder that some workers abandoned longstanding research projects overnight to take up string theory. At the time of writing there is a veritable industry at work on the subject. In the fields of particle physics and gravitation theory scarcely a seminar is given or a journal published that does not refer in some way to strings. ...
>
> Let's suppose that we succeed in understanding the fundamental principle, that we can find a unique solution to the equations. Given that solution we can study the topological properties of that six-dimensional space. From that we can learn the types of particles that would exist at low energy, the ratios of their masses can be worked out just from topological considerations, as well as the strengths with which they interact with each other.
>
> This is the kind of information that one deduces from experiments that are carried out in the lab.
>
> —JOHN SCHWARZ, interviewed in P. C. W. Davies and Julian Brown [1988, pp. 69, 84]

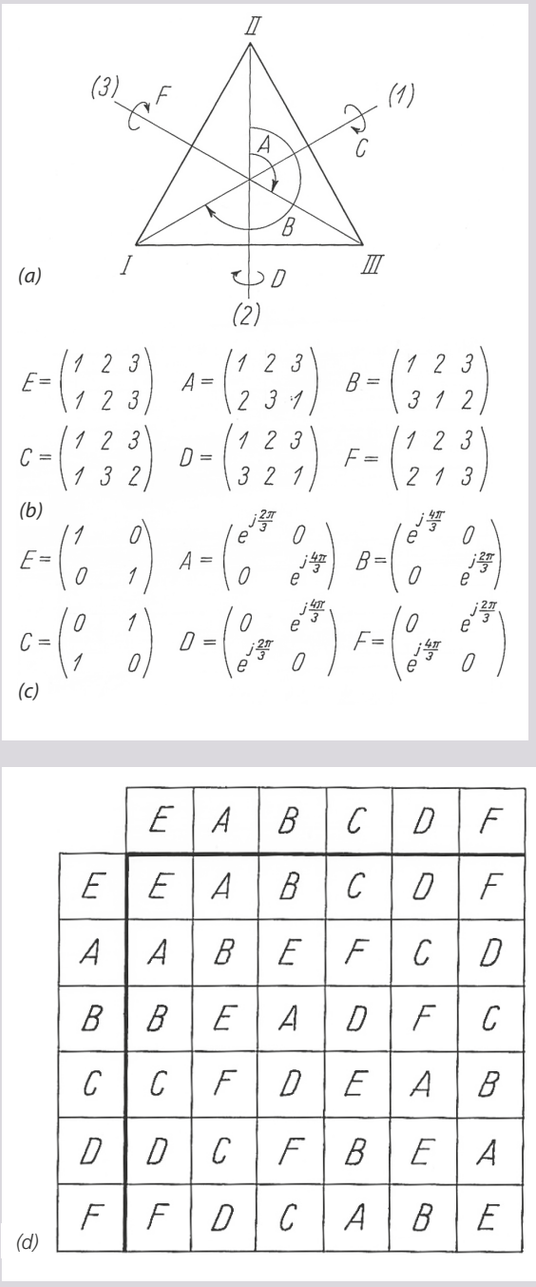
5.7.5 *The Great Laboratory*

Since the conditions that would make a direct proof of the assertions of string theory or superstring theory realizable in a terrestrial laboratory do not exist, the question arises whether such conditions exist anywhere in the universe. Indeed, they were realized in the most distant past, in the PLANCK era, the time interval from 0 to 10^{-43} seconds after the big bang.

Here we see the full force of the fact that cosmology and particle physics can investigate, and perhaps solve, their problems only if they work hand in hand: Without the results of particle physics, the big bang cannot be explained, and without studying cosmological laws, the basic assumptions of particle physics cannot be verified.

Figures 5.195 and 5.196 show the development of the universe from the big bang to the present day. We have also specified the "reach," that is, the limits of validity, of the various theories.

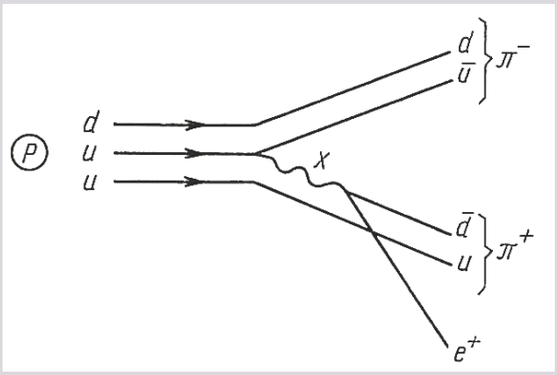
To understand and explain the drama of the lifetime of the universe, which is more exciting than any thriller, one needs all the discoveries of twentieth-century physics, including the entire mathematical machinery. Moreover, as a single, irreproducible object, the universe poses specific new philosophical problems.



(d)

	E	A	B	C	D	F
E	E	A	B	C	D	F
A	A	B	E	F	C	D
B	B	E	A	D	F	C
C	C	F	D	E	A	B
D	D	C	F	B	E	A
F	F	D	C	A	B	E

▲ **Figure 5.193** The three groups **(a)**, **(b)**, **(c)**, consisting of six elements each, are constructed in such a way and the notation so chosen that the multiplication tables of the three groups are identical **(d)**, which is easy to show from the respective operations. Such groups are said to be isomorphic. Thus the study of one of them gives information about the structure of the others. The matrix group **(c)**, which is more amenable to mathematical manipulation, can serve as a representative of the other groups. The other two groups are **(a)** the symmetry operations of an equilateral triangle and **(b)** the permutations of three numbers (permutation group).



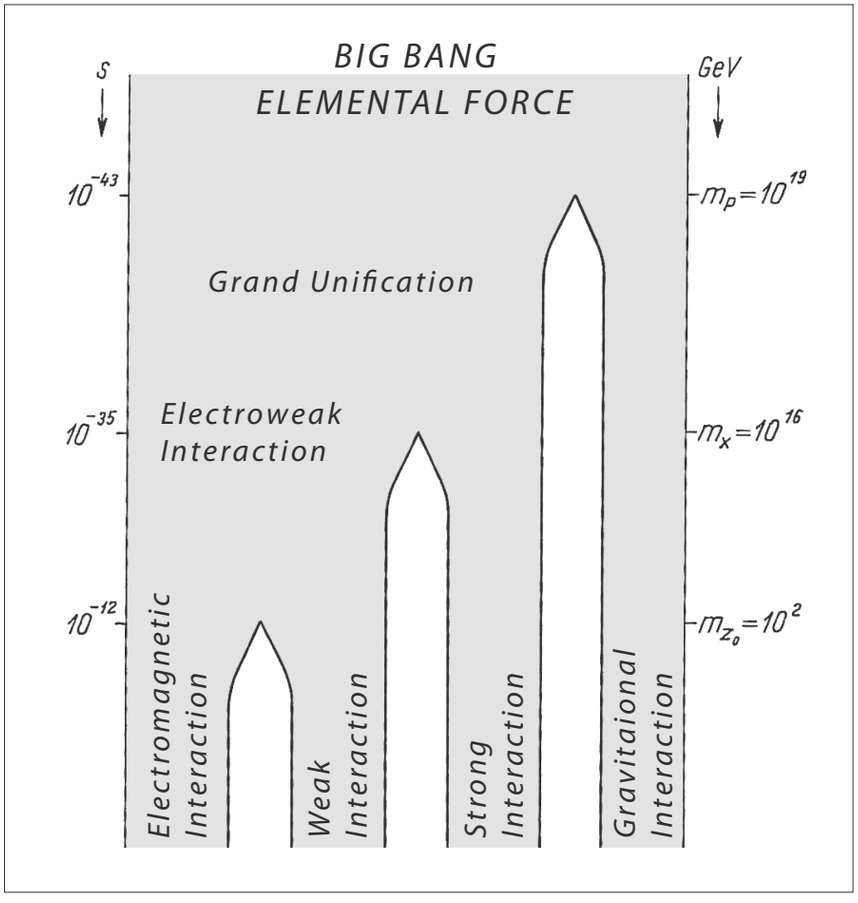
▲ **Figure 5.194** Visualization of the decay of a proton: $p \to \pi^+ + e^+$. The boson *X* mediates between the strong and electroweak interactions and in the process comes up against the conservation of baryon number and of lepton number.

Quotation 5.65

Meanwhile, the historical connection between experimental physics and theory has been lost, as far as superstring theory is concerned. Until the string people can explain and interpret perceived properties of the real world, they are simply not doing physics. Should they be paid by physics departments and be permitted to pervert impressionable students? Will young Ph.D.s whose expertise is limited to superstring theory be employable if and when the string snaps? String thoughts may be more appropriate to departments of mathematics or even to schools of divinity. How many angels can dance on the head of a pin? How many dimensions are there in a compactified manifold thirty powers of ten smaller than a pinhead? Has medieval theology been resurrected?

—Sheldon Glashow, *Interactions: A Journey Through the Mind of a Particle Physicist and the Matter of This World* [p. 335]

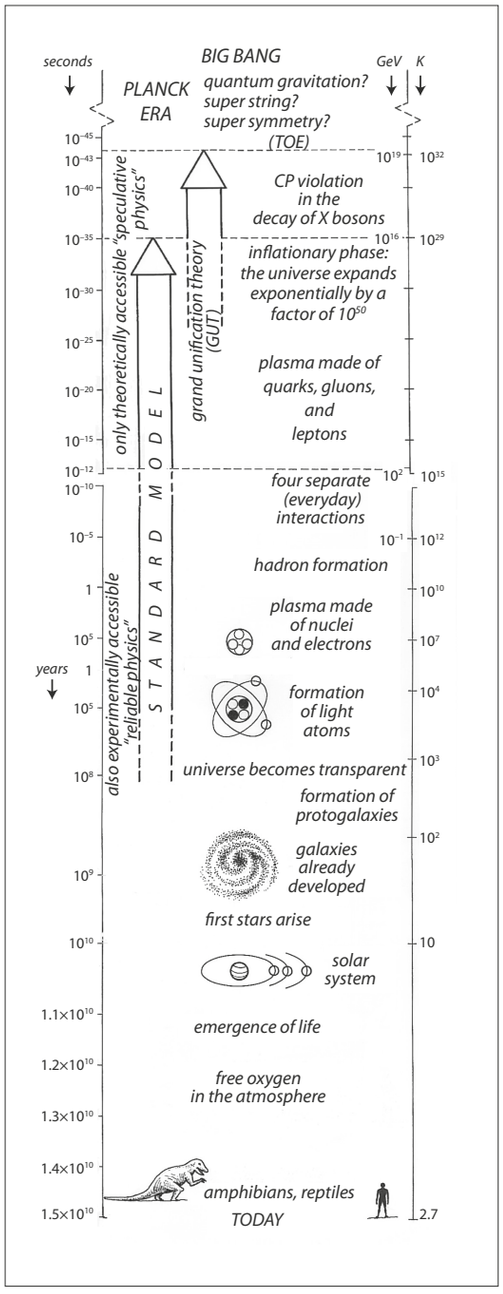
The stage on which this drama is enacted is created by the general theory of relativity. The "stars" are the fundamental particles with their "elective affinities." Through their actions, nuclei and atoms, the known objects of our daily lives, take the stage, and finally there follow the macroscopic objects such as stars, galaxies, and clusters of galaxies. At time 10^{-12} seconds after the big bang, the average energy of particles was about 100 GeV ($\approx 10^{15}$ K). This is the order of magnitude realizable in today's terrestrial laboratories, at least for individual particles, though not with respect to the temperature corresponding to this energy density. All the assertions of any given theory are from then on directly verifiable. The four "everyday" interactions exercise their effects independently and individually, and thus the well-worked-out theories are applicable to all phenomena (reliable physics). The stages of the most important events and processes in the universe are presented somewhat more extensively in Figure 5.196.



▲ **Figure 5.195** The most important events in the microphysical (elementary particle) realm took place in the interval from 0 to 10^{-12} seconds after the big bang. Thereafter, the laws of the everyday world rule with their now separated four fundamental interactions in the energy range 0 to 10^2 GeV.

The relative intensities of the strong, electromagnetic, weak, and gravitational interactions in our everyday world are 1, 1/137, 10^{-5}, and 10^{-39}, respectively. At higher energies, these values approach each other more and more, which suggests the complete melding together in the Planck era.

However, our existence in this world is made possible only through the unequal strengths of these forces. Strong forces ensure the integrity of the nuclei; the electromagnetic forces holding atoms together see to the variability necessary for life accompanied by a relative stability. The weak interaction provides for the appearance of stars as a source of energy and balances the gravitational forces so that the stars will not collapse prematurely. Many studies have pointed out the serious consequences if the ratios of the strengths were different, even by a small about.



◀ **Figure 5.196** From the big bang to today: the history of the universe.

This picture can give only a fleeting overview of events. The details vary depending on the point in time and the researcher.

The CP violation occurring in the decay of *X* bosons can be held accountable for the matter–antimatter asymmetry in today's world. (One assumes, of course, that originally there was symmetry.) This means that the decays that led to ordinary matter are more probable than those whose decay products consisted of antimatter. The difference is only a very small fraction of all of matter (10^{-9}).

It is from this fraction that our world is composed, since the other particles were transformed by annihilation radiation into gamma energy.

The arguments over the origin, the coming into being of our world, are not products of our time. We have already encountered the ideas of Aristotle, Augustine, Maimonides, and Heisenberg in this regard. Here we shall content ourselves with recalling something from the thoughts of Maimonides:

> [A]ccording to us, time is a created and generated thing. … For if you affirm as true the existence of time prior to the world, you are necessarily bound to believe in the eternity (of the world). For time is an accident which necessarily must have a substratum. Accordingly, it follows necessarily that there existed some thing prior to the existence of this world existing now.
>
> —*The Guide of the Perplexed*, II.13

This viewpoint is strengthened by the following experimental facts:

1. The temporal course of the expansion of the universe.
2. The rate of concentration of the lighter atoms (H, He, Li) in the universe.
3. The background radiation corresponding to a temperature of 2.72 K.

This radiation is the consequence of the equilibrium radiation that arose when the age of the universe was 3×10^5 years. Because atoms no longer absorbed photons when the temperature cooled to 3000 K, the cosmos became "transparent." The equilibrium radiation further cooled as a result of the expansion of the universe to today's value of 2.72 K.

These phenomena belong simultaneously to both the areas of gravitation and nuclear physics. Traces of such effects that could say something about the correctness of the theories of particle physics are not immediately to be found.

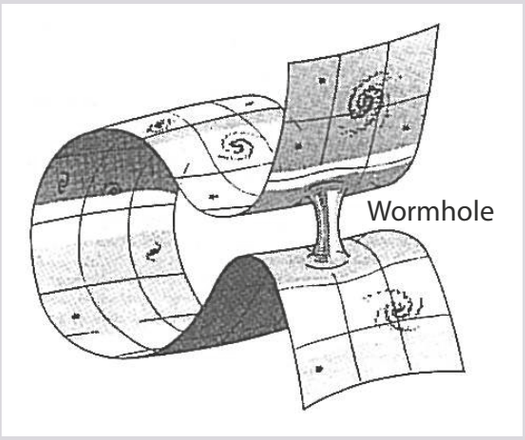
Quotation 5.66

Yet another approach began with a formal analysis of quantum gravity by Abhay Ashtekar at the University of Syracuse in the State of New York. Expanding upon his ideas, the young investigators Lee Smolin and Carlo Rovelli proposed that the fundamental ingredients of space-time are not the points in there but rather the *closed loops*. This looks a bit like what we saw in string theory, but this is an altogether different approach; according to Smolin and Rovelli, the essential thing is that these loops are tied together in *knots*, and that outside the knots there is no space-time at all.

This is an attempt at constructing a new theory that I am following with interest—at last something that looks a bit like what I would want. In this theory, the only thing relevant is the number and the kinds of "knots" linking the loops, and these are matters that in principle could be expressed simply by series of whole numbers. Just like loose points in space-time, except that they are naturally connected! Accidentally, or perhaps not so accidentally, *knot theory* is one of the most difficult branches of modern mathematics. During their excursions in string mathematics, the string supporters had also hit the knot problem. Edward Witten discovered during his analysis several new mathematical theorems about knots which earned him the prestigious Fields Medal.

If you think that these new ideas are no more than day-dreaming and a waste of time, I should disclose that what I have mentioned so far is not nearly the most extravagant attempts and approaches that have been published. In conferences and workshops that sometimes become like brainstorming parties, we are treated to far more bizarre concoctions. People philosophize about quantum jumps from one universe to another (usually through wormholes), about parallel worlds in "quantum cosmology," about worlds in which the constants of Nature differ from ours, but which are connected to us by wormholes, and even about the question of whether you can create a universe inside a test tube.

—Gerard 't Hooft, *In Search of the Ultimate Building Blocks* [p. 174]



▲ **Figure 5.197** Difficulties increase when we are dealing with the topology of multidimensional (abstract) spaces and particularly if we wish to visualize them. Already in 1919, Theodor Kaluza, and somewhat later (1926), Oskar Klein, enlarged the four-dimensional space of general relativity theory to a five-dimensional manifold. We have already mentioned why these additional dimensions are unnoticeable. Namely, a piece of paper can be conceptualized as a two-dimensional space. If one rolls it up into a cylinder, it is still two-dimensional. However, if one chooses the radius of the cylinder to be very small in relation to its length, then the entire construct can be viewed as a line, that is, as a one-dimensional space. The essential principle in the Kaluza–Klein conception is that the components of the gravitational field in the fifth direction satisfy Maxwell's equations of electrodynamics. Einstein was at first enthusiastic, but then he turned against it. However, perhaps these ideas gave him the impulse to work on a unified field theory.

The supersymmetry (SUSY) theory works with 11 (= 3 + 1 + 7) dimensions, while superstring theory began with 26 dimensions (1984) and now manages with 10 (= 9 + 1).

Four-dimensional curved space can be visualized as a warped surface, embedded in the usual three-dimensional space (Figure 5.187).

This surface is subject to a variety of deformations. A black hole can be seen as one such deformation. Stephen Hawking and others assign a particular significance to "wormholes." These link two distant regions with each other or even connect our universe to another universe. The figure (after 't Hooft's book [1997, p. 173]) shows such a wormhole. Einstein's gravitational theory allows such curiosities. However, some physicists maintain that what is allowed must exist, to the great delight and feeling of vindication of authors of science fiction. If one follows 't Hooft's "acknowledgment" further (Quotation 5.66), then one no longer judges so harshly the statements of sociologists about the debates among physicists (Section 5.7.1).

The difficulties come into sharp focus when determining the age of the universe. Various conceptions, combined with differing experimental results, lead to an age between five and twenty billion years. A generally accepted age is 13.2×10^9 years (see Figure 5.181(b)).

In Figure 5.196, 15×10^9 years is assumed.

The biggest mystery, which is at the same time an enticing provocation, is the existence, or the origin, of the big bang itself. Gravitation exercises an attractive force, and so one should expect that a singular point with infinite mass density would be more likely to turn into a black hole than to expand explosively. It is not a comforting assumption for theoreticians that the initial conditions must be seen simply as givens, with kinetic energy as one of these.

Moreover, it is apparent that these conditions are finely tuned among themselves. Just think of the relationship among the coupling constants (Figure 5.195) or the spatial homogeneity of matter. It is highly unlikely that these data could have been somehow reconciled "from without." They must already exist, absorbed somehow into the (correct) theory.

In the further working out of the problems, it is inflation theory (Alan Guth, 1980) that has excited the greatest interest. On the basis of the theory of the grand unification, Guth has attempted to track down the conditions that could lead to a repulsive effect or, expressed more simply, to an "antigravitation."

In the theories based on gauge symmetry, such as GUT, the vacuum state in the quantum-mechanical interpretation possesses some very remarkable properties.

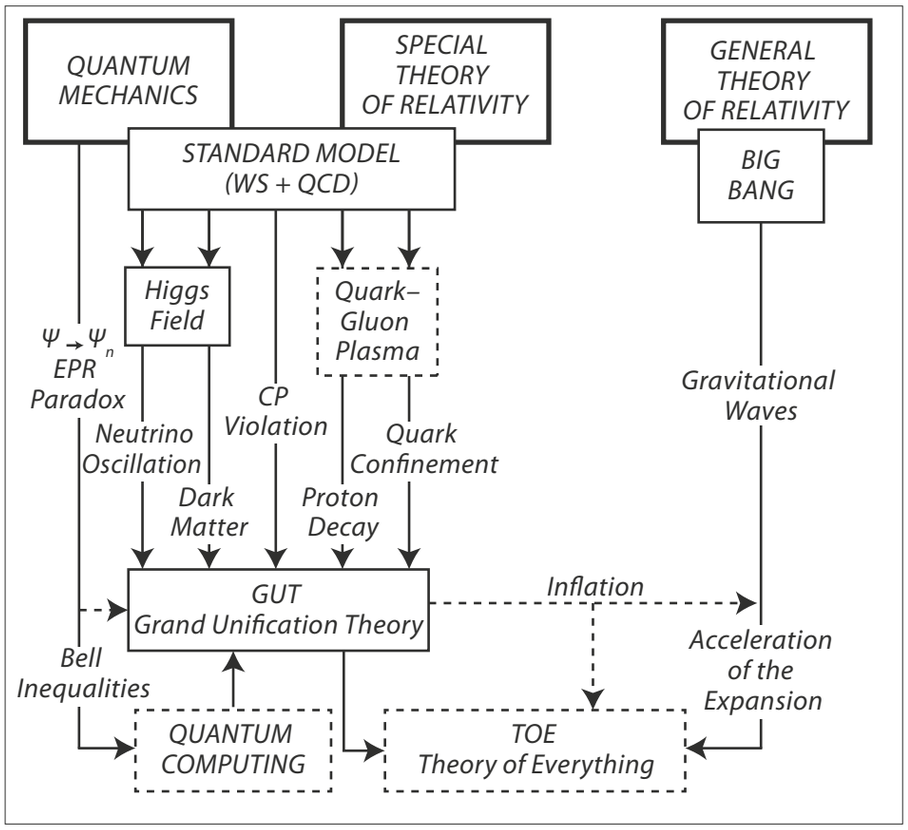
For example, a vacuum can exist in an excited state. (This corresponds to an aggregate state of matter, a phase in solid-state physics.) The excited vacuum is indeed empty in the sense that it contains no particles, yet it possesses an enormous energy and an enormous pressure. However, this pressure exhibits extraordinary behavior: It has a negative value (if it is assumed that the universe is in an unstable state). A result of the negative pressure is antigravitation effects. Guth's train of thought was this: Let us assume that "in the beginning" (*in principio*), the universe was in an excited state. Therefore, a small quantum fluctuation sufficed to end that unstable condition. The "antigravitation," caused by the negative vacuum pressure, causes an exponentially increasing expansion, the inflation phase.

The problem of the initial conditions solves itself in this theory: They can be arbitrary; inflation automatically sets the correct, coordinated values.

Stephen Hawking maintains in his theory that one can completely avoid the initial conditions, since the universe has no origin, even though time in the past is finite.

The critical reader might feel that this cosmology, this "exegesis," is simply empty talk in jargon that sounds authentic while saying nothing, with the goal of masking the ignorance of its proponents, and even more, that it offers the possibility for charlatans to glorify their phantasmagorias with the veneer of science.

We would add two more comments. First, physicists who wrestle with everyday problems, even physicists of the first rank, may share such views. Second, theoretical physicists are prepared after only a two- or three-year education in modern mathematics and physical theories to understand what has been said above in all its detail and even to develop it further if they possess the necessary talent. This explains why the field of superstring theory seems esoteric, because it is accessible only to the initiated, while at the same time it attracts charlatans. The majority of theoretical physicists and perhaps the majority of physicists altogether are, however, fully convinced that human understanding is capable of formulating such abstract, logical thought structures that even if they have no direct connection



◀ **Figure 5.198** If we compare this diagram with Figure 5.1, we do not find here any clashing arrows that would indicate known contradictions (in color in Figure 5.1). The Standard Model is unable to give answers to a number of important questions but its assertions are not in contradiction of any experimental facts. (Diagram is in part after G. Vesztergombi.)

A summary of this sort presents even greater difficulties at the turn of this millennium than at the turn of the previous century. In preparing the earlier figure, we had knowledge of all the new disciplines that would be established over the century ahead. Thus we could define the questions to be answered and problems to be solved. In the leading journals, such as *Annalen der Physik* and *Physical Review*, hundreds of articles have been published in which new and newer problems are continually solved that in turn, however, raise a new set of problems. It is easy to decide what the relevant problems were after the fact, the answers to which led to the establishment of relativity theory and quantum mechanics. However, we don't know for certain—we only suspect or fantasize—about the syntheses to take place in the coming century; we can therefore not maintain that the solution of the selected questions in fact represents the correct path to the establishment of newer fundamental laws.

The twentieth century consigns to the current a fundamental problem that it struggled over for decades: the measurement process itself and the evaluation of results or, formulated quite generally, the "interface" between mankind and nature.

with reality, can nonetheless be used to deal adequately with problems relating to the big bang. By this we mean that they hope to arrive at statements that in fact refer to the real world and are thereby verifiable through experiment.

It is not at all surprising that abstract concepts and theories cannot be expressed in colloquial language. It is much more surprising that one can speak of such things at all (see Bohr's thoughts in this direction, Quotation 0.9 and Figure 5.197).

We need only recall that the conceptual system of Maxwell's theory, the electromagnetic field, made very great demands on the abstraction capacity of the best thinkers of the nineteenth century. The consequences of this theory, electromagnetic waves, have nevertheless long been generally accepted—if more on trust than understanding.

5.7.6 *Questions and Doubts Multiply*

If someday a TOE (theory of everything) is born out of a synthesis of superstring theory and inflation theory, to what degree can it be viewed as the realization of a dream, the dream of a final theory that many physicists desire and even express in the titles of their books (for example, *Dreams of a Final Theory*)? If we compare Figure 5.1 with Figure 5.198, we see that the final result of the former figure represents the starting point for the latter figure. Will the creation of an all-encompassing theory bring an end to this continuity? Or will science be able to make a similar sketch at the beginning of the twenty-second century, where the starting point is given by TOE? There are many who shrink from the word "final." They consider it unacceptable, indeed logically false, that a principle does not assume another, lying at a deeper level, on which it can be based. Others believe that a final theory exists that would be capable of providing a complete explanation of the in-

nermost structure of our universe and that our generation stands at the cusp of such a theory. Even the pessimists admit this, with the proviso that it is not accessible to the human mind. Will the newly created theory—final or provisional—give an answer to the anthropic principle, and exclusively on the basis of physical laws? Or will it, in contrast, use this principle in the selection of competing theories? The anthropic principle states that our world is so finely tuned to the requirements of human existence (Figure 5.195) that this fact demands an explanation. More rigorously formulated, the principle states that the laws of nature must be so constructed that they make possible the existence of a human intelligence that is capable of investigating those laws.

In what follows, we shall speak about some of the problem areas referred to in Figure 5.198.

5.7.6.1 Neutrino Oscillation

In parallel to the development of the technology of neutrino detection, the significance of neutrino astronomy is growing: Neutrinos provide information from the interiors of stars. Thus our ideas about their structure—density, temperature distribution, etc.—can be verified or expanded. Measurements taken since the 1970s with ever increasing precision yielded a stark discrepancy with the results calculated on the basis of the accepted solar model: The number of recorded neutrinos was at most half of the number expected.

After a number of unsuccessful attempts for explanation, the current theory is that electron neutrinos created in the Sun change during their travels into muon or tauon neutrinos. This is called neutrino oscillation. For a neutrino with mass zero, however, such a transformation would be, in principle, forbidden. The Standard Model allows for, but does not explain, a nonzero mass for the neutrinos.

5.7.6.2 CP Violation

The difficulty in investigating CP violation is that, until very recently, this effect has been observed only within the system $K^0(\bar{d}s)$, $\bar{K}^0(\bar{d}s)$. Based on the Standard Model, it was expected that the system consisting of the meson pair $B^0(d\bar{b})$ and $B^0(\bar{d}b)$ would also exhibit CP violation. The operation of the "B factory," at the SLAC (Stanford Linear Accelerator) in California with a production rate of several thousand B mesons per day, provided an affirmative answer.

CP violation is also connected with the asymmetry between matter and antimatter. The difficulty of an explanation based on what has been said above is that in the time period during which the balanced state of equilibrium of matter and antimatter was destroyed, that is, from the big bang to 10^{-36} seconds, stable mesons in the existing "quark–gluon soup" were completely unable to exist.

5.7.6.3 Acceleration of the Expansion of the Universe

As we have seen earlier (Figure 5.185), above a certain critical mass density, gravitational attraction acts as a brake on expansion; therefore, a slowing down, which is to say, a decrease in the speed of expansion, is to be expected. However, recent measurements point to an increase in speed. This was discovered through a systematic study of Type 1a supernovas, where one can measure simultaneously the velocity (via the red shift) and distance (from their characteristic luminosity profiles).

Formally, such an acceleration can be "explained," or better, described, by the introduction of the cosmological constant λ in Einstein's fundamental equation

(Figure 5.185). The fitting of the cosmological constant to the properties of space and the matter that fills it raises fundamental questions. The speed of increase or decrease depends, namely, on the energy density of matter and on the pressure that this matter exerts in all three directions. (Consider that the unit of pressure is force/m^2 = force × m/m^3 = energy/m^3, that is, it agrees with the energy density.) An acceleration is possible only with a negative value of the sum of these two quantities (see, for example, the explanation of inflation in Section 5.7.5). Among the known forms of matter, the result is always a positive value. Thus, a new form of matter (quintessence, *quinta essentia*) may have to be postulated.

5.7.6.4 Quantum Computing

The physical components of computers have become increasingly complex in the course of the development of ever newer branches of physics. The mechanical gears of Pascal (1640) and Leibniz (1671) would later be replaced by electromechanical relays. Indeed, some of the first computers used such devices. The application of electron tubes represented great progress. After the formulation of the theory of stored programs by John von Neumann (1944), computers attained their modern form. Then in the 1950s came transistor-based computers, those with discrete switching elements, and finally with integrated circuits.

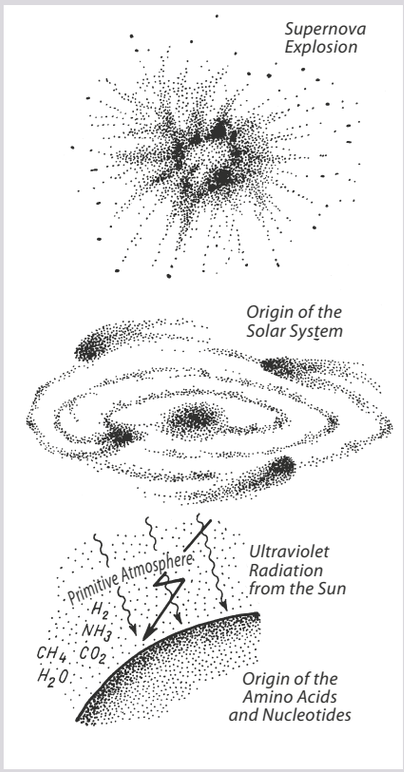
It should be mentioned here that computer technology has successfully discharged its debt to basic research, namely by making it possible, via the Internet, to spread new ideas to the entire world "*in statu nascendi*" and in such a way that these ideas can be built on or rejected.

Today, we are approaching a stage at which the basic computational elements consist of individual atoms. Here, however, the laws of quantum mechanics come into play with their probabilistic statements and uncertainty relations. Does this establish a limit on the possibilities for further development? No. It is, rather, a challenge. As a result of recent interpretations of the debates of the last century on the fundamentals of quantum mechanics and the repetition of basic experiments, the contours of new possibilities for application have come to light.

In 1936, Alan Turing created the general theory of "classical" computers. In 1986, David Deutsch proposed the quantum analogue of the Turing machine. The difference between the two types of computers can be summarized as follows: While conventional computers store data in the form of bits that can take on the values 0 and 1, a quantum computer stores information in two-stage quantum states, such as, for example, the spin of a proton. The crucial point is that these quantum states, called qubits, can become "entangled." This means that N qubits exist in 2^N different states. In such a quantum superposition, all of these states can be processed simultaneously. However, this means, ultimately, that a quantum computer can operate correspondingly faster than a conventional computer (DiVincenzo and Terhal [1998]).

Some possibilities for the realization of a qubit (the term was introduced by Ben Schumacher in 1995) are as follows: for photons, vertical and horizontal polarization; for atoms, electrons, and atomic nuclei, the two spin alignments; and furthermore, for atoms, the two energy levels.

The construction of a quantum computer is straightforward in principle: One simply takes simple quantum-logical elements and combines them into a quantum network. However, with a growing number of elements, difficulties following from quantum laws arise that today still await a solution. The most serious of these is



◀ **Figure 5.199** Critical steps in cosmic evolution that eventually led to the most highly organized form of matter: life.

The heavy elements were created (perhaps) during a supernova explosion. The physical conditions—the narrow temperature and pressure range, appropriate liquid, gaseous, and solid phases of materials—conducive to the *continued existence* of the sensitive, fragile, and extremely complex protein molecules, which combine reproducible stability with the necessary variability, were ensured by our solar system. For the *initial formation* of the building blocks, electrical phenomena in the primitive atmosphere may have been responsible. Once self-reproducing units were formed, the tempo of evolution was dictated by the laws of biology.

the "decoherence" of the quantum state, resulting in loss of information, due to the quantum system's interaction with its environment. Irrespective of whether new practical applications might be successfully implemented, these techniques and the accompanying new experimental facts will create momentum in the investigations of basic quantum-mechanical problems such as the riddle of the measurement process (Figure 5.77) or that of the EPR paradox with Bell's inequalities (Figure 5.78).

5.7.7 *"Between Nothing and Infinity"*

The eternal and unchanging Aristotelian world encompassing the Sun, the Moon, five planets, and the stars embedded in a crystalline sphere was forced to cede its place to a dynamic universe resembling a giant laboratory full of exquisite objects that change in exquisite ways.

From a smallish planet orbiting an average star located at the edge of a galaxy of average size, mankind observes the divine drama of cosmic events. We can confidently say that we belong in the flow of these cosmic processes with every drop of our blood. Biologists trace the origins of life back to protein molecules as initial building blocks; but the hydrogen contained in our bodies also took part in the primordial processes of the universe, and every iron nucleus in human blood originated within a massive star, went through a supernova explosion before it could become, via the development of a new star system, a component of our solid earth and now play an important role our biological function (Figure 5.199).

The grandeur of all of human history is dwarfed, shrivels to nothing, and is lost in the enormous currents of the cosmos (Quotation 5.63), and yet this terrestrial world is not insignificant, for alongside all the marvelous objects in the universe, next to pulsars, quasars, and black holes, the great cosmic creation brought forth in this place the human brain and human consciousness with its capacity to be aware of both the infinitely small and the infinitely large (Figure 5.200).

> Here is where natural knowledge leads us: if it is not true, there is no truth in man; and if it is true, he finds in it a great source of humiliation, forced to humble himself one way or another.
>
> And since he cannot subsist without believing this knowledge, before entering into greater explorations of nature, I want him to consider nature both seriously and at leisure just once, and also to reflect upon himself and to judge, in a comparison of these two objects, whether any proportion holds between them.
>
> Let man then contemplate the whole of nature in its lofty and full majesty, and let him avert his view from the lowly objects around him. Let him behold that brilliant light set like an eternal lamp to illuminate the universe. Let the earth seem to him like a point in comparison with the vast orbit described by that star. And let him be amazed that this vast orbit is itself but a very small point in comparison with the one described by the stars rolling around the firmament. But if our gaze stops here, let our imagination pass beyond. It will sooner tire of conceiving things than nature of producing them. This whole visible world is only an imperceptible trace in the amplitude to nature. No idea approaches it. However much we may inflate our conceptions beyond these imaginable spaces, we give birth only to atoms with respect to the reality of things. It is an infinite sphere whose center is everywhere and circumference nowhere. ...

Let man, returning to himself, consider what he is with respect to what exists. Let him regard himself as lost in this remote corner of nature, and from the little cell in which he finds himself lodged, I mean the universe, let him learn to estimate the just value of the earth, kingdoms, cities, and himself.

What is a man in the infinite?

But to present him with another equally astonishing prodigy, let him examine the most delicate things he knows. ... Let him lose himself in wonders as astonishing in their minuteness as the others are in their extent! For who will not marvel that our body, imperceptible a little while ago in the universe, itself imperceptible inside the totality, should now be a colossus, a world, or rather a whole, with respect to the nothingness beyond our reach? ...

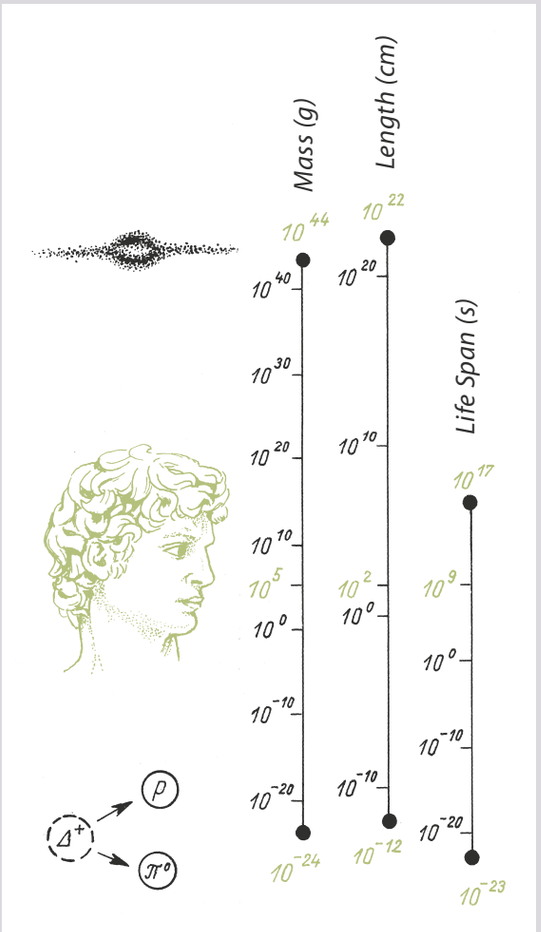
For what, in the end, is man in nature? A nothing compared with the infinite, an everything compared to the nothing, a midpoint between nothing and everything. ...

This is our true state. It is what makes us incapable of certain knowledge or absolute ignorance. We float on a vast ocean, ever uncertain and adrift, blown this way and that. Whenever we think we have some point to which we can cling and fasten ourselves, it shakes free and leaves us behind. And if we follow it, it eludes our grasp, slides away, and escapes forever. Nothing stays still for us. ...

All things, then, are caused and causing, supporting and dependent, mediate and immediate; and all support one another in a natural, though imperceptible chain linking together things most distant and different. So, I hold it is as impossible to know the parts without knowing the whole as to know the whole without knowing the particular parts. ...

Man is to himself the most prodigious object of nature, for he cannot conceive what body is, still less what mind is, and least of all how a body can be unified to a mind. This is the culmination of his difficulties, and yet it is his very being. *The way the spirit is united to the body cannot be understood by man, and yet it is man.* [*Modus quo corporibus adhaerent spiritus comprehendi ab hominibus non potest, et hoc tamen homo est;* Saint Augustine, *City of God*, XXI.10.]

—Blaise Pascal, *Pensées* [pp. 58–62]



▲ **Figure 5.200** The place of mankind between "nothing" and the "infinite."



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Epilogue: Looking Ahead in Physics

By Edward Witten

Before venturing a few guesses about the near future in physics, it is good to step back and remember some of the surprises that have occurred during the thirty-some years that I have been in the field. Among other things, this may help give us some due humility concerning our ability to foresee the future.

I actually had just learned enough of theoretical physics to understand what was going on when the J/Ψ particle was discovered in the fall of 1974. That was quite a bombshell, the biggest in accelerator physics while I have been in the field. The discovery of this particle was a big surprise because it was a strongly interacting particle with strong decays, but by orders of magnitude it was much more long-lived (or "narrow" in energy terms) than any previously observed strongly interacting particle. However, after physicists got over the shock, they realized that the discovery made a lot of sense in terms of the asymptotically free gauge theory of strong interactions and the existence of a new quark, the "charmed quark." Before long, the further exploration of the J/Ψ and related particles had given a big boost to what came to be known as the Standard Model of elementary particle physics.

The J/Ψ was a particularly big bombshell, but the fact is that in the previous 20 years, there had been one highly unexpected discovery after another—hadron resonances, muons, strange particles, the extended nature of the proton, parity violation, CP violation, the eightfold way, scaling in electron scattering—too many to list. So at the time, as a graduate student, I assumed that this sort of thing was part of life as a physicist. However, things took an unexpected turn soon after, and I have to say that this was one of the biggest surprises during my time as a physicist. Once the J/Ψ had been discovered, and a few other important pieces fell into place (including the quarks and leptons of the third generation), it turned out that the biggest surprise coming from accelerator physics in the next 30 years was how well the Standard Model works. I believe that its authors thought that they had for the first time a sensible theory of elementary particles and hoped that it would be a useful approximation, but I seriously doubt that any of them thought in 1974 that the Standard Model would still be holding up so well in 2011, when the power of experimentation has reached so far beyond where it was back then.

I must say that as a student I did not anticipate the richness and beauty of the discoveries in condensed matter physics that have occurred during these years. Naively speaking, how can something simple and beautiful emerge from something as complicated and messy as the Schrödinger equation of electrons and atomic nuclei? The answer in a nutshell is that (as Phil Anderson likes to say) "more is different"—new and simple structures can emerge from cooperative effects of many particles together. There are all kinds of examples, but the ones I tend to care about most are the ones that can be usefully understood using the methods

DOI: 10.1201/9781032697697-7
This chapter has been made available under a CC-BY-NC-ND license.

of quantum field theory that are familiar in particle physics. Striking examples include the quantum Hall effect and the much more recent topological insulators. It is fascinating and often humbling to see how seemingly esoteric effects of quantum field theory can work in an ordinary crystal or some other real-world condensed matter system. At this stage, it is pretty clear that condensed matter physics will continue to be interesting for a long time, in part because of the continuing ability to fabricate new materials and probe old materials in new ways.

Another big surprise—and for the most part closer to my research interests—is how much progress has been made in cosmology. One reason involved theory: inflationary cosmology was an important and fundamental, yet simple, idea. But the biggest single reason that cosmology has made so much progress since my student days is that the cosmic microwave radiation, or CMB—especially its very slight but fundamental anisotropy—proved to be an astonishingly precise and informative probe of the early universe. Probably some people suspected this a long time ago, but it certainly came as a surprise to me when this emerged in the last ten to twenty years.

However, the biggest experimental or observational surprise of all during my years in physics has been a discovery that came from astrophysics and cosmology, but in which the CMB played only a supporting role. This was the discovery of the acceleration of the expansion of the universe, which is most simply interpreted in terms of a cosmological constant, namely, the energy density of a vacuum. To a particle physicist, this was a shock. The natural value of the vacuum energy density is extremely high, so somehow it has been almost entirely canceled. Otherwise, according to Einstein's equations, the whole universe would curl up to atomic or even subatomic dimensions. Until observation showed that the cancellation is incomplete, we generally assumed that there was a perfect cancellation because of some unknown symmetry principle. Now that we know that the cancellation is incomplete, the thinking of many of my colleagues has gone to an opposite extreme—suspecting that the apparent value of the cosmological constant that we see is a local "accident," peculiar to our own surroundings but not representative of the universe as a whole. This is certainly not the vision of physics that any of us had in mind back in my student days. If it is right, I hope that we will be able to get more clues indicating that this is the case. Personally, I have never overcome the shock of learning about the cosmic acceleration. It took several years of improved data before I came to believe that it was real.

Finally, there has been a surprise that mostly had to do with advances in theory—though advances that were certainly accelerated by the hothouse atmosphere of experimental developments in the 1960s and 1970s. This is the emergence of string theory as a route to going beyond quantum field theory in a fundamental way and incorporating gravity in a quantum theory. One thing I am sure of is that I would never have tried to invent something like string theory on purpose. It would have been too big a jump. I am a little skeptical that anyone I know would have seriously gone off looking for something like that—and even more skeptical that anyone would have succeeded. Luckily, string theory came looking for us, in the sense that it emerged as a byproduct of an only partly successful attempt to understand the strong interactions. None of its inventors had any idea, going in, what it was that they were stumbling upon. But string theory has proved to be a remarkable and remarkably powerful generalization of the standard framework of quantum field theory. Its emergence opened the opportunity to acquire a lot of very interesting insights, and I consider myself lucky to have been working in physics during this period.

Having made this brief tour of how things didn't really turn out quite like we would have guessed 35 years ago, let us now venture a few rather cautious observations about the future.

One of the biggest questions certainly is whether the Standard Model will continue to hold up, or whether in the era of the Large Hadron Collider (LHC) we will finally get beyond the Standard Model. Almost certainly the LHC will find the Standard Model Higgs particle. The big question is whether it will be the Higgs particle by itself—a capstone to the Standard Model as the theory of particle physics at accessible energies—or the Higgs particle plus a lot of other things that will enable us to resolve some of the mysteries of the Standard Model. As I write, for example, there is a lot of interest in tentative experimental findings about the possible CP-violating decays of B_s particles.

Should these tentative findings be confirmed, we would finally reach the limits of the Standard Model and the next phase would bring a return to something a little like the situation in the 1950s, when the new particles emerging from accelerators were *terra incognita*. On the other hand, it is equally possible that as more and better data are accumulated, the tentative findings that I mentioned will evaporate. I certainly hope it is the former, but things aren't always as we hope.

In cosmology, the discoveries of the last few decades created something akin to a Standard Model. Observational techniques are continuing to improve on many fronts, and this will go on for quite a while. The big question is whether the standard model of cosmology will hold up, at least for a long time, or whether, as new data come in, there will be more surprises. Personally, I would not be too surprised if the cosmological standard model does hold up. The golden age of cosmology we have lived through has been glorious, but unfortunately nothing like that goes on forever. However, even if the cosmological standard model is going to hold up, there is one really important and clarifying observation that may become possible in the next decade: this would be the observation of a sort of handedness (technically called the B-mode) in the polarization of the CMB. Should this be observed in the form that has been suggested in the context of inflationary cosmology, it would, at least to me, be the clincher confirming that cosmic inflation really happened in our past.

If the cosmological standard model does hold up, in which areas—beyond the possible discovery of the B-mode—will astronomy be exciting in the coming decades? One obvious answer—but certainly not the only one—is the study of extra-solar planets. For at least the next half century, our ability to observe and study them will be in a state of steady growth, and there are bound to be many exciting and fun surprises.

In particle physics, I dearly hope that we are getting close to the limits of the Standard Model. If that isn't the case, we can still hope for exciting advances on a number of fronts. The biggest would be an observation of proton decay. Perhaps more likely is an observation of CP violation in neutrino physics. (The two may go together as the optimal way to study the neutrinos may involve a detector that is also suited for a better study of proton decay.) And it is certainly possible that we will get more stimulating input from cosmology—though it is pretty hard to imagine anything as stimulating as the discovery of the cosmic acceleration.

I have not yet offered even a cautious observation about theory. I am sure string theory will continue to be interesting and exciting. We will continue to learn interesting things from it, and hopefully we will get closer to understanding, in some sense, what string theory means. Can we understand how it works in the real world? This depends on a lot of imponderables, including the nature of the answer, the clues we can get from experiment, and how clever and lucky we will be.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Bibliography

Introduction

D'Alembert, Jean le Rond and Diderot, Denis (eds.). *Encyclopédie,* 35 Vols. Paris: Briasson, David, Le Breton and Durand, 1751–1772.

Andersson, Gunnar. "Presuppositions, Problems, Progress." In *The Structure and Development of Science*, edited by Gerard Radnitzky and Gunnar Andersson, pp. 3–15. New York: Springer, 1979.

Aristotle. *Metaphysics*, translated by Hugh Tredennick. Cambridge, MA: Harvard University Press, 1935.

Barraclough, G. (ed.). *Knaurs Großer Historischer Weltatlas*. Munich: Droemer-Knaur, 1979.

Bernal, J. D. *Science in History*, 4 Vols. New York: Penguin Books, 1969. (Originally published in German as *Die Wissenschaft in der Geschichte*. Berlin: Deutscher Verlag der Wissenschaften, 1961.)

Boekhoff, H. and Winzer, F. (eds.). *Kulturgeschichte der Welt*, 2 Vols. Braunschweig: WestermannVerlag, 1963.

Boltzmann, Ludwig. *Vorlesungen über Maxwells Theorie der Elektricität und des Lichtes*, 2 Vols. Leipzig: J. A. Barth, 1891–1893.

Brentjes, B., Richter, S., and Sonnenmann, R. *Geschichte der Technik*. Leipzig: Verlag Edition, 1978.

Bridgman, P. W. *Reflections of a Physicist*. New York: Philosophical Library, 1950.

Brush, Stephen G. *Resources for the History of Physics*. Hannover, NH: The University Press of New England, 1972.

Brush, Stephen G. *The Temperature of History: Phases of Science and Culture in the Nineteenth Century*. New York: Burt Franklin, 1978.

Bynum, W. F. and Browne, E. J. (eds.). *Dictionary of the History of Science*. London: MacMillan, 1982.

Dugas, René. *A History of Mechanics*. New York: Dover Publications, 1988. (Originally published as *Histoire de la Mécanique*, edited by René Dugas. Neuchâtel: Éditions du Griffon, 1950.)

Davies, P. C. W. and Brown, Julian (eds.). *Superstrings: A Theory of Everything?* New York: Cambridge University Press, 1988.

Dijksterhuis, E. J. *Die Mechanisierung des Weltbildes*. Berlin: Springer-Verlag, 1956.

Dugas, René. *A History of Mechanics*, translated by J. R. Maddox. London: Routledge and Kegan Paul, 1957.

Durant, N. and Durant, A. *The Story of Civilization*, 11 Vols. New York: Simon and Schuster, 1975.

Einstein, Albert. "The Herbert Spencer Lecture" (delivered at Oxford University, Oxford, UK, June 10, 1933). In *Ideas and Opinions*, pp. 264–265. New York: Dell Publishing, 1981.

Feynman, Richard. *The Character of Physical Law*, The 1964 Messenger Lectures. Cambridge, MA: MIT Press, 1967.

Forbes, R. J. and Dijksterhuis, E. J. *A History of Science and Technology*, 2 Vols. Harmondsworth: Penguin Books, 1963.

Gilispie, C. C. *Dictionary of Scientific Biography*, 14 Vols. New York: Schribner's Sons, 1970–1977.

Gingerich, O. (ed.). *The Nature of Scientific Discovery*. Washington, DC: Smithsonian Institute Press, 1975.

Goethe, Johann Wolfgang von. *Maxims and Reflections*, translated by Bailey Saunders. London: Macmillan and Co., 1893.

Goldstein, Martin and Goldstein, Inge F. *How We Know: An Exploration of the Scientific Process*. New York: Plenum Press, 1978.

The Harvard Project Physics Course. New York: Holt, Rinehart and Winston, 1970.

Heisenberg, Werner. *Physics and Beyond: Encounters and Conversations*, translated by Arnold J. Pomerans. New York: Harper & Row, 1971.

Heisenberg, Werner. "Is Physics Coming to an End?" (lecture held in Budapest). *Fizikai Szemle* 2 (1976), pp. 62–64.

Holton, G. and Morison, R. S. *Limits of Scientific Inquiry*. New York: W. W. Norton Co., 1978.

Hund, F. *Geschichte der physikalischen Begriffe*. Mannheim: Bibliographisches Institut, 1972.

Kirby, R. S., Withington, S., Darling, A. B., and Kilgour, F. G. *Engineering in History*. New York: McGraw-Hill, 1956.

Kistner, A. *Geschichte der Physik*, 2 Vols. Leipzig: Sammlung Göschen, 1906.

Kline, M. *Mathematical Thought from Ancient to Modern Times*. Oxford, UK: Oxford University Press, 1972.

Kragh, Helge. *Dirac: A Scientific Bibliography*. Cambridge, UK: Cambridge University Press, 1980.

Kuhn, Thomas. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1967.

Liertz, Rhaban. *Die Naturkunde von der menschlichen Seele nach Albert dem Grossen*. Cologne: J. P. Bachem, 1932.

Lorenz, Konrad. *Civilized Man's Eight Deadly Sins*, translated by Marjorie Kerr Wilson. New York: Harcourt Brace, 1974.

Mittelstaedt, Peter. *Die Sprache der Physik.* Mannheim: Bibliographische Institut-Wissenschaftsverlag, 1972.

Moorhead, J. K. (ed.). *Conversations of Goethe with Johann Peter Eckermann*, translated by John Oxenford. Toronto: Everyman Library, J. M. Dent, 1930.

Newman, J. R. (ed.). *The World of Mathematics,* 4 Vols. New York: Simon and Schuster, 1956.

Owen, G. E. *The Universe of the Mind*. Baltimore: The Johns Hopkins Press, 1971.

Pannekoek, A. *A History of Astronomy*. London: Allen and Unwin, 1961.

Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics*. New York: Oxford University Press, 1989.

Физический Энциклопедический (Physics Encyclopedia). (Chief Ed. Prokhorov). Moscow, 1983.

Poincaré, Henri. *Science and Hypothesis*, translated by W. J. G. [William John Greenstreet]. New York: The Walter Scott Publishing Co., 1905. (Available at http://www.archive.org/details/sciencehypothesi00poin.)

Popper, K. R. *Objektive Erkenntnis*. Hamburg: Hoffmann und Campe, 1973.

Priestley, J. *The History and Present State of Electricity*. London: J. Dodsley, J. Johnson, and T. Cadell, 1775.

Radnitzky, G. and Andersson, G. (eds.). *The Structure and Development of Science*. Dordrecht: Reidel, 1978.

Rilke, Rainer Maria. *Duino Elegies*, translated by A. S. Kline. Available at *Poetry in Translation,* http://poetryintranslation.com/PITBR/German/Rilke.htm, 2001.

Rosenberg, F. *Die Geschichte der Physik*, 3 Vols. Braunschweig: Vieweg, 1882–1890. (Reprinted in 1960.)

Russell, B. *History of Western Philosophy*. London: Allen and Unwin, 1946.

Rybnikov, K. A. История математики (History of Mathematics). Moscow: Moscow University Press, 1960.

Snow, C. P. *The Two Cultures and the Scientific Revolution*. New York: Cambridge University Press, 1959.

Suppe, F. (ed.). *The Structure of Scientific Theories*. Urbana, IL: University of Illinois Press, 1974.

Taton, R. *Historie générale des Sciences*, 4 Vols. Paris: Presse Universitaire de France, 1957–1972.

Wilder, Thornton. *The Bridge of San Luis Rey*. New York: Albert and Charles Boni, 1927.

Wussing, H. (ed.). *Geschichte der Naturwissenschaften*. Leipzig:Verlag Edition, 1983.

Chapter 1

Aeschylus. *Prometheus Bound*, translated by A. J. Podlecki. Oxford, UK: Aris & Phillips, 2005.

Archimedes. *The Sand-Reckoner*. In *The Works of Archimedes*, translated by T. L. Heath. Cambridge, UK: Cambridge University Press, 1897.

Archimedes. *The Method*. In *The Works of Archimedes*, edited and translated by T. L. Heath. Cambridge, UK: Cambridge University Press, 1912. (Reprinted by Dover Publications in 2002.)

Aristophanes. *Clouds*, edited and translated by Jeffrey Henderson. Cambridge, MA: Harvard University Press, 1998.

Aristotle. *Metaphysics*, translated by Hippocrates G. Apostle. Des Moines, IA: The Peripatetic Press, 1979. (Originally published by Indiana University Press in 1966.)

Aristotle. *Metaphysics*, translated by Hugh Tredennick. Cambridge, MA: Harvard University Press, 1935.

Aristotle. *On the Heavens*, translated by W. K. C. Guthrie. Cambridge, MA: Harvard University Press, 1945. (Originally published 1939.)

Aristotle. *On the Heavens*, translated by J. L. Stocks. In *The Complete Works of Aristotle*, Vol. I, edited by Jonathan Barnes. Princeton, NJ: Princeton University Press, 1984.

Aristotle. *Physics*, translated by Hippocrates G. Apostle. Bloomington, IN: Indiana University Press, 1969.

Aristotle. *Physics*, translated by by R. P. Hardie and R. K. Gaye. Available at *The Internet Classics Archive*, http://classics.mit.edu//Aristotle/physics.html, 2009.

Augustine. *Confessions*, translated by Albert C. Outler. Available at *Christian Classics Ethereal Library*, http://www.ccel.org/ccel/augustine/confessions.i.html, 2006.

Augustine. *Confessions*, translated by D. S. Pine-Coffin. London: Penguin Books, 1961.

Boas Hall, Marie. *The Pneumatics of Hero of Alexandria: A Facsimile of the 1951 Woodcroft Edition*. London: McDonald, 1971.

Brecker, K. and Feirtaf, M. (eds.). *Astronomy of the Ancients*. Cambridge, MA: MIT Press, 1979.

Chai, Ch'u and Chai, Winberg. *Confucianism*. Hauppauge, NY: Barron's Educational Series, 1973.

Christianson, G. *This Wild Abyss*. New York: The Free Press, 1978.

Cleomedes. *On the Circular Motion of the Celestial Bodies*. In *The World of Mathematics*, Vol. I, edited by James R. Newman. New York: Simon and Schuster, 1956.

Cohen, M. R. and Drabkin, I. E. *A Source Book in Greek Science*. Cambridge, MA: Harvard University Press, 1958.

Democritus. *Fragments*. In *Die Fragmente der Vorsokratiker,* by Hermann Diels. Berlin: Weidmann, 1935–1937.

Dirac, P. A. M. "The Excellence of Einstein's Theory of Gravitation." Lecture delivered at the UNESCO Symposium on the Impact of Modern Scientific Ideas on Society (Einstein's Centenary), Munich–Ulm, Germany, September 18, 1978.

Durant, Will. *The Story of Civilization: Part I: Our Oriental Heritage*. New York: Simon and Schuster, 1954.

Farrington, Benjamin. *Greek Science: Its Meaning for Us*, 2 Vols. (Vol. I "Thales to Aristotle" and Vol. II "Theophrastus to Galen"). Harmondsworth: Penguin Books, 1949.

Frontinus, Sextus Julius. *The Aqueducts of Rome*, translated by Charles E. Bennett. In *The Stratagems and the Aqueducts of Rome*. Cambridge, MA: Harvard University Press, 1969. (Originally published 1925.)

Geminos of Rhodes. *Introduction to the Phenomena*, translated by James Evans and J. Lennart Berggren. Princeton, NJ: Princeton University Press, 2006.

Gille, B. *Les Mécaniciens grecs: La naissance de la technologie*. Paris: Edition du Seuil, 1980.

Harper, Robert Francis. *The Code of Hammurabi King of Babylon*. Chicago: University of Chicago Press, 1904.

The Harvard Project Physics Course. New York: Holt, Rinehart and Winston, 1970.

Heraclitus. *Fragments,* translated by John Burnet, 1912. Available at http://en.wikisource.org/wiki/Fragments_of_Heraclitus.

Heidel, Alexander, *The Babylonian Genesis.* Chicago: Phoenix Books, University of Chicago Press, 1963. (Originally published 1942.)

Heisenberg, Werner. *Physics and Beyond: Encounters and Conversations*, translated by Arnold J. Pomerans. New York: Harper and Row, 1971.

Hero of Alexandra. *Opera quae supersunt omnia,* edited by W. Schmidt et al. Munich: K. G. Saur Verlag, 1998. (Originally published 1899–1914.)

Heronis Alexandrinus. *Opera,* Vol. III, edited by H. von Schone. Leipzig: Teubner 1903.

Lucretius. *De Rerum Natura,* translated by W. H. D. Rouse, revised by Martin Ferguson Smith. Cambridge, MA: Harvard University Press, 2002. (Originally published 1924.)

Maimonides. *The Guide of the Perplexed*, Vol. II, translated by Shlomo Pines. Chicago: University of Chicago Press, 1963.

Mann, Thomas. *Joseph and His Brothers*, translated by John E. Woods. New York: Everyman's Library, 2005.

Marcus Aurelius. *Meditations*, translated by Gregory Hays. New York: Modern Library, 2002.

Marshack, A. *The Roots of Civilization: The Cognitive Beginnings of Man's First Art, Symbol and Notation*. New York: McGraw-Hill, 1972.

Meissner, B. *Babylonien und Assyrien*, 2 Vols. Heidelberg: Carl Winters Universitas Buchhandlung, 1921–1925.

Neugebauer, O. *The Exact Sciences in Antiquity*. Princeton, NJ: Princeton University Press, 1951.

Neugebauer, O. *Arithmetik und Rechentechnik der Ägypter*, Quellen und Studien zur Geschichte der Mathematik. Berlin: Springer-Verlag, 1931.

Neugebauer, O. *Mathematische Keilschriftexte*, Quellen und Studien zur Geschichte der Mathematik. Berlin: Springer-Verlag, 1935.

Peet, T. E. *The Rhind Mathematical Papyrus*. London: Liverpool University Press, 1923.

Plato. *Phaedo*, translated by G. M. A. Grube. In *Plato: Complete Works*, edited by John M. Cooper and D. S. Hutchinson. Indianapolis: Hackett, 1997.

Plato. *Republic,* translated by G. M. A. Grube, revised by C. D. C. Reeve. In *Plato: Complete Works*, edited by John M. Cooper and D. S. Hutchinson. Indianapolis: Hackett, 1997.

Plato. *The Republic*, translated by Benjamin Jowett, Available at *The Internet Classics Archive*, http://classics.mit.edu/Plato/republic.html, 2009.

Plato. *Gorgias*, translated by Benjamin Jowett. New York: The Macmillan Company, 1892. (Digital text available at http://www.ilt.columbia.edu/publications/projects/digitexts/plato/gorgias/gorgias.html.)

Plato. *Timaeus*, translated by Donald J. Zeyl. In *Plato: Complete Works*, edited by John M. Cooper and D. S. Hutchinson. Indianapolis: Hackett, 1997.

Pliny the Elder. *Natural History: A Selection*, translated by John Healy. London: Penguin Books, 1991.

Plutarch. "The Face on the Moon." In *Moralia*, translated by Harold Cherniss and William C. Helmbold. Cambridge, MA: Harvard University Press, 1957.

Plutarch. *Lives*, translated by Bernadotte Perrin. London: Heinemann, 1955. (Originally published 1917.)

Ptolemy. *Tetrabiblios*, translated by W. G. Waddell. Cambridge, MA: Harvard University Press, 1971. (Originally published 1940.)

Reisch, Gregor. *Margarita philosophica*. Freiburg: Johann Schott, 1503.

Russell, Bertrand. *A History of Western Philosophy*. New York: Simon and Shuster, 1972. (Originally published 1945.)

Sambursky, S. *The Physical World of the Greeks*. London: Routledge and Kegan Paul, 1956.

Sambursky, S. *The Physical World of Late Antiquity*. London: Routledge and Kegan Paul, 1962.

Santillana, Giorgio de. *The Origins of Scientific Thought: From Anaximander to Proclus, 600 BC to 300 AD*. Chicago: University of Chicago Press, 1961.

Sarton, G. *A History of Science: Ancient Science through the Golden Age of Greece*. Oxford, UK: Oxford University Press, 1953.

Sarton, G. *A History of Science: Hellenistic Science and Culture in the Last Three Centuries BC*. Cambridge, MA: Harvard University Press, 1959.

Sommerfeld, Arnold. *Atomic Structure and Spectral Lines*, 1925, Vol. I, translated from the fifth German edition by Henry L. Brose. New York: E. P. Dutton and Company, 1931.

Stückelberger, A. (ed.). *Antike Atomphysik*. Munich: Heimeran, 1979.

Struve, V. V. *Mathematischer Papyrus des Museums in Moskau*, Quellen und Studien zur Geschichte der Mathematik. Berlin: Springer-Verlag, 1930.

Szabó, A. *Anfänge der griechischen Mathematik*. Budapest: Akadémiai Kiadó, 1969.

Taton, R. *Historie générale des Sciences*, 4 Vols. Paris: Presse Universitaire de France, 1957–1972.

Verene, Donald Phillip (ed.). *Symbol, Myth, and Culture: Essays and Lectures of Ernst Cassirer 1935–1945*. New Haven, CT: Yale University Press, 1979.

Vitruvius. *The Ten Books on Architecture*, translated by Morris Hicky Morgan. New York: Dover Publications, 1960. (Originally published 1914.)

Chapter 2

Aristotle. *Nicomachean Ethics*, translated by W. D. Ross. Available at *The Internet Classics Archive*, http://classics.mit.edu//Aristotle/nicomachaen.html, 2009.

Augustine. *Confessions*, translated by D. S. Pine-Coffin. London: Penguin Books, 1961.

Augustine. *The City of God against the Pagans,* edited by R. W. Dyson, Cambridge Texts in the History of Political Thought. Cambridge, UK: Cambridge University Press, 1998.

Bacon, Roger. *De secretis operibus artis et naturae.* In *The Age of Faith*, edited by Will Durant. New York: Simon and Schuster, 1950.

Banniard, M. *Le haut Moyen Age occidental*. Paris: Presses Universitaires de France, 1980.

Bibliotheca Corviniana (edited by Csapodi, C., Csapodiné Gárdonyi K., and Szántó, T.). Ontario: Magyar Helikon, 1967.

Boethius. *Consolation of Philosophy*, translated by Joel C. Relihan. Indianapolis: Hackett, 2001.

Buridan, John. "The Impetus Theory of Projectile Motion," translated by Marshall Clagett. In *A Source Book in Medieval Science*, edited by Edward Grant, pp. 275–279. Cambridge, MA: Harvard University Press, 1974.

Buridan, John. "The Compatibility of the Earth's Diurnal Rotation with Astronomical Phenomen," translated by Marshall Clagett. In *A Source Book in Medieval Science*, edited by Edward Grant, pp. 500–502. Cambridge, MA: Harvard University Press, 1974.

Clagett, Marshall. *Science of Mechanics in the Middle Ages*. Madison: University of Wisconsin Press, 1961.

Clagett, Marshall (ed.). *Nicole Oresme and the Medieval Geometry of Qualities and Motions*. Madison: University of Wisconsin Press, 1968.

Clark, Kenneth. *Leonardo da Vinci: An Account of His Development as an Artist*. Cambridge, UK: Cambridge University Press, 1952.

Corban, A. B. *The Medieval Universities*. London: Methuen and Co., 1975.

Crombie, A. C. *Robert Grosseteste and the Origin of Experimental Science 1100–1700*. Oxford, UK: Clarendon Press, 1953.

Crombie, A. C. *The History of Science from Augustine to Galileo*. New York: Dover Publications, 1995. (Originally published as *Augustine to Galileo: The History of Science A.D. 400–1650*. London: Falcon Press, 1952.)

Dante. *Inferno*. In *The Divine Comedy*, translated by John Ciardi. New York: W. W. Norton, 1970. (Originally published 1954.)

Dante. *Paradiso*. In *The Divine Comedy*, translated by John Ciardi. New York: W. W. Norton, 1970. (Originally published 1954.)

Döbler, H. F. *Kultur- und Sittengeschichte der Welt*, 10 Vols. Gütersloh: Bertelsmann Kunstverlag, 1971.

Dorfman, Y. G. *Vsemirnaya Istoriya Fiziki s Drevneishikh Vremen do Kontsa XVIII Veka* (A World History of Physics since the Most Ancient Times up to the End of the 18th Century). Moscow: Nauka, 1974.

Duby, G. *L'Europe au Moyen Age: Art Roman, Art Gothique*. Paris: Arts et Métiers Graphiques, 1979.

Dugas, René. *A History of Mechanics*. New York: Dover Publications, 1988.

Dürrenmatt, Friedrich. *Romulus the Great*. In *Four Plays: 1957–62,* translated by Gerhard Nellhaus. London: Jonathan Cape, 1964. (Originally published 1958.)

Einhard and Notker the Stammerer. *Two Lives of Charlemagne*, translated with an introduction by Lewis Thorpe. London: Penguin Books, 1969.

Engels, Frederick. *The Origin of the Family, Private Property, and the State in Light of the Researches of Lewis H. Morgan*. New York: International Publishers, 1970. (Originally published 1942.)

Farrington, Benjamin. *Greek Science: Its Meaning for Us*, 2 Vols. (Vol. I "Thales to Aristotle" and Vol. II "Theophrastus to Galen"). Harmondsworth: Penguin Books, 1949.

Galileo Galilei. *Discoveries and Opinions of Galileo*, translated with an introduction and notes by Stillman Drake. New York: Doubleday Anchor, 1957.

Garfagnini, G. C. *Cosmologie medievali*. Torino: Loescher Editore, 1978.

Gibbs-Smith, C. *Die Erfindungen des Leonardo da Vinci*. Stuttgart: Belser, 1978.

Grant, E. *Physical Science in the Middle Ages*. New York: John Wiley, 1979.

Grant, E. (ed.). *A Source Book in Medieval Science*. Cambridge, MA:Harvard University Press, 1974.

Heidelberger, M., and Thressen, S. *Natur und Erfahrung: Von der mittelalterlichen zur neuzeitlichen Naturwissenschaft*. Reinbek bei Hamburg: Deutsches Museum, Rowohlt Verlag, 1981.

Khayyám, Omar. *The Ruba'iyat of Omar Khayyám*, translated by Peter Avery & John Heath-Stubbs. New York: Penguin Books, 1979.

Klemm, L. *Der Beitrag des Mittelalters zur Entwicklung der abendländischen Technik*. Wiesbaden: Steiner Verlag, 1961.

Koyré, Alexandre. *From the Closed World to the Infinite Universe*. Baltimore: Johns Hopkins University Press, 1957.

Lindberg, D. C. (ed.). *Science in the Middle Ages*. Chicago: University of Chicago Press, 1978.

Maimonides. *The Guide of the Perplexed*, translated by Shlomo Pines. Chicago: University of Chicago Press, 1974.

Mason, S. F. *A History of the Sciences*. London: Routledge and Kegan Paul, 1953.

Maurer, A. A. *Medieval Philosophy*. New York: Random House, 1962.

Nasr, S. H. *Islamic Science: An Illustrated Study*. London: World of Islam Festival Publishing Company Ltd., 1976.

Nemorarius, Jordanus. *Liber de Ratione Ponderis,* translated by Ernest A. Moody. In *The Medieval Science of Weights*, edited by Ernest A. Moody and Marshall Clagett. Madison: University of Wisconsin Press, 1952.

Newman, J.R. (ed.). *The World of Mathematics,* 4 Vols. New York: Simon and Schuster, 1956.

Nicholas of Cusa. *Idiota de Staticis Experimentis.* In *Nicholas of Cusa on Wisdom and Knowledge*, translated by Jasper Hopkins. Minneapolis: The Arthur J. Banning Press, 1996.

Oresme, Nicole. *Le Livre du ciel et du monde*, edited and translated by A. D. Menut and A. J. Denomy. Madison: University of Wisconsin Press, 1968.

Oresme, Nicole. "The Configuration of Qualities and Motions, Including a Geometric Proof of the Mean Speed Theorem," translated by Marshall Clagett. In *A Source Book in Medieval Science*, edited by Edward Grant, pp. 243–252. Cambridge, MA: Harvard University Press, 1974.

Peter of les Vaux-de-Cernay. *The History of the Albigensian Crusade*, translated by W. A. and M. D. Sibly. Woodbridge, UK: The Boydell Press, 1998.

Piltz, A. *The World of Medieval Learning*. Totowa, NJ: Barnes and Noble Books, 1981.

Randles W. G. L. *De la terre plate au globe terrestre: Une mutation épistémologique rapide (1480–1520).* Paris: Libraire Armand Colin, 1980.

Saint Francis. "The Song of Brother Sun." In *St. Francis of Assisi: Writings and Early Biographies*, edited by Marion A. Habig, translated by O. F. M. Benen Fahy. Chicago: Franciscan Herald Press, 1973.

Santillana, Giorgio de. *The Origins of Scientific Thought: From Anaximander to Proclus, 600 BC to 300 AD*. Chicago: University of Chicago Press, 1961.

Sarton, George. *Introduction to the History of Science*, Vol. III. Baltimore: Williams & Wilkins Company, 1947.

Strabo. *Geography*, translated by Horace Leonard Jones. Cambridge, MA: Harvard University Press, 1960. (Originally published 1917.)

Swaan, W. *Die großen Kathedralen*. Verlag M. Du Mont Schauberg, Köln, 1969.

Truesdell, C. *Essays in the History of Mechanics*. New York: Springer-Verlag, 1968.

van der Waerden, B. L. *Science Awakening*, Vol. 1. Noordhoff, Groningen, 1954.

Varahamihira. *Alberuni's India*, translated by Edward C. Sachau. London: Kegan Paul, Trench, Trubner & Co. Ltd., 1910.

The Venerable Bede. *Bede's Ecclesiastical History of the English People*, edited by Bertram Colgrave and R. A. B. Mynors. Oxford, UK: Clarendon Press, 1969.

Villon, Francois. "Ballade des Dames du temps jadis." In *Complete Poems*, edited with English translation and commentary by Barbara N. Sargent-Baur, p. 77. Toronto: University of Toronto Press, 1994.

Villon, Francois. "Le Testament." In *Complete Poems*, edited with English translation and commentary by Barbara N. Sargent-Baur, p. 59. Toronto: University of Toronto Press, 1994.

Walsh, James J. *The Thirteenth: Greatest of Centuries*. New York: Catholic Summer School Press, 1912.

Weinberg, J.R. *Nicholas of Autrecourt.* New Haven, CT: Princeton University Press, 1948.

Weimar, P. (ed.). *Die Renaissance der Wissenschaften im 12. Jahrhundert*. Zürich: Artemis Verlag, 1981.

Chapter 3

D'Alembert, Jean le Rond. *Preliminary Discourse to the Encyclopedia of Diderot*, translated by Richard N. Schwab and Walter E. Rex. New York: The Bobbs-Merrill Company, 1963.

Bacon, Franics. "The New Organon." In *The Works of Francis Bacon*, Vol. VIII, collected and edited by James Spedding, Robert Leslie Ellis, and Douglas Denon Heath, pp. 59–350. Boston: Taggard and Thompson, 1863.

Bacon, Francis. *Novum Organum*, translated by William Wood. In *The Physical and Metaphysical Works of Lord Bacon, Including The Advancement of Learning and Nouvum Organum*, edited by Joseph Devey, pp. 380–567. London: G. Bell and Sons, 1904.

Bauer, G. *Giovannoi Battista Benedetti, Vordenker und Wegbereiter der galileischen Physik*. Frankfurt: Verlag Harri Deutsch, 1989.

Bernal, J. D. *Science in History*. London: C. A. Watts & Co., 1965. (Originally published 1954.)

Boas Hall, Marie. *The Scientific Renaissance 1450–1630: The Rise of Modern Science*. New York: Harper & Brothers, 1962.

Bruno, Giordano. "On the Infinite Universe and Worlds." In *Giordano Bruno: His Life and Thought*, edited and translated by Dorothy Waley Singer. New York: Henry Schuman, 1950.

Butterfield, H. *The Origins of Modern Science 1300–1800*. London: G. Bell, 1957.

Caspar, M. and Dyck, W. von. *Johannes Kepler in seinen Briefen*. Munich: Oldenbourg Verlag, 1930.

Christianson, J. R. "Tycho Brahe's German Treatise on the Comet of 1577: A Study in Science and Politics." *Isis* 70:1 (1979), pp. 110–140.

Cohen, I. B. *The Birth of a New Physics*. London: Heinemann Press, 1960.

Cohen, I. B. *The Newtonian Revolution*. Cambridge, UK: Cambridge University Press, 1980.

Copernicus. "Commentariolus." In *Three Copernican Treatises*, translated by Edward Rosen, pp. 55–90. New York: Octagon Books, 1971.

Copernicus. *On the Revolutions of the Heavenly Spheres*, translated by A. M. Duncan. New York: Barnes and Noble Books, 1976.

Deely, John. *Four Ages of Understanding: The First Postmodern Survey of Philosophy from Ancient Times to the Turn of the Twenty-First Century*. Toronto: University of Toronto Press, 2001.

Descartes, René. *Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences*, translated by Donald A. Cress. Indianapolis: Hackett, 1980.

Descartes, René. *Principles of Philosophy*, translated by Valentine Rodger Miller and Reese P. Miller. London: D. Reidel Publishing Company, 1983.

Dijksterhuis, E. J. *The Mechanization of the World Picture*, translated by C. Dikshoorn. New York: Oxford University Press, 1969. (Translation originally published 1961; published in German as *Die Mechanisierung des Weltbildes*. Springer-Verlag, Berlin/Göttingen/Heidelberg.)

Dugas, René. *A History of Mechanics*, translated by J. R. Maddox. London: Routledge and Kegan Paul, 1957.

Einstein, Albert. Foreword to *Dialogue Concerning the Two Chief World Systems* by Galileo Galilei and translated by Stillman Drake. Berkeley: University of California Press, 1953.

Einstein, Albert. "Johannes Kepler." *Frankfurter Zeitung*, November 9, 1930. Reprinted in *Ideas and Opinions*, edited by Cal Seelig and translated by Sonja Bargmann, pp. 262–266. New York: Wings Books, 1954.

Galileo Galilei, *Dialogues Concerning Two New Sciences*, translated by Henry Crew and Anfonso de Salvio. New York: The Macmillan Company, 1914.

Galileo Galilei. *Dialogue Concerning the Two Chief World Systems—Ptolemaic & Copernican,* translated by Stillman Drake. Berkeley: University of California Press, 1953.

Galileo Galilei. "The Assayer." In *Discoveries and Opinions of Galileo*, translated by Stillman Drake, pp. 229–280. New York: Anchor Books, 1957.

Galileo Galilei. "Letter to the Grand Duchess Christina." *In Discoveries and Opinions of Galileo*, translated by Stillman Drake, pp. 173–216. New York: Anchor Books, 1957.

Gilbert (Gilberd), William. *On the Magnet* (*De Magnete*, 1601). New York: Basic Books, 1958.

Hall, A. R. *Die Geburt der naturwissenschaftlichen Methode*. Gütersloh: Bertelsmann Verlag, 1965.

Herivel, J. *The Background to Newton's Principia*. Oxford, UK: Clarendon Press, 1965.

Hund, F. *Geschichte der physikalischen Begriffe*. Mannheim: Bibliographisches Institut, 1972.

Huygens, Christiaan. *Oeuvres Complètes,* 22 Vols. The Hague: Martinus Nijhoff, 1880–1950.

Huygens, Christiaan. *Treatise on Light*, translated by Silvanus P. Thompson. New York: Dover Publications, 1962.

Kepler, Johannes. *Mysterium Cosmographicum: The Secret of the Universe,* translated by A. M. Duncan. New York: Abaris Books, 1981.

Kepler, Johannes. *New Astronomy*, translated by William H. Donahue. New York: Cambridge University Press, 1992.

Kepler, Johannes. *Epitome of Copernican Astronomy and Harmonies of the World*, translated by Charles Glenn Wallis. Amherst, NY: Prometheus Books, 1995.

Kepler, Johannes. *The Harmony of the World*, translated by E. J. Aiton, A. M. Duncan, and J. V. Field. Philadelphia: American Philosophical Society, 1997.

Koestler, A. *The Sleepwalker*. New York: MacMillan, 1959.

Koyré, A. *From the Closed World to the Infinite Universe*. Baltimore: The John Hopkins Press, 1957.

Mach, Ernst. *The Science of Mechanics: A Critical and Historical Account of Its Development*. La Salle, IL: The Open Court Publishing Co., 1960. (Originally published 1893.)

Manuel, F. E. *A Portrait of Isaac Newton*. Cambridge, MA: The Belknap Press of Harvard University Press, 1968.

Newton, Isaac. *Principia*, translated by Andrew Motte, revised by Florian Cajori. Cambridge, UK: Cambridge University Press, 1934.

Norman, Robert. *The Newe Attractive, Shewing the Nature, Propertie, and Manifold Vertues of the Loadstone*, reprint editon. London, 1720. (Originally published 1581.)

Pascal, Blaise. "Preface to the Treatise on Vacuum." In *The Thoughts, Letters, and Opuscules of Blaise Pascal*, translated by Orlando Williams Wight, pp. 543–552. New York: Hurd and Houghton, 1864.

Pemberton, Henry. *A View of Sir Isaac Newton's Philosophy*. London: S. Palmer, 1728.

Rheticus. "Narratio Prima." In *Three Copernican Treatises*, translated by Edward Rosen, pp. 107–196. New York: Octagon Books, 1971.

Santillana, Giorgio de. *The Crime of Galileo*. Chicago: University of Chicago Press, 1955.

Scott, J. F. *The Scientific Work of R. Descartes*. London: Taylor & Francis, 1959.

Stukeley, William. *Memoirs of Sir Isaac Newton's Life,* edited by A. Hastings White. London: Taylor and Francis, 1936.

Struik, D. J. *The Land of Stevin and Huygens*. Dordrecht: D. Reidel Publishing Company, 1981.

Voltaire. "Letters on England" ("Lettres Anglaises") XIV. In *French and English Philosophers*, Harvard Classics, edited by Charles W. Eliot, pp. 110–115. New York: P. F. Collier and Son, 1910.

Westfall, Richard S. *Never at Rest: A Biography of Isaac Newton*. Cambridge, UK: Cambridge University Press, 1980.

Wolf, A. A History of Science, Technology and Philosophy in the 16th and 17th Century. London: Allen and Unwin, 1962.

Zinner, E. "Entstehung und Ausbreitung der Copernicanischen Lehre." Given at the 200th anniversary celebration of the Friedrich–Alexander University in Erlangen, Germany, 1943.

Chapter 4

D'Alembert, Jean le Rond. *Preliminary Discourse to the Encyclopedia of Diderot*, translated by Richard N. Schwab and Walter E. Rex. New York: The Bobbs-Merrill Company, 1963.

Anderson, David L. *The Discovery of the Electron*. Princeton, NJ: Van Nostrand, 1964.

Black, Joseph. *Lectures on the Elements of Chemistry*, edited by John Robinson, Vol. I. Philadelphia: Mathew Carey, 1807.

Brush, S. G. *Kinetic Theory*, Vol. 1 "The Nature of Gases and Heat." Oxford, UK: Pergamon Press, 1965.

Brush, S. G. *Kinetic Theory*, Vol. 2 "Irreversible Processes." Oxford, UK: Pergamon Press, 1966.

Brush, S. G. *The Temperature of History: Phases of Science and Culture in the 19th Century*. New York: Franklin, 1978.

Carnot, Sadi. *Reflexions on the Motive Power of Fire: A Critical Edition with the Surviving Manuscripts*, translated and edited by Robert Fox. New York: Lilian Barber Press, 1986.

Cohen, I. B. Franklin and Newton: *An Inquiry into Speculative Newtonian Experimental Science and Franklin's Work in Electricity as an Example thereof*. Chicago: University of Chicago Press, 1956.

Dugas, René. *A History of Mechanics*, translated by J. R. Maddox. London: Routledge and Kegan Paul, 1957.

Einstein, Albert. *Ideas and Opinions*, translated by Sonja Bargmann. New York: Wings Books, 1954.

Elliott, R. S. *Electromagnetics*. New York: McGraw-Hill, 1966.

Euler, Leonhard. *Letters of Euler on Different Subjects in Natural Philosophy Addressed to a German Princess*, 2 Vols., translated by David Brewster. New York: J. J. Harper, 1833.

Fontenelle, Bernard le Bovier de. *Digression sur les Anciens et les Modernes*. In *Oeuvres diverses*, Vol. III. The Hague: Chez Antoine van Dole, 1736.

Frankel, C. *The Faith of Reason: The Idea of Progress in the French Enlightenment*. New York: King's Crown Press, 1948.

Friedman, Francis L. and Sartori, Leo. *The Classical Atom*. New York: Addison-Wesley, 1965.

Gilbert, William. *On the Magnet*, edited by Derek H. Price. New York: Basic Books, 1958.

Glasstone, Samuel. *Sourcebook on Atomic Energy*, Second Edition. Princeton, NJ: Van Nostrand, 1958.

Goethe, Johann Wolfgang von. *Die Schriften zur Naturwissenschaft*, Vol. 6 "Zur Farbenlehre," edited by D. Kuhn and K. L. Wolff. Weimar: Verlag Hermann Böhlaus Nachf, 1959.

Hall, A. R. *From Galileo to Newton 1630–1720*. Edinburgh: R & R Clark, 1963.

Hall, A. R. *Philosopher at War: The Quarrel between Newton and Leibniz*. Cambridge, UK: Cambridge University Press, 1980.

Harman, P. M. *Energy, Force, Matter: The Conceptual Development of Nineteenth Century Physics.* Cambridge, UK: Cambridge University Press, 1982.

The Harvard Project Physics Course. New York: Holt, Rinehart and Winston, 1970.

Heilbron, J. L. *Electricity in the 17th and 18th Centuries: A Study of Early Modern Physics*. Berkeley: University of California Press, 1979.

Hertz, Heinrich. *The Principles of Mechanics*, translated by D. E. Jones and J. T. Walley. New York: Macmillan and Co., 1899.

Horváth, J. B. *Physica particularis*. Tyrnavia: M. Rieger, 1770.

Hume, David. *A Treatise of Human Nature.* New York: Oxford University Press, 2000.

Huygens, Christiaan. *Treatise on Light*, translated by Silvanus P. Thompson. New York: Dover Publications, 1962.

Kant, Immanuel. *Critique of Pure Reason*, translated by Norman Kemp Smith. London, Macmillan, 1929.

Laforgue, Jules. "Funeral March for the Death of the Earth." In *Poems of J. Laforgue*, translated by Patricia Terry, p. 19. Berkeley: University of California Press, 1958. (Originally published 1881.)

Laplace, Pierre Simon. *A Philosophical Essay on Probabilities*, translated by Frederick Wilson Truscott and Frederick Lincoln Emory. New York: Dover Publications, 1951.

Leibniz, G. W. *Philosophical Essays*, translated by Roger Ariew and Daniel Garber. Indianapolis: Hackett, 1989.

Leibniz, G. W. "Postscript of a Letter to Basnage de Beauval (1696)." In *Philosophical Essays*, translated by Roger Ariew and Daniel Garber, pp. 147–148. Indianapolis: Hackett, 1989.

Locke, John. *An Essay Concerning Human Understanding*, edited by A. D. Woozley. New York: New American Library, 1974.

Lucretius. *On the Nature of Things*, translated by W. H. D. Rouse, revised by Martin Ferguson Smith. Cambridge, MA: Harvard University Press, 1992.

Magie, William Francis. *A Source Book in Physics*. Cambridge, MA: Harvard University Press, 1963. (Originally published 1935.)

Mason, S. F. *A History of the Sciences*. London: Routledge and Kegan Paul, 1953.

Maxwell, James Clerk. *A Treatise on Electricity and Magnetism*, Third Edition, 2 Vols. Oxford, UK: Clarendon Press, 1892.

Maxwell, James Clerk. *A Dynamical Theory of the Electrical Field*. Edinburgh: Scottish Academic Press, 1982.

Mayer, Robert. "Bemerkungen über die Kräfte der unbelebten Natur." *Annalen der Chemie und Pharmacie* 42 (1842), pp. 233–240.

Mendeleev, Dmitri Ivanovich. *The Principles of Chemistry*, Part 4. New York: Collier, 1901.

Mertz, J. T. *A History of European Thought in the Nineteenth Century.* New York: Dover Publications, 1965.

Millikan, R. A., Roller, D., and Warson, E. C. *Mechanics, Molecular Physic, Heat and Sound*. Cambridge, MA: MIT Press, 1965.

Moseley, Henry G. J. "The High-Frequency Spectra of the Elements, Part I" *Philosophical Magazine* 26 (1913), pp. 1024–1034; and "… Part II." *Philosophical Magazine* 27 (1914), pp. 703–713. (Quoted in George Sarton, "Moseley: The Numbering of the Elements." *Isis* 9 (February 1927), pp. 96–111.)

Newton, Isaac. *Opticks*, New York: Dover Publications, 1952.

Poincaré, Henri. *Science and Hypothesis*, translated by W. J. G. [William John Greenstreet]. New York: The Walter Scott Publishing Co., 1905. (Available at http://www.archive.org/details/sciencehypothesi00poin.)

Priestley, J. B. *The History and Present State of Electricity*, 2 Vols. New York: Johnson Reprint Corporation, 1966. (Reprinted from the third edition, London, 1775; originally published 1767.)

Roller, D. *The Early Development of the Concepts of Temperature and Heat*. Cambridge, MA: Harvard University Press, 1950.

Solymar, Laszlo. *Getting the Message: A History of Communications*. Oxford, UK: Oxford University Press, 1999.

Spassky, B. I. История Физики (History of Physics), 2 Vols. Moscow: Moscow University Press, 1977.

Thomson, Benjamin (Count Rumford). "An Inquiry Concerning the Source of the Heat Which Is Excited by Friction." *Philosophical Transactions of the Royal Society of London* 88 (1798), pp. 80–102.

Thomson, Benjamin (Count Rumford). "An Inquiry Concerning the Weight Ascribed to Heat." In *The Early Development of the Concepts of Temperature and Heat: The Rise and Decline of the Caloric Theory*, prepared by Duane Roller, pp. 50–59. Cambridge, MA: Harvard University Press, 1966. (Originally published 1799.)

Thomson, J. J. "On the Velocity of the Cathode-Rays." *Philosophical Magazine* 38 (1894), pp. 358–365.

Thomson, William (Lord Kelvin). *Mathematical and Physical Papers*, Vol. I. Cambridge, UK: Cambridge University Press, 1882.

Tricker, R. A. R. *Early Electrodynamics: The First Law of Circulation*. Oxford, UK: Pergamon Press, 1965.

Tricker, R. A. R. *The Contributions of Faraday and Maxwell to Electrical Science*. Oxford, UK: Pergamon Press, 1966.

Truesdell, C. *Essays in the History of Mechanics.* New York: Springer, 1968.

Volta, Alessandro. "On the Electricity Excited by the Mere Contact of Conducting Substances of Different Kinds: In a Letter from Mr. Alexander Volta, F. R. S. Professor of Natural Philosophy in the University of Pavia, to the Rt. Hon. Sir Joseph Banks, Bart. K. B. P. R. S." *Philosophical Transactions of the Royal Society of London* 90 (1800), pp. 403–431.

Voltaire. "Micromegas." In *The Portable Voltaire*, edited and translated by Ben Ray Redman, pp. 413–435. New York: The Viking Press, 1968. (Originally published 1948.)

Waterston, J. J. "On the Physics of Media that Are Composed of Free and Perfectly Elastic Molecules in a State of Motion." *Philosophical Transactions of the Royal Society of London A* 183 (1892), pp. 1–79.

Whitaker, Sir Edmund. *A History of the Theories of Aether and Electricity*, Vol. I. New York: Thomas Nelson and Sons Ltd., 1958. (Originally published 1910.)

Whittaker, E. *A History of Aether and Electrictty*, Vol. I "The Classical Theories." London: Thomas Nelson and Sons Ltd., 1951.

Whittaker, E. *A History of Aether and Electricity*, Vol. II "The Modern Theories 1900–1926." London: Thomas Nelson and Sons Ltd., 1953.

Williams, L. Pearce. *The Origins of Field Theory.* New York: Random House, 1966.

Winkler, J. H. *Die Stärke der elektrischen Kraft des Wassers in gläsernen Gefäßen, welche durch den Musschenbroek'schen Versuch bekannt geworden*. Leipzig: Breitkopf, 1746.

Wisdom, J. O. *Philosophy and Its Place in Our Culture*. New York: Gordon and Breach, 1975.

Young, Thomas. "An Account of Some Cases of the Production of Colours." *Philosophical Transactions of the Royal Society of London* 92 (July 1802), pp. 387–397. (Quoted in [Elliot 1966, p. 11].)

Chapter 5

Aristotle. *On the Heavens*, translated by J. L. Stocks. In *The Complete Works of Aristotle*, Vol. I, edited by Jonathan Barnes. Princeton, NJ: Princeton University Press, 1984. (Also available in *The Internet Classics Archive*, edited by Daniel C. Stevenson, http://classics.mit.edu//Aristotle/heavens.html, 2009.)

Augustine. *Confessions*, translated by D. S. Pine-Coffin. London: Penguin Books, 1961.

Barrow, J. D. *The Origin of the Universe*. New York: Basic Books, 1994.

Becquerel, Henri. "Sur les radiations invisibles émises par les corps phosphorescents." *Comptes Rendus* 122 (1896), pp. 501–506.

Bennet, C.H., Brassard, G., Crepeau, C., Jozsa, R., Peres, A., and Wootlers, W. "Teleporting an Unknown Quantum State via Dual Classical and EPR Channels." *Phys. Rev. Lett.* 70 (1993), pp. 1895–1899.

Bohr, Niels. "On the Constitution of Atoms and Molecules." *Philosophical Magazine series 6* 26 (1913), pp. 1–25, 476–502, and 857–875. (In three parts.)

Bohr, Niels. *The Theory of Spectra and Atomic Constitution*, translated by A. D. Udden. Cambridge, UK: Cambridge University Press, 1924.

Bohr, Niels. *Atomphysik und menschliche Erkenntnis*. Braunschweig: Vieweg, 1958.

Bohr, Niels. *Essays 1958–1962 on Atomic Physics and Human Knowledge.* New York: Interscience Publishers, 1963.

Born, Max. "Physics and Relativity." In *Physics in My Generation: A Selection of Papers*, by Max Born, pp. 189–206. New York: Pergamon Press, 1956.

Born, Max. "Statistical Interpretation of Quantum Mechanics." In *Physics in My Generation: A Selection of Papers*, by Max Born, pp. 183–184. New York: Pergamon Press, 1956.

Born, Max, Born, Hedwig, and Einstein, Albert. *The Born-Einstein Letters*, translated by Irene Born. New York: Macmillan, 1971.

Boorse, H. A. and Motz, L. *The World of the Atom*. New York: Basic Books, 1966.

Broglie, L. de. *Continu at discontinu en Physique moderne*. Paris: Albin Michel, 1941.

Brooks, M. (ed.). *Quantum Computing and Communications*. New York: Springer, 1999.

Bunge, M. and Shea, W. R. (eds.). *Rutherford and Physics at the Turn of the Century*. New York: Watson, 1979.

Chadwick, Sir James. "Some Personal Notes on the Search for the Neutron." In *Proceedings of the Tenth International Congress of the History of Science* (Ithica, New York, 1962), pp. 159–162. Paris: Hermann, 1964.

Clifford, W. K. "On The Space Theory of Matter." In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Vol. I, edited by William Ewald, pp. 523–524. Oxford, UK: Clarendon Press, 1996.

Condon, Edward U. "60 Years of Quantum Physics." *Physics Today* 15 (October 10, 1962), pp. 37–49. (Quoted in [Jammer 1966, footnote 37, pp. 207–208].)

Csikai, Gyula and Szalay, Sándor. *Proceedings of the International Congress on Nuclear Physics in Paris, 1958*. Paris: Publications Dunod, 1959.

Curie, Ève. *Madame Curie*, translated by Vincent Sheean. Toronto: Doubleday, Doran and Company, 1937.

Curie, P. and Curie, Mme. S. "Sur les corps radio-actifs." *Comptes rendus de l'Académie des Sciences* 134 (1902), pp. 85–87.

Curie, P. and Laborde, A. "Sur la chaleur dégagée spontanément par les sels de radium. " *Comptes Rendus de l'Académie des Sciences* 136 (1903), pp. 673–675.

Davies, P. *Superforce: The Search for a Grand Unification Theory of Nature*. New York: Simon and Schuster, 1985.

Davies, P. (ed.). *The New Physics*. Cambridge, UK: Cambridge University Press, 1989.

Davies, P. C. W. and Brown, Julian (eds.). *Superstrings: A Theory of Everything?* Cambridge, UK: Cambridge University Press, 1988.

DiVincenzo, D. and Terhal, B. "Decoherence: The Obstacle to Quantum Computation." *Physics World:* "Quantum Information" special issue (March 1998), pp. 53–58.

Duhem, P. *Traité d'énergétique ou de thermodynamique générale*, Vol. 1. Paris: Gauthier-Villars, 1911.

Dummer, G. W. A. *Electronic Inventions and Discoveries*. London: Pergamon Press, 1978.

Dürrenmatt, Friedrich. *The Physicists*, translated by James Kirkup. New York: Grove Press, 1964. (Originally published in German in 1952.)

Eddington, Sir Arthur. *Space, Time and Gravitation: An Outoline of the General Relativity Theory*. Cambridge, UK: Cambridge University Press, 1966. (Originally published 1920.)

Elliott, Robert S. *Electromagnetics: History, Theory, and Applications*. New York: McGraw-Hill, 1966.

Einstein, Albert. "Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" (Does the Inertia of a Body Depend on Its Energy Content?) *Annalen der Physik* 18 (1905), 639–641.

Einstein, Albert. "Das Prinzip von der Erhaltung der Schwerpunktsbewegung und die Trägheit der Energie" (The Principle of Conservation of Motion of the Center of Gravity and the Inertia of Energy). *Annalen der Physik* 20 (1906), 627–633.

Einstein, Albert. "The Foundation of the General Theory of Relativity," translated by Alfred Engel. In *The Collected Papers of Albert Einstein*, Vol. 6, edited by A. J. Kox, Martin J. Klein, and Robert Schulmann. Princeton, NJ: Princeton University Press, 1997. (Originally published in *Annalen der Physik* in 1916.)

Einstein, Albert. *Special and General Relativity*, translated by Robert W. Lawson. London: Methuen & Co. Ltd., 1920. (Originally published in German in 1919.)

Einstein, Albert. "Autobiographical Notes," translated by Paul Arthur Schilpp. In *Albert Einstein: Philosopher–Scientist*, by Paul Arthur Schilpp, pp. 1–95. Evanston, IN: The Library of Living Philosophers, Inc., 1949.

Einstein, Albert. *Ideas and Opinions*, translated by Sonja Bargmann. New York: Wings Books, 1954.

Einstein, Albert. "Letter to Max Born, November 7, 1944." In *Physics in My Generation: A Selection of Papers*, by Max Born, p. 90. New York: Pergamon Press, 1956.

Einstein, Albert. *The World As I See It.* London: John Lane The Bodley Head Ltd., 1935.

Einstein, Albert and Haas, W. J. de. "Experimenteller Nachweis der Ampereschen Molekularströme" (Experimental Proof of Ampère's Molecular Currents). *Deutsche Physikalische Gesellschaft Verhandlungen* 17 (1915), pp. 152–170.

Einstein, Albert and Infeld, Leopold. *The Evolution of Physics*. Cambridge, UK: Cambridge University Press, 1938.

Feinberg, G. *What Is the World Made Of? Atoms, Leptons, Quarks and Other Tantalizing Particles*. New York: Anchor Books, 1978.

Feyerabend, P. *Science in a Free Society*. London: NLB, 1973.

Feyerabend, Paul. *Against Method*: *Outline of an Anarchistic Theory of Knowledge*. London: Verso, 1976. (Originally published 1975.)

Feynman, R. P. *QED: The Strange Theory of Light and Matter*. Princeton, NJ: Princeton University Press, 1988.

Fischer, D. and Duerbeck, H. *HUBBLE: A New Window to the Universe*. New York: Copernicus, 1995.

Frisch, Otto. *What Little I Remember*. Cambridge, UK: Cambridge University Press, 1979.

Gale, G. "Science and the Philosophers." *Nature* 312:5994 (1984), pp. 491–495.

Galshow, S. L. *Interactions: A Journey Through the Mind of a Particle Physicist and the Matter of This World.* New York: A Warner Communications Company, 1988.

Gamow, George. *Biography of Physics*. New York: Hutchinson, 1962.

Gamow, George. *Thirty Years That Shook Physics: The Story of Quantum Theory.* New York: Dover Publications, 1985.

Gardner, Martin. *The Ambidextrous Universe.* New York: Basic Books, 1964.

Gell-Mann, M. and Ne'eman, Y. *The Eightfold Way*. New York: W. A. Benjamin, 1964.

Glashow, Sheldon. *Interactions: A Journey through the Mind of a Particle Physicist and the Matter of This World*. New York: Warner Books, 1988.

Gleick, J. *Chaos: Making a New Science*. New York: Penguin Books, 1988.

Graetzer, H. G. and Anderson, D. L. *The Discovery of Nuclear Fission: A Documentary History*. New York: Van Nostrand Reinhold Co., 1971.

Guth, A. H. *The Inflatory Universe*. Reading, MA: Perseus Books, 1997.

Hahn, O. and Strassmann, F. "Über den Nachweis und das Verhalten der bei der Bestrahlung des Urans mittels Neutronen entstehenden Erdalkalimetalle" (On the Detection and Characteristics of the Alkaline Earth Metals Formed by Irradiation of Uranium with Neutrons). *Naturwissenschaften* 27 (1939), pp. 11–15.

Harding, S. *The Science Question in Feminism*. Ithaca, NY: Cornell University Press, 1986.

The Harvard Project Physics Course. New York: Holt, Rinehart and Winston, 1970.

Harwitt, M. *Cosmic Discovery: The Search, Scope and Heritage of Astronomy.* Brighton: The Harvester Press, 1981.

Heisenberg, Werner. "Quantum Theoretical Re-interpretation of Kinematic and Mechanical Relations," translated by B. L. van der Waerden. In *Sources of Quantum Mechanics*, edited by B. L. van der Waerden, pp. 261–276. New York: Dover Publications, 1968. (Originally published in Zeitschrift für Physik 33 (1925), pp. 879–893.)

Heisenberg, Werner. *The Physicist's Conception of Nature,* translated by Arnold J. Pomerans. Westport: Greenwood Press, 1970. (Translation originally published 1958; published in German as *Das Naturbild der heutigen Physik*. Reinbek bei Hamburg: Rowohlt, 1955.)

Heisenberg, Werner. *Physics and Philosophy: The Revolution in Modern Science*. Amherst, NY: Prometheus Books, 1999. (Originally published 1958.)

Heisenberg, Werner. Interview by John L. Heilbron and Thomas S. Kuhn for the Oral History project of the Archive for the History of Quantum Physics, February 15, 1963.

Heisenberg, Werner. *Physics and Beyond: Encounters and Conversations*, translated by Arnold J. Pomerans. New York: Harper & Row, 1971.

Heller, K. D. *Ernst Mach: Wegbereiter der modernen Physik*. Wien: Springer-Verlag, 1964.

Helmholtz, Hermann von. *Treatise on Physiological Optics*, 3 Vols., translated and edited by James P. C. Southall. Menasha, WI: George Banta Publishing Company, for The Optical Society of America, 1924.

Helmholtz, Hermann von. *Philosophische Vorträge und Aufsätze.* Berlin: Academie-Verl, 1971. (Originally published 1855, 1862.)

Hey, A.J.G. (ed.). *Feynman and Computation: Exploring the Limits of Computers.* Boulder, CO: Westview Press, 1999.

Hess, Viktor F. "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten" (Observations of Penetrating Radiation in Seven Free Balloon Flights). *Physikalische Zeitschrift* XIII (1912), pp. 1084–1091.

Hogan, C.J. *The Little Book of the Big Bang*. Copernicus Book. New York: Springer, 1998.

Holton, G. *Thematic Origins of Scientific Thought: Kepler to Einstein*. Cambridge, MA: Harvard University Press, 1973.

Holton, G. (ed.). *The Twentieth-Century Sciences*. New York: W. W. Norton Company Inc., 1972.

Holton, G. *Thematic Origins of Scientific Thought: Kepler to Einstein*, revised edition. Cambridge, MA: Harvard University Press, 1988. (Originally published 1973.)

Holton, G. *Science and Anti-Science*. Cambridge, MA: Harvard University Press, 1997.

Holton, G. and Blanpied, W. (ed.). *Science and Its Public: The Changing Relationship*. Dordrecht: Reidel, 1976.

Hooft, Gerard 't. *In Search of the Ultimate Building Blocks*. Cambridge, UK: Cambridge University Press, 1997.

Hume, David. *A Treatise of Human Nature.* New York: Oxford University Press, 2000. (Originally published 1882.)

Hund, F. *The History of Quantum Theory*, English translation. Edinburugh: Harrap, 1974. (Originally published in German as *Geschichte der Quantentheorie* in 1967.)

"The Hymn to the Sun from El-Amarnah," translated (into German) by Adolf Erman, then translated into English by Aylward M. Blackman. In *The Literature of the Ancient Egyptians: Poems, Narratives, and Manuals of Instruction, from the Third and Second Millennia* B.C., pp. 289–290. London: Methuen & Co., 1927.

Infeld, Leopold. *Quest: The Evolution of a Scientist*. New York: Doubleday, Doran & Co., 1941.

Jammer, Max. *Concepts of Mass in Classical and Modern Physics*. Cambridge, MA: Harvard University Press, 1961.

Jammer, Max. *The Conceptual Development of Quantum Mechanics*. New York: McGraw Hill, 1966.

Jammer, Max. *The Philosophy of Quantum Mechanics in Historical Perspective*. New York: John Wiley, 1974.

Jaspers, Karl. *Über Bedingungen und Möglichkeiten eines neuen Humanismus*. Stuttgart: Reclam, 1962.

Kant, Immanuel. *Critique of Practical Reason*, translated by H. W. Cassirer. Milwaukee: Marquette University Press, 1998.

Kirchhoff, G. *Vorlesungen über mathematische Physik*. Leipzig: Druck and Verlag von B. G. Teubner, 1876.

Lao Tzu. *The Tao Te Ching*, translated by Ralph Alan Dale. London: Watkins Publishing, 2002.

Leibniz, G. W. "Letter to Queen Sophie Charlotte of Prussia, on What Is Independent of Sense and Matter." In *Philosophical Essays*, translated by Roger Ariew and Daniel Garber, pp. 186–192. Indianapolis: Hackett, 1989.

Lorentz, H. A. "Electromagnetic Phenomena in a System Moving with Any Velocity Smaller than That of Light." *Proceedings of the Academy of Science, Amsterdam* 6 (1904), 809–831. (Available in *The Principle of Relativity: A Collection of Original Papers on the Special and General Theory of Relativity,* by Lorentz, H. A., Einstein, A., Minkowski, H. and H. Weyl and translated by Perrett and Jeffery, pp. 9–34. London: Methuen, 1923.)

Lorentz, H. A., Einstein, A., and Minkowski, H. *Das Relativitätsprinzip*. Leipzig: B. G. Teubner Verlag, 1913.

Mach, Ernst. *The Science of Mechanics: A Critical and Historical Account of Its Development*, translated by Thomas J. McCormack. La Salle, IL: The Open Court Publishing Co., 1960. (Originally published 1893.)

Maimonides. *The Guide of the Perplexed*, translated by Shlomo Pines. Chicago: University of Chicago Press, 1963.

Mann, Thomas. *The Magic Mountain*, translated by John E. Woods. New York: Vintage, 1995.

Mason, S. F. *A History of the Sciences*. London: Routledge and Kegan Paul, 1953.

Mehra, J. *Einstein, Hilbert and the Theory of Gravitation*. Dordrecht: D. Reidel Publishing Co., 1974.

Mehra, J. and Rechenberg, H. *The Historical Development of Quantum Theory*, 4 Vols. New York: Springer, 1982.

Mehra, Jagdish. *The Physicist's Conception of Nature*. Dordrecht: Kluwer Academic Publishers, 1973.

Meyers, R. A. (ed.). *Encyclopedia of Modern Physics*. New York: Academic Press, 1990.

Minkowski, H. "Space and Time." Address delivered at the 80th Assembly of German Natural Scientists and Physicians, Cologne, Germany, September 21, 1908. (Available in *The Principle of Relativity: A Collection of Original Papers on the Special and General Theory of Relativity,* by Lorentz, H. A., Einstein, A., Minkowski, H. and H. Weyl and translated by Perrett and Jeffery, pp. 73–96. London: Methuen, 1923.)

Neumann, John von. *The Mathematical Foundations of Quantum Mechanics*, translated from the German edition by Robert T. Beyer. Princeton, NJ: Princeton University Press, 1983. (Translation originally published 1955; published in German as *Mathematische Grundlagen der Quantenmechanik*. Berlin: Verlag Julius Springer, 1932.)

Newton, Isaac. *Opticks*, New York: Dover Publications, 1952.

Nobel Lectures. *Physics*, 4 Vols. Amsterdam: Elsevier Publishing Co., 1972.

Noddack, Ida. "Über das Element 93" (On the Element 93). *Angewandte Chemie* 47:37 (1934), pp. 653–655.

North, J. *History of Astronomy and Cosmology*. New York: Norton, 1994.

Ostwald, W. *Die Energie*. Leipzig: J. A. Barth, 1908.

Pais, Abraham. *Subtle Is the Lord: The Science and Life of Albert Einstein*. Oxford, UK: Oxford University Press, 2005.

Pauli, Wolfgang. *Theory of Relativity*, translated by G. Field. New York: Dover Publications, 1981. New York:. (Translation originally published by Pergamon Press in 1958; published in German as "Relativitätstheorie." In *Enzyklopädie der mathematischen Wissenschaften*, Vol. V19. B. G. Teubner Verlag, Leipzig, 1921.)

Pickering, A. *Constructing Quarks: A Sociological History of Particle Physics.* Edinburgh: Edinburgh University Press, 1984.

Pascal, Blaise. *Pensées,* edited and translated by Roger Ariew. Indianapolis: Hackett, 2005.

Perlmutter, S., Bahcall, N. A., Ostriker, J. P., and Steinhardt, P. J. "The Cosmic Triangle: Assessing the State of the Universe." *Science* 284 (1999), pp. 1481–1488.

Planck, Max. "Zur Theorie des Gesetzes der Energieverteilung in Normalspektrum" (On the Theory of the Law of Energy Distribution in Normal Spectrum). Presentation to the German Physical Society, Berlin, December 14, 1900.

Planck, Max. "Über das Gesetz der Energieverteilung im Normalspectrum" (On the Law of Energy Distribution in the Normal Spectrum). *Annalen der Physik* 4:9 (1901), pp. 564–566.

Planck, Max. *Vorlesungen über die Theorie der Wärmestrahlung* (Lectures on the Theory of Thermal Radiation, in German). Leipzig: Verlag von Johann Ambrosius Barth, 1906.

Planck, Max. *The Universe in the Light of Modern Physics*, translated by W. H. Johnston, London: G. Allen & Unwin, Ltd., 1931.

Planck, Max. *Scientific Autobiography and Other Papers*, translated by Frank Gaynor. New York: Philosophical Library, 1949.

Planck, Max. *Physikalische Abhandlungen und Vorträge*, 3 Vols. Braunschweig: Vieweg, 1958.

Poincaré, Henri. "L'état actuel et l'avenir de la physique mathématique." *Bulletin des sciences mathématiques* 28:2 (1904), pp. 302–324 .

Poincaré, Henri. *Science and Hypothesis*, translated by W. J. G. [William John Greenstreet]. New York: The Walter Scott Publishing Co., 1905. (Available at http://www.archive.org/details/sciencehypothesi00poin.)

Poincaré, Henri. *Science and Method*, translated by Francis Maitland. New York: Cosimo, Inc., 2007. (Originally published 1914.)

Popper, Karl. *Objective Knowledge.* Oxford, UK: Clarendon Press, 1979. (Originally published 1972.)

"Quantum Information." Special Issue of *Physics World* (with Introduction by A. Zeilinger), March 1998.

Rees, M. *Before the Beginning*. Reading, MA: Perseus Books, 1997.

Riemann, Georg Friedrich Bernhard. "On the Hypotheses Which Lie at the Foundations of Geometry." In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Vol. II, edited by William Ewald, pp. 652–661. Oxford, UK: Clarendon Press, 1996.

Romer, A. *The Discovery of Radioactivity*. New York: Dover Publication, 1964.

Ruelle, D. *Chance and Chaos*. Princeton, NJ: Princeton University Press, 1991.

Runes, Dagobert D. (ed.). *Twentieth Century Philosophy.* New York: Philosophical Library, Inc., 1947.

Russell, Bertrand. *Mysticism and Logic*. Nottingham: Spokesman Books, 2007.

Rutherford, E. and Soddy, F. "The Radioactivity of Thorium Compounds." *Journal of the Chemical Society* 81 (1902), pp. 837–860.

Ryder, L. *Elementary Particles and Symmetries*. New York: Gordon and Breach Science Publishers, 1975.

Schilpp, P. A. (ed.). *Albert Einstein: Philosopher-Scientist*, Vol. 1. New York: Harper and Brothers, 1959.

Schweber, Silvan S. *Einstein & Oppenheimer: The Meaning of Genius*. Cambridge, MA: Harvard University Press, 2008.

Seelig, C. *Albert Einstein: A Documentary Biography.* London: Staples, 1956.

Serge, Emilio. "Biographical Introduction." In *Enrico Fermi: Collected Papers*, edited by Emilio Serge, Vol. 1, pp. xvii–xlii. Chicago: University of Chicago Press, 1962.

Shea, William R. *Otto Hahn and the Rise of Nuclear Physics*. New York: Springer, 1983.

Silk, J. *The Big Bang: The Creation and Evolution of the Universe*. San Francisco: W. H. Freeman and Co., 1980.

Snow, C. P. "Einstein." In *Einstein: The First Hundred Years*, edited by Maurice Goldsmith, Alan Mackay, and James Woudhuysen, pp. 3–18. New York: Pergamon Press, 1980.

Solomey, N. *The Elusive Neutrino*. New York: Scientific American Library, 1997.

Sommerfeld, Arnold. "To Albert Einstein's Seventieth Birthday" In *Albert Einstein: Philosopher–Scientist*, by Paul Arthur Schilpp, pp. 97–106. Evanston, IN: The Library of Living Philosophers, Inc., 1949.

Steane, A. M. and Rieffel, E. G. "Beyond Bits: The Future of Quantum Information Processing." *Computer* 33:1 (2000), pp. 38–45.

Thomson, William (Lord Kelvin). "Nineteenth Century Clouds over the Dynamical Theory of Heat and Light." *Philosophical Magazine* 2 (1901), pp. 1–40. (Lecture delivered to the Royal Institution, April 27, 1900.)

United Nations. *Proceedings of the International Conference on the Peaceful Uses of Atomic Energy*, Vol. 4. Geneva: United Nations, 1955.

Vickers, Brian. *Analogy versus Identity in Occult and Scientific Mentalities in the Renaissance*. Cambridge, UK: Cambridge University Press, 1984.

Weinberg, Steven. *The First Three Minutes*. New York: Basic Books, 1977.

Weinberg, Steven. *Dreams of a Final Theory*. New York: Pantheon Books, 1992.

Weisskopf, Victor F. "My Life as a Physicist." In *Physics in the Twentieth Century: Selected Essays,* by Victor F. Weisskopf, pp. 1–21. Cambridge, MA: MIT Press, 1972.

Weizsäcker, C. F. von. *Die Einheit der Natur*. Munich: Hanser Verlag, 1971.

Weyl, Hermann. *Philosophy of Mathematics and Natural Science*, revised and augmented English edition based on a translation by Olaf Helmer. Princeton, NJ: Princeton University Press, 1949.

Whitehead, A. N. *Science and the Modern World.* Cambridge, UK: Cambridge University Press, 1953.

Whittaker, Edmund. *A History of the Theories of Aether and Electricity*, Vol. II "The Modern Theories 1900–1926." New York: Harper, 1960. (Originally published by Longmans, Green, and Co. in 1910.)

Williams, L. P. *Relativity Theory: Its Origins and Impact on Modern Thought.* New York: John Wiley and Sons, 1968.

Witten, Edward. "Magic, Mystery, and Matrix." *Notices of the American Mathematical Society* 45:9 (October 1998), pp. 1124–1129.

Witten, Edward. "Unraveling String Theory." *Nature* 438 (December 22/29, 2005), pp. 1085–1087.

Yang, C. N. *Elementary Particles*. Princeton, NJ: Princeton University Press, 1962.

Color Plates

Adams, Henry. "American Ideals." In *History of the United States of America during the Administrations of Thomas Jefferson*, edited by Earl N. Harbert, pp. 107–125. New York: Library of America, 1986.

Bibliotheca Corviniana (edited by Csapodi, C., Csapodiné Gárdonyi, K., and Szántó, T.). Ontario: Magyar Helikon, 1967.

Brinkmann, Donald. *Mensch und Technik: Grundzüge einer Philosophie der Technik*. Bern: A. Franke, 1946.

Copernicus. *On the Revolutions of the Heavenly Spheres*, translated by A. M. Duncan. New York: Barnes and Noble Books, 1976.

Descartes, René. "To Chanut, On Nicholas Cusa and the Infinite (June 6, 1647)." In *Philosophical Essays and Correspondence*, edited by Roger Ariew, pp. 277–280. Indianapolis: Hackett, 2000.

Donne, John. *Ignatius His Conclave.* Available at *Luminarium: Anthology of English Literature*, http://www.luminarium.org/sevenlit/donne/ignatius.htm, 1999. (Originally published 1611.)

Euripides. *The Suppliant Women*, translated by Frank William Jones. In *The Complete Greek Tragedies*, Vol. IV "Euripides," edited by David Grene and Richmond Lattimore. Chicago: University of Chicago Press, 1958.

Fisher, H. A. L. *A History of Europe*. London: Edward Arnold & Co., 1945. (Originally published 1936.)

Ford, Joseph. "What Is Chaos, that We Should Be Mindful of It?" In *The New Physics*, edited by Paul Davies, pp. 348–372. New York: Cambridge University Press, 1989.

Galileo. *Dialogue Concerning the Two Chief World Systems*, translated by Stillman Drake. Berkeley: University of California Press, 1953.

Hardy, Thomas. "At a Lunar Eclipse." Available at *The DayPoems Poetry Collection*, http://www.daypoems.net/plainpoems/1020.html, accessed 2011.

Heisenberg, August. *Staat und Gesellschaft des byzantinischen Reiches* (Government and Society of the Byzantine Empire), 2 Vols. Berlin: Leipzig, 1923.

Heisenberg, Werner. *Physics and Beyond: Encounters and Conversations*, translated by Arnold J. Pomerans. New York: Harper & Row, 1971.

Jammer, Max. *Concepts of Mass in Classical and Modern Physics.* New York: Harper Torchbooks, 1961.

Kapitsa, P. L. "Some Principles of the Creative Upbringing and Education of Today's Youth." In *Experiment, Theory, Practice: Articles and Addresses*, by P. L. Kapitsa, pp. 204–215. Dordrecht: D. Reidel Publishing Company, 1980. (Originally published as "Некоторые принципы творческого воспитания и образования современной молодежи" (in Russian). *Voprosy filosofii* 7 (1971), 17–24.)

Klemm, Friedrich. *A History of Western Technology*, translated by Dorothy Waley Singer. New York: Charles Scribner's Sons, 1959.

Koestler, Arthur. *The Sleepwalkers: A History of Man's Changing Vision of the Universe*. London: Penguin Books, 1990. (Originally published 1959.)

Leibniz, G. W. "From the Letters to Clarke (1715–16)." In *Readings in Modern Philosophy*, edited by Roger Ariew and Eric Watkins, pp. 304–313. Indianapolis: Hackett, 2000.

Molière. *The Learned Ladies*, translated by Richard Wilbur. New York: Harcourt Brace, 1978. (Originally published 1977.)

Mumford, Lewis. *The City in History: Its Origins, Its Transformations, and Its Prospects*. New York: Harcourt Brace, 1961.

Nasr, Seyyed Hossein. *Islamic Science: An Illustrated Study.* London: World of Islam Festival Publishing Company, 1976.

Newton, Isaac. *A Catalogue of the Portsmouth Collection of Books and Papers Written by or Belonging to Sir Isaac Newton.* Cambridge, UK: Cambridge University Press, 1888.

Newton, Isaac. *Principia*, translated by Andrew Motte, revised by Florian Cajori. Cambridge, UK: Cambridge University Press, 1934.

Newton, Isaac. "Scholium Generale." In *Principia*, second edition, translated by Andrew Motte. Cambridge, UK: Cambridge University Press, 1729. (Originally published 1713.)

Plato. *Phaedo*, translated by G.M.A. Grube. In *Plato: Complete Works*, edited by John M. Cooper and D. S. Hutchinson. Indianapolis: Hackett, 1997.

St. Benedict's Rule for Monasteries, translated by Leonard J. Doyle. Collegeville, MN: Liturgical Press, 1950.

Shockley, W. F. "Transistor Technology Evokes New Physics." Nobel lecture, December 11, 1956. (Available in *Nobel Lectures in Physics 1942–1962*, by the Nobel Foundation, pp. 344–376. Singapore: World Scientific, 1998.)

Sophocles. *Antigone.* Translated by David Grene. In *The Complete Greek Tragedies*, Vol. II "Sophocles," edited by David Grene and Richmond Lattimore. Chicago: University of Chicago Press, 1958.

Strabo. *Geographica,* translated by H. C. Hamilton. London: George Bell & Sons, 1903.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Name Index

Alternate names appear in parentheses; many names exist in a number of variants, all of which may not be included here. The boldface numbers indicate pages on which extensive information can be found, Roman numerals indicate the number of a color plate, and fl (floruit) indicates the period during which a person was most active.

A

Abbé, Ernst Karl (1840–1905), German physicist 401

Abelard, Peter (Pierre Abélard, Petrus Abaelardus) (1079–1142), French theologian, philosopher 147, 152, XII

Abraham, Max (1875–1922), German physicist 410

Adams, Henry Brooks (1838–1918), American historian XVIII

Adelard of Bath (fl. 1116–1142), English Scholastic philosopher, translator 133, 134

Aepinus, Franz Ulrich Maria Theodor (1724–1802), German physicist, active primarily in St. Petersburg 332, 336, 337

Aeschylus (525/524–456/455 bce), Greek tragedian 31, 47, 102, 436

Agricola, Georgius (Georg Bauer) (1494–1555), German physician, mineralogist, metallurgist 175, 332

Ahmed ibn Yusuf (835–912), Arab mathematician XIII

Ahmose (Ahmes) (fl. ca. 1700 bce), Egyptian scribe, mathematician (?), author of the Rhind Papyrus 34

Akbar the Great (Jalaluddin Muhammad, Jalāl ud-Dīn Muhammad Akbar) (1542–1605), Mughal emperor 171

Akhenaten (Amenhotep IV) (reigned ca. 1350–ca. 1335 bce), Egyptian pharaoh 34, 554

Al-Battānī (Albatenius, Abu Abdallah Muhammad ibn Jabir ibn Sinan ar-Raqqi al-Harrani as-Sabi al-Batani) (ca. 858–929), Islamic astronomer, mathematician 128, 177

Albert of Saxony (Albertus de Saxonia, Albertus Parvus) (ca. 1316–1390), German Scholastic philosopher, rector of the University of Paris, founder and first rector of the University of Vienna 159, IX, XIII

Alberti, Leon Battista (1404–1472), Italian architect, poet, mathematician 158, 161

Albertus Magnus (Saint Albert the Great) (ca. 1200–1280), German polymath ("Doctor Universalis") 9, 79, 111, 120, 124, 126, 128, 129, 153, 154, 157, 161, 168, 329, XII

Al-Biruni, Abu Rayhan Muhammad ibn Ahmad (973–1048), Islamic scholar of Chorasmian origin, mathematician, astronomer X

Alcuin of York (Alcuinus, Ealhwine) (ca. 735–804), Anglo-Saxon scholar, friend and teacher of Charlemagne 119, 120, 128, 131

Alembert, Jean Le Rond d' (1717–1783), French physicist, mathematician 3, 216, 301, 302, 306, 307, 308, 310, **311–14**, 321, 322–26, 341

Alexander of Aphrodisias (fl. ca. 200), Greek philosopher, commentator on Aristotle X

Alexander the Great (356–323 bce), Macedonian king 35, 46, 47, 78, 99, 100, 105, 242

Alexander of Hales (ca. 1183–1245), English Scholastic philosopher ("Doctor Irrefragibilis") 120, 126

Al-Farabi (Alpharabius, Abū Nasr Muhammad ibn al-Farakh al-Fārābi) (ca. 873–ca. 950), Turkish-born Islamic mathematician, philosopher 128, 134

Al-Farghani, Ahmad ibn Muhammad ibn Kathīr (Alfraganus) (d. after 861), Persian astronomer 128

Alfonso X the Wise (1221–1284), king of Castile and León, poet, scholar, ordered the preparation of the (astronomical) Alfonsine tables under the direction of Yehudah ben Moshe and Isaac ibn Sid 129

Alfred the Great (Ælfrēd) (ca. 849–899), king of Wessex, military commander, scholar, translator 118, 120

Alfvén, Hannes (1908–1995), Swedish physicist, Nobel Laureate (1970) 494

Algarotti, Count Francesco (1712–1764), Italian philosopher, art critic 274

Alhazen (ibn al-ḥasan) (965–ca. 1039), Islamic natural philosopher, mathematician, physicist, ophthalmologist 49, 139, 226

Al-Iskandari, Ahmad ibn Muhammad Ibn ʿAta Allah (d. 1309), third sheikh of the Shadhili Sufi order X

al-Kāshī (Ghiyāth al-Dīn Jamshīd ibn Masʿūd, Jamshīd Kāshānī) (ca. 1380–1429), Persian mathematician, astronomer 138, 139

al-Khazini, Abd al-Rahman (fl. 1115–1130), Islamic physicist, astronomer 138

al-Khwārizmī (Mohammad ebne Mūsā Khwārazmī) (ca. 780–ca. 850), Persian-Arab mathematician, astronomer 128, 133, 137

al-Kindī, Abū Yūsuf Yaʿqūb ibn Isḥāq (ca. 801–873), Arab polymath 128

Alvarez, Luis Walter (1911–1981), English-born American physicist, Nobel Laureate (1968) 494

Ambler, Ernest (b. 1923), English-born American physicist 542

Ampère, André Marie (1775–1836), French physicist, mathematician 332, 344, 345, 346–50, 352, 353, 357

Anaxagoras (ca. 500–428 bce), Greek natural philosopher 47, 48, 73, 151

Anaximander (ca. 610–ca 546 bce), Greek natural philosopher 47, 48, 60, 477, 546

Anaximenes of Miletus (ca. 585–ca. 525 bce), Greek natural philosopher 47, 48, 60

Andersen, Hans Christian (1805–1875), Danish writer 467

Anderson, Carl David (1905–1991), Swedish-born American physicist, Nobel Laureate (1936) 492, 508, 526, XXI

Anderson, Herbert Lawrence (1914–1988), American physicist 19

Anderson, Philip Warren (b. 1923), American physicist, Nobel Laureate (1977) 494, 581

Andronicus of Rhodes (fl. ca. 70 bce), Greek philosopher, editor of Aristotelian writings 78–79

Ångström, Anders Jonas (1814–1874), Swedish physicist, astronomer 399

Anna Yaroslavna (Anne of Kiev) (ca. 1030–1075), daughter of Yaroslav I, wife of Henry I of France 224

Anselm of Canterbury (1033–1109), Italian-born Scholastic philosopher, theologian, "Father of Scholasticism" 153, 217

Anthemius of Tralles (fl. 6th century bce), Byzantine mathematician, sculptor, architect 131

Antiphon (b. ca. 480 bce), Greek philosopher 96

Antonius, Marcus (Mark Antony) (ca. 82–30 bce), Roman statesman, military commander 105

Apáczai Csere, János (1625–1659), Hungarian pedagogue, encyclopedist 224

Apianus, Petrus (Peter Apian) (1495–1552), German mathematician, astronomer, cartographer 74, 167, 172

Apollodorus of Damascus (fl. 110), Roman engineer, architect 47

Apollonius of Perga (ca. 262–ca. 190 bce), Hellenistic mathematician, astronomer 47, 48, 78, 97, 99, 131, 135, 159, 267, 275

Appleton, Sir Victor Edward (1892–1965), English physicist, Nobel Laureate (1947) 392, 492

Arago, Dominique François Jean (1786–1853), French physicist, astronomer 344, 360, 363

Archelaus (fl. 5th century bce), Greek natural philosopher 56

Archimedes (ca. 287–ca. 212 bce), Greek mathematician, physicist of Syracuse 3, 12, 47, 48, 65, 73, 78, **79–89**, 90, 96, 98, 121, 131, 134, 135, 142, 148, 159, 160, 162, 163, 165, 218, 242, 249, 256, 275, 296, XIII

Archytas of Tarentum (428–347 bce), Greek mathematician, philosopher 47, 48, 83, 95, 97, 159

Aristarchus of Samos (310–ca. 230 bce), Hellenistic astronomer, mathematician 47, 48, 72, 73, 76, **90–92**, 93, 94, 165, 177, 182, 185, XIV

Aristophanes (ca. 446–ca. 386 bce), Greek comedist 47, 56, 57, 102

Aristotle (384–322 bce), polymath, founder (with Socrates and Plato) of the European philosophical tradition 10, 24, 25, 26, 29, 46, 47, 48, 61, 62–63, 63–64, 66, 70–71, 72, 74, **75–79**, 81, 90, 91, 97, 103, 111, 113, 127, 128, 129, 130, 131, 132, 133, 134, 135, 138, 139, 142, 143, 145, 146, 147, 148, 149, 151, 155, 158, 159, 160, 161, 162, 163, 166, 167, 172, 173, 174, 176, 194, 198, 199, 201, 209, 211, 214, 224, 238, 257, 272, 275, 297, 300, 304, 314, 329, 331, 479, 481, 540, 556, 558, 559, 560, 573, VI, IX, X, XIII, XV

Arnaldus de Villanova (Arnaud de Ville-Neuve) (1235–1311), French physician, alchemist of Catalan origin 332

Arrhenius, Svante August (1859–1927), Swedish physicist, chemist XIX

Arsonval Jacques Arsène d' (1851–1940), French physician, inventor XIX

Arsamides. *See* Archimedes

Ashtekar, Abhay (b. 1949), Indian-born American physicist 573

Ashurbanipal (Aššur-bāni-apl) (fl. ca. 650 bce), Assyrian king 34

Aspect, Alain (b. 1947), French physicist 485

Aston, Francis William (1877–1945), English physicist, chemist, Nobel Laureate (chemistry 1922) 392, 496, 506, 511

Atkinson, Robert d'Escourt (1898–1982), English physicist 551, 552

Attila (Attila the Hun) (406–453), king of the Huns, "the Scourge of God" 47, 284

Augustine of Hippo (Aurelius Augustinus Hipponensis) (354–430), Church father, philosopher, theologian 6, 47, **104–9**, 111, 153, 154, 217, 235, 559, 560, 573, 579, XI, XII

Augustus, Gaius Julius Caesar Octavianus (63 bce–14 ce), Roman emperor 47

Averroes (Ibn Rushd) (1126–1198), Islamic philosopher 127, 128, 129, 133, 134, 138, 139, 158, 168, IX

Avicenna (Abu Ali Sina Balkhi) (ca. 980–1037), Islamic polymath of Tajik origin 128, 133, 134, 136, 138, 151

Avogadro (Romano Amedeo Carlo Avogadro di Quaregna e di Cerreto) (1776–1856), Italian physicist 323, 373, 388

B

Babinet, Jacques (1794–1872), French physicist 447

Bach, Johann Sebastian (1685–1750), German composer 5, 16, 323

Back, Ernst (1881–1959), German physicist 491

Bacon, Francis (1561–1626), English philosopher, politician 10, 155, **213–16**, 218, 220, 226, 271, 275, 298, 312, 322, 324, 325, 375, X

Bacon, Roger (ca. 1214–1294), English philosopher, naturalist ("Doctor Mirabilis") 124, 126, 128, 144, 154, 155, 156, 210

Baliani, Giovanni Battista (1582–1666), Italian naturalist 213

Balmer, Johann Jakob (1825–1898), Swiss mathematician, physicist 399–400, 401

Balzac, Honoré de (1799–1850), French writer 323, XVIII

Barberini, Francesco (1597–1679), cardinal, nephew of Pope Urban VIII 433

Barberini, Maffeo (1568–1644), pope (1623–1644) as Urban VIII 196

Bardeen, John (1908–1991), American physicist, Nobel Laureate (1956, 1972) 400, XXI, XXV

Barkla, Charles Glover (1877–1944), English physicist, Nobel Laureate (1917) 491

Barrow, Isaac (1630–1677), English theologian, mathematician, physicist 226, 255, 291, 295, 326, 563

Bartholin, Rasmus (Erasmus Bartholinus) (1625–1698), Danish physician, naturalist 283

Basil of Caesarea (Saint Basil the Great) (ca. 330–379), Greek theologian, encyclopedist 132

Basov, Nikolai Gennadievich (1922–2001), Soviet physicist, Nobel Laureate (1964) 400, 493, XXI, XXVI

Baudot, Jean-Maurice-Émile (1845–1903), French telegraph engineer XIX

Bauer, Edmond Henri (1880–1963), French physicist 487

Bay, Zoltán (1900–1992), Hungarian-born American physicist 447, 508

Beauval, Henri Basnages De (1657–1710), French historian, lexicographer 299

Becher, Johann Joachim (1635–1682), German physician, chemist, economist 240

Becher, Johannes R. (1891–1958), German writer 240

Beckett, Samuel (1906–1989) Irish writer, Nobel Laureate (literature, 1969) 23

Becquerel, Alexandre-Edmond (1820–1891), French physicist, father of Henri Becquerel 323, XIX

Becquerel, Antoine César (1788–1878), French scientist, father of Alexandre-Edmond Becquerel XIX

Becquerel, Henri Antoine (1852–1908), French physicist, Nobel Laureate (1903) 391, 392, 490, 496, **498–500**, 501, 502, 503

Bede (the Venerable Bede) (ca. 672–735), English theologian, chronicler 119, 127, 128, 134

Bednorz, Johannes Georg (b. 1950), German physicist, Nobel Laureate (1987) 495, XXV

Beeckmann, Isaac (1588–1637), Dutch naturalist 211, 212, 234, 257

Beethoven, Ludwig van (1770–1827), German composer XVIII

Bell, John Stewart (1928–1990), British physicist 485

Beltrami, Eugenio (1835–1900), Italian mathematician 402

Benedetti, Giambattista (1530–1590), Italian mathematician, physicist 159–60, 241

Benedict Biscop (Biscop Baducing) (c. 628–690), Anglo-Saxon abbot, founder of Monkwearmouth-Jarrow Priory 128

Benedict of Nursia (Saint Benedict) (480–547), founder of Western monasticism 123, 124

Bentham, Jeremy (1748–1832), English philosopher, founder of utilitarianism 325

Bergson, Henri (1859–1941), French philosopher, Nobel Laureate (literature 1927) 235

Berkeley, George (1685–1753), Irish philosopher, theologian 273, 299, 320

Bernal, John Desmond (1901–1971), English physicist, molecular biologist 197

Bernoulli, Daniel (1700–1782), Swiss mathematician, physicist, botanist, anatomist, son of Johann Bernoulli 237, 256, 296, 305, 375, 387

Bernoulli, Jakob (Jacob, Jacques) (1654–1705), Swiss mathematician, brother of Johann Bernoulli 296, 305, 312, 317, 323

Bernoulli, Johann (Jean) (1667–1748), Swiss mathematician 296, 303, 304, 305, 317, 323

Berthollet, Claude Louis (1748–1822), French chemist 341, 372

Bessarion, Johannes (1403–1472), cardinal, theologian, humanist of Byzantine origin 165

Bessel, Friedrich Wilhelm (1784–1846), German astronomer 436, 552

Besso, Angelo (1873–1955), Swiss engineer, friend of Einstein 422

Bethe, Hans Albrecht (1906–2005), German-born American physicist, Nobel Laureate (1967) 493, 551, 552, 553

Bianchini, Giovanni (Johannes Blanchinus) (1410–ca. 1469), professor of mathematics and astronomy at the University of Ferrara, court astrologer of Leonello d'Este 168

Binnig, Gerd (b. 1947), German physicist, Nobel Laureate (1986) 495, 499

Biot, Jean Baptiste (1774–1862), French physicist 346, 347, 348, 360, 363, 373

Bismarck, Otto von (1815–1898), German statesman 323

Bjorken, James Daniel (b. 1934), American physicist 548

Black, Joseph (1728–1799), Scottish chemist, physicist, physician 365, **366–68**, 369, 370

Blackett, Patrick Maynard Stuart, Lord (1897–1974), English physicist, Nobel Laureate (1948) 397, 487, 492, 507, 526

Bláthy, Ottó Titusz (1860–1939), Hungarian electrical engineer 353, XIX

Bloch, Felix (1905–1983), Swiss-born American physicist, Nobel Laureate (1952) 492, XXV

Block, Martin (b. 1925), American physicist 542

Bloembergen, Nicolaas (b. 1920), Dutch-born American physicist, Nobel Laureate (1981) XXV

Boccaccio, Giovanni (1313–1375), Italian writer 128
Boerhaave, Hermann (1668–1738), Dutch physician, chemist 368
Boethius, Anicius Manlius Severinus (ca. 480–525), Roman philosopher, statesman 47, 78–79, 129–30, 134, 167, XIII
Bogoliubov, Nikolai Nikolaevich (1909–1992), Soviet physicist, mathematician XXV
Bohm, David Joseph (1917–1992), American physicist 482, 485
Bohr, Aage Niels (b. 1922), Danish physicist, son of Niels Bohr, Nobel Laureate (1975) 494, 514
Bohr, Harald (1887–1951), Danish mathematician, brother of Niels Bohr 449
Bohr, Niels Henrik David (1885–1962), Danish physicist, Nobel Laureate (1922) 10, 21, 24, 312, 318, 397, 398, 399, 400, 419, 439, 442, 445, **448–51**, 452, 453, 461, 465, 466–67, 470, 471, 472, 473, 474, 475, 482, 487, 489, 490, 491, 496, 504, 507, 512, 515, 518, 523, XXI, XXIII, XXIV
Bolesław I (the Brave, Bolesław I Chrobry) (967–1025), Polish king 117
Boltzmann, Ludwig Eduard (1844–1906), Austrian physicist 16, 354, 357, 358, 359, 377, 381, 382, 383, 385, 386, 387, 400, 405, 408, 438, 440, 444, 445, 447, 467, 545, 546
Bolyai, Farkas (1775–1856), Hungarian mathematician, father of János Bolyai 401, 429
Bolyai, János (1802–1860), Hungarian mathematician 12, 48, 95, 97, 323, 341, 401, 429, 431
Bonaventura (Giovanni di Fidanza), Italian theologian, doctor of the Church, Franciscan ("Doctor Seraphicus") 126
Born, Max (1882–1970), German physicist, Nobel Laureate (1954) , 403, 418, 453, 454, 456, 457, 462, 464, 465, 466, 470, 474, 482, 493, 522, 525, 546, XXIV, XXV
Borro, Girolamo (Hieronymus Borrius) (1512–1592), Italian natural philosopher 174
Borsos, Miklós (1906–1990), Hungarian sculptor 23
Boscovich, Roger Joseph (Ruđer Josip Bošković) (1711–1787), Croatian scientist, Jesuit 314–15, 353
Bose, Satyendra Nath (1894–1974), Indian physicist 446–47, XXIII
Bothe, Walter Wilhelm Georg Franz (1891–1957), German physicist, Nobel Laureate (1954) 487, 489, 493, 508, 509, 510
Botticelli, Sandro (Alessandro di Mariano di Vanni Filipepi) (1445–1510), Italian painter 152
Boyle, Robert (1627–1691), Irish naturalist 24, 49, 175, 236–37, 238–40, 298, 299, 300, 373
Bradley, James (1692–1762), English astronomer 412
Bradwardine, Thomas (ca. 1290–1349), English mathematician, scholastic philosopher 144–46, 159, XIII
Bragg, Sir William Henry (1862–1942), English physicist, Nobel Laureate (1915) 426, 489, 491, 514, XXV
Bragg, Sir William Lawrence (1890–1971), English physicist, son of William Henry Bragg, Nobel Laureate (1915) 392, 426, 470, 489, 491, 514, XXV
Brahe, Tycho (1546–1601), Danish astronomer 72, 73, 128, 174, **185–88**, 189, 190, 191, 192, 193, 194, 275, 552, 556
Brahmagupta (598–668), Indian mathematician, astronomer 136, 137
Brattain, Walter Houser (1902–1987), American physicist, Nobel Laureate (1956) 400, 492, 493, XXI, XXV
Braun, Karl Ferdinand (1850–1918), German physicist, Nobel Laureate (1909) 361, 490
Breit, Gregory (1899–1981), Russian-born American physicist 539
Brennus (fl. 387 bce), Gallic military commander 47
Breton, André François le (1708–1779), French publisher 322
Breu, Jörg (ca. 1475–1537), German painter VII
Breuil, Henri Éduard (1877–1961), French archeologist 32
Brewster, Sir David (1781–1868), English physicist 360
Bridgman, Percy Williams (1882–1960), American physicist, Nobel Laureate (1946) 16, 17
Brillouin, Léon (1889–1969), French physicist 438, XXV
Brillouin, Marcel (1854–1948), French physicist, father of Léon Brillouin 438
Broglie, Louis Victor Prince de (1892–1981), French physicist, Nobel Laureate (1929) 438, 439, 455, 459–60, 461, 480–81, 487, 490, 491
Broglie, Maurice de (1875–1960), French physicist, brother of Louis de Broglie 439, 459, 470, 487, XXI
Brüche, Ernst Carl Reinhold (1900–1985), German physicist 499
Bruckner, Anton (1824–1896), Austrian composer 382
Brunelleschi, Filippo (1377–1446), Italian architect, sculptor, painter, goldsmith 161
Bruno, Giordano (1548–1600), Italian natural philosopher 24, 153, 156, 157, 162, 172–73, 175, 185, 199, 241, 299, 329
Bruns, Heinrich Ernst (1848–1919), German astronomer, geodesist, mathematician 319
Brush, Stephen G. (b. 1935), American historian of science 17, 20
Bryson of Heraclea (ca. 450–ca. 390 bce), Greek mathematician 96
Bucherer, Alfred Heinrich (1863–1927), German physicist 428
Buddha (Siddhārtha Gautama) (ca. 560–ca. 480 bce), Indian founder of Buddhism 46, 47, 48, 545, XXIV
Buffon, Georges-Louis Leclerc de (1707–1788), French naturalist 322
Bunsen, Robert Wilhelm (1811–1899), German chemist 399, 552
Buridan, Jean (Johannes Buridanus) (ca. 1295–1358), French scholar, rector of the University of Paris 128, 145, **146–48**, 197, IX
Busch, Hans (1884–1973), German physicist 499
Butenandt, Adolf Friedrich Johann (1903–1995), German biochemist, Nobel Laureate (chemistry 1939) 522
Butler, Sir Clifford Charles (1922–1999), English physicist 528

C

Cabrera, Blas (1878–1945), Spanish physicist 487
Caesar, Gaius Julius (100–44 bce), Roman military leader, statesman 40, 47, 104, 105
Callippus (Calippus) (ca. 370–ca. 300 bce) **72**
Campanella, Tommaso (1568–1639), Italian philosopher, heretic, revolutionary, author of the utopian *La città del sole* 185
Canton, John (1718–1772), English naturalist 339
Cantor, Moritz Benedikt (1829–1920), German mathematician, author of the standard work *Vorlesungen über Geschichte der Mathematik* (1880–1908) 37
Cardano, Girolamo (Hieronymus Cardanus) (1501–1576), Italian mathematician, naturalist, physician 128, 159, 160, XIII
Carlisle, Sir Anthony (1768–1842), English surgeon, discoverer of electrolysis XIX
Carlson, David Emil (b. 1942), American physicist 528
Carlyle, Thomas (1795–1881), English writer, chronicler 10
Carnap, Rudolf (1891–1970), German-born American philosopher 563
Carnot, Lazare Nicolas Marguerite (1753–1823), French mathematician, engineer, statesman, father of Sadi Carnot 12
Carnot, Nicolas Léonard Sadi (1796–1832), French engineer, physicist 323, 373, 365, **373–74**, 375, XIX
Cassiodorus, Flavius Magnus Aurelius (ca. 485–ca. 585), Roman statesman, scholar 47, 130, 134
Catherine II, the Great (1729–1796), Russian empress, supporter of the French Enlightenment 329
Cauchy, Augustin Louis (1789–1857), French mathematician 65, 303, 304, 360, 400
Cavendish, Henry (1731–1810), English chemist, physicist, astronomer 240, 264, 323, 338, 339, 340, 392, XVIII
Cayley, Arthur (1821–1895), British mathematician 568
Celsius, Anders (1701–1744), Swedish physicist, astronomer 366
Cervantes Saavedra, Miguel de (1547–1616), Spanish writer 137, 173
Ceulen, Ludolph van (1540–1610), Dutch mathematician 98
Chadwick, Sir James (1891–1974), English physicist, Nobel Laureate (1935) 397, 487, 489, 492, 507, 508–9, 511, 514, 523, XXI
Chaitin, Gregory John (b. 1947), American mathematician, computer scientist XXI

Chamberlain, Owen (1920–2006), American physicist, Nobel Laureate (1959) 493
Chambers, Ephraim (ca. 1680–1740), English encyclopedist 322
Chambre, Marin Cureau de la (1594–1669), French physician, philosopher 248, 249, 251, 254
Champollion, Jean-François (1792–1832), French Egyptologist, deciphered hieroglyphics 362
Chandrasekhar, Subrahmanyan (1910–1995), Indian-born American physicist, Nobel Laureate (1983) 495, 552, 560
Chanut, Pierre Hector (1604–1677), French diplomat XVI
Charlemagne (Carolus Magnus) (742–814), king of the Franks, Roman emperor 117, 118, 119, 120, 128, 129, 131
Charles Martel ("The Hammer," Carolus Martellus) (ca. 688–741), ruler of the Franks, victor over the Arabs at Poitiers 128
Charles the Bald (Charles II) (823–877), West-Frankish king, Roman emperor 117, 118
Charles V (1500–1558), Spanish king, Holy Roman emperor 128, 159
Charpak, Georges (b. 1924), Polish-born French physicist, Nobel Laureate (1992) 533
Chaucer, Geoffrey (1340–1400), English poet, statesman 128
Cherenkov, Pavel Alekseyevich (1904–1990), Soviet physicist, Nobel Laureate (1958) 493, 532, 534
Chesterton, Gilbert Keith (1874–1936), English writer 173
Chopin, Frédéric François (Fryderyk Franciszek) (1810–1849), Polish composer 323
Christina of Sweden (Kristina Augusta) (1626–1689), Swedish queen, daughter of Gustavus Adolphus 225, 274, XVI
Chrysippus of Soli (ca. 280–ca. 207 bce), writer, Greek philosopher, Stoic 47
Cicero, Marcus Tullius (106–46 bce), Roman politician, writer 82, 103, 158, 218, XIV
Cimabue, Cenni di Pepo (Bencivieni di Pepo) (ca. 1240–ca. 1302), Italian painter, precursor of Giotto 126
Clagett, Marshall (1916–2005), American historian of science 140, 142, 145, 148
Clairaut, Alexis Claude (1713–1765), French mathematician, astronomer 270, 308
Clapeyron, Benoît Paul Émil (1799–1864), French engineer, physicist 373
Clarke, Edward M. (1804–1846), Irish-born English engineer 353
Clausius, Rudolf Julius Emanuel (1822–1888), German physicist 360, 377–78, 379–80, 381, 387
Clavius, Christopher (Christoph Klau) (1538–1612), German-born Jesuit astronomer, mathematician 183
Cleanthes of Assos (ca. 330–ca. 230 bce), Stoic philosopher, successor to Zeno 47, 79
Cleghorn, Robert (fl. ca. 1779), English naturalist 367
Cleomedes (fl. between mid first century bce and fourth century ce), Greek astronomer 94–95
Clerselier, Claude (1614–1684), French scholar, publisher of Descartes's works 253, 254
Clifford, William Kingdon (1845–1879), English mathematician, philosopher 430, 434
Clovis I (ca. 466–511), French king, founder of France 47
Cockcroft, John Douglas (1897–1967), English physicist, Nobel Laureate (1951) 397, 487, 492, 501, 508, 509, 514
Cohen, I. Bernard (1914–2003), American historian of science 182
Colbert, Jean-Baptiste (1619–1683), French economist, statesman 243
Columbanus (Columban) (543–615), Irish writer, missionary 119
Columbus, Christopher (1451–1506), Italian mariner, discoverer of America 94, 128, 524
Comenius, John Amos (Jan Amos Komenský) (1592–1670), Czech theologian, pedagogue 225, 241
Compton, Arthur Holly (1892–1962), American physicist, Nobel Laureate (1927) 470, 489, 490, 491, 519
Comte, Auguste (1794–1857), French philosopher 323, 563
Conant, James Bryant (1893–1978), American chemist 519
Condon, Edward Uhler (1902–1974), American physicist 457
Condorcet, Marie Jean Antoine Nicolas de Caritat, Marquis (1743–1794), French mathematician, philosopher, politician 325–26
Confucius (551–479 bce), Chinese philosopher, statesman 46, 47, 48
Constantine I, the Great (Constantinus Flavius Valerius Aurelius Claudius) (ca. 280–337), Roman emperor 115, VI
Constantius, Flavius Valerius (Constantius Chlorus, Constantius I) (ca. 250–306), Father of Constantine the Great, emperor of the Western Roman Empire (305–306) 115
Cook, James (1728–1779), British naval officer, explorer 329
Cooper, Leon N. (b. 1930), American physicist, Nobel Laureate (1972) XXV
Copernicus, Nicolaus (Mikolaj Kopernik) (1473–1543), Polish astronomer, naturalist 5, 17, 18, 26, 49, 54, 72, 73, 82, 90, 128, 149, 156, 162, 165, 174, 175, **177–85**, 188, 190, 191, 198, 199, 225, 275, 330, 433, 552, XIV
Corelli, Arcangelo (1653–1713), Italian composer 323
Cosyns, Max (1906–1998), Belgian scientist 487
Cotes, Roger (1682–1716), English mathematician, astronomer 264
Coulomb, Charles Auguste de (1736–1806), French physicist 337, 338–39, 340, 344, 345
Courant, Richard (1888–1972), German-born American mathematician 458
Cowan, George Arthur (b. 1920), American chemist 527, 532, 535
Crick, Francis Henry Compton (1916–2004), British molecular biologist, physicist, neuroscientist 323
Critchfield, Charles Louis (1910–1994), American physicist 551
Crombie, Alistair Cameron (1915–1996), Australian historian of science 134, 158, 168
Cronin, James Watson (b. 1931), American physicist, Nobel Laureate (1980) 494
Crookes, Sir William (1832–1919), English physicist, chemist 389, 390, 505
Csikai, Gyula (b. 1930), Hungarian physicist 527
Ctesibius (Ktesibios, Tesibius) (fl. 285–222 bce) Greek (Alexandrian) inventor 48, 99
Curie, Ève (1904–2007), French journalist, daughter of Marie and Pierre Curie 500
Curie, Jacques (1855–1941), French physicist, brother of Pierre Curie 501
Curie, Marie (née Skłodowska) (1867–1934), Polish-born French physicist, chemist, Nobel Laureate (1903, chemistry 1911) 391, 438, 447, 470, 487, 490, 496, 500, **501–4**, 514
Curie, Pierre (1859–1906), French physicist, Nobel Laureate (1903) 391, 490, 496, 500, **501–4**, 505, 514, XXV
Cusa, Nicholas of (Nicolaus Cusanus, Nicholas of Kues) (1401–1464), German theologian, naturalist, philosopher 128, 156, 157, 159, 160, 161, 166, 174, XVI
Cuvier, Georges (1769–1832), French naturalist XVIII
Cyrano de Bergerac (1619–1655), French poet, playwright, satirist 187
Cyril (827–869), Greek missionary, brother of Methodius 132

D

Dalén, Nils Gustaf (1869–1937), Swedish engineer, Nobel Laureate (1912) 491, 495
Dalitz, Richard Henry (1925–2006), Australian-born English physicist 542
Dalton, John (1766–1844), English chemist, physicist 323, 373, 388
Dandelin, Germinal Pierre (1794–1847), Belgian mathematician 100
Dante (Dante Alighieri, Durante degli Alighieri) (1265–1321), Italian poet 5, 126, 128, 138, 151, 152
Darwin, Charles Robert (1809–1882), English naturalist 323, 409, 551
Darwin, Erasmus (1731–1802), English physician, naturalist, grandfather of Charles Robert Darwin 323
Darwin, Sir Charles Dalton (1887–1962), British physicist, grandson of Charles Robert Darwin 419, 468, 476
Davis, Robert Houser (b. 1926), American physicist 537, 554
Davisson, Clinton Joseph (1881–1958), American physicist, Nobel Laureate (1937) 396, 459, 460, 488, 492

Davy, Sir Humphry, 1st Baronet (1778–1829), Cornish chemist, physicist 343, 350, 351, 368, 375, 388
Debye, Peter Joseph Willem (1884–1966), Dutch-born American physical chemist, Nobel Laureate (chemistry 1936) 446, 448, 487
Dehmelt, Hans Georg (b. 1922), German-born American physicist, Nobel Laureate (1989) 495
Delacroix, Eugène (1798–1863), French painter 323
Delambre, Jean-Baptiste (1749–1822), French astronomer 447
Demarçay, Eugène-Anatole (1852–1904), French chemist 505
Democritus (ca. 460–ca. 370 bce), Greek philosopher 29, 47, 48, 49, 55, 60, **61**, 87, 101, 151, 237, 238, 271, 387, 389, 552
Dennison, David Matthias (1900–1976), American physicist 491
Deprez, Marcel (1843–1918), French electrical engineer XIX
Derham, William (1657–1735), English clergyman, natural philosopher 341
Déri, Miksa (1854–1938), Hungarian engineer 353, XIX
Desaguliers, Jean Théophile (1683–1744), French-born English naturalist 332
Desargues, Girard (1593–1662), French mathematician, engineer 161, 241
Descartes, René (Cartesius) (1596–1650), French philosopher, mathematician, naturalist 24, 95, 104, 144, 153, 154, 156, 189, 212, 213, 214, **216–23**, 224–25, 226–30, 232–33, 234, 235, 238, 240, 242, 246, 247, 248, 251, 254, 255, 258, 270, 271, 273, 274, 275, 276–77, 281, 282, 283, 284, 285, 287, 288, 289, 290, 292, 296, **297–99**, 300, 302, 310, 314, 322, 324, 325, 330, 333, 358, 400, 406, 407, 552, 563, XIII, XV, XVI
Deutsch, David Elieser (b. 1953), Israeli physicist 577
Dewar, Sir James (1842–1923), English chemist, physicist 488
Dickens, Charles (1812–1870), English writer 323
Diderot, Denis (1713–1784), French writer, philosopher, encyclopedist 25, 215, 216, 321, 322, 325, 326, XX
Digges, Thomas (ca. 1546–1595), English mathematician, astronomer 184
Dijksterhuis, Eduard Jan (1892–1965), Dutch historian of science 191, 192, 273
Diocletian (Gaius Aurelius Valerius Diocletianus) (ca. 244–311), Roman emperor of Illyrian origin 115
Dionysius II (the Younger) (ca. 397– 343 bce), ruler of Syracuse 57
Diophantus of Alexandria (fl. ca. 250 ce), Hellenic mathematician 47, 48, 100
Dioscorides, Pedanius (ca. 40–90), Greek physician, pharmacologist, botanist 151
Dirac, Paul Adrien Maurice (1902–1984), English physicist, Nobel Laureate (1933) 23, 59, 319, 364, 399, 456, 457, 467, 470, 471, 474, 475, 476, 477, 478, 487, 490, 492, 525, 526, XXI
Dirichlet, Peter Gustav Lejeune (1805–1859), French-born German mathematician 419
DiVincenzo, David P. (b. 1959), American physicist 577
Dolivo-Dobrovolsky, Mikhail Osipovich (1862–1919), Russian electrical engineer XIX
Dollond, John (1706–1761), English optician 286
Domagk, Gerhard (1895–1964), German pathologist, bacteriologist, Nobel Laureate (physiology or medicine 1939) 522
Dominis, Marco Antonio de (Markantun de Dominis) (1566–1624), Italian mathematician, physicist of Dalmatian origin, archbishop of Dalmatia and Croatia 285
Donder, Théophile Ernest de (1872–1957), Belgian mathematician, physicist 487
Donne, John (1572–1631), English poet XIV
Doppler, Christian Johann (1803–1853), Austrian physicist, mathematician 364
Dostoevsky, Fedor (1821–1881), Russian writer 323
Drude, Paul Karl Ludwig (1863–1906), German physicist 410
Dufay, Charles François de Cisternay (du Fay) (1698–1739), French botanist, physicist 332–33, 335, 336
Dufy, Raoul (1877–1953), French painter XIX
Dugas, René (1897–1957), French historian of science 12, 150, 245, 248, 249, 251, 253, 254, 257, 259, 308, 309, 310
Duhem, Pierre (1861–1916), French physicist, historian of science 140, 163, 406, 409, IX
Dumesnil, Pierre Louis (1698–1781), French painter XVI
Duns Scotus, John (ca. 1266–1308), Scottish Scholastic ("Doctor Subtilis") 126, 128, 159, 160
Dürer, Albrecht (1471–1528), German artist 99, 128, 136, 157, 161, 162, 176
Dürrenmatt, Friedrich (1921–1990), Swiss writer 521
Dushman, Saul (1883–1954), Russian-born American physical chemist 490

E

Eckermann, Johann Peter (1792–1854) German poet and writer 24
Ecphantus (fl. 4th century bce), Greek philosopher XIV
Eddington, Arthur Stanley (1882–1944), English astronomer, physicist 429, 431, 551, 552
Eddy, Mary Baker (1821–1910), founder of the Christian Science movement 49
Edison, Thomas Alva (1847–1931), American inventor, industrialist 323, XIX, XX
Ehrenfest, Paul (1880–1933), Austrian-born Dutch physicist 450, 471
Einstein, Albert (1879–1955), Swiss, German, and American physicist, Nobel Laureate (1921) 5, 13, 17, 18, 19, 20, 22, 23, 49, 59, 194, 210, 212, 272, 273, 277, 328, 349, 363, 364, 400, 403, 405, 409–10, **416–21**, 422, 426, 427–29, 431, 432, **433–37**, 438, 439, 446, 447, **448**, 449, 450, 451–52, 453, 454, **457–58**, 459–60, 465, 466, 470, 471, 472–73, 475, 482, 485, 486, 489, 490, 491, 495, 496, 510, 521, 523, 538–39, 547, 552, 557, 561, 562, 563, 570, 574, 576, XIX, XXI, XXII, XXIII, XXV
Elizabeth I (1533–1603), English queen, daughter of Henry VIII 329
Ellis, Charles Drummond (1895–1980), English physicist 487
Elsasser, Walter Maurice (1904–1991), German-born American physicist, geophysicist 511
Empedocles (ca. 490–430 bce), Greek philosopher, writer, physician 47, 48, 60–61, 62, 336
Engels, Friedrich (1820–1895), German philosopher, politician 113, 409
Eötvös, Loránd (Vásárosnaményi Báró Eötvös Loránd, Baron Roland von Eötvös) (1848–1919) Hungarian physicist 431, 435, 436
Epictetus (55–135), Greek Stoic philosopher 47, 100
Epicurus (341–270 bce), Greek philosopher 47, **101**, XV
Erasmus of Rotterdam, Desiderius (1466 or 1469–1536), Dutch humanist 171
Eratosthenes of Cyrene (276–194 bce), Greek mathematician, geographer 47, 48, 83, 87, 92, **93–94**, 95, 167
Esaki, Leo (Reona) (b. 1925), Japanese physicist, Nobel Laureate (1973) 494
Escher, Maurits Cornelis (1898–1972), Dutch graphic artist 402
Euclid of Alexandria (ca. 365–ca. 300 bce), Greek mathematician 5, 12, 13, 47, 48, 53, 65, 78, **96–97**, 99, 129, 130, 131, 133, 135, 142, 151, 159, 160, 167, 256, 428, 429, 431, 437, X, XIII, XXV
Eudemus of Rhodes (ca. 370–ca. 300 bce), pupil of Aristotle 73–74, 78
Eudoxus of Cnidus (410 or 408–347 bce), Greek mathematician, astronomer 47, 48, **72**, 74, 75, 83, 95, 96
Euler, Leonhard (1707–1783), Swiss mathematician, physicist, astronomer 65, 100, 286, 291, 296, 298, **301–4**, **305–9**, 310–11, 312, 315, 316, 317, 319, 329, 335, 375, XXIII
Euler-Chelpin, Hans Karl August Simon von (1873–1964), German-born Swedish chemist, Nobel Laureate (chemistry 1929) 522
Eupalinus of Megara (fl. ca. 530 bce), Greek engineer 98
Euripides (480–406 bce), Greek tragedian 47, 436, II
Eysenck, Hans Jürgen (1916–1997), German-born British psychologist 63
Ezekiel (fl. 6th century bce), Israelite prophet 46

F

Fabry, Maurice Paul Auguste Charles (1867–1945), French physicist 447, 542

Fajans, Kazimierz (1887–1975), Polish-born American physical chemist 504
Faller, James Elliot (b. 1934), American physicist 340
Faraday, Michael (1791–1867), English physicist 5, 18, 25, 331, 332, 339, 345, 348, **349–54**, 355, 356, 357, 358, 368, 388, 391, 397, 400, 534, XIX, XX
Fedorov, Ivan (1510–1583), Russian printer 224, 225, XI
Ferdinand II de' Medici (1610–1670), grand duke of Tuscany, funded the Accademia del Cimento 366
Fermat, Pierre de (1601–1665), French jurist, mathematician, physicist 230–33, 242, 248, 249, 251, 253, 254, 277, 283, 288, 289–90, 291, 292–93, 303, 309, 310, 311, 452–53
Fermi, Enrico (1901–1954), Italian physicist, Nobel Laureate (1938) 400, 457, 475, 487, 490, 492, 504, 506, 508, 515, **516–17**, 518, 519, 523, 526, 527, 528, 530, 539, XXI
Ferraris, Galileo (1847–1897), Italian physicist, electrical engineer XIX
Feshbach, Herman (1917–2000), American physicist 514
Feyerabend, Paul Karl (1924–1994), Austrian-born American philosopher 25, 563, 564
Feynman, Richard Phillips (1918–1988), American physicist, Nobel Laureate (1965) 4, 20, 21, 477, 478, 493, 524, 535, 542
Fibonacci (Leonardo of Pisa) (ca. 1170–ca. 1250), Italian mathematician 40, 49, 128, **140–42**, XIII
Fichte, Johann Gottlieb (1762–1814), German philosopher XVIII
Ficinus, Marsilius (Masilio Ficino) (1433–1499), Italian philosopher, physician 158
Fisher, Herbert Albert Laurens (1865–1940), English historian XI
Fitch, Val Logsdon (b. 1923), American physicist, Nobel Laureate (1980) 494
Fitzgerald, George Francis (1851–1901), Irish physicist 416
Fizeau, Armand Hippolyte Louis (1819–1896), French physicist 361, 362, 363–64, 412, 413
Fock, Vladimir Alexandrovich (1898–1974), Soviet physicist, mathematician 475, 477
Fontenelle, Bernard le Bovier de (1657–1757), French writer, philosopher 274, 277, 281
Ford, Joseph (1927–1995), American physicist XXI
Foucault, Jean Bernard Léon (1819–1868), French physicist, astronomer 363, 364
Fourier, Jean Baptiste Joseph (1768–1830), French mathematician, physicist 344, 365, **369–72**, 400, 454
Fowler, Ralph Howard (1889–1944), English mathematician, physicist 467–68
Fowler, William Alfred (1911–1995), American physicist, Nobel Laureate (1983) 495
Francesca, Piero della (ca. 1420–1492), Italian artist, geometer 161
Francis of Assisi (Giovanni Francesco Bernardone) (1182–1226), Italian founder of the Franciscan order 124
Franck, James (1882–1964), German-born American physicist, Nobel Laureate (1925) 443, 455, 491, 523
Frank, Ilya Mikhailovich (1908–1990), Soviet physicist, Nobel Laureate (1958) 493, 532, 534
Franklin, Benjamin (1706–1790), American natural scientist, statesman 226, 332, **333–38**, 339, XVIII
Fraunhofer, Joseph von (1787–1826), German physicist, optician, astronomer 364, 398, 552
Frederick II (1534–1588), king of Denmark and Norway 185
Frederick II, the Great (1712–1786), Prussian king XX
Frederick III, the Simple (1341–1377), king of Sicily from 1355 to 1377 141, 165
Frege, Gottlob (1848–1925), German philosopher, logician 2
Frenkel, Yakov I. (1894–1952), Soviet physicist XXV
Fresnel, Augustin Jean (1788–1827), French engineer, physicist 360, 363
Friedman, Alexander Alexandrovich (1888–1925), Russian mathematician, physicist 552, 557
Friedman, Jerome Isaac (b. 1930), American physicist, Nobel Laureate (1990) 495
Frisch, Otto Robert (1904–1979), Austrian physicist 518, 519, 524
Fritzsch, Harald (b. 1943), German theoretical physicist 566
Frontinus, Sextus Julius (40–113), Roman builder, engineer 47, 101
Fuller, Richard Buckminster (1895–1983), American architect, futurist, inventor 65

G

Gabor, Sir Dennis (1900–1979), Hungarian-born English electrical engineer, Nobel Laureate (1971) 400, 494, 496, XXI
Gaddi, Taddeo (ca. 1300–1366), Italian painter, father of Agnolo Gaddi, assistant to Giotto 126
Gaius (full name unknown) (fl. ca. 140–ca. 180), Roman jurist 47
Galen, Claudius (Galenus) (129–200), Roman physician of Greek origin 47, 62, 111, 133, 151, 168, 174, 211, 332, X
Galerius (Gaius Galerius Valerius Maximianus) (ca. 260–311), Roman emperor 115
Galilei, Vincenzo (ca. 1520 –1591), Italian musician, father of Galileo Galilei 195, 198
Galileo (Galileo Galilei) (1564–1642), Italian physicist, mathematician, astronomer 2, 10, 13, 24, 49, 143, 144, 148, 149, 150, 156, 162, 175, 185, **194–213**, 214, 217, 219, 226, 227, 230, 233, 234, 237, 238, 240, 241, 242, 243, 244, 245, 247, 255, 257, 263, 265, 266, 270, 271, 275, 291, 299, 303, 329, 366, 409, 433, 434, 436, 552, 559, XV, XX
Galois, Évariste (1811–1832), French mathematician 568
Galvani, Luigi Aloisio (1737–1798) Italian anatomist, natural scientist 342–43
Gamow, George (1909–1968), Russian-born American physicist 269, 433, 437, 467, 487, 507, 509, 510, 524, 552, 557, XXI
Garrick, David (1716–1779), English actor 329
Gassendi, Pierre (1592–1655), French philosopher, physicist 62, 238, 239, 299, 341
Gaulard, Lucien (1850–1888), French inventor XIX
Gauss, Carl Friedrich (1777–1855), German mathematician, physicist, astronomer 16, 17, 97, 340, 341, 400, 401, 429, 430, 431, 560, 563, XIX, XXIII
Gay-Lussac, Louis Joseph (1778–1850), French chemist, physicist 346, 347, 372, 373, 388
Geiger, Hans (1882–1945), German physicist 397, 505, 508
Geissler, Heinrich (1814–1879), German inventor, physicist 388
Gell-Mann, Murray (b. 1929), American physicist, Nobel Laureate (1969) 494, 524, 535, 537, 540, 545, 547, 566, 569, XXI
Geminus of Rhodes (fl. 1st century bce), Greek philosopher, astronomer 66
Gennes, Pierre-Gilles de (1932–2007), French physicist, Nobel Laureate (1991) 495
Georgi, Howard Mason, III (b. 1947), American physicist 567, 569
Gerard of Cremona (Gerardo da Cremona, Gerardus Cremonensis) (ca. 1114–1187), Italian scholar, translator of Arabic works 92, 133
Gerbert d'Aurillac (Pope Sylvester II) (ca. 946–1003), French priest, theologian 141, XII
Gerhaert, Nikolaus (Nikolaus Gerhaert von Leyden) (ca. 1420–1473), Dutch sculptor 140
Gerlach, Walter (1889–1979), German physicist 491
Germer, Lester Halbert (1896–1971), American physicist 460
Gesner, Conrad (1516–1565), Swiss natural scientist, philosopher, linguist 173, 174
Geulincx, Arnold (1624–1669), Belgian philosopher 297, 298
Ghiorso, Albert (b. 1915), American chemist 514
Giaever, Ivar (Glæver) (b. 1929), Norwegian-born American physicist, Nobel Laureate (1973) 494
Gibbs, Josiah Willard (1839–1903), American mathematician, physical chemist, engineer 380, 386–87, XIX
Gilbert, William (1544–1603), English court physician, naturalist 175, 211, 216, 275, **329–30**, 331
Ginzburg, Vitaly L. (b. 1916), Soviet physicist, Nobel Laureate (2003) 534, XXV
Giotto di Bondone (1266–1337), Italian painter, architect 125, 126
Glaser, Donald Arthur (b. 1926), American physicist, Nobel Laureate (1960) 493, 532
Glashow, Sheldon Lee (b. 1932), American physicist, Nobel Laureate (1979) 494, 540, 548, 549, 566, 567, 569, 570, 572

Glauber, Johann Rudolf (1604 (?)–1670), German–Dutch alchemist, chemist 370
Gödel, Kurt (1906–1978), Austrian-born American logician, mathematician 22
Godunov, Boris Fedorovich (1551–1605), Russian boyar, regent, and later tsar 171
Goeppert-Mayer, Maria (1906–1972), German-born American physicist, Nobel Laureate (1963) 512
Goethe, Johann Wolfgang von (1749–1832), German writer, naturalist 5, 9, 24, 26, 79, 288, 321, XVIII
Goldhaber, Maurice (b. 1911), Austrian-born American physicist 543
Goldstein, Eugen (1850–1930), German physicist 389, 390, 391, 392
Goodman, Nelson (1906–1998), American philosopher, linguist 15
Gordon, Walter (1893–1939), German physicist 475, 476
Gorgias of Leontini (ca. 487–376 bce), Greek Sophist 47, 55–56
Gorky, Maxim (Aleksey Maksimovich Peshkov, Алексей Максимович Пешков) (1868–1936), Russian writer 497
Goudsmit, Samuel Abraham (1902–1978), Dutch-born American Physicist 451, 470, 474, 491
Goya y Lucientes, Francisco José de (1746–1828), Spanish painter 323
Graaff, Robert Jemison van de. *See* Van de Graaff
Gramme, Zénobe Théophile (1826–1901), Belgian electrical engineer 353
Grassmann, Hermann Günther (1809–1877), German mathematician, philologist 401
Gray, Stephen (1666–1736), English naturalist 322, 335
Green, George (1793–1841), English mathematician, physicist 340
Green, Michael Boris (b. 1946), British physicist 570, 571
Greenberg, Oscar Wallace (b. 1932), American physicist 548
Gregoras, Nicephorus (ca. 1295–1360), Byzantine historian 131–32
Gregory, James (1638–1675), Scottish mathematician, astronomer 291, 294, XVII
Gregory IX (Ugolino di Conti) (papacy 1227–1241), Pope 129
Grimaldi, Francesco Maria (1618–1663), Italian mathematician, physicist 242, 285
Groot, Jan Cornets de (1554–1640), Dutch intellectual, father of Hugo Grotius 210
Gross, David Jonathan (b. 1941), American particle physicist, string theorist, Nobel Laureate (2004) 567
Grosseteste, Robert (ca. 1175–1253), English scholar, Scholastic 124, 128, 154–55, 158
Grossmann, Marcel (1878–1936), Hungarian-born Swiss mathematician 432
Grotius, Hugo (Huig de Groot) (1583–1645), Dutch jurist, statesman, humanist 210
Guericke, Otto von (1602–1686), German physicist, engineer, mayor of Magdeburg 236–37, 330, 333
Guidobaldo I (Guido Ubaldo) (1472–1508), duke of Urbino, 1482 to 1508 XIII
Guillaume, Charles Édouard (1861–1938), Swiss physicist, Nobel Laureate (1920) 491
Gustavus Adolphus (Gustav II Adolf) (1594–1632), king of Sweden 242, XVI
Gutenberg, Johannes Gensfleisch zur Laden zum (ca. 1398–1468), German inventor of printing with movable type 165, XI, XIII
Guth, Alan Harvey (b. 1947), American theoretical physicist 574
Gyarmati, István (1929–2002), Hungarian physicist 380

H

Haas, Wander Johannes de (1878–1960), Dutch physicist 491
Hadrian (Publius Aelius Hadrianus) (76–138), Roman emperor 47, 92
Hahn, Otto (1879–1968), German chemist, Nobel Laureate (chemistry 1944) 397, 504, 516–18, 522, XXI
Halban, Hans (1908–1964), German physicist 518
Hall, Edwin Herbert (1855–1938), American physicist 498
Halley, Edmond (1656–1742), English astronomer 118, 226, 256, 263, 270, 328, 552
Hallwachs, Wilhelm Ludwig Franz (1859–1922), German physicist XIX
Hals, Frans (ca. 1580–1666), Flemish painter 242
Hamilton, Sir William Rowan (1805–1865), Irish mathematician, physicist, astronomer 304, 311, 316, 319, 323, 460, 464, 472
Hammurabi (d. 1750 bce), Babylonian king 34, 44
Handel, George Frideric (1685–1759), German-born English composer 241, 323
Hannibal (246–183 bce), Carthaginian general 47
Harold Godwinson (Harold II) (d. 1066), last Anglo-Saxon king 118, 129
Harriot, Thomas (1560–1621), English mathematician, astronomer 226
Harrison, John (1693–1776), English inventor, manufacturer of precision clocks 329
Harun al-Rashid (763–809), caliph of Baghdad 128, 129
Harvey, William (1578–1657), English anatomist, physicist, discoverer of the circulation of blood 216, 241
Hasdale, Martin (1571–1630), member of Emperor Rudolph II's court 199
Hasenöhrl, Friedrich (1874–1915), Austrian physicist 427, 438, 462
Hawking, Stephen William (b. 1942), British physicist, astronomer 21, 570, 574
Haydn, Franz Joseph (1732–1809), Austrian composer XVIII
Heaviside, Oliver (1850–1925), English physicist, mathematician 344, 358
Hegel, Georg Wilhelm Friedrich (1770–1831), German philosopher 32, 299, 409, XVIII
Heiberg, Johan Ludvig (1854–1928), Danish philologist, historian 82, 82–83
Heine, Heinrich (1797–1856), German poet XVIII
Heisenberg, August (1869–1930), German historian, Byzantinist, father of Werner Heisenberg 132
Heisenberg, Werner Karl (1901–1976), German physicist, Nobel Laureate (1932) 2, 4, 5, 10, 20, 21, 24–25, 61, 62, 63, 132, 304, 312, 319, 399, 400, 403, 439, **452–58**, **465–70**, 471, 477, 479, 480, 485, 486, 489, 490, 509, 522, 524, 526, 539, 546–47, 561, 568, VI, XXI, XXIV, XXV
Heitler, Walter Heinrich (1904–1981), Swiss physicist XXV
Helm, Georg Ferdinand (1851–1923), German mathematician 408
Helmholtz, Hermann Ludwig Ferdinand von (1821–1894), German physicist, physiologist 328, 344, 360, 376, 377, 379, 380, 387, 391, 399, 441, 478, 481, 551, 552, 558, XIX
Helvetius, Claude Adrien (1715–1771), French philosopher 321, 325
Henderson, Thomas James Alan (1798–1844), Scottish astronomer 552
Henriot, Émile (1885–1961), French chemist 487
Henry, Joseph (1797–1878), American physicist 353
Henry I (1008–1060), king of France 224
Henry IV (Henri IV) (1553–1610), king of Navarre and of France (the first Bourbon) 171
Heraclides Ponticus (Herakleides) (387–312 bce), Greek astronomer 47, 48, 72, 73, 74, 149, XIV
Heraclitus of Ephesus (ca. 535–475 bce), Greek philosopher 47, 48, 54–55, 60, 151, 557
Herapath, John (1790–1868), English naturalist 375, 377
Herivel, John W. (b. 1918), British historian of science 263
Hermann, Grete (1901–1984), German philosopher, mathematician 403
Hermann, Jakob (1678–1733), Swiss mathematician 375
Hermes Trismegistus, fictional author of astrological writings, Greek name of the Egyptian god Thoth 103, 109
Hero of Alexandria (ca. 10–ca. 70), Greek engineer, mathematician 47, 48, 98, 100, 101, 103
Herodotus of Halicarnassus (ca. 484–ca. 425 bce), Greek historian 31, 47
Herrad of Landsberg (ca. 1130–1195), German abbess 127
Herschel, Sir William (Friedrich Wilhelm) (1738–1822), German-born English astronomer 398, XIX
Hertz, Gustav Ludwig (1887–1975), German physicist, nephew of Heinrich Rudolf Hertz, Nobel Laureate (1925) 455, 491

Hertz, Heinrich Rudolf (1857–1894), German physicist 328, 359, 360, 361, 362, 389, 390, 400, 417, 443, 448, 455

Hertzsprung, Ejnar (1873–1967), Danish astronomer 552, 554, 555, 558

Herzberg, Gerhard (1904–1999), German-born Canadian physicist 453

Herzen, Édouard (1877–1936), Belgian chemist 487

Hess, Victor Franz (1883–1964), Austrian-born American physicist, Nobel Laureate (1936) 492, 526, 529, 530

Hevelius, Johannes (Johannes Hewel, Johann Hewelke, Johannes Höwelcke, Jan Heweliusz) (1611–1687), German astronomer 139, 177

Hevesy, George de (György) (1885–1966), Hungarian-born Physical chemist, active in Budapest, Freiburg, Copenhagen, and Stockholm, Nobel Laureate (chemistry 1943) 397, 572

Hewish, Sir Antony (b. 1924), English astronomer, Nobel Laureate (1974) 494, 552

Hey, James Stanley (1909–2000), English physicist 550

Heytesbury, William (ca. 1313–ca. 1372), English logician, naturalist 144, 150, 159

Hiero II (ruled 270–215 bce), king of Syracuse 79, 82, 83–84

Higgs, Peter Ware (b. 1929), British physicist 566, 567, 575, 583, XXI

Hilbert, David (1862–1943), German mathematician 432, 457

Hill, Henry Allen (b. 1933), American physicist 340

Hindemith, Paul (1895–1963), German composer 193

Hipparchus (ca. 190–ca. 120 bce), Greek astronomer 47, 48, 72, 74, 92, **93–94**, 131, 167, 290

Hippias of Elis (fl. 5th century bce), Greek philosopher, mathematician 96

Hippocrates of Chios (ca. 470–ca. 410 bce), Greek mathematician 47, 48, 95, 96, 98

Hippocrates of Cos (ca. 460–ca. 370 bce), Greek physician, naturalist 47, 62, 95, 111, 168, 174, 211, 324

Hitler, Adolf (1889–1945), German politician, chancellor 426, 440, 455, 488, 490, 523

Hittorf, Johann Wilhelm (1824–1914), German physical chemist 389

Hobbes, Thomas (1588–1679), English philosopher 298, 320

Høffding, Harald (1843–1931), Danish philosopher 467

Hofstadter, Robert (1915–1990), American physicist, Nobel Laureate (1961) 493

Hogben, Lancelot Thomas (1895–1975), British experimental zoologist, medical statistician, popularizer of mathematics 102

Holbach, Paul-Henri Thiry, Baron d' (Paul Heinrich Dietrich) (1723–1789), German-born French philosopher 321, 325, 326

Holbein, Hans (the Younger) (1497–1543), German painter 122, 176

Homer (fl. ca. 8th century bce), Ancient Greek legendary epicist, author of the *Iliad* and the *Odyssey* 45

Honnecourt, Villard de (fl. ca. 1245), French architect 124

Hooft, Gerardus 't (b. 1946), Dutch physicist, Nobel Laureate (1999) 567, 573, 574

Hooke, Robert (1635–1703), English naturalist 236, 240, 241, 242, 282, 285, 287

Horace (Quintus Horatius Flaccus) (65–8 bce), Roman poet 18, 47

Hospital, Guillaume François Antoine, Marquis de l' (de L'Hôpital) (1661–1704), French mathematician 289

Houston, William Vermillion (1900–1968), American physicist XXV

Houtermans, Friedrich Georg (1903–1966), Polish-born American physicist 551, 552

Hoyle, Sir Fred (1915–2001) English astronomer 555

Hubble, Edwin Powell (1889–1953), American astronomer 552, 557

Hugh of Saint Victor (1096–1141), French theologian 126, 127

Hugo, Victor (1802–1885), French poet, writer 323

Humboldt, Baron Alexander von (1769–1859), German naturalist, explorer, polymath 372

Hume, David (1711–1776), Scottish philosopher, historian 155, 299, 320, 323, 326, 478, 547, 548

Hunayn ibn Ishaq (Johannitius) (809–877), physician, director of the translation group in Baghdad 128

Hund, Friedrich (1896–1997), German physicist 8, 236, 268, 439, 443, 452, 453

Huxley, Aldous (1894–1963), English writer XVII

Huygens, Christiaan (1629–1695), Dutch mathematician, astronomer, physicist 156, 175, 220, 222, 223, 226, **240–55**, 256, 258, 259, 260, 265, 272, 273, 275, 282–84, 285, 286, 287, 288, 296, 297, 298, 307, 312, 314, 325, 358, 359, 360, 363

I

Ignatius Loyola (Ignacio López de Loyola) (1491–1556), Spanish nobleman, founder of the Jesuit order XIV

Infeld, Leopold (1898–1968), Polish physicist 418, 433

Ioffe, Abram Fedorovich (1880–1960), Soviet physicist 487

Isabella I (1451–1504), queen of Castille, wife of Ferdinand II of Aragon 128

Isidore of Miletus (fl. 6th century), Byzantine architect 131

Isidore of Seville (ca. 560–636), archbishop of Seville, encyclopedic writer 128, 130

Ivan IV Vasilyevich (Ivan the Terrible) (1530–1584), first Russian tsar 225

Ivanenko, Dmitri D. (1904–1994), Soviet physicist 422

J

Jacobi, Karl Gustav Jakob (1804–1851), German mathematician 304, 319, 321, 460

Jacobi, Moritz Hermann (Boris Semyonovich) von (1801–1874), Prussian engineer, physicist, brother of Karl Gustav Jacobi XIX

Jacopo de Barbari (ca. 1440–before 1516), Italian painter, printmaker XIII

Jadwiga (Hedwig) (1370–1399), queen of Poland, wife of the Lithuanian grand duke Jogaila, king of Poland as Władysław II 225

James I (1566–1625), king of Scots as James VI, king of England and of Ireland as James I 215

Jammer, Max (b. 1915), German-born Israeli physicist, philosopher of physics 450, XXII

Jansen, Cornelius (Cornelius Jansenius) (1585–1638), Dutch theologian, bishop of Ypres 235

Jansky, Karl (1905–1950), American electrical engineer 550, 552

Janssen, Zacharias (1580–ca. 1638), inventor of the microscope, built the first telescope together with Hans Lippershey in 1608 241

Janus Pannonius (János Csezmicei) (1434–1472), Hungarian poet (in the Latin language), humanist 128

Jaspers, Karl (1883–1969), German philosopher 409

Javan, Ali (b. 1926), Iranian-born American physicist 414

Jeans, James Hopwood (1877–1946), English physicist, mathematician, astronomer 438, 441, 444, 447

Jedlik, Ányos István (1800–1895), Hungarian physicist, inventor 323, 353, XIX

Jensen, Johannes Hans Daniel (1907–1973), German physicist, Nobel Laureate (1963) 493, 512

Jesus of Nazareth (7 to 2 bce–26 to 36 ce), central figure of Christianity 47, 141

Jogaila (Jagiełło) (1348–1434), grand duke of Lithuania, founder of the Jagiellon branch of the Lithuanian Gediminids dynasty 225

Johannes de Sacrobosco (Sacro Bosco, John of Holywood) (ca. 1195–ca. 1256), English-born mathematician, professor at the University of Paris, author of many textbooks 168

Johannes Scotus Eriugena (John the Irishman) (ca. 815–877) Irish Scholastic, philosopher 120, 128

John of Damascus (ca. 676–749), Byzantine theologian 132

John XXI. *See* Petrus Hispanus

Joliot-Curie, Frédéric (1900–1958), French physicist, Nobel Laureate (chemistry 1935) 502, 509, 510, 515, 516, 526

Joliot-Curie, Irène (1897–1956), French physicist, Nobel Laureate (chemistry 1935) 487, 502, 509, 510, 515, 516, 526

Joos, Georg Jakob Christof (1894–1959), German theoretical physicist 411

Jordan, Ernst Pascual (1902–1980), German physicist 319, 454, 456, 477

Jordanus Nemorarius (Jordanus de Nemore) (fl. ca. 1220), author of a variety of works on geometry, arithmetic, mechanics (likely not the same individual as Jordan of Saxony, deacon of the Dominican Order in Paris) 128, **142–43**, 144, 208, XIII

Josephson, Brian David (b. 1940), English physicist, Nobel Laureate (1973) 494

Joubert, Jules-François (1834–1910), French scientist XIX

Joule, James Prescott (1818–1889), English physicist 323, 375, 376–77, 378, 379, 381, 387, XIX

Julian the Apostate (Julianus Apostata, Flavius Claudius Julianus) (331–363), Roman emperor, Neo-Platonist 102–3

Justinian I (Flavius Petrus Sabbatius Iustinianus) (482–565), Byzantine emperor, sponsor of the *Corpus Juris Civilis* 47, 116, 124, 133, VI

K

Kallmann, Hartmut (1896–1978), German physical chemist 532

Kaluza, Theodor Franz Eduard (1885–1954), German mathematician, physicist 574

Kamerlingh-Onnes, Heike (1853–1926), Dutch physicist, Nobel Laureate (1913) XXV

Kant, Immanuel (1724–1804), German philosopher, naturalist 78, 96, 178, 299, 320, 323, **326–30**, 403, 478, 540, 548, 551, 552, 560, 563, XVIII

Kapitsa, Pyotr Leonidovich (1894–1984), Soviet physicist, Nobel Laureate (1978) 494, XXIII

Kármán, Theodor von (Tódor Kármán) (1881–1963), Hungarian-born American physicist XXV

Kastler, Alfred (1902–1984), French physicist, Nobel Laureate (1966) 493

Kaufmann, Walter (1871–1947), German physicist 428

Kelvin, William Thomson, 1st Baron Kelvin (Lord Kelvin) (1824–1907), British physicist 222, 323, 339, 358, 373, 377, 378, 381, 382, 400, 405, 406, 545, 546, 558

Kendall, Henry Way (1926–1999), American physicist, Nobel Laureate (1990) 495

Kepler, Johannes (1571–1630), German astronomer, mathematician 13, 24, 42, 48, 62, 82, 153, 169, 175, 176, 185, 187, 188, **189–94**, 195, 197, 198, 199, 211, 212, 226, 240, 242, 256, 258, 260, 263, 264, 268, 272, 275, 289, 292, 302, 325, 330, 546, 552, 556, 563, XVII

Kerr, John (1824–1907), Scottish physicist XIX

Kerst, Donald William (1911–1993), American physicist 531

Keuffel, Jack Warren (1919–1974), American physicist 532

Kierkegaard, Søren (1813–1855), Danish philosopher 467

Kirchhoff, Gustav Robert (1824–1887), German physicist 344, 358, 399, 400, 407, 438, 440, 478

Klein, Felix Christian (1849–1925), German mathematician 402, 467, 475, 476, 535

Klein, Oscar Benjamin (1894–1977), Swedish physicist 574

Kleist, Ewald Jürgen von (1700–1748), German jurist, cleric, physicist 333

Klitzing, Klaus von (b. 1943), German physicist, Nobel Laureate (1985) 495, 499

Kneller, Sir Godfrey (Gottfried Kniller) (1646–1723), German-born English painter XVII

Knoll, Max (1897–1969), German electrical engineer 499

Knudsen, Martin (1871–1948), Danish physicist, oceanographer 439

Koch, Robert (1843–1910), German bacteriologist 323

Kohlrausch, Rudolf Hermann Arndt (1809–1858), German physicist 361

Konopinski, Emil Jan (1911–1990), American physicist 537

Kowarski, Lev (1907–1979), Russian-born French physicist 518

Koyré, Alexandre (1892–1964), French historian of science 159, IX

Kramers, Hendrik Anthony (1894–1952), Dutch physicist 456

Krönig, August Karl (1822–1879), German physicist 377

Kugler, Franz Xaver (1862–1929), German mathematician, chemist, Assyriologist, Jesuit priest 41, 42

Kuhn, Richard (1900–1967), German–Austrian chemist, Nobel Laureate (chemistry 1938) 522

Kuhn, Thomas Samuel (1922–1996), American historian of science, philosopher 17, 18, 21, 565

Kurlbaum, Ferdinand (1857–1927), German physicist 442, 443

Kusch, Polykarp (1911–1993), American physicist, Nobel Laureate (1955) 477, 493

L

Laborde, Albert (1878–1968), French physical chemist 503, 505

Lagrange, Joseph Louis, Comte de (1736–1813), French mathematician, physicist, astronomer 162, 296, 304, 311, 316, 318, 319, 352, 370, 372, 400, 406, 464

Lalande, Joseph Jérôme (1732–1807), French astronomer 174

Lamarck, Jean Baptiste de Monet (1744–1829), French naturalist 323

Lamb, Willis Eugene (1913–2008), American physicist, Nobel Laureate (1955) 477, 493

Landau, Lev Davidovich (1908–1968), Soviet physicist, Nobel Laureate (1962) 467, 493, 499, 543, 556, XXV

Landé, Alfred (1888–1975), German-born American physicist 491

Langevin, Paul (1872–1946), French physicist 392, 429, 511, XXV

Lao Tse (Laozi) (fl. ca. 5th century bce), Chinese philosopher, founder of Taoism 545

Laplace, Pierre zSimon, Marquis de (1749–1827), French mathematician, astronomer, physicist 25, 316, 329, 340, **341**, 346, 352, 370, 372–73, 374, 400, 407, 480, 486, 552

Laporte, Otto (1902–1971), American physicist 540

Lattes, César Mansueto Giulio (1924–2005), Brazilian physicist 528

Laue, Max von (1879–1960), German physicist, Nobel Laureate (1914) 147, 426, 427, 450, 472, 486, 489, 491, 523, XXV

Laurentius de Voltolina (fl. 2nd half 14th century), Italian artist IX

Lavoisier, Antoine Laurent (1743–1794), French chemist 18, 240, 367, 387, 388, 496

Lawrence, Ernest Orlando (1901–1958), American physicist, Nobel Laureate (1939) 487, 492, 509, 511, 521, 539

Lederman, Leon Max (b. 1922), American physicist, Nobel Laureate (1988) 495, 537

Lee, Tsung-Dao (b. 1926), Chinese-born American physicist, Nobel Laureate (1957) 493, 528, 535, 539, 540, 542, 543, 544, 545, XXI

Leeuwenhoek, Antonie Philips van (1632–1723), Dutch naturalist 241, 242

Legendre, Adrien Marie (1752–1833), French mathematician 370

Leibniz, Gottfried Wilhelm, Freiherr von (1646–1716), German polymath 242, 245, 256, 272, 273, 275, 282, 290, 291, 294, **295–96**, 297–98, 299, 300, 302, 303, 306, 311, 317, 329, 365, 400, 416, 437, 445, 540, 547, 577, XVII

Lemaître, Georges Henri Joseph Édouard (1894–1966), Belgian astrophysicist 557

Lenard, Philipp Eduard Anton (1862–1947), German physicist, Nobel Laureate (1905) 360, 390, 439, 443, 448–49, 490

Lenin, Vladimir Ilyich (1870–1924), Soviet politician 407

Lenz, Heinrich Friedrich Emil (1804–1865), Baltic German physicist 353

Leonardo da Vinci (1452–1519), Italian artist, naturalist 38, 128, 139, 157, **162–64**, 165, 166, 168, 176, 409, XIII, XX

Leontius of Byzantium (Leontios) (d. 705), Byzantine emperor from 695 to 698 VI

Lesser, Aleksander (1814–1884), Polish painter XIV

Leucippus of Miletus (fl. ca. 460 bce), Greek philosopher 47, 48, 61, 552

Leutwyler, Heinrich (b. 1938), Swiss theoretical physicist 566

Lewis, Gilbert Newton (1875–1946), American physical chemist 422

Libby, Willard Frank (1908–1980), American chemist, Nobel Laureate (chemistry, 1960) 514

Lichtenberg, Georg Christoph (1742–1799), German physicist, writer 329

Liebig, Justus Freiherr von (1803–1873), German chemist 376

Lifschitz, Evgeny Mikhailovich (1915–1985), Soviet physicist XXV

Lincoln, Abraham (1809–1965), American statesman, president (1860–1865) 323
Lindemann, Carl Louis Ferdinand von (1852–1939), German mathematician 96
Linnaeus, Carl (Carolus Linnaeus, Carl von Linné) (1707–1778), Swedish naturalist 323, 328
Lipmann, Fritz Albert (1899–1986), German-born American biochemist, Nobel Laureate (physiology or medicine, 1953) 522
Lippmann, Gabriel (1845–1921), French physicist, Nobel Laureate (1908) 490
Liszt, Franz (Ferenc) (1811–1886), Hungarian pianist, composer 323
Livy (Titus Livius) (59 bce–17 ce), Roman historian 47
Llull, Ramon (Raymond Lully, Raymundus Lullus or Lullius, Raimundo Lulio) (1232–1315), Spanish philosopher, theologian 126
Lobachevsky, Nikolai Ivanovich (1792–1856), Russian mathematician 95, 97, 341, 401–2, 429, 430, 431
Locke, John (1632–1704), English philosopher 298–301, 312, 320, 325, 326
Lodge, Sir Oliver Joseph (1851–1940), English physicist 399
Lomonosov, Mikhail Vasilyevich (1711–1765), Russian polymath 225, **314–15**, 316, 323
London, Fritz Wolfgang (1900–1954), German physicist, brother of Heinz London XXV
London, Heinz (1907–1970), German-born English physicist XXV
Lorentz, Hendrik Antoon (1853–1928), Dutch physicist, Nobel Laureate (1902) 362, **363–64**, 398, 409, **416–22**, 423, 424, 425, 428, 433, 447, 453, 460, 467, 470, 475, 490, XIX, XXI
Lorenz, Edward Norton (1917–2008), American meteorologist XXI
Lorenz, Konrad Zacharias (1903–1989), Austrian ethologist, Nobel Laureate (physiology or medicine, 1973) 24
Loschmidt, Joseph (1821–1895), Austrian physicist 386
Lothar I (795–855), emperor of the Romans 117, 118
Louis I, the Great, of Hungary (1326–1382), king of Hungary and Poland, father of Jadwiga 221, 225
Louis the German (Ludwig, Lewis, Louis the Bavarian) (806–876), king of Bavaria and of East Francia 117, 118
Louis the Pious (the Fair, the Debonair) (778–840), son of Charlemagne, king of Aquitaine from 781, coemperor (as Louis I) and king of the Franks with Charlemagne 117
Louis XIV (1638–1715), king of France, "Sun King" 241
Lucretius (Titus Lucretius Carus) (ca. 99–ca. 55 bce), Roman poet, philosopher 4, 47, 49, 61–62, **102**, 107–8, 130, 134, 158, 237, 330, 389, 552, V
Lully, Jean Baptiste (Giovanni Battista di Lulli) (1632–1687), Italian-born French composer 242
Lully, Raymond. *See* Llull
Lummer, Otto (1860–1925), German physicist 442, 443
Luther, Martin (1483–1546), German religious reformer, founder of German Protestantism 128, 171, 177

M

Mach, Ernst (1838–1916), Austrian physicist, philosopher 18, 156, 163, 273, 274, 328, **407–8**, 416, 417, 434, 458, 467, 478, 479, 505, 563
Maclaurin, Colin (1698–1746), English mathematician 372
Maestlin, Michael (Mästlin, Möstlin, Moestlin) (1550–1631), German astronomer, mentor of Kepler 188
Magnus, Albertus. *See* Albertus Magnus
Maier, Annelise (1905–1971), German historian of science 140
Maimonides, Moses (Rabbi Moses ben Maimon, Musa ibn Maymun) (1135–1204), physician, scholar, most important Jewish religious philosopher of the Middle Ages 66–67, 132–35, 559, 573
Malebranche, Nicolas (1638–1715), French philosopher 297, 300
Malesherbes, Chrétien Guillaume de Lamoignon de (1721–1794), French statesman, supporter of the *Encylopédie*, great-grandfather of Alexis de Tocqueville 321
Malpighi, Marcello (1628–1694), Italian naturalist 241
Malthus, Thomas Robert (1766–1834), British economist, social philosopher 326
Malus, Étienne Louis (1775–1812), French physicist 360
Mandelbrot, Benoît (1924–2010), Polish-born Mathematician, engineer, active in France and the United States 484
Mann, Thomas (1875–1955), German writer 31, 409, 411
Marcellus, Marcus Claudius (ca. 268–208 bce), Roman general 81–82, 84
Marci, Jan Marek (Johannes Marcus Marci) (1595–1677), Bohemian naturalist 285
Marconi, Guglielmo Marchese (1874–1937), Italian engineer, physicist, Nobel Laureate (1909) 361, 490
Marcus Aurelius (Marcus Aurelius Antoninus Augustus) (121–180), Roman emperor, Stoic philosopher 47, 100, 104–5
Maricourt, Peter of (Petrus Peregrinus de Maricourt, Pierre Pelerin de Maricourt) (fl. ca. 1250), French scholar, naturalist 128, 155, **329–30**, 331
Mariotte, Edme (ca. 1620–1684), French physicist 24, 236, 237, 253, 254, 259, 373
Mark Antony. *See* Antony, Mark
Marsden, Ernest (1889–1970), English physicist 397
Marshak, Robert (1916–1992), American physicist 535
Marx, György (1927–2002), Hungarian physicist 537
Marx, Karl (1818–1883), German philosopher, politician 299, 323, 409
Matilda of Flanders (c. 1031–1083), wife of William the Conqueror 118
Matthews, Paul Taunton (1919–1987), English physicist 550
Matthias Corvinus (Matthias the Just) (1443–1490), king of Hungary, humanist, patron of art and science 128, 158, 165, 224
Maupertuis, Pierre Louis Moreau de (1698–1759), French mathematician 149, 156, 233, 274, 302, 303, 306, **309–11**, 323
Maximian (Marcus Aurelius Valerius Maximianus Herculius) (ca. 250–ca. 310), Roman emperor, coruler with Diocletian 115
Maxwell, James Clerk (1831–1879), British physicist 4, 5, 8, 12, 25, 26, 323, 331, 332, 340, 346, 347, 348, 351, 352, **354–59**, 360, 362, 363, 366, 377, 378, 381, 382, 385, 386, 387, 392, 400, 405, 406, 411, 417, 418, 421, 442, 447, 482, 546, 568, 575, XIX
Mayer, Julius Robert von (1814–1878), German physician, naturalist 376, 377, 380
McKinsey, John Charles Chenoweth (1908–1953), English physicist 13, 14
McMillan, Edwin Mattison (1907–1953), American physical chemist, Nobel Laureate (chemistry, 1951) 511, 531
Méchain, Pierre François André (1744–1804), French astronomer 447
Medici, Cosimo di Giovanni de (1389–1464), Florentine banker, founder of the Medici dynasty, founder of the Platonic Academy 132
Meer, Simon van der (b. 1925), Dutch physicist, Nobel Laureate (1984) 495
Mehra, Jagdish (b. 1931), Indian-born American physicist, historian of science 432
Meitner, Lise (1878–1968), Austrian-born Swedish physicist 487, 504, 516, 518, 519
Menaechmus (380–320 bce), Greek mathematician 95
Mencke, Otto (1644–1707), German philosopher, scientist 282
Mendel, Gregor Johann (1822–1884), Austrian naturalist, abbot of the Brunn Augustinian monastery, founder of modern genetics 323
Mendeleev, Dmitri (1834–1907), Russian chemist 323, 393, 394–95, 396, 497, XXIII
Menelaus of Alexandria (Μενελαος) (fl. ca. 100), Hellenistic mathematician 97
Mengs, Anton Raphael (1728–1779), German painter 329
Mercator, Gerardus (Gheert Cremer) (1512–1594), Flemish geographer, cartographer 162
Merovingians (482–751), Frankish dynasty 117
Mersenne, Marin (1588–1648), French mathematician, music theorist, scientist 218, 230, 241, 275, 341

Methodius (826–885), Greek missionary to the Slavs, brother of Cyril 132
Meton of Athens (Μετων) (fl. ca. 430 bce), Greek astronomer, mathematician 41
Meyer, Julius Lothar (1830–1895), German chemist 394, 395, 396
Michelangelo (Michelangelo di Lodovico Buonarroti Simoni) (1475–1564), Italian sculptor, painter, architect, poet 25, 128, 157
Michell, John (1724–1793), English geologist, inventor of the torsion balance 338
Michelson, Albert Abraham (1852–1931), American physicist, Nobel Laureate (1907) 360, 405, 411, 412, 416, 421, 443, 447, 490
Miller, Daniel Weber (b. 1926), American physicist 537
Millikan, Robert Andrews (1868–1953) American physicist, Nobel Laureate (1923) 69, 347, 396, 491
Minkowski, Hermann (1864–1909), German mathematician, physicist 418, 421, 426, 428
Mohammed (ca. 570–632), founder of Islam 116, 128, 141
Molière (Jean-Baptiste Poquelin) (1622–1673), French comedist 242, XV
Monet, Claude (1840–1926), French painter 302
Monge, Gaspard (1746–1818), French mathematician 161, XX
Montesquieu, Charles de Secondat (1689–1755), French writer, political theorist 302, 322
Montgolfier, Joseph-Michel (1740–1810) and Jacques-Étienne (1745–1799), French inventors, pioneers in flight 328
Morley, Edward Williams (1838–1923), American physical chemist 411, 412, 416
Morse, Samuel Finley Breese (1791–1872), American painter, inventor XIX
Moseley, Henry Gwyn Jeffrey (1887–1915), English physicist 398, 400, 504
Mössbauer, Rudolf (b. 1929), German physicist, Nobel Laureate (1961) 493
Mott, Sir Nevill Francis (1905–1996), British physicist, Nobel Laureate (1977) 487, 494
Mottelson, Ben Roy (1926–1981), Danish physicist, Nobel Laureate (1975) 494, 514
Mozart, Wolfgang Amadeus (Johann Chrysostom Wolfgang Amadeus Mozart) (1756–1791), Austrian composer 323, XVIII
Muller, Hermann Joseph (1890–1967), American biologist, Nobel Laureate (physiology or medicine 1946) 522
Müller, Karl Alexander (b. 1927), Swiss physicist, Nobel Laureate (1987) 495, XXV
Müller, Paul Hermann (1899–1965), Swiss chemist, Nobel Laureate (physiology or medicine 1948) 522
Mulliken Robert Sanderson (1896–1986), American physicist, chemist, Nobel Laureate (chemistry, 1966) 453
Mumford, Lewis (1895–1990), American historian of technology and science, literary critic 43
Murray, Francis Joseph (1911–1996), American mathematician 414, 483
Musschenbroek, Pieter van (1692–1761), Dutch naturalist 333, 334, 335, 337

N

Napoleon I, Bonaparte (1763–1821), emperor of France 341, 343, 369, XVIII
Nasr, Seyyed Hossein (b. 1933), Iranian-born American philosopher of religion X
Navier, Claude Louis Marie Henri (1785–1836), French engineer 303, 304
Nebuchadnezzar II (ca. 630–562 bce), ruler of Babylon (Chaldean dynasty) 47
Neddermeyer, Seth Henry (1907–1988), American physicist 527
Néel, Louis Eugène Félix (1904–2000), French physicist, Nobel Laureate (1970) 494
Nelson, Horatio (1758–1805), English admiral XVIII
Nernst, Hermann Walther (1864–1941), German physical chemist 380
Neumann, Franz Ernst (1798–1895), German mineralogist, physicist (inductivity formula) 360
Neumann, John von (János) (1903–1957), Hungarian-born American mathematician 386, 457, 458, 471, **481–83**, 489, 577, XXI
Newcomen, Thomas (1664–1729), English ironmonger, developed the steam engine 367
Newton, Sir Isaac (1643–1727), English mathematician, physicist, astronomer 4, 5, 8, 17, 18, 24, 26, 66, 71, 76, 96, 149, 164, 174, 185, 193, 194, 212, 213, 219, 226, 228, 237, 240, 241, 242, 243, 245, 252, **255–77**, 282, 283, **284–89**, 290, 291, 292, **294–96**, 298, 299, 300, 301, 302, 303, 304, 305, 306, 308, 309, 310, 311, 314, 316, 317, 322, 323, 325–26, 328, 329, 330, 331–32, 333, 335, 337, 338, 341, 346, 359, 360, 363, 365, 398, 407, 409, 410, 416, 431, 436, 437, 445, 447, 464, 475, 482, 538, 547, 550, 552, 563, XVII, XX, XXIII
Nicephorus Blemmydes (1197–1272), Byzantine polymath 131, VI
Nicholas of Autrecourt (Nicolaus, Nicholas d'Autrecourt) (ca. 1299–ca. 1369), French philosopher, theologian 155, 158, IX
Nicholas of Kues (Nicolaus Cusanus, Nicholas of Cusa) (1401–1464), German theologian, naturalist, philosopher 128, 156–57, 159, 160, 161, 166, 174, XVI
Nicole Oresme (Nicolas Oresme, Nicholas Oresme, Nicolas d'Oresme) (ca. 1323–1382), French scholar **143–44**, 145, 146, 148, **149–51**, 197, 291
Nifo, Agostino (ca. 1473–1538), Italian philosopher 158
Nishijima, Kazuhiko (b. 1926), Japanese physicist 537, 540, 545
Nishina, Yoshio (1890–1951), Japanese physicist 476
Noble, Sir Andrew (1831–1915), Scottish physicist 414, 415, 421
Noddack, Eva Ida (Tacke-Noddack) (1896–1978), German chemist 515
Noël, Étienne (1581–1659), French natural philosopher, Jesuit 235
Noether, Amalie "Emmy" (1882–1935), German mathematician 539
Nollet, Jean Antoine (1700–1770), French physicist 333, 334
Nordheim, Lothar Wolfgang (1899–1985), German physicist 458
Norman, Robert (fl. ca. 1590), English mariner, compass builder, hydrographer 175
Norwood, Richard (1590–1675), English mathematician, astronomer 261
Nuttall, John Mitchell (1890–1958), English physicist 507, 508, 510

O

Occhialini, Giuseppe (1907–1993), Italian physicist 526, 528
Ockham, William of (William of Occam) (ca. 1288–ca. 1348), English theologian, philosopher 128, 156, IX, XII
Odoacer (Odovacar) (435–493), Germanic general, ruler of Italy 116
Ohm, Georg Simon (1789–1854), German physicist 339, 343–44
Olbers, Heinrich Wilhelm Matthäus (1758–1840), German physician, astronomer 562
Omar Khayyám (1048–1122), Persian astronomer, mathematician, philosopher, poet 128, 138
Onsager, Lars (1903–1976), Norwegian-born American chemist, Nobel Laureate (chemistry, 1968) 380
Oppenheimer, Robert (1904–1967), American physicist 476, 556
Origen (Origen Adamantius) (ca. 185–ca. 254), Greek theologian X
Ørsted, Hans Christian (1777–1851), Danish physicist, chemist 8, 18, 323, 334, 345–46, 357, 363
Osiander, Andreas (Andreas Hosemann) (1498–1552), German Lutheran theologian 178, 179, 183
Ostwald, Wilhelm Friedrich (1853–1932), German physical chemist 394, 408, 505
Otto I, the Great (912–973), Holy Roman emperor 117, 118
Ovid (Publius Ovidius Naso) (43 bce–17 or 18 ce), Roman poet 47
Owens, Robert Bowie (1870–1940), American electrical engineer 503

P

Pacinotti, Antonio (1841–1912), Italian physicist, electrical engineer 353
Pacioli, Fra Luca Bartolomeo de (ca. 1446–1517), Italian mathematician 65, 142, XIII

Palestrina, Giovanni Pierluigi da (1525 or 1526–1594), Italian composer 128

Pappus of Alexandria (ca. 290–ca. 350), Greek mathematician, geographer 47, 48, 142, 218

Paracelsus, Philippus Aureolus (Theophrastus Bombastus von Hohenheim) (1493–1541), Swiss physician, iatrochemist of Swabian origin 63, 238

Parmenides of Elea (fl. ca. 500 bce), Greek philosopher 32, 47, 48, 49, 55, 60, 61, 62, 102, 387

Pascal, Blaise (1623–1662), French mathematician, physicist, philosopher 97, 100, 234, **235–36**, 237, 242, 257, 293, 296, 297, 324, 549, 577, 579

Pasteur, Louis (1822–1895), French chemist, microbiologist 323, 541, XIX

Patrick (Naomh Pádraig, Maewyn Succat, Magonus Succetus) (fl. ca. 450), Roman Britain–born Christian missionary, patron saint of Ireland 119

Paul III (Alessandro Farnese) (1468–1549), pope 178

Paul, Wolfgang (1913–1993), German physicist, Nobel Laureate (1989) 495

Paul of Venice (Paulus Venetus, Paolo Nicoletti) (ca. 1368–1428), Italian Scholastic 152

Pauli, Wolfgang (1900–1958), Austrian-born Swiss physicist, Nobel Laureate (1945) 400, 409, 418, 422, 428, 451, 453, 457, 467, 470, **474–75**, 477, 482, 487, 490, 491, 523, 524, 525, 527, 542, XXI, XXV

Pázmány, Péter (1570–1637), Hungarian cardinal, founder of the university at Nagyszombat (later Budapest University) 224

Peierls, Sir Rudolf Ernst (1907–1995), German-born English physicist 487, 515, 523, XXV

Pemberton, Henry (1694–1771), English physicist 264

Penzias, Arno Allen (b. 1933), German-born American astrophysicist, Nobel Laureate (1978) 550

Perepelkin, Yuri Yakovlevich (1903–1982), Russian mathematician, Egyptologist 37

Pericles (Perikles) (ca. 495–429 bce), Athenian statesman 5, 46, 47, 48, II

Périer, Florin (1605–1702), French physicist, brother-in-law of Blaise Pascal 236

Perlmutter, Saul (born 1959), American astrophysicist 563

Pérot, Jean-Baptiste Gaspard (1863–1925), French physicist 447

Perrin, Jean-Baptiste (1870–1942), French physical chemist, Nobel Laureate (1926) 491, 505

Peter Damian (Petrus Damiani, Pietro Damiani, Pier Damiani) (ca. 1007–1072), Italian Scholastic, doctor of the Church 153

Peter I, the Great (1672–1725), Russian tsar 329

Peter Lombard (Petrus Lombardus) (ca. 1100–1160), Italian Scholastic theologian, bishop of Paris, "Magister Sententiarum" 144, 167

Peter of Vaux de Cernay (fl. ca.1215), Cistercian monk, chronicler of the Albigensian Crusade 129

Petrarch (Francesco Petrarca) (1304–1374), Italian poet, humanist 128

Petrus Hispanus (Pedro Julião, Pope John XXI) (ca. 1215–1277), professor of medicine in Siena, physician, scholar 152, 153

Peuerbach, Georg von (Purbach, Peurbach, Purbachius) (1423–1461), Austrian mathematician, astronomer 128, 165, 166

Phidias (Pheidias) (ca. 480–ca. 430 bce), Greek sculptor 47

Phidias (fl. 4th to 3rd century bce), father of Archimedes, an astronomer about whom otherwise nothing is known 82

Philip II (1527–1598), Spanish king 171

Philip II of Macedon (382–336 bce), Macedonian king, father of Alexander the Great 96

Phillips, John Gardner (1917–2001), American astronomer 550

Philo of Alexandria (Philo Judaeus) (20 bce–50 ce), Hellenistic philosopher of Jewish background 47, 48, 102

Philolaus of Croton (ca. 480–ca. 385 bce), Greek philosopher 48, 51, 72, 73

Philoponus, John (Johannes Philoponos) (ca. 490–ca. 570), Late-Greek theologian, physicist, commentator on Aristotle 47, 48, 131, 146

Photios I of Constantinople (ca. 810–ca. 893), Byzantine scholar, patriarch of Constantinople 132

Piaget, Jean (1896–1980), Swiss psychologist 18

Piccard, Auguste Antoine (1884–1962), Swiss physicist 487

Piccolomini, Enea Silvio (Aeneas Sylvius) (1405–1464), pope as Pius II XIV

Pickering, Andrew (b. 1948), English historian, philosopher, sociologist of science 564

Pictet, Marcus Auguste (1752–1825), Swiss-born French physicist 344

Pixii, Hyppolite (1808–1835), French electrical engineer, inventor 353

Planck, Max (1858–1947), German physicist, Nobel Laureate (1918) 26, 328, 355, 364, 380, 405, 406, 417, 421, 422, 426, 428, 438, 439, 440, **441–44**, 445, 446, 447, 448, 451, 460, 465, 470, 479–80, 482, 490, 491, 495, 570, 571, 572, XXI, XXII

Planté, Gaston (1834–1889), French physicist XIX

Plato (427–347 bce), Greek philosopher 2, 5, 29, 46, 47, 48, 51, 54, **56–60**, 61, **62**, **63–65**, 66, 74, 78, 79, 95, 96, 102, 130, 131, 132, 134, 149, 151, 153, 160, 173, 174, 176, 194, 331, 545, 563, 565, I, VI, X, XII, XIII

Pletho, George Gemistus (Plethon) (1355–1452), Byzantine philosopher 132, VI

Pliny the Elder (Gaius Plinius Secundus) (23–79), Roman official, writer, author of the book series *Naturalis Historia* 47, 92, 102, 106

Pliny the Younger (Gaius Plinius Caecilius Secundus) (61 or 62–113), Roman politician, writer 106

Plotinus (ca. 204–270), Greek philosopher, founder of Neo-Platonism 47, **102–3**

Plücker, Julius (1801–1868), German mathematician, physicist 388

Plutarch (Lucius Mestrius Plutarchus) (ca. 46–120), Greek historian, moral philosopher 47, 55, 76, 79–81, 83–84, 90, XIV

Podolsky, Boris (1896–1966), Russian-born American physicist 485

Pogány, Béla (1887–1943), Hungarian physicist 411

Poincaré, Jules Henri (1854–1912), French mathematician, physicist, philosopher 10, 99, 358, 386, 400, 401, 409, 416, 418, **419–22**, 426–27, 428, 430, 432, 438, 453, 460, 486, 498, 539, 563, XIX, XXI

Poinsot, Louis (1777–1859), French mathematician, physicist 65

Poisson, Siméon Denis (1781–1840), French mathematician, physicist 304, 319, 323, 339, 340, 352

Politzer, Hugh David (b. 1949), American theoretical physicist, Nobel Laureate (2004) 567

Polo, Marco (1254–1324 or 1325), Venetian traveler, merchant, explorer 128

Pompadour, Jeanne Antoinette Poisson, Marquise de (1721–1764), mistress of Louis XV, patron of art, science, and literature XVIII

Popov, Alexander Stepanovich (1859–1906), Russian physicist, electrical engineer 361, XIX

Popper, Sir Karl Raimund (1902–1994), Austrian-born English philosopher 3, 565

Porphyry of Tyre (ca. 233–ca. 309), Greek Neo-Platonic philosopher 78–79, 126

Porter, Sir George, Baron Porter of Luddenham (1920–2002), English physical chemist, Nobel Laureate (chemistry, 1967) 514

Posidonius of Apameia (ca. 135–51 bce), Greek philosopher, astronomer 93, 94, 103

Poussin, Nicolas (1594–1665), French painter 242

Powell, Cecil Frank (1903–1969), English physicist, Nobel Laureate (1950) 492, 528, 532

Prandtl, Ludwig (1875–1953), German physicist, "father of aerodynamics" 303

Praxiteles (fl. ca. 350 bce), Greek sculptor 47

Price, Bartholomew (1818–1862), English mathematician 339

Price, Derek John de Solla (1922–1983), physicist, historian of science, information scientist 102

Priestley, Joseph (1733–1804), British naturalist, chemist, electrical engineer 4, 332, 335, 336, 338, 339, 341, 349

Prigogine, Ilya, Viscount (Илья Романович Пригожин) (1917–2003), Russian-born Belgian physical chemist 380

Pringsheim, Ernest (1859–1917), German physicist 442, 443

Proclus Lycaeus (412–485), Greek philosopher, representative of Neo-Platonism 47, 53, 97, 415

Prokhorov, Alexander Mikhailovich (1916–2002), Soviet physicist, Nobel Laureate (1964) 493, XXV

Protagoras (ca. 490–420 bce), Greek philosopher, Sophist 47, 55, 563

Proust, Joseph Louis (1754–1826), French chemist 388

Prout, William (1785–1850), British physician, chemist 393, 506

Psellus, Michael (Psellos) (1017 or 1018–after 1078), Byzantine polymath VI

Ptolemy (Claudius Ptolemaeus) (after 83–ca. 168), Alexandrian astronomer, geographer 35, 40, 47, 48, 72, 74, **89–90**, **92**, 93, 94, 97, 108–9

Ptolemy II Philadelphus (309–246 BCE), founder of the Library of Alexandria 105

Ptolemy I Soter, "The Savior" (ca. 366–283 BCE), general under Alexander the Great, king of (Hellenistic) Egypt, founder of the Ptolemaic dynasty 105

Purcell, Edward Mills (1912–1997), American physicist, Nobel Laureate (1952) 492

Pushkin, Alexander Sergeevich (Александр Сергеевич Пушкин) (1799–1837), Russian poet 323

Pyrrho of Elis (ca. 360–ca. 270 BCE), Greek Skeptic philosopher 100

Pythagoras of Samos (ca. 580–ca. 500 BCE), Greek philosopher, founder of the religious–political society of the Pythagoreans 29, 47, 48, **50–54**, 72, 96, 162, 167, 176, 272, X

Q

Quesnay, François (1694–1774), French economist 322

R

Rabanus Maurus Magnentius (Hrabanus) (ca. 780–856), German scholar (Primus Germaniae Praeceptor) 119, 128

Rabi, Isidor Isaac (1898–1988), Polish-born American physicist, Nobel Laureate (1944) 491–92

Rainwater, Leo James (1917–1986), American physicist, Nobel Laureate (1975) 494

Raman, Sir Chandrasekhara Venkata (1888–1970), Indian physicist, Nobel Laureate (1930) 491

Ramsey, Norman Foster, Jr. (b. 1915), American physicist, Nobel Laureate (1989) 495, 542

Rankine, William John Macquorn (1820–1872), British engineer 381, XX

Raphael (Raffaello Sanzio) (1483–1520), Italian painter, architect 157

Rayleigh, John William Strutt, 3rd Baron (Lord Rayleigh) (1842–1919), British physicist 339, 375, 376, 392, 441–42, 444, 446, 490

Réaumur, René Antoine Ferchault de (1683–1757), French naturalist 336

Regiomontanus (Johannes Müller) (1436–1476), German astronomer, mathematician 128, 165, 168, 177

Regius, Henricus (1598–1679), Dutch philosopher 298

Reines, Frederick (1918–1998), American physicist, Nobel Laureate (1995) 527, 532, 535

Reisch, Gregor (ca. 1467–1525), German polymath 92, 167

Rembrandt (Rembrandt Harmenszoon van Rijn) (1606–1669), Dutch painter 242

Retherford, Robert Curtis (1912–1981), American physicist 477

Reynolds, Osborne (1842–1912), British physicist, engineer 303

Rheticus (Georg Joachim von Lauchen) (1514–1574), German astronomer, mathematician 176, 177, 178

Rhind, Alexander Henry (1833–1863), Scottish archeologist 35

Ricci, Michelangelo (1619–1682), Italian mathematician, Roman Catholic cardinal 256

Riccioli, Giovanni Battista (1598–1671), Italian theologian, astronomer 187

Richardson, Sir Owen Williams (1879–1959), English physicist, Nobel Laureate (1928) 400, 487, 490, 491, 496

Richman, Georg Wilhelm (Георг Вильгельм Рихман) (1711–1753), German-born Russian physicist 333, 336

Richter, Burton (b. 1931), American physicist, Nobel Laureate (1976) 494, 548

Riemann, Georg Friedrich Bernhard (1826–1866), German mathematician 349, 429, 430, 433, 436

Ritter, Johann Wilhelm (1776–1810), German physicist, chemist 398

Ritz, Walter (1878–1909), German mathematician, physicist 400, 439, 517

Robert of Chester (Robertus Castrensis) (fl. ca. 1150), English Arabist 134

Roberval, Gilles Personne de (1602–1675), French mathematician 291, 294

Robison, John (1739–1805), English naturalist, publisher of Black's lectures 338, 366, 367

Robson, James Michael (James Michael "Rab" Rabinovich) (1900–1982), Belgian-born Canadian geneticist, physicist, of Russian–Jewish origin 537

Rochester, George Dixon (1908–2001), English physicist 528

Roentgen, Wilhelm Conrad (Röntgen) (1845–1923), German physicist, Nobel Laureate (1901) 323, 391–92, 393, 498

Rohrer, Heinrich (b. 1933), Swiss physicist, Nobel Laureate (1986) 495, 499

Rømer, Ole Christensen (Römer) (1644–1710), Danish astronomer 230, 283, 284

Romulus Augustus (Romulus Augustulus) (fl. 463–476), the last Western Roman emperor 116

Roosevelt, Franklin Delano (1882–1945), president of the United States of America (1933–1945) 523

Roscellinus (Roscelin of Compiègne) (ca. 1050–ca. 1120), French theologian, philosopher XII

Rosen, Nathan (1909–1995), American-born Israeli physicist 485

Rosenfeld, Leon (1904–1974), Belgian physicist 487

Rousseau, Jean-Jacques (1712–1778), Swiss-born French writer, moral philosopher, social theorist 235, 321, 323

Rovelli, Carlo (b. 1956), Italian physicist 573

Rubbia, Carlo (b. 1934), Italian physicist, Nobel Laureate (1984) 566

Rubens, Heinrich (1865–1922), German physicist 442, 443

Rudolf I (Rudolph of Habsburg) (1218–1291), king of the Romans 129

Rudolf II (1552–1612), king of Hungary and Bohemia, Holy Roman emperor 188, 189

Ruhmkorff, Heinrich Daniel (1803–1877), German maker of scientific instruments XIX

Rumford, Sir Benjamin Thompson, Count (1753–1814), Anglo-American and German military advisor, organizer of social and scientific institutions **368–69**, 371, 373

Ruska, Ernst A. F. (1906–1988), German electrical engineer, Nobel Laureate (1986) 495, 499

Russell, Bertrand Arthur William, Lord (1872–1970), British mathematician, philosopher, Nobel Laureate (literature 1950) 2, 49, 96, 325, 563, 564, 565

Russell, Henry Norris (1877–1957), American astronomer 552, 554, 555, 558

Rutherford, Ernest (Lord Rutherford of Nelson) (1871–1937), British physicist, Nobel Laureate (chemistry 1908) 388, 396, 400, 439, 449, 450, 487, 489, 490, 496, 498, **501–9**, 514, 520, 532, XIX, XXIII

Ružička, Leopold (1887–1976), Croatian-born Swiss chemist, Nobel Laureate (chemistry 1939) 522

Rydberg, Janne Robert (1854–1919), Swedish physicist 399–400, 439

Ryle, Sir Martyn (1918–1984), British astrophysicist, Nobel Laureate (1974) 494

S

Sacrobosco. *See* Johannes de Sacrobosco

Sakurai, Jun John (1933–1982), Japanese-born American physicist 535

Salam, Abdus (1926–1996), Pakistani-born English physicist, Nobel Laureate (1979) 494, 543, 548, 566, 569

Salpeter, Edwin Ernest (1924–2008), American astrophysicist 554, 556

Sandage, Allan Rex (b. 1926), American astronomer 550

Santillana, Giorgio Diaz de (1902–1974), Italian-born American philosopher, historian of science 74, 103, 113, 199

Sarton, George Alfred Leon (1884–1956), American historian of science 3, 128

Sauveur, Joseph (1653–1716), French mathematician, physicist 241

Savart, Félix (1791–1841), French physicist 346

Savić, Pavle (1909–1994), Serbian physical chemist 516

Scheiner, Christoph (1579–1650), German physicist, astronomer, Jesuit 195, 227, 229

Schelling, Friedrich Wilhelm Joseph von (1775–1854), German philosopher 345
Schiller, Johann Christoph Friedrich von (1759–1805), German writer 323, XVIII
Schlick, Moritz (1882–1936), German philosopher 563
Schmidt, Gerhard Karl (1865–1949), German chemist 501, 504
Schmidt, Maarten (b. 1929), Dutch-born American astronomer 550
Schooten, Franz van (ca. 1615–1660), Dutch mathematician 290
Schopenhauer, Arthur (1788–1860), German philosopher 102, 381
Schott, Gaspar (Casparus) (1608–1666), German mathematician, author of popular-scientific books 236, XI
Schrieffer, John Robert (b. 1931), American physicist, Nobel Laureate (1972) XXV
Schrödinger, Erwin (1887–1961), Austrian physicist, Nobel Laureate (1933) 4, 304, 312, 319, 399, 439, 445, 457, **458–65**, 467, 470, 471, 475, 477, 482, 484, 486, 490, XXI, XXIII
Schubert, Franz (1797–1828), Austrian composer 323
Schumacher, Benjamin Wade (b. 1960), American physicist 577
Schuster, Sir Franz Arthur Friedrich (1850–1934), German-born British electrical engineer, inventor, industrialist 390
Schwalow, Arthur Leonard (1921–1999), American physicist, Nobel Laureate (1981) 494
Schwartz, Melvin (1932–2006), American physicist, Nobel Laureate (1988) 495, 537
Schwarz, John Henry (b. 1941), American physicist 570, 571
Schwarzschild, Karl (1873–1916), German astronomer 431
Schwinger, Julian Seymour (1918–1994), American physicist, Nobel Laureate (1965) 364, 477, 524
Scot, Michael (Scotus) (1175–1232?), Scottish scholar, translator 134
Scott, Sir Walter (1771–1832), Scottish poet, novelist XVIII
Seaborg, Glenn Theodore (1912–1999), American chemist, Nobel Laureate (chemistry, 1951) 514
Seebeck, Thomas Johann (1770–1831), German physicist 344
Segrè, Emilio Gino (1905–1989), Italian-born American physicist, Nobel Laureate (1959) 457, 523
Seidel, Ludwig Philip von (1821–1896), German mathematician, astronomer 229
Selényi, Pál (1884–1954), Hungarian physicist 459
Seleucus I (Nicator) (ca. 358–281 bce), Macedonian officer under Alexander the Great, founder of the Seleucid dynasty 42
Seneca, Lucius Annaeus (Seneca the Younger) (ca. 4 bce–65 ce), Roman philosopher, writer, politician 47, 100, 101, 128, 130, 134, 151, 158, 167
Sextus Empiricus (ca. 160–210), Alexandrian, Roman, or Athenian physician, philosopher 55, 100
Shakespeare, William (1564–1616), English poet, dramatist 171
Shaw, George Bernard (1856–1950) Anglo-Irish dramatist, Nobel Laureate (literature, 1925) 323
Shelley, Mary (née Wollstonecraft Godwin) (1797–1851), British writer 562
Shelley, Percy Bysshe (1792–1822), English poet 323, XVIII
Shockley, William Bradford (1910–1989), American physicist, Nobel Laureate (1956) 400, 492, 493, XXI, XXV
Siegbahn, Kai Manne Börje (1918–2007), Swedish physicist, son of Karl Manne Georg Siegbahn, Nobel Laureate (1981) 494, 514
Siegbahn, Karl Manne Georg (1886–1978), Swedish physicist, Nobel Laureate (1924) 494, 514
Siemens, Werner von (1816–1892), German electrical engineer, inventor, industrialist 323, 353, 400
Siger of Brabant (ca. 1235–ca. 1282), Scholastic philosopher 153
Simeon I, the Great (864–927), ruler of Bulgaria 168
Simplicius of Cilicia (ca. 490–ca. 560), Neo-Platonist philosopher 198
Skobeltsyn, Dmitri Vladimirovich (1892–1990), Soviet physicist 532
Slater, John Clarke (1900–1976), American physicist 474, 489
Sluse, René-François de (Slusius) (1622–1685), French mathematician XVII
Smith, Adam (1723–1790), Scottish economist, moral philosopher 323
Smolin, Lee (b. 1955), American theoretical physicist 573
Smoluchowski, Marian von (1872–1917), Polish physicist 448
Snell, Arthur H. (1909–1989), American physicist 537
Snell, Willebrord (Willebrord Snellius) (1580–1626), Dutch mathematician, physicist, cartographer 24, 227, 248
Snow, Lord Charles Percy (1905–1980), English writer, politician, physicist 1, 2, 418
Socrates (ca. 470–399 bce), Greek philosopher 46, 47, 48, **56–60**, 102, 151, I
Soddy, Frederick (1877–1956), British physical chemist, Nobel Laureate (chemistry 1921) 496, 503, 504, 506, XXII
Soleil, Madame Germaine (1913–1996), French astrologer 109
Solvay Gaston Joseph (1838–1922), Belgian chemist, industrialist 438
Sommerfeld, Arnold (1868–1951), German mathematician, physicist 50, 408, 439, 450, 453, 471, XXV
Sophocles (ca. 496–406 bce), Greek tragedian 47, 180, II
Sorbon, Robert de (1201–1274), French canon, the name "Sorbonne" can be traced to the school founded by him in 1257 126, 128
Sosigenes (fl. ca. 70 bce), Alexandrian astronomer, mathematician 40
Soto, Domingo de (Domenicus de Soto) (1495–1560), Spanish theologian, philosopher 159, 163, 174
Spartacus (ca. 109–71 bce), Thracian-born Roman slave, leader of the largest slave revolt in the Roman Empire 47
Speusippus (407–339 bce), Greek philosopher, Plato's nephew 57
Spinoza, Baruch (Benedictus) de (1632–1677), Dutch philosopher 12, 78, 242, 298, 299, 419
Stahl, Georg Ernst (1660–1734), German physician, chemist 240
Stanley, Wendell Meredith (1904–1971), American biochemist, Nobel Laureate (chemistry, 1946) 522
Stark, Johannes (1874–1957), German physicist, Nobel Laureate (1919) 449, 456, 491
Staudinger, Hermann (1881–1965), German chemist, Nobel Laureate (chemistry 1953) 522
Stefan, Joseph (1835–1893), Austrian physicist 383, 439–40
Stefan, Uroš IV Dušan, the Mighty (ca. 1308–1355), emperor of the Serbian–Greek Empire 226
Steinberger, Jack (b. 1921), American physicist, Nobel Laureate (1988) 537
Steinmetz, Charles Proteus (1865–1923), German-born American electrical engineer 344
Stephen I (ca. 975–1038), first king of Hungary 117
Stephen III, the Great (ca. 1433–1504), ruler of Moldavia 226
Stephenson, George (1781–1848), British engineer XX
Stern, Otto (1888–1969), German-born American physicist, Nobel Laureate (1943) 491, 492, 525
Stevin, Simon (1548–1620), Dutch engineer, mathematician, physicist 143, **210–12**
Stirling, James (1692–1770), Scottish mathematician 384, 446
Stokes, Sir George Gabriel (1819–1903), British mathematician, physicist 69, 303, 304, 361, 442, XXIII
Stoney, George Johnstone (1826–1911), British physicist 391
Strabo (64 or 63 bce–ca. 24 ce), Greek geographer, historiographer 114, 132, 158, I
Strassmaier, Johann Nepomuk (1846–1920), Assyriologist, Jesuit priest 41
Strassmann, Fritz (1902–1980), German physical chemist 516, 517, 518
Strato of Lampascus (Straton) (ca. 340–269 or 270 bce), Greek philosopher 76–77
Stravinsky, Igor Fyodorovich (1882–1971), Russian-born American composer 497
Stueckelberg, Ernst Karl Gerlach (1905–1984), Swiss physicist 537
Stukeley, William (1687–1765), English physician, friend of Newton 260
Sudarshan, Ennackal Chandy George (b. 1931), Indian-born American physicist 535
Sulzer, Johann Georg (1720–1779), Swiss philosopher, pedagogue 343

Suppes, Patrick Colonel (b. 1922), American pedagogue, philosopher 13, 14
Swift, Jonathan (1667–1745), Anglo-Irish writer 302
Swineshead, Richard (fl. ca. 1350), English mathematician, naturalist 144, 159, 160
Sydenham, Thomas (1624–1689), English physician, "the English Hippocrates" 300
Sylvester II (Silvester, Gerbert d'Aurillac) (ca. 946–1003), pope 117, XII
Syrlin, Jörg, the Elder (ca. 1425–1491), German sculptor 1
Szilárd, Leo (1898–1964), Hungarian-born American physicist 518, 519, 523, 538

T

Tait, Peter Guthrie (1831–1901), English mathematician, physicist 390
Tamm, Igor Yevgenevich (1895–1971), Soviet physicist, Nobel Laureate (1958) 476, 493, 509, 532, 534
Tartaglia (Niccolò Fontana) (ca. 1500–1557), Italian mathematician 159, 160, XIII
Taylor, Richard Edward (b. 1929), Canadian physicist, Nobel Laureate (1990) 495
Tchaikovsky, Peter Ilich (1840–1893), Russian composer 323
Teller, Edward (Ede) (1908–2003), Hungarian-born American physicist 521, 539, 552
Terhal, Barbara M. (b. 1969), Dutch-born American physicist 577
Tertullian (Quintus Septimius Florens Tertullianus) (ca. 160–ca. 220), early Christian theologian, creator of the Latin ecclesiastical language 153
Tesla, Nikola (1856–1943), Croatian-born American physicist, electrical engineer 353
Thābit ibn Qurra (Al-ṣābiʾ Thābit ibn Qurra al-ḥarrānī) (836–901), Syrian-born Islamic naturalist, astronomer, philosopher, translator 142
Thales of Miletus (ca. 625–ca. 547 bce), Greek mathematician, founder of Ionian natural philosophy 46, 47, 48, 49, **60**, 95, 151, X, XIX
Theaetetus (ca. 410–ca. 368 bce), Greek mathematician 65
Thénard, Louis Jacques (1777–1857), French chemist 373
Theoderic the Great (Flavius Theodoricus) (454–526), king of the Ostrogoths 116, 129
Theodore Metochites (1270–1332), Byzantine politician, philologist, encyclopedist 131
Theodosius I, the Great (Flavius Theodosius) (347–395), last emperor of the unified Roman Empire 116, 131
Theophrastus (371–ca. 287 bce), Greek philosopher 78, 211
Thibaut, Jean (1901–1960), French physicist 526
Thomas Aquinas (ca. 1225–1274), Scholastic philosopher, theologian of Italian origin ("Doctor Angelicus") 6, 21, 24, 25, 26, 75, 79, 111, 120, 124, 126, 128, 129, 133–34, 138, 152, 153–54, 158, 168, 329, IX, XII
Thompson, Sir Benjamin. *See* Rumford
Thomson, Elihu (1853–1937), English-born American engineer, inventor XIX
Thomson, Sir George Paget (1892–1975), English physicist, Nobel Laureate (1937), son of J. J. Thomson 396, 460, 492, 514
Thomson, Sir Joseph John (1856–1940), English physicist, Nobel Laureate (1906) 379, **388–93**, 395–96, 397, 398, 426, 439, 449, 490, 502, 506, 514, XXI
Thomson, William. *See* Kelvin
Thutmose III (Thutmosis) (ruled 1479–1425 bce), Egyptian pharaoh 35
Ting, Samuel Chao Chung (b. 1936), American physicist, Nobel Laureate (1976) 494, 548
Tisserand, François Félix (1845–1896), French astronomer 349
Titian (Tiziano Vecellio) (ca. 1485–1576), Italian painter 128
Tolman, Richard Chace (1881–1948), American physical chemist, physicist 422
Tolstoy, Lev Nikolaevich (1828–1910), Russian writer 323
Tomislav I (ruled ca. 910–928), first king of Croatia 117
Tomonaga, Sin-Itiro (1906–1979), Japanese physicist, Nobel Laureate (1965) 477, 478, 493
Torricelli, Evangelista (1608–1647), Italian physicist, mathematician 234, 237, 244, 256, 291
Toulmin, Stephen Edelston (b. 1922), English philosopher, historian of science 17
Tour, Maurice Quentin de la (1704–1788), French painter XVIII
Townes, Charles Hard (b. 1915), American physicist, Nobel Laureate (1964) 400, 414, 493, XXV
Trajan (Marcus Ulpius Nerva Traianus) (53–117), Roman emperor 105
Tribonianus (d. ca. 545), highest official of the Byzantine emperor Justinian I, jurist 47
Trouton, Frederick Thomas (1863–1922), English physicist 414, 415, 421
Truesdell III, Clifford Ambrose (1919–2000), American historian of science 163, 168, 307, 378
Turgot, Anne Robert Jacques, Baron de Laune (1727–1781), French economist, statesman 325, 326
Turner, Joseph Mallord William (1775–1851), English painter XVIII
Tyndall, John (1820–1893), Irish physicist 377, 439, 440

U

Uhlenbeck, George Eugene (1900–1988), Dutch-born American physicist 451
Ulpian (Domitius Ulpianus) (ca. 170–223), Roman jurist, statesman 47
Ulugh Beg (1394–1449), Uzbek ruler, astronomer 139, 185
Umar, the Great (Omar) (ca. 581–644), caliph, companion of Mohammed, statesman, conqueror 105, 132
Urban VIII. *See* Barberini, Maffeo

V

Van de Graaff, Robert Jemison (1901–1967) American physicist 509
Varāhamihira, Daivajna (Varaha, Mihira) (505–587), Indian astronomer, mathematician, philosopher 136
Varignon, Pierre (1654–1722), French physicist, mathematician 85, 303, 304–5
Varley, Cromwell (1828–1883), English physicist 389
Vasari, Giorgio (1511–1574), Italian painter, architect, art historian 161
Vavilov, Sergei Ivanovich (1891–1951), Soviet physicist 534
Veksler, Vladimir Iosifovich (1907–1966), Soviet physicist 511, 531
Velázquez, Diego Rodríguez de Silva y (1599–1660), Spanish painter 242
Veneziano, Gabriele (b. 1942), Italian theoretical physicist 570
Verdi, Giuseppe (1813–1901), Italian composer 323
Vermeer, Jan (Johannes) (1632–1675), Dutch painter 242
Verschaffelt, Jules-Émile (1870–1955), Belgian physicist 487
Vesalius, Andreas (1514–1564), German-born Flemish physician 82, 128
Vespasian (Titus Flavius Vespasianus) (9–79), Roman emperor 105
Vesztergombi, György (b. 1943), Hungarian physicist 575
Vico, Giovanni Battista (1668–1744), Italian philosopher of history 299
Viète, François (Franciscus Vieta) (1540–1603), French mathematician 128, 289, 290, XIII
Villalpando, Francisco de (1552–1608), Italian architect, structural engineer 160, 161
Villard de Honnecourt (fl. ca. 1245), French architect 124
Villard, Paul (1860–1934), French physicist 503
Villon, François (1431–after 1462), French poet 127, 128, 129, 130, 147
Vincent of Beauvais (Vincentius Bellovacensis, Vincentius Burgundus) (ca. 1190–ca. 1264), French pedagogue, Dominican friar, author of the primary encyclopedia of the Middle Ages, *Speculum Majus* 126
Virgil (Vergil, Publius Vergilius Maro) (70–19 bce), Roman poet 4, 47
Vitruvius (Marcus Vitruvius Pollio) (fl. ca. 50 bce), Roman architect, military engineer 29, 47, 79, 90–91, 130, 134
Vivaldi, Antonio (1678–1741) Italian composer 323

Vives, Juan Luis (1492–1540), Spanish humanist, philosopher, pedagogue 163
Viviani, Vincenzo (1622–1703), Italian mathematician 234
Vleck, John Hasbrouck van (1899–1980), American physicist, Nobel Laureate (1977) 494
Voigt, Woldemar (1850–1919), German physicist 415, 416, 426
Volta, Alessandro (1745–1827), Italian physicist 323, 331, **342–43**, 344
Voltaire (François Marie Arouet) (1694–1778), French writer, philosopher 149, 274, 277, 296, **299–301**, 302, 321, 323, 325, XX
Vossius, Isaac (1618–1689), Dutch scholar 227, 248

W

Waals, Jan Diderik van der (1837–1923), Dutch physicist, Nobel Laureate (1910) 488, 490
Wagner, Richard (1813–1883), German composer 323
Wallenstein, Albrecht Wenzel von (1583–1634), General and statesman in the Thirty Years' War 242
Wallis, John (1616–1703) English mathematician 193, 194, 195, 290–91
Walther von der Vogelweide (ca. 1170–ca. 1230), Middle High German poet 128
Walton, Ernest Thomas Sinton (1903–1995), Irish physicist, Nobel Laureate (1951) 487, 492, 501, 508, 509, 514
Wanli (Zhu Yijun) (1563–1620), Chinese emperor 171
Wantzel, Pierre Laurent (1814–1848), French mathematician 96
Warburg, Emil (1846–1931), German physicist 439
Waterston, John James (1811–1883), English physicist 375–76, 378, 387
Watson, James Dewey (b. 1928), American molecular biologist 323
Watt, James (1736–1819), English engineer, inventor 114, 323, 366, 367, 373, 374, XVIII
Weber, Wilhelm Eduard (1804–1891), German physicist 348–49, 353, 361, 366
Webster, David Locke (1888–1976), American physicist 508
Weinberg, Steven (b. 1933), Austrian-born American physicist, Nobel Laureate (1979) 548, 562, 563, 565, 566, 570
Weiss, Pierre Ernest (1865–1940), French physicist 500, XXV
Weisskopf, Victor Frederick (1908–2002), Austrian-born American physicist 471, 514, 523, 524, 525, 542, 547
Weizsäcker, Karl Friedrich Freiherr von (1912–2007), German physicist 454, 470, 512, 551, 552
Weyl, Hermann (1885–1955), German-born American physicist 540, 547
Wheeler, John Archibald (1911–2008), American physicist 535
Whitehead, Alfred North (1861–1947), English mathematician, philosopher 2, 437, XX
Whitman, Walt (1819–1892), American poet 323
Whittaker, Sir Edmund Taylor (1872–1956), English mathematician, astronomer, historian of science 409
Wideröe, Rolf (1902–1996), Norwegian physicist, engineer 531
Wiedemann, Gustav (1826–1899), German physicist 389
Wien, Wilhelm (1864–1928), German physicist, Nobel Laureate (1911) 393, 439, 440, 441, 444, 454, 469, 491
Wiener, Norbert (1894–1964), American mathematician 464, 465
Wigner, Eugene Paul (Pál Jenő Wigner) (1902–1995), Hungarian-born American physicist, Nobel Laureate (1963) 432, 493, 519, 537, 538–39, 540, 563
Wilczek, Frank Anthony (b. 1951), American theoretical physicist, Nobel Laureate (2004) 567
Wilder, Thornton (1897–1975), American writer 22
Wiles, Sir Andrew John (b. 1953), British mathematician 453
William I (William the Conqueror) (1027–1087), English king 118, 129
William IV (William the Wise) (1532–1592), landgrave of Hesse-Kassel, astronomer 177
William of Moerbeke (Willem van Moerbeke) (ca. 1215–1286), Flemish Dominican, scholar, translator 134, 135
Wilson, Charles Thomson Rees (1869–1959), British physicist, Nobel Laureate (1927) 392, 491, 496
Wilson, Kenneth Geddes (b. 1936), American physicist, Nobel Laureate (1982) 495
Wilson, Robert Woodrow (b. 1936), American physicist, radio astronomer, Nobel Laureate (1978) 494, 550
Winckelmann, Johann Joachim (1717–1768), German archeologist, art historian 329
Winkler, Johann Heinrich (1703–1770), German physicist 335
Wisdom, Arthur John Terence Dibben (1904–1993), British philosopher 329
Witten, Edward (b. 1951), American theoretical physicist 569, 570, 573, 581–83
Wittgenstein, Ludwig (1889–1951), Austrian philosopher 563
Wolff, Christian Freiherr von (1679–1754), German philosopher 314
Wolfskehl, Paul Friedrich (1856–1906), German mathematician, established the Wolfskehl prize 452
Wollaston, William Hyde (1766–1828), British naturalist 398
Wordsworth, William (1770–1850), English poet XVIII
Wu, Chien-Shiung (1912–1997), Chinese-born American physicist 528, 539, 541, 542

X

Xenophanes of Colophon (ca. 565–ca. 470 bce), Greek philosopher, poet 47, 48

Y

Yablochkov, Pavel Nikolaevich (1847–1894), Russian electrical engineer 353, XIX
Yang, Chen-Ning Franklin (b. 1922), Chinese-born American physicist, Nobel Laureate (1957) 493, 528, 539, 542, 543, 552
Yaroslav I the Wise (978–1054), grand prince of Kiev 224
Yermak Timofeyevich (Ermak) (d. 1585), Cossack leader 171
Young, Thomas (1773–1829), English physician, physicist, linguist 288, 359–60, 362, 364, 365
Yukawa, Hideki (1907–1981), Japanese physicist, Nobel Laureate (1949) 340, 492, 522, 527

Z

Zeeman, Pieter (1865–1943), Dutch physicist, Nobel Laureate (1902) 364, 398
Zeldovich, Yakov Borisovich (1914–1987), Soviet physicist 537, 557
Zeno of Citium (ca. 336–ca. 264 bce), Greek philosopher 47, **100–101**
Zeno of Elea (ca. 490–ca. 430 bce), Greek philosopher, mathematician 47, 48, 54, **55**
Zermelo, Ernst (1871–1953), German mathematician 386
Zernike, Frits (1888–1966), Dutch physicist, Nobel Laureate (1953) 493
Zinn, Walter Henry (1906–2000), American physicist 518
Zipernowsky, Károly (1853–1942), Hungarian electrical engineer 353, XIX
Zweig, George (b. 1937), Russian-born

Subject Index

absolute space 410–16, 419
absolute temperature scale 366
abstraction 9–10, 12–14, 31–32
acceleration. *See also* free fall
  circular 205, 259
  dynamics and 212–13
  in free fall 201, 255
  by superposition 243–44
accelerators, particle. *See* particle accelerators
Achilles and the tortoise 54, 55
actinum 516
Adding Machine, physicist as 562
affluvium 333
agriculture, medieval 122
Ahmes papyrus. *See* Rhind papyrus
air pressure 234–37
air pump, first 236
Alembert's principle, d' 311–14
algebra 137
algorithms 137
alpha rays 503, 504, 507
analytic judgments 327
angular momentum, conservation of 536, 538
anode rays 392–93
antecedents 6
antichthon 54
antigravitation, big bang and 574
Antikythera apparatus 98, 102
antiparticles 543–45
antisensationalism 55
apeiron 60, 546–48
Arabic numerals 137, 140–41
Arab world 116, 132–33, 137–39, X
Archimedean solids 65
argumentation 56
Aristotelian dynamics 65–71, 75, 145–46
Aristotelian elements 60–61, 62–65, 238. *See also* matter
arithmogeometry 51–52
art 76–78, 158, 161–64
Aston whole-number rule 506
astrolabe, Byzantine 131
astrology 103–5, 108–9, 164–65, 193
astronomy. *See also* planetary motion
  advancements in 549–50
  Egyptian 40
  Greek 54, 66–67, 71–75, 79–81, 89–94
  medieval views of 147–48
  Mesopotamian 41–44
  Renaissance study of 164–65, 177–201
atomism, Democritus's 61–62
atoms
  chemistry and 237–40, 408
  electron and 387–400, 448–51
  elementary particles and 525–29, 546–48
  Greek views of 60–65
  neutrons and 508–9
  Newtonian 271–72
  nucleus of 496–98, 505–7, 509–14
  models of 395, 396–98, 398–400, 448–51, 505–7
axes of inertia 307
axioms 12–14, 97

background radiation. *See* cosmic microwave background radiation (CMB)
balance scale 10
ballistics. *See* projectile motion
Balmer's formula 399–400
barothermoscope 366
baryons 528, 537
Basel school 375
batteries, galvanic 343
beauty, physics and 23, 581
Bell inequalities 485, 578
Benedictine order 123–24
beta rays 503, 504, 508, 527
betatron 531
big bang 556–59, 571–75
blackbody radiation 438–48
black holes 556
Bohr atomic model 398–400, 448–51
Bohr frequency condition 450
Bolyai-Lobachevsky geometry 401–2
bosons 528, 566, 567, XXVII
Boyle-Mariotte law 236, 237
brachistochrone 207
bubble chamber 532
bucket analogy, Newton's 273
buckminsterfullerene 65
Byzantine Empire 116, 131–32, VI

calculus 264–65, 291–96, 309–11
calendars 40, 71, 549
caloricum 365, 367–68, 369–74
Carnot cycle 373–74
Cartesian coordinates 289–91
categories, Kantian 328
cathedrals, medieval 121, 140, 158
cathode rays 388–93
causality problem 477–83
cause, effective 67–71, 78. *See also* dynamics
cave
  illusions of the 214–15
  Plato's 57–60
celestial motion. *See* planetary motion
Celsius scale 366
central force 258, 265–68
centrifugal force 254–55
centripetal force 259
CERN project 562, XXVII
chain reaction, nuclear 518–22
change, Greek views of 54–55, 61
characteristic values. *See* eigenvalues
charge, electrical 334–38, 426–27, 528, 534, 536–37
charge conjugations 543–45
chemistry 237–40, 343, 387–400, 408
Cherenkov counter 532
Chernobyl XXVIII
Christianity 102–3, 104, 105–9
circle
  motion in a 206–7, 219, 245–51, 254–55, 257–60
  squaring of 96, 98
cities, medieval 121
classification 4–7, 10–12, 31–32
clocks 98–99, 102, 123, 203
cloud chamber 507, 532
CMB. *See* cosmic microwave radiation (CMB)
Cogito, ergo sum 217–19
collective model, nuclear 514
collision, laws of 218–20, 252–54, 255, 256–60, 314–15
color 284–88, 548, 566–67
compass, magnetic 122–23, 329–30
Compton effect 489–90
computing, quantum 577–78
conduction, heat 369–72
conic sections 97, 99–100
conservation laws, elementary particles and 536–43
conservation of charge 337–38, 534, 536–37
Consolation of Philosophy (Boethius) 129–30
contact, immediate 314–15
continuity equation 306
coordinates, geometric 289–91
Copenhagen interpretation 465–70, 484–86

Copernican revolution 177–85
corpuscular theory, Boyle's 239
cosmic microwave background radiation (CMB) 550, 559, 573, 582
cosmic rays 529–30, 550
cosmological constant 582
cosmos
  Cartesian 220–23
  conceptions of 51, 72, 74–78, 561–65
  Newtonian 270–75
Coulomb interaction. *See* electromagnetic interaction
Coulomb's law 350, 476
counter-earth, Pythagorean 54
CP violation 543–45, 576
creativity 16–17
Critique of Pure Reason (Kant) 326–29
Crookes tubes 389
crystal ball model of nucleus 514
cube, doubling of 95, 97
current, electrical 340–49
curvature, Gaussian 430
cycloid, pendular motion and 246–49
cyclotron 509, 531

decimal systems 35–36, 136–37
deductive method 7
De rerum natura (Lucretius) 102, 107–8
detectors, particle 505, 531–32
Dialogo (Galileo) 197–201
dielectrics 339, 352, 357
diffusion chamber 532
diopter 98, 100
Dirac equation 475–77, 526–27
Discorsi (Galileo) 201
displacement, to measure volume 79–80
displacement current 357
displacement law, Soddy-Fajans 504
displacement law, Wien's 440
distance, measurement of 421–25
divisibility, limits of 63
division 34, 36
Dominican order 154
double refraction 283–84, 287–88, 360
drag coefficient 412
duality, Cartesian 218, 297–99
dynamics
  Galileo and 212–13
  Huygens and 242–55
  Newtonian 68–69, 255–60, 302–4, 305–9, 314–15
  peripatetic 65–71, 75, 145–46

eclipses 34, 92, 94
ecstasy, cognition as 103
effluvium, electromagnetic 330, 333
Egypt, Ancient 32–37, 40
eigenstates, quantum 471–77
eigenvalues 463–64
eight-fold way model 545
Eleatics, the 54, 55, 61
electrodynamics 329–64, 417, 426–27, 451–52, 475–77
electrolysis, laws of 351–52
electromagnetic interaction 533–35, 548–49
electron
  atomic structure and 395–400, 448–51, 506–7, 507–8
  Lorentz's theory of 363–64
  periodic table and 393–95
  radiation 388–93, 503, 504, 508, 527
electron microscopy 499–500
electrostatics 330, 332–40
electroweak interaction 566
elementary particles
  antiparticles and 543–45
  conservation laws of 536–43
  in fundamental interactions 534–36
  quantum computers and 577–78
  types of 525–29, 545–46, 548–49, 565–67, XXVII
elements, Aristotelian 60–61, 62–65, 238. *See also* matter
Elements (Euclid) 96–97, 99
elliptical orbits. *See* Kepler, laws of
empiricism 154–55. *See also* sensory data
Encyclopédie 321–26
energeticism 408
energy
  atomic theory and 408
  conservation of 365–66, 373–74, 376–77, 536–38
  heat as 368–69
  mass and 425–29, 510–11
  nuclear 518–23, 550–54, XXVIII
  quantum theory and 438–48
Enlightenment, the 319–21
entropy 379–87, 441–44
Epicureanism 101, 102
epistemology. *See* knowledge
epochs, scientific 3–10, 24–27
EPR (Einstein–Podolsky–Rosen) paradox 485, 578
equants 90, 177, 182, 190
equipartition principle 367, 406, 441
equivalence law 429
ether
  Aristotle and 75
  electrodynamics and 282–83, 360–61, 363–64
  nature of 405
  relativity and 410–16, 419
ethics, scientific 20–21, 23, 298, 523–24, 561–65
Euclidian geometry 96–97
Europe, birth of 115–21
evolute 249
exclusion principle 451
expansion, accelerating universal 576–77, 582
experience, Aristotle and 76–78. *See also* sensory data
explanandum 6
explanans 6
exponential law of radioactive decay 504
extent, physical 218–19
extremal principle, kinetic 309–11
extreme value calculations 292–93
eye, study of 227
eyeglasses, development of 123, 229

Fahrenheit scale 366
falsification 565
Faraday dark space 388
Faraday rotation 352
Faraday's laws 351–52
Fermat's last theorem 452–53
Fermat's principle 230–32
fermions 528, 566, XXVII
Fermi's theory of beta decay 508, 527
Fertile Crescent 32–45
feudalism 120–21
Fibonacci numbers 141–42
field, electromagnetic 353–59, 408, 475–77, 534
figurate numbers 51–52
final cause 310
fission, nuclear 514–22, 524
flavor, quark 548
fluents, calculus of 295
fluid flow 306
fluorescence 140–41, 498–501
fluvial valley cultures 32–45
fluxions, calculus of 264–65, 295
force. *See* dynamics
Fourier series 370
fractals 484
fractions 35–36
Franciscan order 124
Frankenstein, physicist as 562
Fraunhofer lines 398
free fall
  Huygens and 244–45, 255
  medieval views of 147
  Renaissance study of 159–60, 201–6, 208–9, 210–12
friction 68–69, 330, 333, 337, 368–69
frustrum, pyramidal 36, 37, 38
fusion, nuclear 522–23, 550–54

Galilean principle of relativity 197
galvanic batteries 343
gamma rays 503, 504
Gamow's theory of alpha decay 507
gases, kinetic theory of 377–78, 388
Gaussian curvature 430
Geiger-Müller counter 505, 532

Geiger-Nuttall rule 507
Geissler tubes 388, 399
general relativity 429–33, 556–59, 570, 572
generans 70
generators, electrical 352–53
genius, century of 241–42, 275
Geographia (Eratosthenes) 94
geography, Greek 90–95
geometric astronomy 71–75
geometry
  analytic 289–91
  in art 161–62
  astronomy and 71–75
  axioms of 12
  Greek 31, 51–54, 95–97
  non-Euclidean 401–2, 429–33
glass electricity 333, 335–37
gnosis 565
golden ratio 54
Goodman paradox 15
grand unification theory (GUT) 567, 569–71
gravitation 221, 256, 260–64, 432–33, 533–34
gravitational interaction 533–34
gravitational mass 213
Greece, history of 45–49, 113–14, II
Greek, translation from 133–35
Greek fire 131
group, mathematical 418, 539, 567–69
GUT. *See* grand unification theory (GUT)

hadrons 535, 566–67
Hall effect 498–99
Hamiltonian equations of motion 455
Hammurabi, Code of 44
harmony 50–51, 53, 75
harness, horse 122
heat
  caloricum theory of 366–68, 369, 369–74
  conservation of energy and 376–77
  entropy and 379–87
  inductive method and 215
  kinetic theory of 368–69, 374–76, 377–78, 408
  measurement of 365–66
  transfer of 316
heat death theory 380–81
heliocentrism 177–85
heliostatic 182
Hellenism 46, 78, 99–109
Hertzsprung-Russell diagram 554–55
hidden parameters 480–83
hieroglyphics, Egyptian 34, 35
Higgs boson 567, 583
historian, task of 524–25
holography 496
horror vacui 71, 233–37
human body 61
humanists, Renaissance 158–59
humors, four 62–63
hydrostatics 80, 88–89
hyperons 528

Iceland spar 283–84, 287–88, 360
icosahedron 54, 62, 64–5
ideal, Platonic 54–57
idealization. *See* abstraction
idler wheels, as model of electromagnetic field 354–57
idola fori 215
idola specus 214–15
idola theatri 215
idola tribus 214
illusions, Baconian 214–15
impetus 146–47
inclined plane 201–6, 208–9, 244–45
independent particle model 511–12, 514
India 136–37
induction theorem, Faraday's 349–51
inductive method 7, 14–15, 213–16
inertia 243, 255, 306–7
inertial mass 213
inflation theory 574, 582
interactions, fundamental 532–36, 565–67
interference lines 485–86
Investiture Controversy 121
involute, cycloidal 249
ionization 502–3
ionization chamber 532
irrational numbers 52–53
Islam. *See* Arab world
isospin 539–40
isotopes 498, 504, 506

Judaism 132–35
Julian calendar 40

Kantian categories 328
kaons 528
Kelvin scale 366
Kepler, laws of 191–93, 258, 563
kinematics, descriptive. *See* mechanics
kinetic energy 365–66, 368–69, 374–79, 381–87. *See also* energy
Kirchhoff laws 344
Klein model, geometric 402
knowledge
  classification of 4–7, 10–12, 31–32
  medieval recovery of 129–35
  reliability of 21–22, 562–65
  sources of 297–99, 311–12, 326–29

labor, human
  in ancient times 43
  in the Middle Ages 113–14, 121, 122, 123–24, VII
lanthanum 516
Large Hadron Collider (LHC) 562, 583, XXVII
laser, operation of 493
latent heat 367, 368, 369
law of areas 258. *See also* Kepler, laws of
length, early measurement of 42
leptons 528, 536, 537, XXVII
lever 70–71, 80–81, 84–88, 142–43
Leyden jar 333
LHC. *See* Large Hadron Collider (LHC)
Library of Alexandria 105
life-cycle, stellar 554–56
light
  electromagnetics and 359–63, 395–96, 398–400
  gravitation and 432–33
  nature of 281–89
  as particle 288–89, 359–60, 458–65
  quantification of 448–51, 458–65
  refraction of 226–32, 283–88, 310
  velocity of 228–32, 310, 361–62, 411–16, 421–25
  as wave 282–84, 288, 359–63, 458–65
line, concept of 31
line spectra, atomic 398–400, 448–51
liquid drop model 512–14
living force 365
local time 416, 417
logic 56, 151–54
Lorentz group 418
Lorentz transformation 409, 418, 426, 475–76
lune of Hippocrates 96, 98

Mach number 407
magic numbers, nucleonic 512
magnetism 329–30, 345–46, 348–49, 352–63, 414. *See also* electrodynamics
Magyar Encyklopédia (Csere) 224
Mandarin, physicist as 562
Mandelbrot set 484
marketplace, illusions of the 215
Marxism 18–19
mass. *See also* matter
  center of 306–7
  energy and 212–13, 425–29
  measurement of 42
  Newtonian 265, 270
  point 302, 305
  space and 429–33
  subatomic 509–11, 528
mass defect 509–10
mathematics
  calculus 264–65, 291–96, 309–11
  coordinates in 289–91
  Fourier series 370
  group theory 418, 539, 567–69
  of pendulums 245–46
  state equations 372–73
  truth and 216–19, 299

matrix mechanics, Heisenberg's 452–58
matter. *See also* mass
  chemistry and 237–40
  conservation of 306
  Greek views of 60–65, 75
  kinetic theory of 377–78, 387–400, 406
Maupertuis's principle 302–3, 309–11
mean, arithmetic 53
Mécanique Analitique (Lagrange) 304
mechanics. *See also* motion
  Archimedes and 80–81, 84–88
  electromagnetism and 358–59
  medieval 126–27, 143–44, VIII
  Newtonian 255–60, 302–4, 305–9, 314–15, 486–89
  optics and 360–61
  positivist 407–9
  quantum 452–57, 458–70, 477–86, 488–90, 507–8
  relativity and 426
  Renaissance 159–60, 162–63, 201–8, 242–55
  thermodynamics and 386
megamachine, Mumford's 43
Merton rule 144
meson, mu. *See* muon
meson, pi. *See* pions
mesons 528
Mesopotamia 32–35, 36–45
Metonic cycle 41
microscope 400–401
microstate, molecular 383–84
mills, water 122
minima naturalia 63
Minkowski "world" 426
mirages, Baconian 214–15
mirror symmetry 539, 540–45, XXVI
modeling, scientific 8–10. *See also* abstraction
moment of inertia 306–7
momentum, conservation of 536, 538–39
monadism 298
monasteries 119, 123–24
monopoles, magnetic 364
Moscow papyrus 34, 35, 36
Mössbauer effect 495
motion. *See also* mechanics; planetary motion
  Cartesian 218–20
  circular 245–51, 254–55
  extremal 309–11
  Greek views of 54–55, 65–78
  heat as 215, 368–69, 374–76
  of light 281–89, 410–16, 421–25, 458–65
  medieval study of 143–44, 146–47
  Newton's laws of 71, 256–60
  point mass 305–9
  Renaissance study of 159–60, 191–93, 201–12, 244–45
  superposition of 243–44
motor accidentalis 70
motor conjunctus 67, 70
multiplication 34, 36, 40
muon (mu meson) 527
music. *See* harmony
Mysterium cosmographicum (Kepler) 189
mythology 32–33, 43–44, 102–5, 107–8

needle-radiation hypothesis 459
negative numbers 136
Neo-Platonism 102–3
Nestorians 133
neutrino 527, 536–37, 543, 556, 576
neutrino astronomy 550
neutron
  discovery of 498, 506–7, 508–9
  fission and 515, 517, 518–21
  nuclear mass and 509–10
neutron stars 556
Newtonian dynamics. *See* dynamics
nonconvex regular solids 65
nonelectric materials 332
nuclear shell model 511–12, 514
nuclear structure 505–7
number
  concept of 31, 50–54
  notation of 35, 37–41, 53, 97, 100
  systems 34, 35–36, 38–39
numerology 176–77, 193

observation, quantum theory and 467–70
Ockham's razor 156
Ohm's law 343–44
On Method (Archimedes) 82–88
On the Sizes and Distances of the Sun and the Moon (Aristarchus) 91–92
operators, quantum 471–77
optical model, nuclear 514
optics 226–32, 281–89, 359–63, 458–65. *See also* light
Orpheus 50, 51
oscillation, neutrino 576
oval, Cartesian 229
overregulation, cybernetic 216
Oxford School 154–55

paradigm shifts 17, 565
paradoxes
  Eleatic 54, 55, 61
  EPR 485, 578
  and inductive method 15
parallax 184
parallels, problem of 401
parameters, hidden 480–83
parity violation 524, 528, 540–43
particle, light as 288–89, 359–60, 458–65
particle accelerators 498, 509, 530–31, 561–62, XXVII
Pauli exclusion principle 451, 511
pendulum 206–7, 245–51, 312
penetration, cylindrical 269–70
pentagram 54
perfect numbers 51
periodic table 393–95, 396, 504
peripatetic dynamics 65–71, 75, 145–46
peripheries, cultural 223–26
perspective, laws of 161
phase space, molecular 383
phenomena 55. *See also* sensory data
phenomenology. *See* sensory data
philology 157
philosophy. *See also* Greece, history of
  medieval 151–54
  Newtonian 270–75
  quantum theory and 477–79, 486
  of science 10–15, 561–65
phlogiston theory 240
Phoenecians 44
photons 448, 485–86. *See also* light
pi 36, 38, 96, 98
pion (pi meson) 527–28
Planck era 571–75
Planck law 442–48, 451–52
plane, inclined 201–6, 208–9, 244–45
planetary motion. *See also* astronomy
  Greek views of 54, 66–67, 71–75, 79–81, 89–94
  medieval views of 147–48, 164–65
  Newtonian 260–64, 270–71
  observation of 34, 41–43
  relativity and 433
  Renaissance study of 177–201
Platonic solids 54, 62–65, 189
plow, development of 122
plum pudding model, atomic 395
Poincaré model, geometric 402
point mass 302, 305, 306–9
polarization, optical 360
poles, magnetic 329–30
position, atomistic 61
positional notation 136
positivism, philosophical 407, 478–79, 563
positron 526–27
prediction, causal 479–80
Principia (Newton) 256, 264–70, 271, 272, 302
printing press 165–68, XI
probability 381–87, 465–70
projectile motion 207–8, 270
proportion, Pythagorean 53
protons 506–7, 535–36

pure air, Aristotelian 235
pyramids, volume of 36, 37, 38
Pythagorean theorem 36–38, 52–53, 453quadratic equation 38
quadratix 96, 97
quantification
  abstraction and 9–10
  of energy 338–40, 366–68, 444–48
  of motion 143–44, 452–65
quantum chromodynamics 566
quantum field theory 439, 570, 582
quantum jump 469
quantum theory
  computers and 577–78
  Copenhagen interpretation of 465–70
  development of 437–40
  electrodynamics and 451–52, 475–77
  energy and 444–48
  entropy and 441–44
  mechanics and 452–57, 458–65, 477–86, 488–90, 507–8
  operators and 471–77
quarks 528, 545–46, 548–49, 566–67, XXVII
qubit 577

radiation
  atomic structure and 396
  blackbody 438–48
  cosmic 529–30, 550, 559, 573, 582
  nuclear 498, 501–4, XXVIII
  thermal 369
  x-ray 391–92, 499–500
radio astronomy 550, 559
radioactive decay, exponential law of 504
rainbows 229
raven paradox 15
Rayleigh-Jeans law 441–44, 446
reaction cross section 510
reason. *See* logic
Reason, Age of 319–21
reason, pure 326–29
reciprocals 35–36
red giant stars 555
reductio ad absurdum 82–88
Reformation, the 158
refraction 226–32, 283–88, 310
regular solids 54, 62–65, 189
relativism 55, 562–64
relativity
  Galilean principle of 197
  general 429–33, 556–59, 570, 572
  mechanics and 486–88
  special 409–29, 510–11, XXII
Renaissance, the 157–68
res cogitans 218
res extensa 218
resinous electricity 333, 335–36
resistance 67–71
retrograde motion, planetary 74
revelation, divine 153–54, 155
reversibility, motion's 245
rhetoric 56
Rhind papyrus 34, 35–36, 40
Richardson's law 490
rings, Newton's 286–87
river civilizations 32–45
Roman Empire 115–16
rotation, Faraday 352
rotational symmetry 539
Royal Society 281–82
rudder, stern-mounted 122
Rutherford atomic model 396–98
Rutherford-Bohr atomic model 505–7

saltpeter process 554
scale, balance 10
scanning tunneling microscope (STM) 499–500
Sceptical Chymist, The (Boyle) 238–39
Skepticism 100, 102
Schrödinger equation 460–63
scintillation counter 532
sensory data
  doubt and 217–18
  method and 298–99
  positivism and 320, 407
  pure reason and 326–29
  truth and 55–60, 61, 100–101, 154–55
set theory, number and 31
sexagismal number system 34, 38–39
shape, atomistic 61
shifts, paradigm 17, 565
Sidereus nuncius (Galileo) 195–97
single photon experiment 485–86
slavery 113–14
Snell-Descartes law 226–30
soccer-ball molecule 65
social process 18–19
Soddy-Fajans displacement law 504
solids
  Archimedean 65
  Platonic 54, 62–65, 189
  regular 54, 62–65, 189
Solvay Congress 438–39, 487, 489–90
Sommerfeld ellipse model 450–51
Sophists, Greek 56
Sothis cycle 40
soul 61, 66–67
space 61, 410–16, 419, 429–33
spar, Iceland 283–84, 287–88, 360
spark chamber 532
special relativity 409–29, 510–11, XXII
specific heat 366–67, 406
spectra 284–88, 398–400, 448–51
spectral analysis 398–99, 549–50
speed of light 228–32, 310, 361–62, 411–16, 421–25
spheres, celestial 66–67, 183, 186, 195–201
spin, atomic 451, 528, 542–43
spontaneous symmetry breaking 570
SPS. *See* Super Proton Synchrotron (SPS)
square root 39–40
squaring the circle 96, 98
SSC. *See* Superconducting Super Collider (SSC)
Standard Model 565–67, 581–83, XXVII
stars 550–56
state equations 372–73, 471–77
static electricity 330, 332–40
statics 142–43, 160, 304–5
statistics. *See* probability
Stefan-Boltzmann law 439–40, 446
stellar development 554–56
STM. *See* scanning tunneling microscope (STM)
Stoicism 100–101, 102, 104–5
Stokes's law of friction 69
strangeness, conservation of 537
string theory 570–71, 582
strong interaction 535
subatomic particles. *See* elementary particles
substance. *See* matter
substance, heat. *See caloricum*
Superconducting Super Collider (SSC) 561–62
superconductivity XXV
supernovas 555–56
superposition of motions 243–44
Super Proton Synchrotron (SPS) XXVII
superstring theory 570–71, 582
SUSY (supersymmetry) 569–70
symmetry 31–32, 61–62, 539–46, 567–69, XXVI
synchrocyclotrons 531
synthesis, theoretical 7–8
synthetic judgments 327

tabula rasa 298–99
tangents 291–92, 294
technology, medieval revolution in 121–23
temperature 316, 365–68
tensors 426
terrestial motion. *See* peripatetic dynamics
tetractys, holy 51, 545
theater, illusions of the 215
theology 153–54, 155
theory, reality and 59
theory of everything (TOE) 570–71, 575–76
thermodynamics. *See* heat
theta-tau puzzle 524, 542
thorium 503–4
three-field agriculture 122

time
  absolute 410–11, 415–16, 419
  Augustine on 105–9
  causality problem and 479–80
  measurement of 42, 421–25
time-reversal symmetry 545
TOE. *See* theory of everything (TOE)
toppling, statics of 160
torque, motion and 309
Torricelli void 234–35
torsion balance 338–39
transformers, electrical 352–53
transistor, discovery of 492, XXV
transmutation of elements 498, 503–4, 507, 515
transuranium elements 515
tribe, illusions of the 214
triboelectric. *See* electrostatics
trigonometry, Ptolemy and 93
trisection, angular 96, 97
tunnel effect, quantum mechanical 507

uncertainty principle, Heisenberg's 454, 464, 465–70, 471, 485–86
unified field theory 546–48
uniform background radiation 550, 559, 573, 582
universities 124–27, 166, IX, XXIII
uranium 498–501, 515, 516–17

vacuum 61, 71, 75–77, 233–37, 574
Van de Graaff generator 509
variations, calculus of 309–11
velocity
  dynamics and 212–13
  in free fall 201, 204–6, 208, 211–12
  of light 228–32, 410–16, 421–25
  mass and 426
velocity, virtual 71
vibrations, optical 285
viewpoints, Kantian 328
virtual displacement 71
virtual work 304–5
virtus movens 70
volume
  displacement and 79–80, 85
  of pyramidal frustrum 36, 37, 38
  spherical 88
vortices 221, 268–69, 288, 352, 354–57

waste book, Newton's 256
water clocks 99, 102, 203
water power 122
wave, light as 282–84, 288, 359–63, 458–65
wave mechanics, Schrödinger's 458–65
waves, electromagnetic 357–58
weak interaction 535, 542–43, 548–49
white dwarf stars 555–56
Wien-Planck law 441–44, 446
Wien's displacement law 440
work, virtual 304–5
world line 424

x-rays 391–92, 498–500, 504

Zeeman effect 501
zero, origins of 40, 136
zodiac, signs of 41, 42