Greenwood Guides to Great Ideas in Science

The Quantum Revolution

A Historical Perspective

Kent A. Peacock

THE QUANTUM REVOLUTION

Titles in Greenwood Guides to Great Ideas in Science
Brian Baigrie, Series Editor

Electricity and Magnetism: A Historical Perspective
Brian Baigrie

Evolution: A Historical Perspective
Bryson Brown

The Chemical Element: A Historical Perspective
Andrew Ede

The Gene: A Historical Perspective
Ted Everson

The Cosmos: A Historical Perspective
Craig G. Fraser

Planetary Motions: A Historical Perspective
Norriss S. Hetherington

Heat and Thermodynamics: A Historical Perspective
Christopher J. T. Lewis

The Quantum Revolution: A Historical Perspective
Kent A. Peacock

Forces in Physics: A Historical Perspective
Steven Shore

THE QUANTUM REVOLUTION

A Historical Perspective

Kent A. Peacock

Greenwood Guides to Great Ideas in Science
Brian Baigrie, Series Editor

GREENWOOD PRESS
Westport, Connecticut • London

Library of Congress Cataloging-in-Publication Data

Peacock, Kent A., 1952–
The quantum revolution : a historical perspective / Kent A. Peacock.
p. cm. — (Greenwood guides to great ideas in science,
ISSN 1559–5374)
Includes bibliographical references and index.
ISBN-13: 978–0–313–33448–1 (alk. paper). 1. Quantum theory—History—Popular works. I. Title.
QC173.98.P43 2008
530.1209—dc22 2007039786

British Library Cataloguing in Publication Data is available.

Library of Congress Catalog Card Number: 2007039786
ISBN-13: 978–0–313–33448–1
ISSN: 1559–5374

First published in 2008

Greenwood Press, 88 Post Road West, Westport, CT 06881
An imprint of Greenwood Publishing Group, Inc.
www.greenwood.com

Printed in the United States of America

The paper used in this book complies with the Permanent Paper Standard issued by the National Information Standards Organization (Z39.48–1984).

10 9 8 7 6 5 4 3 2 1

CONTENTS

LIST OF ILLUSTRATIONS

SERIES FOREWORD

The volumes in this series are devoted to concepts that are fundamental to different branches of the natural sciences—the gene, the quantum, geological cycles, planetary motion, evolution, the cosmos, and forces in nature, to name just a few. Although these volumes focus on the historical development of scientific ideas, the underlying hope of this series is that the reader will gain a deeper understanding of the process and spirit of scientific practice. In particular, in an age in which students and the public have been caught up in debates about controversial scientific ideas, it is hoped that readers of these volumes will better appreciate the provisional character of scientific truths by discovering the manner in which these truths were established.

The history of science as a distinctive field of inquiry can be traced to the early seventeenth century when scientists began to compose histories of their own fields. As early as 1601, the astronomer and mathematician Johannes Kepler composed a rich account of the use of hypotheses in astronomy. During the ensuing three centuries, these histories were increasingly integrated into elementary textbooks, the chief purpose of which was to pinpoint the dates of discoveries as a way of stamping out all too frequent propriety disputes, and to highlight the errors of predecessors and contemporaries. Indeed, historical introductions in scientific textbooks continued to be common well into the twentieth century. Scientists also increasingly wrote histories of their disciplines—separate from those that appeared in textbooks—to explain to a broad popular audience the basic concepts of their science.

The history of science remained under the auspices of scientists until the establishment of the field as a distinct professional activity in the middle of the twentieth century. As academic historians assumed control of history of science writing, they expended enormous energies in the attempt to forge a distinct and autonomous discipline. The result of this struggle to position the history of science as an intellectual endeavor that was valuable in its own right, and not merely in consequence of its ties to science, was that historical studies of the natural sciences were no longer composed with an eye toward educating a wide audience that included nonscientists, but instead were composed with the aim of being consumed by other professional historians of science. And as historical breadth was sacrificed for technical detail, the literature became increasingly daunting in its technical detail. While this scholarly work increased our understanding of the nature of science, the technical demands imposed on the reader had the unfortunate consequence of leaving behind the general reader.

As Series Editor, my ambition for these volumes is that they will combine the best of these two types of writing about the history of science. In step with the general introductions that we associate with historical writing by scientists, the purpose of these volumes is educational—they have been authored with the aim of making these concepts accessible to students—high school, college, and university—and to the general public. However, the scholars who have written these volumes are not only able to impart genuine enthusiasm for the science discussed in the volumes of this series, they can use the research and analytic skills that are the staples of any professional historian and philosopher of science to trace the development of these fundamental concepts. My hope is that a reader of these volumes will share some of the excitement of these scholars—for both science, and its history.

Brian Baigrie
University of Toronto
Series Editor

PREFACE

This book is a short version of the story of quantum mechanics. It is meant for anyone who wants to know more about this strange and fascinating theory that continues to transform our view of the physical world. To set forth quantum physics in all its glorious detail takes a lot of mathematics, some of it quite complicated and abstract, but it is possible to get a pretty accurate feeling for the subject from a story well told in words and pictures. There are almost no mathematical formulas in this book, and what few there are can be skimmed without seriously taking away from the storyline. If you would like to learn more about quantum mechanics, the books and Web pages I describe in "Further Reading" can lead you as far into the depths of the subject as you wish to go.

One thing this book does not do is to present a systematic account of all of the *interpretations* that have been offered of quantum mechanics. That would take another book at least as long. However, certain influential interpretations of quantum theory (such as the Copenhagen Interpretation, the causal interpretation, and the many-world theory) are sketched because of their historical importance.

Quantum mechanics is often said to be the most successful physical theory of all time, and there is much justification for this claim. But, as we shall see, it remains beset with deep mysteries and apparent contradictions. Despite its tremendous success, it remains a piece of unfinished business. It is the young people of today who will have to solve the profound puzzles that still remain, and this little work is dedicated to them and their spirit of inquiry.

ACKNOWLEDGMENTS

My own research in foundations of quantum mechanics has been supported by the Social Sciences and Humanities Research Council of Canada, the University of Lethbridge and the University of Western Ontario. For valuable discussions, suggestions, guidance, and support in various ways I thank Brian Baigrie, Bryson Brown, James Robert Brown, Jed Buchwald, Kevin deLaplante, Kevin Downing, Brian Hepburn, Jordan Maclay, Ralph Pollock, and (especially) Sharon Simmers.

INTRODUCTION: WHY LEARN THE HISTORY OF QUANTUM MECHANICS?

This book tells the story of quantum mechanics. But what is quantum mechanics? There are very precise and technical answers to this question, but they are not very helpful to the beginner. Worse, even the experts disagree about exactly what the essence of quantum theory really is. Roughly speaking, quantum mechanics is the branch of physical science that deals with the very small—the atoms and elementary particles that make up our physical world. But even that description is not quite right, since there is increasing evidence that quantum mechanical effects can occur at any size scale. There is even good reason to think that we cannot understand the origins of the universe itself without quantum theory. It is more accurate, although still not quite right, to say that quantum mechanics is something that *started* as a theory of the smallest bits of matter and energy. However, the message of this book is that the growth of quantum mechanics is not finished, and therefore in a very important sense we still do not know what it really is. Quantum mechanics is revolutionary because it overturned scientific concepts that seemed to be so obvious and so well confirmed by experience that they were beyond reasonable question, but it is an incomplete revolution because we still do not know precisely where quantum mechanics will lead us—nor even why it must be true!

The history of a major branch of science like quantum physics can be viewed in several ways. The most basic approach to see the history of quantum mechanics is as the story of the discovery of a body of interrelated facts (whatever a "fact" is), but we can also view our story as a history of the concepts of the theory, a history of beautiful though sometimes strange mathematical equations, a history of scientific papers, a history of crucial experiments and measurements, and a history of physical models. But science is also a profoundly human enterprise; its development is conditioned by the trends and accidents of history, and by the abilities, upbringing, and quirks of its creators. The history of science is not just a smooth progression of problems being solved one after the other by highly competent technicians, who all agree with each other about how their work should be done. It is by no means clear that it is inevitable that we would have arrived where we are now if the history of science could be rerun. Politics, prejudice, and the accidents of history play their part (as we shall see, for instance, in the dramatic story of David Bohm). Thus, the history of quantum mechanics is also the story of the people who made it, and along the way I will sketch brief portraits of some of these brilliant and complex individuals.

Quantum mechanics is one of the high points in humanity's ongoing attempt to understand and cope with the vast and mysterious universe in which we find ourselves, and the history of modern physics—with its failures and triumphant insights—is one of the great stories of human accomplishment of our time.

WHY WOULD ANYONE BE INTERESTED IN HISTORY OF SCIENCE?

Learning a little history of science is one of the most interesting and painless ways of learning a little of the science itself, and knowing something about the people who created a branch of science helps to put a human face on the succession of abstract scientific concepts.

Furthermore, knowing at least the broad outlines of the history of science is simply part of general cultural literacy, since we live in a world that is influenced deeply by science. Everyone needs to know something about what science is and how it developed. But the history of modern physics, especially quantum physics, presents an especially interesting puzzle to the historian. In the brief period from 1900 to 1935 there occurred one of the most astonishing outbursts of scientific creativity in all of history. Of course, much has been done in science since then, but with the perspective of hindsight it seems that no other historical era has crammed so much scientific creativity, so many discoveries of new ideas and techniques, into so few years. Although a few outstanding individuals dominate—Albert Einstein (of course!), Niels Bohr, Werner Heisenberg, Wolfgang Pauli, Paul Dirac, and Erwin Schrödinger stand out in particular—they were assisted in their work by an army of highly talented scientists and technicians.

This constellation of talented people arose precisely at a time when their societies were ready to provide them with the resources they needed to do their work, and also ready to accept the advances in knowledge that they delivered. The scientists who created quantum theory were (mostly) not embattled heretics like Galileo, because they did not have to be—their work usually was supported, encouraged, and welcomed by their societies (even if their societies were at times a bit puzzled as to what that work meant). The period in which quantum mechanics was created is thus comparable to a handful of other brilliant episodes in history—such as ancient Athens in her glory, or the England of Elizabeth I—when a multitude of historical factors somehow combined to allow the most talented people to do the best work of which they were capable.

Exactly why do these amazing outbursts of creativity occur? And what could we do to make them happen more regularly? These questions certainly can't be answered in this modest book, but the history of quantum mechanics is an outstanding case study for this large and very important problem.

WHY SHOULD SCIENTISTS LEARN HISTORY OF SCIENCE?

For the general public, history of science is an important part of culture; for the scientist, history of science is itself a sometimes neglected research tool (Feyerabend 1978). It may seem odd to suggest that knowing the history of a science can aid research in that science. But the history of science has particular value as a research tool precisely because it allows us to see that some of the assumptions on which present-day science is based might have been otherwise—and perhaps, in some cases, should have been. Sometimes, when science is presented in elementary textbooks and taught in high school or college, one is given the impression that every step along the way was inevitable and logical. In fact, science often has advanced by fits and starts, with numerous wrong turns, dead ends, missed opportunities, and arbitrary assumptions. Retracing the development of science might allow us to come at presently insoluble problems from a different angle. We might realize that somewhere along the line we got off track, and if we were to go back to that point and start over we might avoid the problems we have now. Science is no different than any other sort of problem-solving activity in that, if one is stuck, there often can be no more effective way of getting around the logjam than going back and rethinking the whole problem from the beginning.

The history of science also helps to teach modern-day scientists a certain degree of humility. It is sobering to learn that scientific claims that are now treated as near-dogma (for instance, the theory of continental drift or the fact that meteors are actual rocks falling from the sky) were once laughed at by conventional science, while theories such as Newtonian mechanics that were once regarded as unquestionable are now understood to be merely approximately correct, if not completely wrong for some applications. Many of the new ideas of quantum mechanics were found to be literally unbelievable, even by their creators, and in the end they were accepted not because we understood them or were comfortable with them, but because nature *told us* that they were true.

The history of quantum theory can also teach us much about the process of scientific discovery. How did Planck, Schrödinger, Heisenberg, or Dirac arrive at their beautiful equations? It may seem surprising to someone not familiar with theoretical physics to realize that there is no way of *deducing* the key equations of new theories from facts about the phenomena or from previously accepted theories. Rather, many of the most important developments in modern physics started with what physicists call an *Ansatz,* a German word that literally means "a start," but which in physics can also be taken as an inspired insight or lucky guess. The new formulas are accepted because they allow a unified deduction of facts that had previously been considered to be unrelated and because they lead to new predictions that get confirmed by experiment. So we often end up with a scientific law expressed in mathematical form that works very well in the sense that we can learn how to use it to predict what will happen in concrete physical situations, but we do not understand *why* it can make those predictions. It just works, so we keep using it and hope that some day we will understand it better.

We now have a branch of physics, quantum mechanics, which is the most powerful and effective theory of physics ever developed in the sense that it gives unprecedented powers of prediction and intervention in nature. Yet it remains mysterious, for despite the great success of quantum mechanics, we must admit in all humility that we don't know why it must be true, and many of its predictions seem to defy what most people think of as "common sense." Quantum mechanics was, as this history will show, a *surprise* sprung on us by nature. To the story of how this monumental surprise unfolded we now turn.

1

THE TWILIGHT OF CERTAINTY

MAX CHOOSES A CAREER

The time had come for Max Planck to make a career choice. He was fascinated by physics, but a well-meaning professor at the University of Munich told him that he should turn to music as a profession because there were no more important discoveries to be made in physics. The year was 1875.

Young Max was an exceptionally talented pianist, and the advice that he should become a musician seemed reasonable. But he stubbornly chose physics anyway. Max was motivated not so much by a yearning to make great discoveries, as an aspiring young scientist might be today, but rather by an almost religious desire to understand the laws of nature more deeply. Perhaps this motivation had something to do with his upbringing, for his ancestors included pastors and jurists, and his father was a professor of law at the University of Kiel.

As a student he was especially impressed by the recently discovered First Law of Thermodynamics, which states that the energy books must always balance—the total amount of energy in a physical system never changes even though that energy can appear in many different forms. To Planck, the First Law seemed to express the ideal of science in its purest form, for it was a law that did not seem (to him!) to be a mere descriptive convenience for humans, but rather something that held true exactly, universally, and without qualification. It is ironic that the deeply conservative Planck would become the one to trigger quantum mechanics, the most revolutionary of all scientific developments. As we shall see, however, Planck was also possessed of unusual intellectual integrity, and the great discovery he was eventually to make had much to do with the fact that he was among those relatively rare people who can *change their minds* when the evidence demands it.

AN AGE OF COMPLACENCY NEARS ITS END

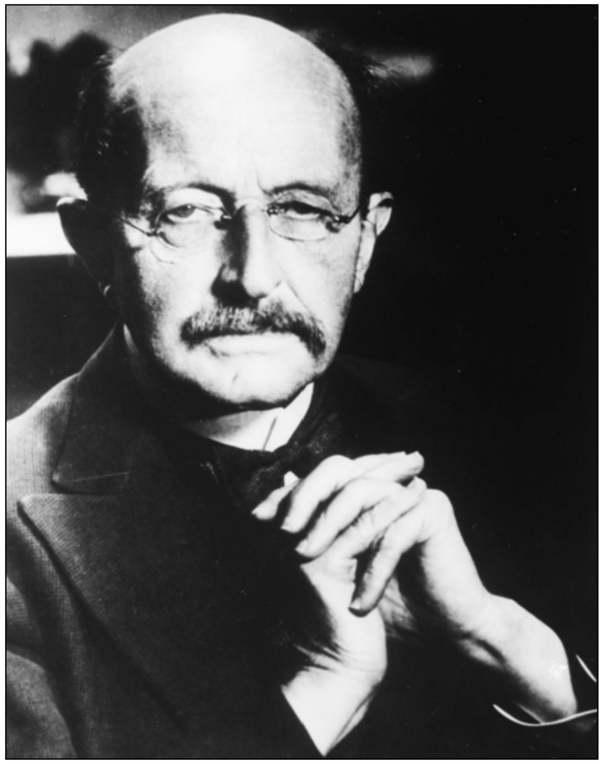

Figure 1.1: Max Planck. AIP Emilio Segre Visual Archives.

Before we describe Planck's discovery of the quantum, we should try to understand why his advisor was as satisfied as he was with the way things were in 1875.

The complacency at the end of the nineteenth century was both scientific and political. After the final defeat of Napoleon in 1815, Western Europe had enjoyed a long run of relative peace and prosperity, marred only by the Franco-Prussian war of 1870–1871. From this conflict Germany had emerged triumphant and unified, proud France humiliated. The British Empire continued to grow in strength throughout the last decades of the century, although it was challenged by rival colonial powers like Germany, France, and Belgium. The brash new nation of the United States was healing from a terrible civil war, flexing its muscles and gaining in confidence, but it seemed unimaginable that the great empires of Europe could ever lose their power.

Meanwhile, things were not so nice for many people who were not European. The prosperity of Europe was bought at the expense of subjugated peoples in Africa, India, and the Americas, who had almost no defense in the face of modern weapons such as machine guns, rapid fire rifles, artillery, the steamship, and the telegraph wire. Eventually Europeans would turn these weapons on each other, but the horrors of World War I lay 40 years in the future when young Max Planck began to study physics.

Science and technology in the nineteenth century had enjoyed unprecedented growth and success. The world was being changed by innumerable innovations such as the steam engine, the telegraph, and later the telephone. Medicine made huge advances (so that by the end of the nineteenth century one could have a reasonable hope of actually surviving a surgical operation), and there was a tremendous expansion of what we now call "infrastructure" such as highways, railways, canals, shipping, and sewers.

The technology of the nineteenth century was underpinned by a great increase in the explanatory and predictive power of scientific theory. Mathematics, chemistry, astronomy, and geology leaped ahead, and all of biology appeared in a new light with Darwin's theory of evolution by natural selection. To many scientists of the time it seemed that there were just a few loose ends to be tied up. As we shall see, tugging on those loose ends unraveled the whole overconfident fabric of nineteenth century physics.

PHYSICS IN THE NINETEENTH CENTURY

The Foundation

Physics investigates the most general principles that govern nature, and expresses those laws in mathematical form. Theoretical physics at the end of the nineteenth century rested on the massive foundation of the mechanics of Sir Isaac Newton (1644–1727), an Englishman who had published his great book *The Mathematical Principles of Natural Philosophy* in 1687. Newton showed how his system of mechanics (which included a theory of gravitation) could be applied to the solution of many long-standing problems in astronomy, physics, and engineering. Newton also was coinventor (with the German Gottfried Leibniz, 1646–1716) of the calculus, the powerful mathematical tool which, more than any other advance in mathematics, made modern physics possible. (Newton, who was somewhat paranoid, accused Leibniz of having poached the calculus from him, and the two geniuses engaged in a long and pointless dispute over priority.)

Newtonian mechanics was deepened and generalized by several brilliant mathematical physicists throughout the eighteenth and nineteenth centuries, notably Leonard Euler (1707–1783), Joseph Louis Lagrange (1736–1813), Pierre Simon de Laplace (1749–1827), and Sir William Rowan Hamilton (1805–1865). By the late nineteenth century it not only allowed for accurate predictions of astronomical motions, but it had evolved into an apparently universal system of mechanics which described the behavior of matter under the influence of any possible forces. Most physicists in late 1800s (including the young Max Planck) took it for granted that any future physical theories would have to be set within the framework of Newtonian mechanics.

Electrodynamics

It is hard for us now to picture that up until almost the middle of the nineteenth century, electricity and magnetism were considered to be entirely distinct phenomena. Electrodynamics is the science that resulted when a number of scientists in the early to mid-nineteenth century, notably Hans Christian Oersted (1777–1851), Michael Faraday (1791–1867), and André Marie Ampère (1775–1836), discovered that electricity and magnetism are different aspects of the same underlying entity, the electromagnetic field. Faraday was a skilled and ingenious experimenter who explained his results in terms of an intuitive model in which electrified and magnetized bodies were connected by graceful lines of force, invisible to the eye but traceable by their effects on compass needles and iron filings.

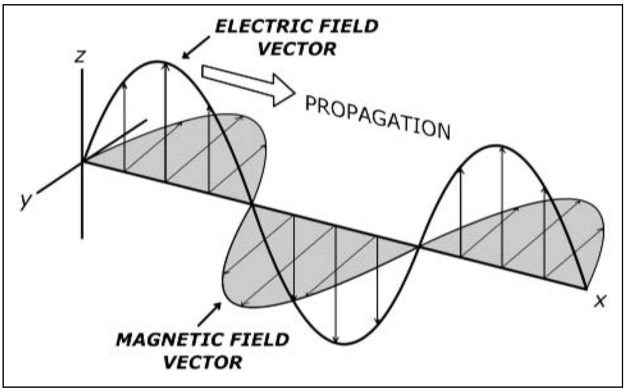

Figure 1.2: Light Waves. Maxwell and Hertz showed that light and other forms of electromagnetic radiation consist of alternating electric and magnetic fields. Illustration by Kevin deLaplante.

Faraday may have been the last great discoverer in physics who did not express his insights in mathematical form. The Scottish mathematical physicist James Clerk Maxwell (1831–1879) unified the known laws of electricity and magnetism into an elegant and powerful mathematical picture of the electromagnetic field that Faraday had visualized intuitively. Maxwell published the first version of his field equations in 1861. He achieved one of the most outstanding examples in physics of a successful unification, in which phenomena that had been thought to be of quite different natures were suddenly seen to be merely different aspects of a single entity. Maxwell's field equations are still used today, and they remain the most accurate and complete description of the electromagnetic field when quantum and gravitational effects can be ignored.

One of the most important predictions of electromagnetic theory is the existence of electromagnetic waves, alternating patterns of electric and magnetic fields vibrating through space at the speed of light. In 1888 the German physicist Heinrich Hertz (1857–1894) detected electromagnetic waves with a series of delicate and ingenious experiments in which he created what were, in effect, the first radio transmitters and receivers. It was soon realized that light itself is simply a flood of electromagnetic waves that happen to be visible to the human eye. Different types of electromagnetic waves may be distinguished by their frequencies or their wavelengths. (Wavelength is inverse to frequency, meaning that as the frequency goes up the wavelength goes down.) The frequency expresses how fast the wave is vibrating and is usually given in cycles per second. The wavelength is the length of the wave from crest to crest. Electromagnetic waves are *transverse,* meaning that they vibrate in a direction perpendicular to their direction of motion, while sound waves and other pressure waves are *longitudinal,* meaning that they vibrate more or less in the direction of motion. The *polarization* of electromagnetic waves is a measure of the direction in which they vibrate.

Electromagnetic waves can vary from radio waves many meters long, to the deadly high energy gamma rays produced by nuclear reactions which have wavelengths less than 1/5000 that of visible light. Visible light itself has wavelengths from about 400 billionths of a meter (violet) to about 700 billionths of a meter (red). The range of observed frequencies of light is called the *spectrum.* We shall have much to say about spectra, which will play a central role in the history of quantum mechanics.

Maxwell's theory was highly abstract, and it took several years before its importance was generally apparent to the scientific community. But by the end of the nineteenth century the best-informed physicists (including Planck) regarded Maxwellian electrodynamics as one of the pillars on which theoretical physics must rest, on a par with the mechanics of Newton. In fact there were deep inconsistencies between the electromagnetic theory of Maxwell and Newtonian mechanics, but few thinkers grasped this fact, apart from an obscure patent clerk in Switzerland whom we shall meet in the next chapter.

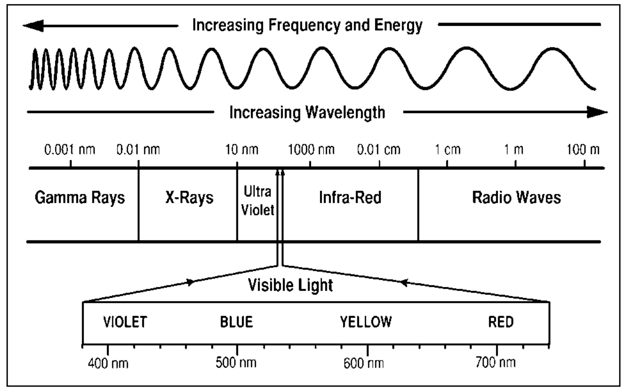

Figure 1.3: The Electromagnetic Spectrum. Electromagnetic waves exist in a spectrum running from low-energy, long-wavelength radio waves to very high-energy, short-wavelength gamma rays. For all such waves the energy is related to the frequency by $E = h\nu$, where ν (Greek letter nu) is the frequency, and h is Planck's constant of action. Illustration by Kevin deLaplante.

Thermodynamics

More than any other branch of physics, thermodynamics, the science of heat, had its origins in practical engineering. In 1824, a brilliant young French engineer, Sadi Carnot (1796–1832), published a groundbreaking analysis of the limitations of the efficiency of heat engines, which are devices such as the steam engine that convert heat released by the combustion of fuel to useful mechanical energy. Following Carnot, several pioneering investigators in the mid-nineteenth century developed the central concepts of what we now call classical thermodynamics. These include temperature, energy, the equivalence of heat and mechanical energy, the concept of an absolute zero (a lowest possible temperature), the First Law of Thermodynamics (which states that energy cannot be created or destroyed, but only converted from one form to another), and the basic relationships between temperature, pressure, and volume in so-called ideal gasses.

The mysterious concept of *entropy* made its first explicit appearance in the work of the German Rudolph Clausius (1822–1888). Clausius defined entropy as the ratio of the change in heat energy to the temperature and coined the term "entropy" from the Greek root *tropé,* transformation. He showed that entropy must always increase for *irreversible* processes. A reversible process is a cycle in which a physical system returns to its precise initial conditions, whereas in an irreversible process order gets lost along the way and the system cannot return to its initial state without some external source of energy. It is precisely the increase in entropy that distinguishes reversible from irreversible cycles. Clausius postulated that the entropy of the universe must tend to a maximum value. This was one of the first clear statements of the Second Law of Thermodynamics, which can also be taken to say that it is impossible to transfer heat from a colder to a hotter body without expending at least as much energy as is transferred. We are still learning how to interpret and use the Second Law.

The concept of irreversibility is familiar from daily life: it is all too easy to accidentally smash a glass of wine on the floor, and exceedingly difficult to put it together again. And yet the laws of Newtonian dynamics say that all physical processes are reversible, meaning that any solution of Newton's laws of dynamics is still valid if we reverse the sign of time in the equations. It ought to be possible for the scattered shards of glass and drops of wine to be tweaked by the molecules of the floor and air in just the right way for them to fly together and reconstitute the glass of wine. Why doesn't this happen? If we believe in Newtonian mechanics, the only possible answer is that it *could* happen, but that it has never been seen to happen because it is so enormously *improbable.* And this suggests that the increase in entropy has something to do with probability, a view that seems obvious now but that was not at all obvious in the mid-nineteenth century.

Clausius himself had (in the 1860s) suggested that entropy might be a measure of the degree to which the particles of a system were disordered or disorganized, but (like most other physicists of the era) he was reluctant to take such speculation seriously. In the classical thermodynamics of Clausius, entropy and other quantities such as temperature, pressure, and heat are *state functions,* which means that they are treated mathematically as continuous quantities obeying exact, exception-free laws.

Unlike electrodynamics, which seemed to have been perfected by Maxwell, thermodynamics therefore remained in an incomplete condition, and its troubles centered on the mysteries of entropy, irreversibility, and the Second Law of Thermodynamics. Planck himself tried for many years to find a way of explaining the apparently exception-free, universal increase of entropy as a consequence of the reversible laws of Newtonian and Maxwellian theory. But the brilliant Austrian physicist Ludwig Boltzmann (1844–1906) showed that there is an entirely different way to think about entropy. Other people (notably James Clerk Maxwell) had explored the notion that heat is the kinetic energy of the myriad particles of matter, but Boltzmann rewrote all of classical thermodynamics as a theory of the large-scale statistics of atoms and molecules, thereby creating the subject now known as *statistical mechanics.*

In statistical mechanics we distinguish *macroscopic* matter, which is at the scale that humans can perceive, from *microscopic* matter at the atomic or particulate level. On this view, entropy becomes a measure of disorder at the microscopic level. Macroscopic order masks microscopic disorder. If a physical system is left to itself, its entropy will increase to a maximum value, at which point the system is said to be in *equilibrium.* At equilibrium, the system undergoes no further macroscopically apparent changes; if it is a gas, for instance, its temperature and pressure are equalized throughout. The apparent inevitability of many thermodynamic processes (such as the way a gas will spread uniformly throughout a container) is due merely to the huge numbers of individual molecules involved. It is not mathematically inevitable, but merely overwhelmingly probable, that gas molecules released in a container will rapidly spread around until all pressure differences disappear.

Could there be exceptions to the Second Law? According to the statistical interpretation, it is not strictly impossible to pipe usable energy from a lower temperature to a higher—it is merely, in general, highly improbable. A pot of water could boil if placed on a block of ice—but we're going to have to wait a very (*very!*) long time to see it happen.

Boltzmann's statistical mechanics ran into strong opposition from a number of scientists. Some, the "energeticists," headed by Wilhelm Ostwald (1853–1932), maintained that all matter consists of continuous fields of energy, so that no fundamental theory should be based on such things as discrete atoms or molecules. Severe criticism also came from the positivist Ernst Mach (1838–1916), who insisted that atoms were not to be taken seriously because they had never been directly observed. (Positivism is a view according to which concepts are not meaningful unless they can be expressed in terms of possible observations.) Mach's influence on physics was both good and bad; while he impeded the acceptance of statistical mechanics, his penetrating criticism of classical concepts of space and time stimulated the young Einstein. Mach also did important work in gas dynamics, and the concept of Mach number (the ratio of the speed of an aircraft to the speed of sound) is named after him.

A major barrier to the acceptance of the statistical interpretation of thermodynamics was the fact that thermodynamic quantities such as pressure, temperature, heat energy, and entropy were first studied as properties of ordinary matter. On the scale of human perception, matter appears to be continuously divisible. We are now accustomed to thinking of the heat content of a volume of a gas as the total *kinetic energy* (energy of motion) of the molecules of which the gas is composed, but heat was first studied as an *apparently* continuously distributed property of smooth matter. In fact, up until about the mid-1800s, heat was still thought of as a sort of fluid, called *caloric.* It therefore seemed reasonable that thermodynamic quantities such as heat or temperature should obey mathematical laws that were as exact as Newton's Laws or Maxwell's field equations, and it was very difficult for most physicists to accept the notion that the laws of thermodynamics were merely descriptions of *average* behavior.

Most important, there was still no irrefutable theoretical argument or direct experimental evidence for the existence of *atoms.* The concept of the atom goes back to the ancient Greek thinker Democritus (ca. 450 B.C.), and the term "atom" itself comes from a Greek word meaning "indivisible." By the nineteenth century the atomic theory was a mainstay of physics and chemistry, but it was regarded by many theoretical physicists as nothing more than a useful calculational device that allowed chemists to work out the correct amounts of substances to be mixed in order to achieve various reactions. There seemed to be no phenomenon that had *no* reasonable explanation except in terms of atoms. Demonstrations that there are such phenomena would be provided in the years 1900–1910 by a number of people, including Einstein himself.

Boltzmann suffered from severe depression, possibly aggravated by the endless debates he was forced to engage in to defend the statistical view, and he committed suicide in 1906. On his gravestone (in Vienna) is engraved the equation that bears his name: $S = k \ln W$ (the entropy of a state is proportional to the natural logarithm of the probability of that state). Had Boltzmann lived a few more years, he would have witnessed the complete vindication of his ideas.

What Is Classical about Classical Physics?

Newtonian mechanics and Maxwellian electrodynamics together made up what we now call *classical physics*. The three defining characteristics of classical physics are *determinism, continuity,* and *locality*; all are challenged by quantum mechanics.

In order to understand what determinism means, we need to know a little about the sort of mathematics used in classical physics. It was taken for granted that physics is nothing other than the mathematical description of processes occurring in space and time. (Later on even this assumption would be challenged by quantum physics.) From the time of Newton onward, most laws of physics are expressed in the form of *differential equations,* one of the most useful offshoots of the calculus invented by Newton and Leibniz. Such equations describe the ways in which physical quantities (such as electrical field strength) vary with respect to other quantities (such as position or time). Newton's First Law of dynamics, for instance, states that the rate of change of momentum of a physical object equals the force impressed upon it. Physical laws expressed in differential equations are applicable to indefinitely many situations. All possible electromagnetic fields obey Maxwell's field equations, for instance. To apply the equations we *solve* them for specific situations; this means that we use *initial* and *boundary* conditions (which describe a given physical scenario) to calculate a mathematical curve or surface that will represent the behavior of the system in that scenario. For example, if we know the initial position and velocity of a moving particle, and the forces acting on it, we can use Newton's First Law to calculate its trajectory over time. The sorts of differential equations used in classical physics are such that (in most cases) fully specified initial and boundary conditions imply unique solutions. In other words, in classical physics the future is uniquely and exactly determined by the past, and this is just what we mean by determinism.

The belief in continuity was often expressed in the phrase "Nature makes no jumps." It was assumed by almost all physicists from the time of Newton onward that matter moves along smooth, unbroken paths through space and time. This view was only reinforced by the success of the Faraday-Maxwell theory of the electromagnetic field, which explained electrical and magnetic forces as a result of a force field continuously connecting all electrically charged bodies. On the field view, the appearance of disconnection between particles of matter is merely that—an appearance. Mathematically, the assumption that all physical processes are continuous required that physics be written in the language of differential equations whose solutions are continuous, differentiable ("smooth") functions.

By the late nineteenth century, many physicists (led by Maxwell and Boltzmann) were using statistical methods, which are *indeterministic* in the sense that a full specification of the macroscopic state of a system at a given time is consistent with innumerable possible microscopic futures. However, the classical physicists of the nineteenth century believed that probabilistic methods were needed only for practical reasons. If one were analyzing the behavior of a gas, for instance, one can only talk about the average behavior of the molecules. It would be totally impossible to know the exact initial positions and velocities of all the gas molecules, and even if one could have these numbers the resulting calculation to predict their exact motions would be impossibly complex. Nevertheless, each individual molecule surely had a position and velocity, and a future that was in principle predictable, even if it was practically impossible to know these things. Planck and his contemporaries in the 1890s would have found it incredible that by the late 1920s it would be reasonable to question the apparently obvious belief that the parts of matter *had* a precise position and momentum before an experimenter interacts with them.

Later in our story we shall have much to say about locality, and the circumstances under which quantum physics forces it to break down. For now, we will just take locality, as it would have been understood in Planck's younger years, to mean that all physical influences propagate at a *finite* speed (and in a continuous manner) from one point to another. According to the doctrine of locality there is no such thing as *action at a distance*—a direct influence of one body on another distant body with no time delay and no physical vehicle by means of which the force was propagated. Most physicists from Newton onward felt that the very notion of action at a distance was irrational; recently, quantum mechanics has forced us to rethink the very meaning of rationality.

Too Many Loose Ends

Physics in the late nineteenth century was an apparently tightly knit tapestry consisting of Newtonian mechanics supplemented with Maxwell's theory of the electromagnetic field and the new science of thermodynamics. Up to roughly 1905 most physicists were convinced that any electromagnetic and thermal phenomena that could not yet be explained could be shoehorned into the Newtonian framework with just a little more technical cleverness. However, in the period from about 1880 to 1905 there were awkward gaps and inconsistencies in theory (mostly connected with the nature of light), and a few odd phenomena such as radioactivity that could not be explained at all. In 1896, Henri Becquerel (1852–1909) noticed that uranium salts would expose photographic film even if the film was shielded from ordinary light. This was absolutely incomprehensible from the viewpoint of nineteenth-century physics, since it involves energy, a lot of it, coming out of an apparently inert lump of ore. The intellectual complacency of the late nineteenth century, like the confident empires that sheltered it, did not have long to last.

BLACKBODY RADIATION AND THE THERMODYNAMICS OF LIGHT

Before Planck

There is an old joke (probably first told by a biologist) that to a physicist a chicken is a uniform sphere of mass *M*. The joke has a grain of truth in it, for physics is often written in terms of idealized models such as perfectly smooth planes or point particles that seem to have little to do with roughhewn reality. Such models are surprisingly useful, however, for a well-chosen idealization behaves enough like things in the real world to allow us to predict the behavior of real things from the theoretical behavior of the models they resemble.

One of the most useful idealized models is the *blackbody,* which was defined by Gustav Kirchhoff (1824–1887) in 1859 simply as any object that absorbs all of the electromagnetic radiation that falls upon it. (It doesn't matter what material the object is made of, so long as it fulfills this defining condition.) Physicists from the 1850s onward began to get very interested in the properties of blackbodies, since many objects in the real world come close to exhibiting near-perfect blackbody behavior. While perfect blackbodies *reflect* no radiation at all, Kirchhoff proved that they must *emit* radiation with a spectral distribution that is a function mainly (or in the case of an ideal blackbody, *only*) of their temperatures. (When we use the term "radiation" here, we are talking about electromagnetic radiation—light and heat—not nuclear radiation. And by spectral distribution, we mean a curve that gives the amount of energy emitted as a function of frequency or wavelength.) Steelmakers can estimate the temperature of molten steel very accurately just by its color. A near-blackbody at room temperature will have a spectral peak in the infrared (so-called heat radiation). The peak will shift to higher frequencies in step with increasing temperature; this is known as Wien's Displacement Law, after Wilhelm Wien (1864–1928). Around 550°C objects begin to glow dull red, while an electric arc around 10,000°C is dazzling blue-white.

Blackbody radiation is also known as *cavity* radiation. To understand this term, consider an object (which could be made of any material that reflects radiation) with a cavity hollowed out inside it. Suppose that the only way into the cavity is through a very small hole. Almost all light that falls on the hole will enter the cavity without being reflected back out, because it will bounce around inside until it is absorbed by the walls of the cavity. The small hole will therefore behave like the surface of a blackbody. Now suppose that the object containing the cavity is heated to a uniform temperature. The walls of the cavity will emit radiation, which Kirchhoff showed would have a spectrum that depended only on the temperature and which would be independent of the material of the walls, once the radiation in the cavity had come to a condition of equilibrium with the walls. (Equilibrium in this case means a condition of balance in which the amount of radiant energy being absorbed by the walls equals the amount being emitted.) The radiation coming *out* of the little hole will therefore be very nearly pure blackbody radiation. Because the spectrum of blackbody radiation depends only on the temperature it is also often called the *normal spectrum*.

In the last 40 years of the nineteenth century the pieces of the blackbody puzzle were assembled one by one. As noted, Kirchhoff was able to show by general thermodynamic considerations that the function that gave the blackbody spectrum depended only on temperature, but he had no way to determine the shape of the curve. In 1879 Josef Stefan (1835–1893) estimated on the basis of experimental data that the total amount of energy radiated by an object is proportional to its temperature raised to the fourth power (a very rapidly increasing function). Stefan's law (now known as the Stefan-Boltzmann Law) gives the total area under the spectral distribution curve. Boltzmann, in 1884, found a theoretical derivation of the law and showed that it is only obeyed exactly by ideal blackbodies.

Research on blackbodies was spurred in the 1890s not only by the great theoretical importance of the problem, but by the possibility that a better understanding of how matter radiates light would be of value to the rapidly growing electrical lighting industry. In 1893, Wien showed that the spectral distribution function had to depend on the *ratio* of frequency to temperature, and in 1896 he conjectured an exponential distribution law (Wien's Law) that at first seemed to work fairly well. In the same period, experimenters were devising increasingly accurate methods of measuring blackbody radiation, using radiation cavities and a new device called the *bolometer,* which can measure the intensity of incoming radiation. (The bolometer was invented by the American Samuel P. Langley (1834–1906) around 1879, and it had its first applications in astronomy.) Deviations from Wien's Law were soon found at lower frequencies (in the infrared), where it gave too *low* an intensity. It is at this point that Planck enters the story.

Planck's Inspired Interpolation

Planck had been Professor of Physics at the University of Berlin since 1892 and had done a great deal of distinguished work on the applications of the Second Law of Thermodynamics and the concept of entropy to various problems in physics and chemistry. Throughout this work, his aim was to reconcile the Second Law of Thermodynamics with Newtonian mechanics. The sticking point was irreversibility. Boltzmann's statistical account of irreversibility as a result of ever-increasing disorder was natural and immediate, but, as we have noted, it implied that the Second Law was not exact and exception-free.

Although Planck had great personal respect for Boltzmann, he questioned Boltzmann's statistical approach in two ways. First, the increase of entropy with time seemed *inevitable* and therefore, Planck believed, should be governed by rigorous laws. He did not want a result that was merely probabilistic since it was virtually an article of faith for Planck that the most general physical laws had to be exact and deterministic. Second, Planck wanted to rely as little as possible on models based on the possible properties of discrete particles, because their existence remained largely speculative.

At first, Planck explored the possibility that the route to lawlike irreversibility could be found in electromagnetic theory. He tried to show that the scattering of light, which had to obey Maxwell's Equations, was irreversible. However, in 1898 Boltzmann proved that Maxwell's electromagnetic field theory, like Newtonian mechanics, was time-reversal invariant. This fact had not been explicitly demonstrated before, and it torpedoed Planck's attempt to find the roots of irreversibility in electromagnetic theory.

Planck became interested in the blackbody problem in the mid-1890s for a number of reasons. First, many of his colleagues were working on it at the time. More important, he was very impressed by the fact that the blackbody spectrum is a function only of temperature; it was, therefore, something that had a universal character, and this was just the sort of problem that interested Planck the most. Also, he thought it very likely that understanding how radiation and matter come into equilibrium with each other would lead to a clearer picture of irreversibility. Planck devised a model of the blackbody cavity in which the walls were made of myriad tiny resonators, vibrating objects that exchanged radiant energy with the electromagnetic fields within the cavity. He established Wien's formula in a way that probably came as close as anyone could ever come to deriving an irreversible approach to equilibrium from pure electromagnetic theory, but without quite succeeding.

Around the same time as Planck and other German scientists were struggling to understand the blackbody spectrum, Lord Rayleigh (1842–1919), probably the most senior British physicist of his generation, took an entirely different approach to the problem. He derived an expression for the number of possible modes of vibration of the electromagnetic waves within the cavity, and then applied a rule known as the *equipartition theorem,* a democratic-sounding principle that claimed that energy should be distributed evenly among all possible states of motion of any physical system. This led to a spectral formula that seemed to be roughly accurate at lower frequencies (in the infrared). However, there is no limit to the number of times a classical wave can be subdivided into higher and higher frequencies, and assuming that each of the infinitely many possible higher harmonics had the same energy led to the *ultraviolet catastrophe*—the divergence ("blow-up") of the total energy of the cavity to infinity. Rayleigh multiplied his formula by an exponential "fudge factor" which damped out the divergence, but which still did not lead to a very accurate result.

As soon as it was found that Wien's Law failed experimentally at lower frequencies, Planck threw himself into the task of finding a more accurate formula for Kirchhoff's elusive spectral function. He had most of the pieces of the puzzle at hand, but he had to find how the entropy of his resonators was related to their energy of vibration. He had an expression for this function derived from Wien's Law and that was therefore roughly valid for high frequencies, and he had a somewhat different expression for this function derived from Rayleigh's Law and that was therefore roughly valid for low frequencies. By sheer mathematical skill Planck found a way to combine these two formulas into one new radiation formula that approximated the Rayleigh Law and Wien's Law at either end of the range of frequencies, but that also filled in the gap in the middle. Planck presented his new radiation law to the German Physical Society on October 19, 1900. By then, it had already been confirmed within the limits of experimental error, and no deviations have been found from it since.

Planck had achieved his goal of finding a statement of natural law that was about as close to absolute as any person can probably hope to achieve, but his formula was largely the product of inspired mathematical guesswork, and he still did not know why it was true. An explanation of his new law had to be found.

The Quantum Is Born

Boltzmann had argued that the entropy of any state is a function (the logarithm) of the probability of that state, but Planck had long resisted this statistical interpretation of entropy. He finally came to grasp that he had to make a huge concession and, in desperation, try Boltzmann's methods.

This meant that he had to determine the *probability* of a given distribution of energy among the cavity resonators. In order to calculate a probability, the possible energy distributions had to be *countable,* and so Planck divided the energies of the resonators into discrete chunks that he called *quanta,* Latin for "amount" or "quantity." He then worked out a combinatorial expression that stated how many ways there are for these quanta to be distributed among all the resonators. (It was later shown by Einstein and others that Planck's combinatorial formula was itself not much better than a wild guess—but it was a guess that gave the right answer!) The logarithm of this number gave him the expression for entropy he needed. There was one further twist: in order to arrive at the formula that experiment told him was correct, the size of these quanta had to be proportional to the frequencies of the resonators. These last pieces of the puzzle allowed him to arrive, finally, at a derivation of the formula for the distribution of energy among frequencies as a function of temperature. He announced his derivation on December 14, 1900, a date that is often taken to be the birthday of quantum mechanics.

Planck was inspired by a calculation that Boltzmann had carried out in gas theory, in which Boltzmann also had taken energy to be broken up into small, discrete chunks. Boltzmann had taken the quantization of energy to be merely a calculational device that allowed him to apply probabilistic formulas, and the size of his energy units dropped out of his final result. This worked for classical gasses, where quantum phenomena do not make an explicit appearance. But Planck found that if he tried to allow his quanta to become indefinitely small, his beautiful and highly accurate formula fell apart. If he wanted the right answer, he had to keep the quanta.

The constant of proportionality between energy and frequency was calculated by Planck from experimental data, and he arrived at a value barely 1% off the modern accepted value, which is 6.626×10^{-27} erg.seconds. (The erg is a unit of energy.) This new constant of nature is now called *Planck's constant of action* and is usually symbolized with the letter h. Action, energy multiplied by time, is a puzzling physical quantity; it is not something like mass or distance that we can sense or picture, and yet it plays a crucial role in theoretical physics. *Why* action must be quantized, and *why* the quantum of

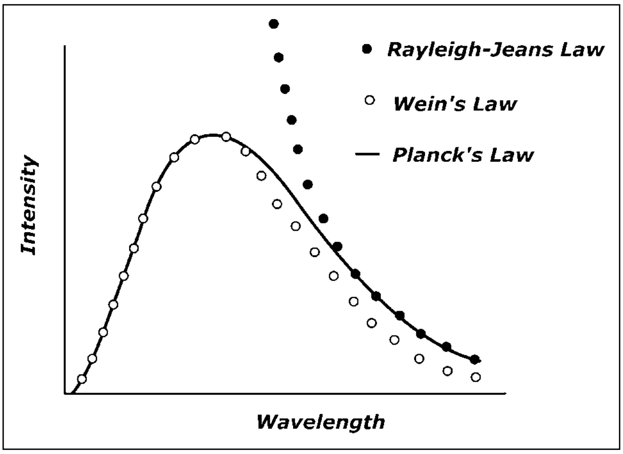

Figure 1.4: Planck's Law. The Rayleigh-Jeans Law fits the experimental curve at long wavelengths, Wien's Law fits the curve well at short wavelengths, and Planck's formula fits the curve at all wavelengths. Illustration by Kevin deLaplante.

action should have the particular value that it has been measured to have, remain complete mysteries to this day.

Historians of physics still debate the exact extent to which Planck in 1900 accepted the reality of energy quanta, and to what extent he still thought of them as a mathematical trick. In later years he tried without success to explain light quanta in classical terms. However, the indisputable fact remains that in 1900 he reluctantly adopted Boltzmann's statistical methods, despite the philosophical and scientific objections he had had towards them for many years, when he at last grasped that they were the only way of getting the result that he knew had to be correct. Planck's outstanding intellectual honesty rewarded him with what he described to his young son as a "discovery as important as that of Newton" (Cropper 1970, p. 7).

2

EINSTEIN AND LIGHT

THE PATENT CLERK FROM BERN

Physics today stands on two great pillars, quantum theory and the theory of relativity. Quantum theory was the work of many scientists, but the foundations of the special and general theories of relativity were almost entirely the work of one person, Albert Einstein. Einstein also played a large role in the creation of quantum mechanics, especially in its early stages; furthermore, he was among the first to grasp the extent to which quantum mechanics contradicts the classical view of the world. Later in his life he sought to replace quantum mechanics with what he thought would be a more rational picture of how nature works, but he did not succeed. He once said that he wanted above all else to understand the nature of light.

Einstein was born of a Jewish family in Ulm, Germany, in 1879. His performance as a student was uneven, but he independently taught himself enough mathematics and physics that he was able to do advanced research by the time he was in his early 20s. He graduated from the Polytechnical Institute of Zurich, Switzerland, in 1900 with the equivalent of a bachelor's degree, although he had a bad habit of skipping classes and got through the final exams only with the help of notes borrowed from a fellow student, Marcel Grossman (1878–1936). Einstein was unable to find an academic or research position and eked out a living with odd jobs, mostly tutoring and part-time teaching. Grossman again came to the rescue and through connections got Einstein an interview with the Swiss Patent Office in Bern. Einstein seems to have impressed the director of the office with his remarkable knowledge of electromagnetic theory, and in 1902 he was hired as a patent examiner-in-training, Technical Expert Third Class.

The Patent Office suited Einstein perfectly. Here he found a quiet haven (and a modest but steady paycheck) that allowed him to think in peace. The technical work of reviewing inventions to see if they were patentable was stimulating. In the years 1901–1904 he published five papers in *Annalen der Physik* (Annals of Physics), one of the most prestigious German research journals. All had to do with the theory and applications of statistical mechanics. One of his major aims in these papers was to find arguments that established without a doubt the actual existence of molecules, but he was also assembling the tools that would allow him to attack the deepest problems in physics. In three of his papers in this period, Einstein single-handedly reconstructed the statistical interpretation of thermodynamics, even though he had at that time no more than a sketchy acquaintance with the work of Boltzmann or the American J. W. Gibbs (1839–1903, the other great pioneer of statistical mechanics). Einstein's version of statistical mechanics added little to what Boltzmann and Gibbs had already done, but it was an extraordinary accomplishment for a young unknown who was working almost entirely alone. Furthermore, it gave him a mastery of statistical methods that he was to employ very effectively in his later work on quantum theory.

THE YEAR OF MIRACLES

The year 1905 is often referred to as Einstein's *annus mirabilis* (year of miracles). He published five papers: one was his belated doctoral thesis—an influential but not earth-shattering piece of work on the "determination of molecular dimensions"—while the other four changed physics forever. One showed that the jiggling motion of small particles suspended in liquids could be used to prove the existence of molecules; one laid the foundations of the theory of relativity; one paper took Planck's infant quantum theory to its adolescence in a single leap; and one short paper (apparently an afterthought) established the equivalence of mass and energy. These papers, written and published within a few months of each other, represent one of the most astonishing outbursts of individual creativity in the history of science.

Brownian Motion

In one of his great papers of 1905 Einstein studied the "motion of small particles suspended in liquids" from the viewpoint of the "molecular-kinetic theory of heat" (i.e., statistical mechanics). This paper does not directly involve quantum mechanical questions, but it is important to the quantum story in that it was one of several lines of investigation in the period 1900–1910 (carried out by a number of scientists) that established beyond a reasonable doubt that atoms—minute particles capable of moving independently of each other—really do exist.

Suppose that there are particles of matter (such as pollen grains), suspended in a liquid, invisible or almost invisible to the naked eye but still a lot larger than the typical dimensions of the molecules of the liquid. The molecules of the liquid bounce around at random and collide with the suspended particles. Einstein realized that the statistics of such collisions could be analyzed using the same methods that are used to analyze gasses. A single collision cannot cause the particle to move detectably, but every so often (and Einstein showed how to calculate how often) a *fluctuation* will occur in which a cluster of molecules will just happen to hit the particle more or less in unison, and they will transfer enough momentum to make the particle jump. Over time the particles will zigzag about in a random "drunkard's walk." This is known as Brownian motion, after the English botanist Robert Brown (1773–1858), who in 1827 observed pollen grains and dust particles mysteriously jittering about when suspended in water.

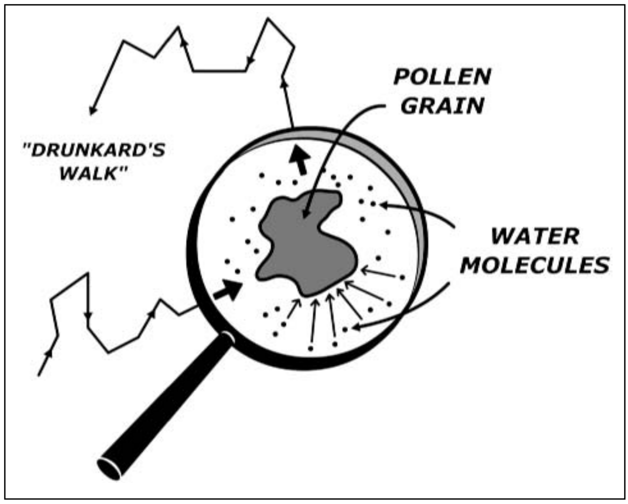

Figure 2.1: Fluctuations and Brownian Motion. Random fluctuations in the jittering motion of the water molecules can cause the much heavier pollen grain to change direction. This amounts to a localized violation of the Second Law of Thermodynamics. Illustration by Kevin deLaplante.

Einstein derived a formula for the mean-square (average) distance the particles jump. Amazingly, his formula allows one to calculate a key quantity known as Avogadro's Number, the number of molecules in a mole of a chemical substance. The Polish physicist Marian Smoluchowsky (1872–1917) had independently obtained almost the same results as Einstein, and their formulas soon passed experimental tests carried out by the French physicist Jean Perrin (1870–1942). This confirmation of the Einstein-Smoluchowsky picture provided one of the most convincing proofs that had yet been obtained of the reality of discrete atoms and molecules. It was also a demonstration that Boltzmann had been right in saying that the Second Law of Thermodynamics was only a statistical statement that holds on average, for a Brownian fluctuation amounts to a small, localized event in which entropy by chance has momentarily decreased.

Einstein's work on Brownian motion demonstrated his remarkable knack for finding an elegant, powerful, and novel result based on the consistent application of general physical principles.

Special Relativity

Relativity theory is not the main topic of this book, but it is impossible to understand quantum theory, and especially the challenges it still faces today, without knowing some basics of the theory that is most commonly linked to Einstein's name.

The problem that Einstein set himself in his great paper of 1905, "On the Electrodynamics of Moving Bodies," was to unify mechanics with the electrodynamics of Maxwell. Theorists of the late nineteenth century argued that one could make no sense of electromagnetic waves unless there was some fluid-like stuff which *waved*, and they imagined that the vacuum of space was filled with an invisible substance called the *ether*. If it existed, the ether had to have very odd properties, since it had to be extremely rigid (given the very high velocity of light) but at the same time infinitely slippery (since the vacuum offers no resistance to motion).

The failure (in the 1880s and later) of all attempts to directly detect the motion of the Earth with respect to the ether is often cited as one of the factors that led to Einstein's theory of relativity. Although Einstein was aware of these observations, they do not seem to have played a major role in his thinking. What really seems to have bothered him was the sheer lack of conceptual elegance that resulted from trying to force electrodynamics to be consistent with Newtonian mechanics. At the beginning of his 1905 paper, he points out that according to the electrodynamics of his time, the motion of a conductor with respect to a magnetic field had a different description than the motion of a magnet with respect to a conductor, despite the fact that only the *relative* motion of the two makes any difference to the actual phenomena observed. He argued that it should make no difference to the laws of physics whether they are described from the viewpoint of an observer in uniform (steady) motion or at rest. This is called the *Principle of Special Relativity.*

To this he added the assumption, which he admitted might seem at first to contradict the Principle of Relativity, that the speed of light in vacuum must be the same (an *invariant*) for all observers regardless of their state of motion. In effect, he was insisting that we should take the observable value of the speed of light in vacuum to be a law of nature. (The qualification "in vacuum" is needed because light usually slows down when it passes through various forms of matter.) At the age of 16 Einstein had puzzled over the following question: what would happen if an observer could chase a beam of light? If he caught up with it, would the light disappear, to be replaced by a pattern of static electrical and magnetic fields? Nothing in Maxwell's theory allows for this, a fact that led Einstein to the postulate that *light is always light*—for everyone, regardless of their state of motion.

All of the startling consequences of the theory of relativity follow from the mathematical requirement that positions and time must transform from one state of motion to another in such a way as to maintain the invariance of the speed of light. Einstein expressed the speed of light in terms of the space and time coordinates of two observers moving with respect to each other at a constant velocity. He then set the resulting expressions for the speed of light for the two observers equal to each other and derived a set of equations that allow us to calculate distances and times for one observer given the distances and times for the other. These formulas are called the *Lorentz* transformations, after the Dutch physicist H. A. Lorentz (1853–1928). (Lorentz had been one of those who had tried without success to find a Newtonian framework for electrodynamics.) According to the Lorentz transformations, clocks run more slowly in moving frames (this is called time *dilation*), lengths contract, and mass approaches infinity (or *diverges*) at the speed of light. But these effects are relative, since any observer in a relativistic universe is entitled to take him or herself to be at rest. For instance, if two people fly past each other at an appreciable fraction of the speed of light, each will calculate the other's clock to be running slow. This contradicts the Newtonian picture, in which time is absolute—the same for all observers. The relativistic deviations from Newton's predictions are very small at low relative velocities (although they can now be detected with sensitive atomic clocks) but become dominant at the near-light speeds of elementary particles.

There are certain quantities, called *proper* quantities, which are not relative, however; hence the term "theory of relativity" is misleading because it does not say that *everything* is relative. As Einstein himself once noted, his theory could more accurately be termed the *theory of invariants*, because its main aim is to distinguish those quantities that are relative (such as lengths and times) from those that are not (such as proper times and rest masses).

An important example of an invariant is elapsed proper time, which is the accumulated local time recorded by an observer on the watch he carries with him. The elapsed proper time of a moving person or particle determines the rate at which the person or particle ages, but it is *path dependent,* meaning that its value depends upon the particular trajectory that the person or particle has taken through spacetime. This is the basis of the much-debated *twin paradox* (not a paradox at all), according to which twins who have different acceleration histories will be found to have aged differently when brought back together again. This prediction has been tested and confirmed to a high degree of accuracy with atomic clocks and elementary particles.

In 1908 the mathematician Hermann Minkowski (1864–1909) showed that special relativity could be expressed with great clarity in terms of a mathematical construct he called *spacetime* (now often called *Minkowski Space*), a four-dimensional geometrical structure in which space and time disappear as separate entities. (There are three spatial dimensions in spacetime, but it can be represented by a perspective drawing which ignores one of the spatial directions.) This is just a distance-time diagram with units of distance and time chosen so that the speed of light equals 1. Locations in spacetime are called *worldpoints* or *events*, and trajectories in spacetime are called *worldlines*. A worldline is the whole four-dimensional history of a particle or a person. The central feature of Minkowski space is the *light cone,* which is an invariant structure (that is, the same for all observers). The light cone at a worldpoint *O* defines the absolute (invariant) past and future for that point. The *forward cone* is the spacetime path of a flash of light emitted at *O,* while the *past cone* is the path of all influences from the past that could possibly influence *O*, on the assumption that no influences can reach *O* from points outside the light cone. (Such influences would be *superluminal,* traveling faster than light.) One of the central assumptions behind modern quantum mechanics and particle physics is that the *light cone* defines the *causal structure* of events in space and time;

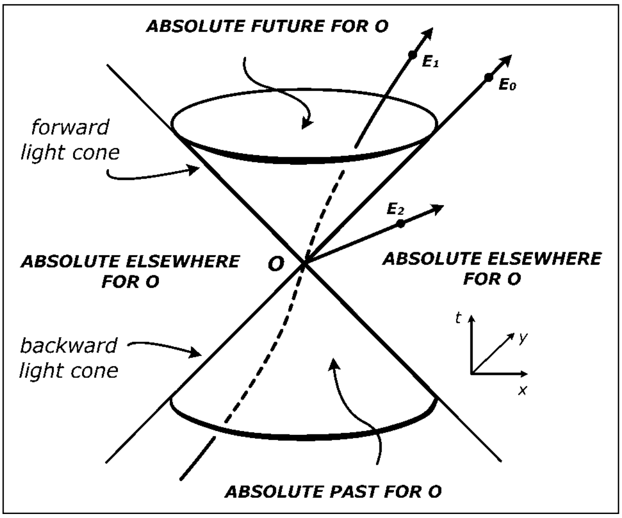

Figure 2.2: Spacetime According to Minkowski. The curve OE_1 is the path of an ordinary particle with mass, such as an electron. OE_0 is the path of a light ray emitted from O, and it is the same for all frames of reference. OE_2 is the path of a hypothetical tachyon (faster than light particle). Illustration by Kevin deLaplante.

that is, it delineates the set of all possible points in spacetime that could either influence or be influenced by a given point such as *O*. We will see that this assumption is called into question by *quantum nonlocality,* which we shall introduce in a later chapter.

Minkowski showed how relativistic physics, including Maxwell's electromagnetic theory, could be described geometrically in spacetime. An irony is that Minkowski had been one of Einstein's mathematics professors at the Zurich Polytechnic, but he had not, at that time, been very impressed with Einstein's diligence.

One of the most impressive consequences of special relativity is the famous equivalence of mass and energy. Physicists now commonly speak of *mass-energy,* because mass and energy are really the same thing. Any form of energy (including the energy of light) has a mass found by dividing the energy by the square of the speed of light. This is a very large number, so the mass equivalent of ordinary radiant energy is small. Conversely, matter has energy content given by multiplying its mass m by the square of the speed of light; that is, $E=mc^2$. This means that the hidden energy content of apparently solid matter is very high—a fact demonstrated with horrifying efficiency in August 1945, when the conversion to energy of barely one-tenth of one percent of a few kilograms of uranium and plutonium was sufficient to obliterate two Japanese cities, Hiroshima and Nagasaki.

It is commonly held that the theory of relativity proves that no form of matter, energy, or information can be transmitted faster than the speed of light. Einstein himself certainly believed this, and in his paper of 1905 he cited the divergence (blow-up) of mass to infinity at the speed of light as evidence for this view. Whether or not Einstein was right about this is a controversial issue that turns out to be crucial to our understanding of quantum mechanics. Some people hold that we can explain certain odd quantum phenomena (to be described later) only if we assume that information is somehow being passed between particles faster than light; others hotly deny this conclusion. The mathematical derivation of the Lorentz transformations and other central results of special relativity depends only on the assumption that the speed of light is an *invariant* (the same for all observers in all possible states of motion), not a *limit.* Some physicists suspect that the existence of hypothetical *tachyons* (particles that travel faster than light) is consistent with the basic postulates of relativity (although they have never been detected), and whether or not any form of energy or information could travel faster than light remains an open question that quantum mechanics will not allow us to ignore.

The Quantum of Light

Einstein's 1905 paper on the light quantum is entitled "On a Heuristic Viewpoint Concerning the Generation and Transformation of Light." To understand Einstein's use of the word "heuristic" we have to say a little about the history of light.

Newton carried out pioneering experiments in which he showed that light from the sun can be spread out by a prism into the now-familiar spectrum (thereby explaining rainbows). In his *Opticks* (1704), Newton speculated that light moves in the form of tiny particles or corpuscles. The English polymath Thomas Young (1773–1829) showed, however, that the phenomena of interference and diffraction make sense only if light takes the form of waves. If light is shone through pinholes comparable in size to the wavelength of the light, it spreads out in ripples in much the way that water waves do when they pass through a gap in a seawall.

Young's views were reinforced by Maxwell's field theory, which showed that light can be mathematically interpreted as alternating waves of electric and magnetic fields. The wave theory became the dominant theory of light in the nineteenth century, and the corpuscular theory was judged a historical curiosity, a rare case where the great Newton had gotten something wrong.

Einstein's brash "heuristic hypothesis" of 1905 was that Newton was correct and light is made of particles after all. To say that a hypothesis is heuristic is to say that it gets useful results but that it cannot be justified by, or may even contradict, other accepted principles. Thus it is something that we would not accept unconditionally, but rather with a grain of salt, hoping that in time we will understand the situation more completely. Einstein taught us something not only about light, but about scientific method: if you find a fruitful hypothesis, do not reject it out of hand because it clashes with what you think you already "know." Rather, learn as much from it as you can, even if you are not yet sure where it fits into the grand scheme of things. Einstein knew perfectly well that there was abundant evidence for the wave interpretation of light. However, he showed that the hypothesis that light moves in the form of discrete, localized particles, or quanta, could explain some things about light that the wave theory could not explain, and would lead to the prediction of new phenomena that would not otherwise have been predictable.

The starting point of Einstein's argument in his 1905 paper on the light quantum was Wien's blackbody energy distribution law. Even though Wien's formula had been superseded by Planck's Law, Einstein knew that it is still fairly accurate for high frequencies and low radiation density. Planck had worked forward from an expression for entropy to his energy distribution law; in contrast, Einstein worked backwards from Wien's law to an expression for the entropy of high-frequency radiation in equilibrium in a blackbody. Einstein demonstrated the remarkable fact that this quantity has the same mathematical form as the entropy of a *gas,* a system of independent, small, localized particles bouncing around freely in space. At least when Wien's law can be used, therefore, light behaves not only like waves, but also very much like a gas of particles whose energy happens to be given by the product of Planck's mysterious constant of action and the frequency of the light waves.

Einstein applied this result to three types of phenomena where light interacts with matter. The most important historically was the *photoelectric effect.* It had been known for some years that light shone on a metal surface causes the emission of electrons from the metal, but the laws governing the process were not well understood. Suppose that light interacts with the electrons as a continuous wave. If the frequency of the light were held constant, a steady increase in the intensity (brightness) of the light would increase both the number and energy of the electrons knocked out of the metallic surface because the light would transmit its energy continuously to the electrons. However, if light interacts with the electrons in the form of discrete particles whose energy is given by the Planck Law, then whether or not the light has enough energy to knock the electrons loose would depend strictly on the frequency of the light. An increase in intensity would only increase the *number* of electrons emitted, not their energies. A test of this prediction of Einstein's would therefore be a crucial test of the light particle hypothesis.

Einstein's theory of light quanta was regarded as very radical. Even Planck, who tirelessly championed Einstein's work, thought that Einstein had overreached. The American physicist Robert Millikan (1868–1953) set out to disprove Einstein's wild hypothesis, but to his surprise his careful experiments (carried out around 1915) confirmed Einstein's theory of the photoelectric effect. The term *photon* for the light particle was introduced in 1926 by the American chemist Gilbert Lewis (1875–1946), and it quickly caught on.

In the story of the photon we see the first clear sign of a phenomenon that would plague quantum theory from this point onward, the necessity of living with points of view that seemed to be in outright contradiction to each other. Is light really made of waves (which are continuous), or particles (which are discrete)? It was already apparent to Einstein in 1905 that we must accept the *wave-particle duality,* which is the fact that light somehow must be *both* continuous and discrete. In his 1905 paper Einstein sketched (nonmathematically) a possible resolution of the anomaly. The key, he pointed out, is that there is a distinction between what applies to averages and what applies to individual cases. (Think of the old joke that the average American family has 2.4 children.) Einstein argued that while we can only understand the interactions of light with matter if we think of light as particulate, the statistics of very large numbers of light particles *averages out* to the wave-like behavior implicit in Maxwell's Equations. The challenge would then be to find the rules that related the behavior of light as individual quanta to their large-scale wave-like behavior. The only thing that Einstein could be sure of in 1905 was the Planck relation between energy and frequency: light quanta have an energy that is a product of Planck's constant of action and the frequency of the light wave with which the quanta are "associated." However, the nature of that association remained utterly mysterious.

EINSTEIN ARRIVES

Specific Heats

In the years immediately following Einstein's *annus mirabilis,* few apart from Planck understood Einstein's theory of relativity, and even Planck did not fully grasp Einstein's insights about the light quantum. However, in 1907 Einstein published a paper in which he applied the quantum hypothesis to another unsolved problem, the puzzle of specific heats. This paper, in effect, founded *solid state physics,* the quantum mechanical study of solid matter. The *specific heat* of a substance (solid, liquid, or gas) is the amount of energy required to raise its temperature by one degree. It can also be thought of as the *heat capacity* of the substance, the amount of heat that can be absorbed by a given amount of the substance at a certain temperature or pressure; equivalently, it is the amount of heat that would be given off if the substance's temperature were lowered by one degree.

The puzzle was that classical mechanics predicted that all solids should have exactly the same specific heat (about six calories per mole.degree). This is called the Dulong-Petit Law, after P. Dulong (1785–1838), and A. Petit (1791–1820), who established experimentally that their law works fairly well at room temperatures or higher. By the late nineteenth century it became technically possible to measure specific heat at lower temperatures, and it was soon clear that below room temperature most solids have much smaller specific heats than the Dulong-Petit Law says they should. The classical Dulong-Petit result follows from the assumption of the *equipartition of energy* (the idea that the energy of the system is shared equally among all possible vibration modes), the same assumption that tripped up Lord Rayleigh in the blackbody story. Einstein applied a quantum correction factor from blackbody theory to the classical Dulong-Petit result and arrived at a formula for specific heats that predicted the drop-off at low temperatures. Experimental work by Walther Nernst (1864–1941) showed, by 1910, that Einstein's formula came quite close to predicting the actual measured values for many materials. In 1912 the Dutch physicist Peter Debye (1884–1966) filled in some gaps in Einstein's theory and arrived at a formula that predicted the specific heats of most substances with remarkable accuracy. The success of the quantum theory of specific heats convinced most physicists that the quantum—and Einstein—had to be taken seriously.

The Wave-Particle Duality

By 1908 Einstein was recognized by Planck and a few other senior physicists as an up-and-coming star, even if they were not always sure what he was talking about. It was time for Einstein to move on from the Patent Office. Getting an academic position was no simple matter, however, even for someone with a Ph.D. and the amazing publication record that Einstein had already built up by 1908. He had to serve time as a *Privatdozent*—an unsalaried lecturer—at the University of Bern, before he secured a professorship at the University of Zurich in 1909 and could finally resign from the Patent Office. Einstein's friend Friedrich Adler (1879–1966) had been recommended for the position because his father was a powerful political figure. However, Adler, a person of selfless generosity, insisted that there was no comparison between his abilities as a physicist and Einstein's and insisted that Einstein should have the job instead.

We have already noted that our story can be viewed in several ways: as a history of ideas, a history of great papers, or a history of decisive experiments. It can also be viewed as a history of conferences. In 1909 Einstein delivered an influential paper at a conference in Salzburg in which he argued for the reality of light quanta and the necessity of accepting the wave-particle duality. In his work in 1909 Einstein applied his mastery of fluctuation theory to light itself. Light fluctuations are the short-lived deviations from average energy that occur in a radiation field. Einstein showed that the expression for fluctuations must consist of *two* parts, a contribution from the wave-like behavior of light, and one from its particle-like behavior. (The latter is similar to the fluctuations that cause Brownian motion.) It was another powerful argument for the wave-particle duality, and another demonstration that the quantum was not going to be explained away merely by fine-tuning the classical wave theory of light.

Einstein was a central figure in the First Solvay Conference, held in Brussels, Belgium, in 1911. This conference brought many of the leading physicists of Europe together to debate the challenge posed to physics by the quantum. The Solvay Conferences on Physics, founded by the Belgian industrialist Ernst Solvay (1838–1921), are held every three years and continue to provide an important stimulus to the development of physics.

Einstein rapidly moved through a succession of increasingly senior academic positions until, in 1913, he was appointed professor at the University of Berlin. He refused to support the German war effort and worked obsessively on his research through the dark years of World War I. In the period from 1917 to 1925 he made further decisive contributions to quantum mechanics, but these will be more easily understood after we describe the work of Niels Bohr and others in the next chapter.

EINSTEIN MAPS THE UNIVERSE

Almost as soon as he had laid the groundwork of the theory of relativity in 1905, Einstein puzzled over ways to describe gravitation in a way that would be consistent with the Principle of Relativity. Newton's Law of Gravitation makes no reference to the *time* it takes for the gravitational interaction to propagate from one mass to another; in other words, it describes gravity as a force that acts instantaneously at a distance. Newton himself was very uncomfortable with this picture (since he thought it obvious that unmediated action at a distance was absurd), but he could find no better explanation of gravity. Einstein thought that it should be possible to describe gravity as a field, similar to Maxwell's electromagnetic field although perhaps with a more complex structure, but none of the apparently obvious ways of doing this worked.

The breakthrough came in 1907, when (sitting at his desk in the Patent Office) Einstein had what he described as the "happiest thought of my life" (Pais 1982, p. 178). It simply struck him that a person falling freely feels no gravitational force. We only feel the "force" of gravity when free fall is opposed by something; for example, when we stand on the surface of the Earth. This insight led Einstein to his *Equivalence Principle:* suppose an astronaut is in a rocket with no windows, and suppose that she experiences a constant uniform acceleration. There is no way she can tell whether the rocket motor is burning in such a way that the spacecraft accelerates constantly in a given direction, or whether it is sitting on the launch pad but subject to a uniform gravitational field. All motion, including the accelerated motion caused by the gravitational field, is relative.

Gravitation is therefore remarkably like an inertial force such as the centrifugal force that tries to throw us off a rotating merry-go-round. The centrifugal force is merely a consequence of the mass of our bodies trying to obey Newton's First Law and keep moving ahead in a straight line against the resistance of the merry-go-round, which is trying to hold us to a circular path. Similarly, Einstein realized that gravitation would be merely the consequence of our bodies trying to move along inertial paths, *if inertial paths were curved by the presence of matter*. The only sign that there is a gravitational field present would be the deviation from parallelism of the trajectories of freely falling bodies. For instance, two weights released side by side will move toward each other as they fall toward the center of the Earth.

Generalizing relativity theory so that it would describe all possible accelerated motions would therefore give a relativistic description of gravitation, but it would require that space and time be *curved.* Einstein needed to find a way of describing objects in arbitrarily curved geometric spaces. Again, his friend Marcel Grossman helped, by pointing out that the mathematics he needed, the *tensor calculus,* had recently been invented. After some false starts and much hard work, Einstein published a monumental paper in 1916 that set forth his *general* theory of relativity. His relativity theory of 1905 is, by contrast, called the *special* theory of relativity because it is concerned only with relative motions in flat (uncurved) spacetime.

General relativity describes gravitation by means of a set of field equations, which state, very roughly speaking, that the shape of space is determined by the distribution of mass-energy in space. There are many possible solutions of Einstein's field equations; each valid solution is a *metric* that describes a possible spacetime geometry. It has taken mathematical physicists many years to find, classify, and study the solutions of Einstein's equations, and it is still not clear that all physically interesting solutions have been found. However, Einstein was able to show that the light from a distant star would be bent by a certain amount by the gravitational field of the sun. This was verified in 1919, when a total eclipse of the sun allowed astronomers to measure the variation from the Newtonian prediction of starlight passing very close to the edge of the sun. Einstein also showed that his theory could explain a small deviation in the orbit of Mercury that no one had been able to account for. With these successful predictions, Einstein was hailed as the successor to Newton, and he spent the rest of his life coping with the sort of adulation that usually only movie stars are subjected to.

General relativity allows us to describe a vast range of possible universes, whose geometries are determined by the way mass-energy is distributed. As Misner, Wheeler, and Thorne (1973) put it, "Space tells matter how to move, and matter tells space how to curve." Some solutions of the field equations have bizarre properties. For instance, in 1935 Einstein and his young coworker Nathan Rosen (1909–1995) found a solution of the field equations that described *wormholes* (or "bridges") that could apparently link distant points in spacetime. (No one knows whether wormholes exist or whether they are merely a mathematical curiosity.) And in 1948 the mathematician Kurt Gödel showed that Einstein's theory allowed for a rotating universe in which it is possible to travel backwards in time.

In 1917, Einstein was startled to discover that his equations apparently predict that the whole universe is unstable—it must either expand or contract, but it cannot remain static. He thought that this had to be a mistake, and inserted a "fudge factor" (the *cosmological constant*) into his equations in order to stabilize the universe. However, in the 1920s Edwin P. Hubble (1889–1953) and other astronomers, using powerful new telescopes, discovered that the universe does indeed expand. Einstein later declared that his introduction of the cosmological constant had been the greatest scientific mistake of his life. However, recent work seems to show that the cosmological constant may be non-zero after all and is such that it tends to cause an *acceleration* of the expansion of the universe. Astronomers are still trying to measure the actual expansion rate of the universe as precisely as possible.

Einstein's general theory of relativity has so far stood up to every experimental test that astronomers have been able to devise. It is clear that general relativity is by far our most accurate description of the large-scale structure of spacetime, even though we still do not know *why* mass-energy should curve spacetime, or which of the myriad possible universes described by Einstein's field equations is the one we actually inhabit. As we shall see, one of the major theoretical challenges of our time is to make general relativity consistent with quantum mechanics, the other great theory that Einstein did so much to create.

EINSTEIN'S LATER YEARS

Einstein was forced to flee Germany in 1933 when Hitler took power. He settled in Princeton, New Jersey, as the star attraction at the new Institute for

Advanced Studies. Over the next few years, he devoted much effort to helping other refugees from Nazi persecution find positions in the free world.

Although Einstein had played a major role in the early years of quantum mechanics, he eventually became the leading critic of quantum mechanics rather than a contributor to it. In 1935, in collaboration with Boris Podolsky and Nathan Rosen, Einstein published an enigmatic article called "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" This paper (usually dubbed the EPR paper, for Einstein, Podolsky, and Rosen) is now recognized as one of Einstein's most influential contributions to physics —although not exactly in the way he had intended! The EPR paper, the debate surrounding it, and its implications will be described in Chapter 6.

In 1939, Einstein signed a letter drafted by his friend the Hungarian physicist Leo Szilard (1898–1964), urging President Franklin Delano Roosevelt to develop the atomic bomb before the Nazis did. Einstein forever regretted his role in triggering the development of nuclear weapons, and, in collaboration with other intellectuals such as the British philosopher Bertrand Russell (1872–1970), campaigned for nuclear disarmament. Throughout the last 30 years of his life he made repeated attempts to construct a field theory that would unify gravitation and electromagnetism in much the way that Maxwell had unified magnetism and electricity. His great hope was to show that particles could be seen as knot-like structures within some sort of universal field, thereby reattaching quantum mechanics and particle theory to a solid classical footing. He never succeeded, although he made many interesting suggestions that are still being digested today. He died in Princeton in 1955, the most revered scientist of modern times.

3

THE BOHR ATOM AND OLD QUANTUM THEORY

THINGS GET COMPLICATED

The quantum mechanics of 1900 to 1925 is often referred to as the Old Quantum Theory, to distinguish it from the modern quantum mechanics that will emerge suddenly in the mid-1920s. The earliest years of the Old Quantum Theory, 1900 to about 1910, were dominated by two figures, Planck and Einstein. After this point the development of quantum mechanics starts to get complicated as more and more physicists get involved and new lines of investigation branch out. If there could be one person to whom the period from 1913–1925 belongs, it would be the Danish physicist Niels Bohr (1885–1962). However, before we can describe what Bohr did we need to understand the two convergent streams of investigation, spectroscopy and nuclear physics, which made his work possible.

SPECTROSCOPY

Take a chemically pure sample of any element and heat it to incandescence —that is, to the point at which it glows. Pass the light through a narrow slit and then through a prism. The prism will fan the light rays out into a band with frequency increasing uniformly from one side to the other (or equivalently, wavelength decreasing). A series of bright, sharp lines will be observed—the *emission spectrum* of that particular element. It was discovered in the 1860s that each element has a unique spectrum like a fingerprint, which allows it to be identified regardless of how it is combined chemically with other elements. If the element is immersed in a very hot gaseous medium (such as the atmosphere of a star) its spectrum takes the form of a series of dark lines called an *absorption spectrum.* These dark lines made it possible to determine what elements were present in distant stars, a feat that some philosophers had thought would be impossible. Spectroscopy (the study of spectra) shows us that any matter in the universe that is hot enough to emit light is composed of the same familiar elements that we find on earth, and it quickly became one of the most powerful tools of the astronomer.

The line spectra of the various elements exhibit regular patterns, with the lines generally being spaced more closely together at higher energies. The puzzle was to understand the rule that governed the patterns of the spectral lines. Physicists in the nineteenth century realized that the spectral fingerprints had to be a clue to the inner structure and dynamics of the atoms that emitted them, but they were unable to find an atomic model that correctly predicted the observed spectral lines. A breakthrough came in 1884, when a Swiss schoolteacher named Johann Balmer (1825–1898) discovered an elegant formula that expressed the wavelengths of the hydrogen lines, the simplest spectrum, in terms of certain combinations of integers. Similar formulas were discovered by several other investigators, and by the late 1890s spectroscopists could identify many known elements by their spectra. Improvements in experimental technique allowed observation of atomic spectra from the infrared to the ultraviolet, and spectral series in these frequency ranges were discovered and described as well.

In 1908 W. Ritz (1878–1909), building on earlier work by J. R. Rydberg (1854–1919), showed that complex spectral lines are arithmetical combinations of simpler lines, called terms. (This is the Rydberg-Ritz *combination principle.*) Their formula depended on a number, which became known as the *Rydberg constant,* which could be estimated from spectral measurements. Frustratingly, however, there seemed to be no way to derive the value of the Rydberg constant theoretically, and no way of inferring atomic structure from the formulas of Balmer, Rydberg, and others. To go any further, something needed to be known about what went on inside the atom itself.

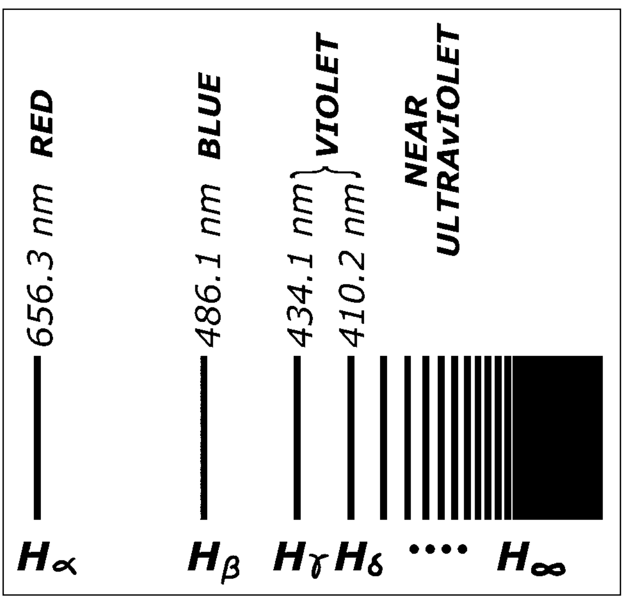

Figure 3.1: Spectral Lines. H∞ is the beginning of the continuum, the energy range where the electron is separate from the atom and can move freely. Illustration by Kevin deLaplante.

THE ELECTRON

As we have seen, up until the first decade of the twentieth century many physicists and chemists believed that matter was continuous and discrete atoms did not really exist. The age of elementary particles began in 1897, when British physicist J. J. Thomson

(1856–1940) discovered the electron. Thomson had been studying cathode rays, which are beams of energetic electrified matter emitted by the cathode (the negative electrode) in a vacuum tube. Thomson showed that cathode rays had to be made of something that was negatively charged, and he argued that the simplest explanation for the behavior of cathode rays was that they were made up of tiny negative particles, or corpuscles ("little bodies") of electric charge. In 1891 G. J. Stoney (1826–1911) had coined the term "electron" for hypothetical particles of electricity, and it soon came to be applied to Thomson's corpuscles. Thomson also showed that his corpuscles must be very small, less than one thousandth of the mass of the hydrogen atom.

Just as Planck and Einstein were about to learn that electromagnetic energy comes in discrete quanta, Thomson had shown that electricity—at least, negatively charged electricity—also occurs only in chunks of a definite size. The charge of the electron is now taken as the basic unit of electrical charge, and the electron became the first of the hundreds of so-called "elementary" particles that were to be identified. But we have to be careful when we say that Thomson *discovered* the electron. Physicists do not discover particles the way that archeologists discover the long-lost tombs of Egyptian pharaohs. Nobody has ever seen an electron or any other elementary particle with their bare eyes (although we can come fairly close to doing that with devices such as the cloud chamber, which show the tracks of charged particles). It is more accurate to say that a particle is "discovered" when it is shown that its existence is the only reasonable explanation for certain observed phenomena.

Thomson's corpuscular hypothesis was at first met with much skepticism, but within the next few years it became clear that it had to be right. Another source of confirmation for the electron hypothesis was the Zeeman Effect, discovered in 1896 by the Dutch physicist P. Zeeman (1865–1943). The Zeeman Effect is the splitting of spectral lines into small groups of closely separated lines of slightly different energies or frequencies under the influence of a strong magnetic field. H. A. Lorentz soon showed that it could be explained if the light from the hot atoms was emitted by negatively charged particles with a mass around one-thousandth that of a hydrogen atom, and if those particles have a *magnetic moment,* meaning that they are like tiny magnets that can interact with an external magnetic field. This was not only further evidence for the existence of electrons, but showed that electrons had to be present within the atom itself as well as in the cathode rays studied by Thomson. Lorentz's model was entirely classical, and it was several years before physicists realized that it had worked so well only because in the correct quantum description of the electrons interacting with the magnetic field, the value of Planck's constant neatly cancels out of the equations!

By not long after 1900 it was clear to physicists that the spectra of atoms had to be a sign of the periodic motions of the electrons within them. But what laws of motion did the electrons follow? Before it would be possible to begin to answer this question, more had to be known about the inner structure of the atom.

THE NUCLEUS

Rays and Radioactivity

In the last decade of the nineteenth century and the first decade of the twentieth century several extraordinary new types of radiation were discovered that could not be understood using the concepts of classical physics.

In 1895 the German physicist Wilhelm Roentgen (1845–1923) discovered and named X-rays. (The name "X-ray" was meant by Roentgen to suggest an unknown, as in algebra.) Although several other investigators had accidentally produced what we now know were X-rays, Roentgen was the first to realize that they were a novel phenomenon. X-rays are produced by the process of accelerating cathode rays through an electrical field (in the range of a few thousand volts) and allowing them to collide with a metal target. Most materials, if bombarded with sufficiently energetic electrons, will emit X-rays, and it was eventually found that all elements had a characteristic X-ray spectrum. Exactly how the cathode rays force a target to emit X-rays, however, would not be properly understood until the advent of quantum mechanics. Roentgen and others suspected almost immediately that X-rays are a form of electromagnetic radiation much more energetic than ultraviolet light, but it would not be until 1912 that this was confirmed when Max von Laue (1879–1960) showed that X-rays would diffract, just like light, with a very short wavelength when shone through a crystalline lattice. Roentgen also noticed that X-rays would image the bones of the human hand, and by 1896 X-ray devices were being used to diagnose broken bones and find bullets in wounded soldiers. Roentgen would be the first recipient of the Nobel Prize in Physics, in 1901.

Nuclear physics began in 1896, when the French physicist Henri Becquerel (1852–1908) discovered that salts of uranium would fog a photographic plate even if they were wrapped in heavy black paper. The term "radioactivity" itself was coined around 1898 by Marie Sklodowska Curie (1867–1934), a brilliant young student of physics at the Sorbonne in Paris who married the French physicist Pierre Curie (1859–1906) in 1895. Marie Curie discovered that thorium was also radioactive, and (with Pierre's help) carried out an exhausting chemical analysis of uranium tailings, which led in 1898 to the isolation of two new elements, polonium and radium, that were hundreds of times more radioactive than uranium itself.

Meanwhile, in 1902, Ernest Rutherford (1871–1937) and his assistant Frederick Soddy (1877–1956), working at McGill University in Montreal, grasped that radioactive decay consists of the actual transmutation of one element into another, as radioactive elements break down into other elements of lighter atomic weight. Rutherford was also the first to state the law of radioactive decay: unstable elements such as radium break down exponentially, at a rate proportional to the quantity of the element present, with a characteristic half-life. For example, a given quantity of radium will decay to half the original amount in about 1,600 years. Although it was not immediately apparent at the time, the law of radioactive decay hinted at one of the most profound mysteries of the quantum world—its inherently probabilistic nature. Why does one radium atom rather than another break down at a given time, even though we have absolutely no way of detecting any difference between them? We still have no clear answer to this question, although we can now say a lot more than could anyone in Rutherford's time about *why* we cannot answer this question.

The early investigators of radioactivity were astonished at the huge amounts of energy that poured out of radioactive atoms in the form of the mysterious "uranic rays." The energy of radioactivity can be a million times or more the energy released by the most powerful chemical reactions. It was remotely possible to imagine a classical explanation for the production of X-rays, since Maxwellian electrodynamics predicts that if a charged particle such as an electron is slowed down by collision with matter it must emit electromagnetic radiation. However, there was absolutely no way to account for the prodigious amounts of energy that was emitted by whatever lurked within the heart of the atom. Becquerel and Marie Curie at first explored the notion that radioactive elements might be somehow picking up some sort of previously undetected field of energy spread throughout space, or possibly absorbing energy from the sun; and Curie even toyed with the notion that the law of conservation of energy was being violated. (This would not be the last time that bizarre quantal phenomena would force physicists to speculate that Planck's sacred first law of thermodynamics was violated at the quantum level.) However, Marie Curie and most other physicists soon realized that the energy was more likely coming from inside the atoms themselves, and she argued (also correctly) that there had to be a slow but measurable loss of mass from radioactive atoms. As early as 1914 the science fiction novelist H. G. Wells (1866–1946) predicted that the energy inside the atom could be used to create weapons of appalling power, and he coined the very term "atomic bomb."

It was soon found that Becquerel's uranic rays consisted of three components, which were dubbed by Rutherford alpha, beta, and gamma radiation. These are quite different in character.

In 1908, after a decade of painstaking investigation, Rutherford demonstrated that the alpha particle is an ion of helium, with a positive charge of two and an atomic weight of roughly four times the weight of the hydrogen atom. It was spectroscopy that clinched the identification of the alpha particle as a form of helium, but no one at this time had any clear idea why it should have the charge and mass it had. Every high school student today knows that the helium nucleus is made up of two protons and two neutrons, and some will even know that there is another isotope of helium with only one neutron. However, in 1905 neither protons nor neutrons had been identified and the concept of an isotope (same atomic number, different numbers of neutrons) was unknown; nor was it yet established (although some suspected) that electrons orbit around a compact nucleus or that the number of electrons in a neutral atom equals the atomic number. It is challenging to describe what physicists in the first decade of the twentieth century knew of the structure of the atom without carelessly lapsing into modern language. One can easily forget how much hard work by men and women of extraordinary talent went into establishing the "obvious" facts that we take for granted today. Their work was often physical as much as intellectual; the early students of the atom such as Becquerel, the Curies, and Rutherford often had to build virtually all of their apparatus by hand, and the Curies labored like coal miners to extract radium from uranium ore, dangerously exposing themselves to radiation in the process.

Beta rays were shown by several investigators to be none other than Thomson's negative corpuscles—electrons—although possessing extraordinarily high energy. This was demonstrated as early as 1902, but beta radiation had certain mysterious properties that would take decades to fully understand.

Gamma radiation was discovered by Paul Villard (1860–1934) in 1900, but it was not until 1914 that Rutherford and E. N. Andrade (1887–1971) conclusively showed that it is a form of electromagnetic radiation, no different from light, radio waves, or X-rays, but with by far the highest energies ever observed.

These three types of atomic radiation, together with X-rays, are sometimes called *ionizing radiation,* since they have enough energy to knock electrons out of the atoms they hit and thereby produce *ions*—charged atomic fragments. Nuclear radiation can shatter delicate biomolecules like a rifle bullet hitting the motherboard of a computer. It took many tragic deaths before the medical dangers of radioactivity were recognized. Marie Curie herself very likely died from the aftereffects of radiation exposure. But the powerful rays emitted from radioactive substances also gave scientists, for the first time, a tool with which they could probe the heart of the atom itself.

Search for a Model

Several physicists from about 1900 to 1912 proposed atomic models with a central positive core and electrons either orbiting around it like planets around the sun or like the rings of Saturn. But none of these models could explain either the stability of the atom (the fact that it does not immediately collapse in a flash of light) or the observed rules of the spectral lines.

The most widely accepted theory of atomic structure up to about 1913 was J. J. Thomson's "plum pudding" model. Thomson imagined that the electrons were embedded like plums throughout a pudding of uniform positive charge, with the number of electrons equal to the atomic number of the atom, so that the overall charge would add up to zero. At this point there was no clear evidence that *positive* charge comes in discrete corpuscles. Thomson imagined the electrons to be swarming around in shells or rings in such a way that their electrostatic repulsions balanced, and he was encouraged by the fact that with some mathematical fudging he could get the structure of the shells to suggest the layout of the periodic table of the elements. The model was quite unable to predict the frequencies of observed spectral lines, however, and some physicists were beginning to fear that this would forever be an impossible task. But Thomson's model did at least give some account of the stability of the atom, and in the period of about 1904 to 1909, with still no direct evidence for the existence of a central positive core, the Thomson model seemed like the best bet—although likely few physicists (including Thomson himself) really believed in it.

Rutherford Probes the Atom

Ernest Rutherford, who in 1895 earned his passage from his homeland, New Zealand, to J. J. Thomson's laboratory at Cambridge on the strength of scholarships and hard work, was an energetic and ingenious experimentalist who could learn enough mathematics to do theory when he had to. Rutherford was the epitome of scientific leadership; many of the key discoveries that founded nuclear physics emerged from his students and coworkers, and he had a powerful physical intuition that could quickly grasp the meaning of an unexpected experimental result.

At the University of Manchester, England, in 1909, Rutherford, Hans Geiger (1882–1945), and Ernest Marsden (1889–1970) carried out a series of experiments that established the existence of the atomic nucleus. They did this by firing alpha particles from radium through a thin layer of gold foil. Gold was chosen because it is dense (thus more likely to interact with the alpha particles) and malleable. The way that particles scatter off a target can give a lot of information about its structure, and scattering experiments, of which Rutherford's was the prototype, would become one of the most important tools of modern physics.

Calculations showed that if Thomson's plum pudding model was correct, alpha particles would be scattered by only a few degrees as they plowed through the gold atoms. Instead, Rutherford was amazed to discover that some alpha particles were scattered away at angles of 90 degrees or greater, with a few even rebounding almost straight backwards. Rutherford later said that this was as surprising as if an artillery shell had bounced off a piece of tissue paper. Rutherford published a pivotal article in 1911 in which he showed that this wide-angle scattering could be explained mathematically if all the positive charge in the atom was concentrated in a tiny volume in or near its center, with the negative electrons orbiting this central core. The force that caused the alpha particles to scatter was simply the electrostatic or Coulomb repulsion between the positively charged alphas and the positive atomic core. This picture implied that most of the atom had to be empty space, a conclusion that is in sharp contradiction to our commonsense beliefs about the solidity of matter. In 1912 Rutherford coined the term *nucleus* for the dense, positively charged nugget of matter at the heart of the atom.

Rutherford's nuclear model of the atom was met with skepticism by Thomson, whose refusal to give up his plum pudding model may seem like obstinacy in the face of Rutherford's scattering experiments. But Rutherford's nuclear model had a huge liability. Maxwell's theory states that an accelerating electrical charge must radiate electromagnetic waves continuously. If negative electrons really do orbit around a positive nucleus, they would quickly radiate away all of their energy and spiral into the nucleus. All of the matter in the universe should have long ago collapsed in a great flash of light. Furthermore, according to classical electromagnetic theory, moving electrons should give off radiation in a continuous range of frequencies, not in the sharp bursts that produce the observed line spectra. If Rutherford's model was correct, then classical electromagnetic theory was in contradiction with both the existence of line spectra and the simple fact that ordinary matter is stable. Thomson's plum pudding model seems implausible now, but it was in fact a sensible attempt to account for the stability of matter using the tools of classical physics. And yet, the nuclear model was the only possible explanation of the wide-angle scattering of Rutherford's alpha particles. As Niels Bohr was soon to show, the missing ingredient was the quantum.

BOHR

The Danish physicist Niels Bohr (1885–1962) is second only to Einstein in his influence on modern physics. Bohr finished his Ph.D. in Copenhagen in 1911, and then traveled to England to work with J. J. Thomson at the Cavendish Laboratory in Cambridge. Thomson did not approve of the fact that Rutherford's radical views about atomic structure had caught the imagination of the young Dane, and in 1912 Bohr moved to Rutherford's lab in Manchester, where he received a more sympathetic reception. (The vigorous Rutherford was also impressed with the fact that Bohr played soccer.)

Figure 3.2: Niels Bohr. AIP Emilio Segre Visual Archives, Margrethe Bohr Collection.

Bohr was soon to solve the problem of the spectra, at least for the hydrogen atom. However, his resolution of the spectral puzzle was an unexpected bonus, because around 1912 to 1913 he was mainly concerned with understanding the puzzle of the mechanical stability of the atom—why the electrons do not spiral helplessly into the nucleus.

Before any progress could be made on this very large problem, it was necessary to get straight on the number of electrons in the hydrogen atom. This is, again, one of those things that seems "obvious" now, but that required a lot of hard work to settle. Building on research by G. C. Darwin (1887–1962) (grand-son of the great biologist Charles Darwin), Bohr in 1913 argued that the hydrogen atom almost certainly contained only one electron. This implied that the nucleus (whatever it might be made of) could have a charge only of +1. The hydrogen atom, therefore, probably had the simplest structure of any of the elements, and so the first test of any new atomic theory had to be whether or not it could explain the properties of hydrogen.

In 1913 Bohr published three brilliant papers in which he applied the quantum theory of Planck and Einstein to Rutherford's planetary atomic model. The gist of Bohr's approach was to take seriously the message delivered by Planck and Einstein, which was that matter can emit and absorb electromagnetic energy only in discrete amounts $E = h\nu$, where ν (the Greek letter *nu*) is a frequency. Since the emission of electromagnetic energy from the atom must correspond to a change of mechanical energy of the electrons in the atom, and since electromagnetic energy is quantized, Bohr thought it only reasonable to suppose that the *mechanical* energies of the electrons were quantized as well. He therefore proposed that electrons can orbit around the nucleus only in certain *stationary states,* each possessing a sharp, well-defined energy. As the electron orbits in a stationary state, its inward electrostatic attraction toward the positive nucleus is balanced by the outward centrifugal force due to its orbital motion, just as the gravitational attraction of planets toward the sun is balanced by the centrifugal force due to their orbital motions. Each possible stationary state can be identified by an integer, which became known as its *principal quantum number.* As George Gamow (1966) put it, Bohr in effect proposed that the atom is like the gearbox of a car that can run in first, second, or third gear and so on, but nothing in between.

The idea of stationary states was the key that allowed Bohr to decipher the spectrum of hydrogen. Before Bohr, physicists had assumed that spectral lines were due to vibration modes of the electrons within the atom. The classical theory of vibrations says that the spectral frequencies ought to be multiples, or *harmonics,* of a basic frequency, like the vibrations of a guitar string. It ought to have been possible to use a powerful mathematical tool called *Fourier analysis,* developed in the nineteenth century by Joseph Fourier (1768–1830), to analyze these harmonics. (Fourier analysis is based on the mathematical fact that complicated waveforms can be represented as sums, or superpositions, of simpler sinusoidal vibrations.) As we have seen, however, the spectral lines of hydrogen and all other elements in fact obey the Combination Principle, which states that every spectral frequency can be expressed as a function of the *difference* between two whole numbers. This is completely incomprehensible in the classical view. Bohr showed that the combination principle finds a natural explanation if we assume that the electrons do not radiate a quantum of energy when they are in a stationary state, but only when they *jump* from one stationary state to another. The energy of a spectral line is then due not to the energy of any one stationary state, but is the *difference in energies* between stationary states (or "waiting places," as Bohr sometimes called them) and is therefore a function of the quantum numbers of both states. When the electron emits a quantum of light it jumps down to an orbit (a stationary state) with a lower energy, and the energy of the quantum emitted will be the difference in energies between the energies of the initial and final stationary states. When the electron absorbs a quantum of radiation (as in stellar absorption lines) it jumps up to an orbit of higher energy. So according to Bohr, Maxwell's theory is just wrong when it predicts that electrons orbiting in the atom will radiate continuously. This brash conclusion was in the spirit of Einstein's youthful suggestion that classical electrodynamics is not an exact theory, but rather something that holds only on average for very large numbers of light quanta.

Bohr's model thus gave an explanation of the stability of the hydrogen atom that was consistent with the principle of the quantization of energy, and that made perfect sense so long as one was willing to admit that classical electrodynamics was not quite right. But Bohr was able to do more: by assuming that the centrifugal force of the electron had to be balanced by the inward electrostatic attraction, and that the angular momentum of the electron in the atom was, like energy, also quantized, Bohr was able to *derive* a formula for the Rydberg constant of spectroscopy. The value he arrived at agreed closely with the observed value, and so he had thus produced a theoretical derivation of the hydrogen Balmer series. Furthermore, he correctly predicted the spectrum of hydrogen in the ultraviolet and infrared. Bohr had thus solved the spectrum of the hydrogen atom, in the sense that he had constructed a model that predicted the precise energies of the spectral lines of the hydrogen atom. This coup immediately established Bohr's reputation as one of the leading physicists of Europe and gave a huge boost to the new quantum mechanics.

Near the end of his life, Einstein described Bohr's atomic theory as "a miracle" (Pais 1982, p. 416) and what he meant by this was that Bohr, by a sort of intellectual sleight of hand, had guessed the right answer to the atomic puzzle using contradictory and incomplete information. From the information available to Bohr in 1913 there was no way to *logically* derive his quantum theory of the hydrogen atom. Rather, it sprang from his intui-

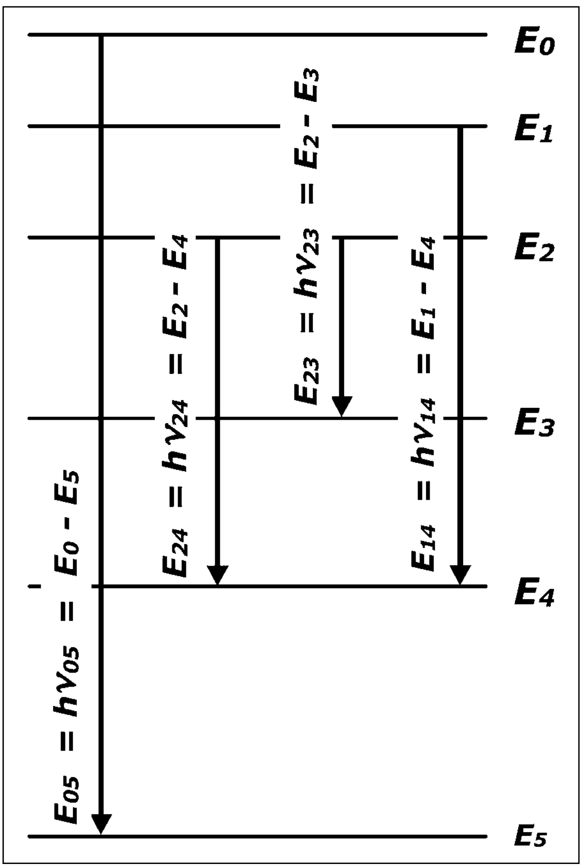

Figure 3.3: Energy Levels in the Bohr Atom. Bohr showed that the energy of an emission line is determined by the difference in energies between electron orbits. Illustration by Kevin deLaplante.

tive ability to know not only what was important (for instance, the combination principle) but what to ignore for the time being (such as the puzzle of what the electrons actually did between emissions of quanta). In fact, many scientific discoveries have been made because a creative thinker was able to guess correctly what discrepancies could be *ignored* in the development of a new theory. It was now impossible to doubt that there had to be something right about the quantum, if only one could figure out just what that could be.

THE NEW THEORY SPREADS ROOTS

Bohr's methods were quickly adopted by several physicists, notably Arnold Sommerfeld (1868–1961) of Munich. Sommerfeld was one of the most versatile mathematical physicists of the early part of the twentieth century, and it is odd that he did not win a Nobel Prize for his many contributions. His ready acceptance of the Bohr theory played a crucial role in its development. Sommerfeld quickly extended the Bohr theory to explain the phenomenon of *fine structure.* If spectral lines were examined closely using new and more accurate instruments that were becoming available, it could be seen that some spectral lines were split, meaning that they were actually combinations of lines very closely spaced in energy. This splitting of spectral lines is called *fine structure.* Bohr's theory of 1913 applied only to circular orbits in the hydrogen atom; Sommerfeld extended Bohr's methods to elliptical orbits. When an object moves on an elliptical orbit it has to move faster when it is closer to the focus. Sommerfeld showed that in many cases the electrons would move fast enough for relativistic effects to be important. The relativistic deviation of spectral lines from Bohr's predictions were a function of a new number, the fine structure constant, approximately equal to 1/137. This mysterious number would prove central to later developments in quantum electrodynamics.

Sommerfeld was also an important teacher (among his doctoral students were the future Nobel winners Werner Heisenberg, Wolfgang Pauli, Hans Bethe, and Peter Debye) and author of textbooks. His book *Atomic Structure and Spectral Lines* appeared in 1919 and went through several editions over the next few years, expanding rapidly each time, and was up until the mid-1920s the "bible" from which most physicists learned their quantum mechanics.

OTHER DEVELOPMENTS AFTER BOHR

Moseley and Atomic Numbers

Another crucial piece of the atomic puzzle was filled in by the young British physicist Henry G. Moseley (1887–1915). He joined Rutherford, who quickly recognized his talent, at Manchester in 1910. Every element has a characteristic X-ray spectrum, just as it has a spectrum in visible light. Moseley showed that the energy of certain X-ray spectral lines varies with respect to the atomic number of the elements according to a simple formula. Before Moseley the meaning of atomic number was not well understood; it was simply an integer that roughly but not exactly corresponded to the atomic weight. Moseley argued (correctly) that the atomic number could be understood simply as the positive charge on the nucleus. (The proton would not be isolated and named by Rutherford until 1918.) Moseley thus confirmed the guess made by Antonius van den Broek (1870–1926) in 1911. (Van den Broek, a lawyer and amateur physicist with a gift for numbers, made numerous helpful suggestions about atomic structure.) Moseley also noticed that assuming his formula held for all elements, there were four missing series when he studied the spectra of known elements. From this he inferred that there had to be four undiscovered chemical elements (in the rare earth group); all were eventually isolated and named. Moseley was killed at the battle of Gallipoli in 1915 at the age of 27.

The Franck-Hertz Experiment

A vivid illustration of the quantization of energy was provided by a pivotal experiment performed by James Franck (1882–1964) and Gustav Hertz (1887–1975, a nephew of Heinrich) in 1914. Their method was simple enough that it is now carried out regularly by undergraduates. They accelerated electrons through a low-pressure gas of mercury. When the electrons' energy reached the energy of a stationary state of mercury they gave up a quantum of energy to the mercury, resulting in a stepwise shape to the curve of current through the apparatus. This demonstrated that atoms could absorb energy only in discrete amounts.

EINSTEIN SHEDS FURTHER LIGHT ON LIGHT, 1916–1917

As if having announced the general theory of relativity in 1916 was not enough, Einstein in 1916 and 1917 published three further radical papers on the quantum theory of radiation. These papers were by far the most general treatments of the statistics of light quanta to that point, and they again demonstrated Einstein's uncanny ability to arrive at powerful conclusions using a few well-chosen assumptions. Like Planck, Einstein set out to describe the conditions that would have to apply for there to be equilibrium between matter and a radiation field. However, by 1916 Einstein had the discoveries of Rutherford and Bohr to work with. Einstein introduced two new concepts: spontaneous and stimulated emission of radiation. In spontaneous emission, an atom that is in an *excited* state (which means that its electrons have been energized to orbits of higher energy) will emit light quanta at random intervals and decay to its *ground* state (lowest energy orbit, having the smallest principal quantum number) according to a probabilistic law that looks exactly like Rutherford's law of radioactive decay. In stimulated emission, a passing photon triggers the emission of a photon from an atom. Miraculously, the emitted photon is a clone of the triggering photon, having exactly the same energy, phase, polarization, and direction of motion as the triggering photon. However, the trigger photon acts like a catalyst in that it is not destroyed or absorbed in the process. Stimulated emission thus allows the amplification of a pulse of electromagnetic radiation, and this is the theoretical basis of the laser ("light amplification by stimulated emission of radiation"). Einstein found that he had to take both types of emission into account in order to correctly describe the balance between absorption and emission when radiation and matter interact. His statistical approach allowed him to derive Planck's radiation law and Bohr's quantum rules all in one stroke.

Einstein's radiation theory of 1916 had another important feature. In 1905 Einstein had treated the energy of light as quantized; in 1916–1917 he showed that light quanta had to have a definite *momentum,* which had to be quantized as well. He imagined a small mirror suspended in a radiation bath and derived a formula for the fluctuations (small random variations) in the momentum of the mirror as a result of fluctuations in the radiation hitting it. He could not get the right answer without assuming that the light quanta hit the mirror with some definite momentum, just as if they were ordinary particles. Since momentum is something that has direction, Einstein's light quanta became known as *needle rays.* This picture gave further support to Einstein's radical suggestion of 1905, that light is not only emitted and absorbed in discrete quanta, but *moves* precisely as if it were made up of particles with definite momenta and energy. Einstein's conclusion would spark the first of several debates that were to occur between him and Bohr about the meaning of quantum mechanics.

THE CORRESPONDENCE PRINCIPLE

Bohr returned to Copenhagen in 1916, and in 1920 became director of his own Institute for Theoretical Physics. Bohr's Institute (funded by a major Danish brewery) was to become a nerve center for the exciting developments in quantum physics that were to burst on the world in the 1920s.

Bohr published the first statement of what he called the Correspondence Principle in 1920, although it had been used implicitly by him and other quantum physicists for several years beforehand. The Correspondence Principle is not a general physical law, but rather a recipe (or *heuristic*) for constructing quantum mechanical models of various systems, one at a time. It is a form of reasoning by analogy, and it demands considerable mathematical skill and good physical judgment to be employed without producing nonsense. Bohr arrived at more than one formulation of the Correspondence Principle, but the gist of it was that in some limit one can assume that a quantum-mechanical system will approximate a classical system with known behavior. The limit could be large quantum numbers (i.e., very high orbits), very small differences in frequency between spectral lines, or found by pretending that the constant of action is near-zero. The practical value of this rough rule is that a physicist who wanted to find the rules for the quantum mechanical behavior of a system could first work out the behavior of a similar system in the classical limit, and then try to guess what quantum mechanical description would *converge to* that classical description in the limit.

The Correspondence Principle was a very useful trick for finding a workable quantum mechanical description of systems that could be approximated classically, and it is still used by physicists today. However, it left two glaring questions unanswered. First, would it ever be possible to find general laws of quantum mechanics that would apply to all possible physical systems, so that a problem could be solved without guessing? Second, could there not be any physical systems that are *entirely* quantum mechanical in the sense that they do not have a large-scale behavior that approximates to a classical picture? It was obvious that physicists did not have the right to assume that all quantum mechanical systems would conveniently have a classical limit that could be used as a guide to their quantum behavior; and yet, in 1920 it was impossible to imagine what a quantum mechanical system with no classical analog could be like, since there were still no general laws of the dynamics of quantum systems. The Correspondence Principle, as useful as it was, could only be a stopgap.

THE GREAT DEBATE BEGINS

Bohr met Einstein in 1920 and they immediately became close friends, but this does not mean that they agreed on all matters to do with physics. An epochal debate began between the two about the meaning of quantum mechanics. In this long dialogue, which was to continue for the rest of their lives, Bohr and Einstein raised several questions that would spark important advances in quantum mechanics—although not always in the way that either Bohr or Einstein themselves had intended!

One of the cornerstones of Einstein's treatment of the light quantum had always been the view that light quanta behave in many respects like ordinary particles with definite positions and momenta. Both Einstein and Bohr were deeply troubled by the contradiction between the concept of light as particulate and the fact that so many optical phenomena, especially interference, only made sense if light traveled as a continuous wave. Einstein (probably more than Bohr) was also very unhappy with the inherently probabilistic way that light quanta want to behave—the fact that the best that physics seemed to be able to do was to describe the average behavior of many quanta over long periods of time, while there was no way to describe or predict the exact behavior of an individual light quantum (if that notion even made sense). But Bohr and Einstein responded to these puzzles in very different ways. Einstein felt that there was no choice but to accept the particulate nature of light even though this had to lead to the consequence that Maxwell's theory was a statistical approximation; Bohr on the other hand surprisingly took a highly conservative line and tried every means he could think of to reject Einstein's view that light quanta are particles, since he did not want to give up the idea that Maxwell's Equations are *exactly* valid at the classical level. In other words, Bohr was willing to give up virtually every vestige of classical behavior at the quantal level (including causation), in return for upholding classical physics at the macroscopic level of ordinary experience; Einstein, by contrast, sought a unified view of nature whereby the same principles, whatever they might be, would hold at both the classical and quantum levels.

THE COMPTON EFFECT: EINSTEIN WINS ROUND ONE

In 1924 Bohr published an enigmatic paper on the quantum mechanics of the electromagnetic field, in collaboration with his younger research associates, the Dutchman Hendrik Kramers (1894–1952) and the American John C. Slater (1900–1976). The Bohr-Kramers-Slater or BKS paper is unusual in that it contains almost no mathematics at all. It somehow managed to be very influential despite its lack of clarity. One of the key tenets of the BKS prototheory was that the conservation of energy is something that holds for electromagnetic interactions only on average. According to BKS, atoms interact without actually interchanging energy or momentum, through a "virtual radiation field" that mysteriously guides their statistical behavior. At the level of individual light quanta the energy books need not balance, so long as everything adds up in the end. Bohr thought it should be this way so that the classical theory of the electromagnetic field would hold exactly, and not merely approximately, at the classical level.

Experiment was not to be kind to the BKS theory. In 1923, the American physicist Arthur H. Compton (1892–1962) carried out relativistic calculations (assuming detailed conservation of energy and momentum on a quantum-by-quantum basis) that predicted that energetic X-ray or gamma-ray quanta, when scattered off of electrons, would lose a certain amount of energy. Experiments by Compton himself and others soon confirmed his prediction. (Kramers had done essentially the same calculation as Compton, but Bohr persuaded him that he was wrong and Kramers discarded his notes—thereby losing his chance at the Nobel Prize that eventually went to Compton.) Bohr resisted for a while, but eventually it was clear to all physicists that the Compton Effect had confirmed Einstein's view that the quantum of light interacted like a discrete particle. Einstein had thus won round one of his long debate with Bohr over the meaning of quantum theory—but more rounds were to follow.

A COMPLETELY CONTRADICTORY PICTURE

By the early 1920s quantum mechanics had evolved far beyond Bohr's simple but powerful atomic model of 1913. Bohr, Sommerfeld, and others had produced a quantum theory of atomic structure that came fairly close to explaining the structure of the periodic table of the elements, and they were able to calculate (by methods that are now considered very inefficient and indirect) the spectra of several atoms and ions possessing relatively few electrons. However, the uncomfortable fact remained that the Bohr-Sommerfeld (as it was by then called) version of quantum mechanics was still little better than a process of inspired guesswork guided by the Correspondence Principle. Physicists of enormous skill painstakingly cooked up quantum mechanical models for each separate atom or ion without any clear idea of the general rules that such models should obey. They would opportunistically use whatever approximations they needed to get a good prediction even if there were contradictions with other models. There was no general method by which the spectra of an atom with any number of electrons could be predicted. Also, there was no general way of predicting the intensities and polarizations of the spectral lines of any atom, and the splitting of spectral lines by magnetic fields was still poorly understood. The question of intensity was especially important: spectral lines are more intense if the transition that produces them is more *probable,* and the problem of expressing quantum laws in terms of probabilities was soon going to take center stage.

At a deeper theoretical level, there was still no notion of how to resolve the great contradiction between the quantization of electromagnetic energy and classical electrodynamics. Utterly lacking was a general theory of quantum mechanics from which one could deduce the quantum behavior of matter and energy from first principles. The surprising outlines of such a theory were about to emerge suddenly through the work of several extraordinarily talented young scientists.

4

UNCERTAIN SYNTHESIS

BOSE COUNTS QUANTA

The heroic years of the 1920s are dominated by three figures: Werner Heisenberg (1901–1976), Erwin Schrödinger (1887–1961), and Paul A. M. Dirac (1902–1984). However, a good place to begin the story of this revolutionary period is the receipt of a letter by Einstein, in 1924, from an unknown young Indian physicist named Satyendra Nath Bose (1894–1974). Bose had written a short paper in which he had presented a remarkable new derivation of Planck's radiation law. His paper (written in English) had been rejected by a leading British physics journal, whose editors thought that Bose had simply made a mistake, and he sought the help of Einstein, whom he revered. Einstein recognized the worth of Bose's paper, and personally translated it into German and had it published. In those days a paper could be accepted for publication simply on the recommendation of a senior scientist; it did not have to run the gauntlet of skeptical referees eager to find any reason at all for rejection. It is quite possible that many of the most revolutionary physics papers of the 1920s might never have seen the light of day if they had to go through the sort of refereeing process that new work now faces.

What troubled Bose was that all derivations of Planck's Law that had been done to that date used, at a crucial point in the calculation, a *classical* expression for the relation between radiation energy density and average oscillator energy. Not only did this seem like cheating—a quantum result should be based on purely quantum principles—but it might even be invalid, because Bose felt that there was no guarantee that classical results (which might only be true on average) were fully correct in the quantal realm.

Bose showed that he could derive Planck's Law using a new trick for counting the number of possible arrangements of light quanta (still not called photons in 1924) with a given energy. (It was also essential to Bose's method that he, like Einstein, treated the light quantum as an object with momentum as well as position.) The essence of his trick, although this was only made clear by Dirac a few years later, was that Bose treated light quanta as if they were objects possessing what is now called *permutation invariance.* This means that any permutation (rearrangement of the order) of light quanta with the same energy would be indistinguishable from each other, and would therefore have to be counted as one. Compare this with how we would calculate the permutations of ordinary objects like pennies. There are, for instance, six permutations of three pennies, since each penny, no matter how perfectly manufactured, has small differences from the others that enable it to be tracked when moved around. But since light quanta are *indistinguishable* there is just *one* permutation of three quanta. Bose had demonstrated that there is something very odd about the way elementary particles such as photons have to be counted. From the point of view of modern quantum mechanics, Bose still did not quite know what he was doing—but like Planck 24 years earlier, he had somehow guessed the right answer by means of a penetrating mathematical insight.

Einstein quickly published three more papers in which he applied Bose's statistical methods to gasses of particles. In 1905 Einstein had treated radiation as if it were a gas. Now he turned the reasoning around in typical Einsteinian fashion and treated molecular gasses like gasses of radiation quanta. Einstein predicted that a gas that obeyed Bose's peculiar counting rule would, below a certain critical temperature, condense into a collection of particles all in a single quantum state. His prediction turned out to be correct. Such states, now called Bose-Einstein condensates (BECs), were first observed when several investigators in the late 1920s and 1930s discovered the phenomenon of superfluidity. (This will be described in more detail later on.) Gaseous BECs were first produced in 1995.

The gaseous BECs so far created in the lab consist only of a few hundred thousand atoms at most, but there is no theoretical limit to the size of a BEC in either liquid or gas form. Bose and Einstein had correctly predicted the existence of a form of matter that is *purely* quantum in its behavior even on the macroscopic scale. It is a form of matter to which the Correspondence Principle cannot apply, and which could therefore be expected to have strongly nonclassical behavior on arbitrarily large scales. This does not mean that one could not go on using the Correspondence Principle as a rough-and-ready guide to the construction of models, but after the work of Bose and Einstein one no longer had the right to assume that the Correspondence Principle will always work.

There was another nonclassical feature of Einstein's quantum gasses that would turn out to be very important, but that no one understood in 1924 (and that is perhaps not fully understood today). In a classical gas that can be treated by Boltzmann's nineteenth-century methods, the molecules are statistically independent like successive tosses of dice; in a quantum gas and many other quantum systems there are correlations between the properties of the particles no matter how far apart they may be, and this fact deeply troubled Einstein.

He was later to ironically describe this mysterious mutual influence between elementary particles as "spooky action at a distance" (Isaacson 2007, p. 450) and he never reconciled himself to its existence.

PAULI'S EXCLUSION PRINCIPLE

One of the most talented of the constellation of young physicists who created quantum mechanics in the 1920s was the Austrian-born physicist Wolfgang Pauli (1900–1958). He contributed several key advances in quantum physics and was rewarded with the Nobel Prize in 1945. However, Pauli was also a sharp and relentless critic of work (including his own) that he thought was not clear enough, and some historians of physics have argued that Pauli, especially in his later years, exerted a detrimental effect on the advancement of physics through his scathing criticism of the not-yet-perfectly formulated ideas of younger physicists.

Perhaps the last triumph of the Old Quantum Theory, and one of its most enduring legacies, was the Exclusion Principle, formulated by Pauli in 1924. In its simplest form, the Pauli Exclusion Principle states that no two electrons in an atom can have exactly the same quantum numbers. This means that the electrons in a given atom have to fill up all the possible orbits in the atom from the lowest energy orbits upward, and in one stroke this gave an explanation of the structure of the Periodic Table of the Elements, and many facts about chemical bonding and structure. It also explained the simple fact that there is such a thing as solid matter, for it is the Exclusion Principle that prevents matter from collapsing into a single quantum state (like a Bose-Einstein condensate) and which may even be responsible for the existence of space itself.

But how many quantum numbers are there? By the early 1920s, Bohr's atomic theory recognized three quantum numbers for electrons in the atom (representing the diameter, orientation, and eccentricity of an electron's orbit); jumps between the different possible quantum numbers determined the energies of the various spectral series. The trouble was that atoms subjected to a magnetic field demonstrated a slight splitting of their spectral lines, called the *anomalous Zeeman effect*, that could not be accounted for by Bohr's theory. Physicists in this period tried many ways of constructing quantum numbers that could account for the enormous variety of line splitting that occurred in multi-electron atoms, and they sarcastically referred to their efforts as "Term Zoology and Zeeman Botany." Pauli, Alfred Landé (1888–1976), and a brilliant 19-year-old assistant of Sommerfeld's named Werner Heisenberg proposed that there is an additional quantum number that could take on precisely two values. Thus, for every set of three possible values of the ordinary quantum numbers there were two extra "slots" that could be occupied by an electron. This showed promise in explaining the splitting of spectral lines in a magnetic field. Oddly, though, if the other quantum numbers came in units of 1, 2, 3, . . . , the new number had to be half-integral: it could only take on values of +/– 1/2.

The apparently obvious interpretation of the new quantum number was that it somehow labeled possible states of angular momentum of the electrons; in other words, its existence seemed to show that electrons were spinning objects. If a charged object rotates about its own axis, it generates a *magnetic moment,* meaning that it acts like a magnet with a definite strength. The most obvious way to account for the magnetic moment of the electron and other particles, therefore, was that they rotate about their own axes. The observed splitting of spectral lines would occur because the intrinsic magnetic moment of the electron would interact with the magnetic moment generated by the electron's orbit about the nucleus. (A current loop generates a magnetic field.) However, Pauli at first insisted on describing this additional quantum number merely as a "classically indescribable two-valuedness" (Jammer 1989, p. 138), and he kept using it mainly because without an extra quantum number the Exclusion Principle could not explain the Periodic Table. Paili thought that it was necessary to avoid trying to find any physical picture or interpretation of the new quantum number, except that it had to be a label that is intrinsic to the electron and not merely a consequence of its motion in the atom. He especially was critical of any suggestion that the electron had an intrinsic angular momentum, because a rotating electron could not be anything like an ordinary spinning object. Given the known mass of the electron, and the fact that its radius, if it has a definite radius at all, is so small that it still cannot be observed today, Pauli calculated that the electron would have to be spinning much faster than the speed of light in order to have the angular momentum it was observed to have. And this, he believed, was absurd because it would violate the theory of relativity. But the notion of electron spin could not be so easily dismissed.

THE DISCOVERY OF ELECTRON SPIN

The fact that objects on the atomic scale have an intrinsic spin had already been demonstrated experimentally in 1922 by Walter Gerlach (1889–1979) and Otto Stern (1888–1969).

Stern and Gerlach shot silver atoms through a nonuniform magnetic field that increased in a certain direction. In such an arrangement the particles will deflect due to the interaction of their own magnetic moments with the external field. Allow the particles to hit a detector screen, such as a photographic plate. If the orientation of the rotation of the particles varies randomly then classical physics says that the deflected particles will make a continuous smear on the detector screen. What Stern and Gerlach and others found, however, is that for atoms and elementary particles such as the electron, there will be a small number of discrete spots on the detection screen; for the electron there will be just two. It seems that the intrinsic angular momentum of atoms and the elementary particles of which they are composed is *quantized,* meaning that it can take on only certain discrete values. The Stern-Gerlach experiment was hailed as the direct observation of the quantization of a physical property *other* than radiant energy. (While the Bohr theory depended on the quantization of orbital parameters, this could only be indirectly inferred through the existence of spectral lines.)

Another strikingly nonclassical feature of the electron's intrinsic angular momentum is that the electrons will split into two directions regardless of the orientation of the applied magnetic field in the detector; this is called *space quantization.*

Despite Pauli's objections a number of physicists, including Compton and Ralph Kronig (1904–1995), had speculated that the electron might have an intrinsic spin, but the idea was first developed quantitatively in 1925 by two Dutch graduate students, George Uhlenbeck (1900–1988) and Samuel Goudsmit (1902–1978). Their theory of spin gave an excellent account of the magnetic behavior of electrons in the atom, except that its estimate of line splitting was out by a factor of two. In 1926 Llewellyn H. Thomas (1903–1992) explained the discrepancy in terms of an effect now called Thomas precession, which is due to the relativistic time dilation of the electrons. Spin causes line splitting because the magnetic moment of the electron either adds to or subtracts from its orbital magnetic moment, depending on the value of the spin. Pauli finally accepted that the concept of spin was valid. (Although Pauli was often a sharp critic of the work of others, he was also not afraid to admit he had been wrong.) The profound difficulty he had raised, that if the electron really was a spinning object then it was like no classical object at all, was not resolved but merely set aside. Once again physicists ignored a glaring discrepancy when doing so allowed them to make progress on another front. Eventually Pauli himself would create a formal theory of spin that would be consistent with fundamental quantum mechanics. But a lot had to happen before that would be possible.

MATRIX MECHANICS

Modern quantum mechanics burst rapidly on the scene in the years 1925 to 1927 through at least three major lines of investigation: matrix mechanics, wave mechanics, and the "transformation" theory of Paul Dirac. At first these approaches seemed very different, but they were eventually shown to be different ways of saying the same thing, or almost the same thing. These three approaches will be described separately to begin with, but this is a risky oversimplification since these developments occurred in parallel over a few years and strongly influenced each other in crucial ways.

Heisenberg's Sunrise

Werner Heisenberg (1901–1976) was born in Würzburg, Germany, and studied physics under Sommerfeld at the University of Munich in the disordered years immediately after Germany's defeat in World War I. One of his strongest intellectual influences was his youthful reading of the Athenian philosopher Plato's *Timaeus.* This long dialogue, written before 350 B.C., is a rambling speculation on the origin and nature of the cosmos. Heisenberg was fascinated by the crude but imaginative atomic theory sketched by Plato, who suggested that the structure and properties of matter could be understood in terms of the five regular or "Platonic" solids. Heisenberg recognized that there was no scientific basis for Plato's speculations; however, he was inspired by Plato's vision that it should be possible to understand the physical world in terms of mathematical symmetries.

Heisenberg was gifted with exceptional intellectual quickness, and under Sommerfeld's steadying guidance he was publishing useful contributions to the Old Quantum Theory before the age of 21. In 1922 he met Bohr at a conference, the "Bohr Festival," held in Göttingen, Germany. A long conversation convinced Bohr that Heisenberg had unusual talent, and he invited Heisenberg to come to Copenhagen. First, however, Heisenberg went to Göttingen and worked with Max Born (1882–1970). Born had studied under the great Göttingen mathematician David Hilbert (1862–1943) and for Born it was very important to be clear and rigorous about the mathematics used in physics. Born was to display intellectual leadership that was crucial to the development of quantum mechanics in the following years, and he got Heisenberg involved in exploring the idea that the continuum methods of classical physics should be replaced by a mathematics written in terms of finite differences rather than differential quantities. This was inspired by the fact that energy and angular momentum had already been shown to be quantized. By now it was clear to Born (and Heisenberg) that an entirely new approach was needed in order to have any hope of understanding the quantum. Born was already calling it "quantum mechanics," though no one yet knew what that could mean.

By 1925 Heisenberg had spent time working in Copenhagen and had published papers with Kramers, stretching the Correspondence Principle as far as it could go. Much of their work involved the attempt to model atomic systems with virtual oscillators, just as Planck had done in 1900, although with much more powerful mathematical tools. In June Heisenberg was back in Göttingen trying to find the right quantum description of certain kinds of oscillators, but he was nearly incapacitated by hay fever. He took a vacation on the treeless island of Heligoland overlooking the North Sea, where he hoped that the sea air would give him some relief from pollen. He kept working, and began to see a new approach.

The key was that he decided he should work only with observable quantities. The Bohr theory had depended on classically describable electron orbits, but these orbits are unobservable, like the unobservable ether of the old prerelativistic electrodynamics that Einstein had shown was irrelevant. Heisenberg's idea, which had occurred to no one else, was to assume that classical equations of motion such as Newton's Law were correct but to reinterpret the quantities in terms of which they were written (position, momentum) as a special kind of sum called a *Fourier series,* with the terms of the series expressed as functions of the observable intensity and polarization of the light emitted by the atom.

Fourier analysis had been invented by Joseph Fourier (1768–1830) in the early nineteenth century, and it had become one of the most useful tools of theoretical physics by the time Heisenberg studied it with Born in Göttingen. (Fourier, incidentally, was the first scientist to predict global warming as a result of carbon dioxide emissions.) Fourier showed that virtually all mathematical functions that could be useful in physics or engineering can be represented as a sum, or *superposition,* of sine and cosine waves with the right amplitudes and phase factors.

Figure 4.1: Werner Heisenberg. Photograph by Freidrich Hund, AIP Emilio Segre Visual Archives.

Heisenberg set out to write position and momentum as Fourier sums of complex-valued quantities that Born came to call *transition amplitudes.* Each such amplitude is a function of two integers, the quantum numbers of the states that the virtual oscillator jumps or transitions between. The result was that position and momentum were represented as square arrays of complex numbers. Heisenberg had to work out the algebra of these arrays, so that he would know how to write the dynamical equations in terms of them. To his surprise he found that he could only get the right equations if the arrays had the very odd property that they failed to *commute*—that is, the product xy of the arrays for x and y would not equal yx.

Using his strange arrays of complex numbers, Heisenberg derived the existence of the zero point energy (a minimum energy that all systems have even in their ground states) for which no clear explanation had been given before, and showed the energy levels have to be quantized. He calculated through the night, and finally (using the assumption of conservation of energy) derived energy levels that agreed with experiment. He climbed a promontory on the island, watched the sunrise, and "was happy" (van der Waerden 1967, p. 25) because he knew he had found the key to Born's elusive "quantum mechanics."

The Three-Man Work

Heisenberg's paper "On the Quantum-Theoretical Reinterpretation of Kinematic and Mechanical Relations" was soon in circulation, and Born had (with some difficulty) realized that Heisenberg's odd arrays of numbers were nothing other than *matrices.* It may seem odd that Born and Heisenberg, who were highly trained in mathematical physics, did not immediately realize this fact, since matrix algebra is now often taught in high school. Matrix theory had been developed in the nineteenth century by the British mathematician Arthur Cayley (1821–1895), but in the early 1920s it was still merely an abstruse topic in pure mathematics that was not normally learned by physicists. Born was only able to recognize what Heisenberg had done because of a brief encounter with matrices in his student days. However, he was not skilled in matrix manipulation. Fortunately his assistant, 22-year old Pascual Jordan (1902–1980), was expert in matrix algebra and soon became an important contributor to quantum mechanics in his own right. Born and Jordan quickly published a paper together, extending Heisenberg's methods, and then all three collaborated on a monumental paper, published early in 1926. In this paper, often called the "three-man work," they further developed matrix theory and applied it to several key problems. This paper definitively set forth matrix mechanics, which is the version of quantum mechanics based on the algebraic manipulation of matrices that represent observable quantities such as position, momentum, and energy. Detailed calculations showed that the new matrix mechanics was very successful in predicting the anomalous Zeeman Effect, other forms of line splitting, and line intensities. The three authors even produced a new derivation of Planck's Law, taking advantage of Bose's counting rules.

Shortly after the publication of the three-man paper, Pauli used the new mechanics to rederive the entire Bohr theory of the hydrogen atom. In 1927 he showed how to construct a *spin operator,* which describes the spin of the electron in three-dimensional space. This operator is built up of four simple two-by-two matrices now called the *Pauli spin matrices,* and these have very wide application throughout quantum mechanics and, most recently, in quantum computation.

There was no question that the new matrix mechanics was enormously effective. It had swept aside almost all the difficulties that had plagued the old Bohr-Sommerfeld approach. But its founders were aware that they had almost entirely lost contact with the physical picture of what might be going on inside the atom. Quantum mechanics was starting to look as if it was nothing more than a highly effective (although complicated) mathematical formalism for calculating observable results such as probabilities and energies, with little or no way of telling what that formalism actually meant—or, if a picture of what underlies the quantum rules could ever be uncovered, it would be unlike anything that the classical mind had ever dreamed of.

WAVE MECHANICS

Louis de Broglie: If Waves Are Particles Then Particles Are Waves

Prince Louis Victor de Broglie (pronounced roughly "de Broi") (1892–1987) was a scion of an old aristocratic French family. As a young man he explored several fields; his interest in physics was sparked by reading a report on the First Solvay Conference of 1911. He luckily escaped the deadly trenches of World War I and instead served his country as an electronics technician, which must have stimulated his thinking about electromagnetism. His older brother Maurice de Broglie (1875–1960) was an accomplished experimental physicist. Under his brother's guidance, Louis became familiar with X-rays and the photoelectric effect.

Louis's greatest talent was a gift for spotting the obvious—or rather, for spotting what would eventually become obvious to everyone else. He thought deeply about the mysterious wave-particle duality that had been pointed out by Einstein as early as 1909, and arrived at a question that must have seemed almost childish at the time: if electromagnetic waves are also particles, then might not particles of seemingly solid matter (such as electrons) be somehow also waves? If this were so, then beams of electrons would have a wavelength and a frequency just like light, and this would offer a ready explanation of Bohr's quantization conditions: an electron in its orbit about the nucleus would be like a standing wave, such as the wave in a plucked guitar string, and only an integral number of wavelengths could fit into an orbit. This very simple insight could have been seen (but was not) by anyone from about 1913 onwards.

Using facts about momentum and the behavior of waves from Einstein's special relativity, de Broglie derived simple but elegant formulas for the wave-like properties of particles, and also found yet another a new derivation of Planck's blackbody law. He showed that any object with momentum p has a *de Broglie wavelength* λ, equal to h/p, where h is Planck's constant. The quantum mechanical properties of matter become important when the de Broglie wavelength of an object is comparable to or larger than its size, while the wavelength for a classical object such as a car is utterly negligible.

De Broglie published his ideas in three short papers in 1923 and then collected them together in his doctoral thesis of 1924. His work was praised warmly by Einstein, who said that de Broglie had "lifted a corner of the great veil" (Isaacson 2007, p. 327). (As with Bose, Einstein also helped to get de Broglie's highly unconventional work published.) De Broglie's insight brought the quantum story full circle, but in a way that only compounded the mystery of the wave-particle duality: all forms of matter and energy (be it light or electrons) are both particle and wave. But how could matter possibly be both particle and wave? And what were the laws that governed the structure and behavior of matter waves?

If de Broglie's picture of electrons as waves was correct, then electrons that were fired through small openings comparable to their wavelengths should exhibit wavelike diffraction and interference phenomena. These predictions were not directly confirmed until 1927, when the Americans C. Davisson (1881–1958) and L. Germer (1896–1971) demonstrated diffraction effects when they scattered electrons off a polished nickel crystal. Similar results were obtained by the British experimenter G. P. Thomson (1892–1975), the son of J. J. Thomson. Louis de Broglie was awarded the Nobel Prize in Physics in 1929, by which time the complete duality of wave and particle was accepted as a cornerstone of quantum theory.

Schrödinger: Music of the Orbitals

Erwin Schrödinger (1887–1961) was born in Austria and studied at the University of Vienna. By the time Heisenberg took his vacation on Heligoland, Schrödinger was professor of physics in Zurich, Switzerland. Schrödinger was a thinker of wide-ranging interests, and in his later years he made important contributions to biology.

Schrödinger found de Broglie's insights about matter waves to be intuitively appealing. However, the problem was that de Broglie had only told half of the story, in that he had only described the *kinematics* of quantum waves. This means that he had shown in general terms how matter waves can be described in spacetime but had not given their *dynamics.* In other words, de Broglie had not said anything about what produced his matter waves. Schrödinger knew that if there are waves then there has to be a governing wave equation, which typically takes the form of a partial differential equation whose solutions are possible wave forms.

Schrödinger was highly skilled in the mathematics of classical mechanics, and he set to work to find the right equation. He repaired to a cottage in Arosa, Switzerland, in the company of an unknown young woman. The mysterious lady of Arosa seems to have stimulated his creativity, and, like Heisenberg, he made one of those quantum leaps of the theoretical imagination that are so hard to analyze or explain.

Figure 4.2: Erwin Schrödinger. Photograph by Francis Simon, courtesy AIP Emilio Segre Visual Archives.

Schrödinger took an important cue from the work of the nineteenth-century mathematician William Rowan Hamilton (1805–1865), one of the greatest of Ireland's sons. Hamilton had shown that Newton's Laws of mechanics could be rewritten in a form that made them look remarkably like the laws of optics. Schrödinger constructed a wave equation that he essentially guessed by analogy with Hamilton's nineteenth-century version of Newtonian mechanics. Schrödinger's Equation can be written in many forms, depending on the structure of the problem it is applied to. The functions that satisfy Schrödinger's wave equation are called *wave functions* or sometimes ψ-functions, since they are usually written using the Greek letter ψ (psi). The most general time-dependent Schrödinger Equation says, roughly, that the rate of change of the wave function with respect to time is proportional to the result of the *Hamiltonian* operating on the wave function. The Hamiltonian (borrowed and generalized from Hamilton) is an operator that represents the energy of the system. An operator is a mathematical machine that transforms functions into functions. Applying the Schrödinger Equation to a particular problem is largely a question of knowing what Hamiltonian to apply, and this is often a matter of inspired guesswork. Thus it is wrong to say that modern quantum mechanics eliminates all of the guesswork that was endemic to the Old Quantum theory; however, it concentrates the need for guesswork to a much smaller area.

In January 1926 Schrödinger published the first of a series of papers with the title "Quantum Mechanics as an Eigenvalue Problem." To find the eigenvalues of a vibrating system is to find its characteristic modes of vibration; like a guitar string, any vibrating system will have a certain basic and harmonic frequencies. Bohr was supposed to have shown that no such model could apply to the atom; however, Schrödinger suggested that the energies of the emission lines corresponded to *beat frequencies* between the characteristic vibration modes of the electrons in the atom. The attraction of this conception to Schrödinger was that he thought it would get rid of the need for what he later called "this damned quantum jumping" (Heisenberg 1971, p. 79), which offended his classically trained sensibilities. Schrödinger thought he had shown that a quantum "jump" would simply be a continuous (although no doubt very rapid) transition from one vibration mode to the next. While eigenvalues played an important role in the matrix mechanics of Heisenberg, Born, and Jordan, Schrödinger was the first to give them a possible physical meaning.

Schrödinger showed that with his methods several problems in quantum mechanics could be solved, including the spectrum of the hydrogen atom. The way it is done is to write the Hamiltonian for the electron in the atom; this is simply a quantum-mechanical version of the classical expression for the total energy of the electron, expressed as a sum of the kinetic energy of the electron and the potential energy it possesses due to its electrostatic interaction with the nucleus. The resulting partial differential equation can be solved by a class of eigenfunctions called *spherical harmonics,* which are the normal modes, or natural modes of vibration, of an elastic sphere. The various possible values of the spherical harmonics are the *eigenfunctions* or *eigenmodes* of the Hamiltonian, and they give the familiar orbitals, such as the *s* and *p* orbitals, of chemistry. For a while Schrödinger believed that the waveforms he had described gave a classical and continuous distribution of electrical charge around the nucleus; this hopeful interpretation would not last long.

He reproduced all of the results that had been obtained so laboriously in the Bohr theory, and so obscurely in matrix mechanics. In principle his equation can be used to calculate the orbital structures of any atom or molecule at

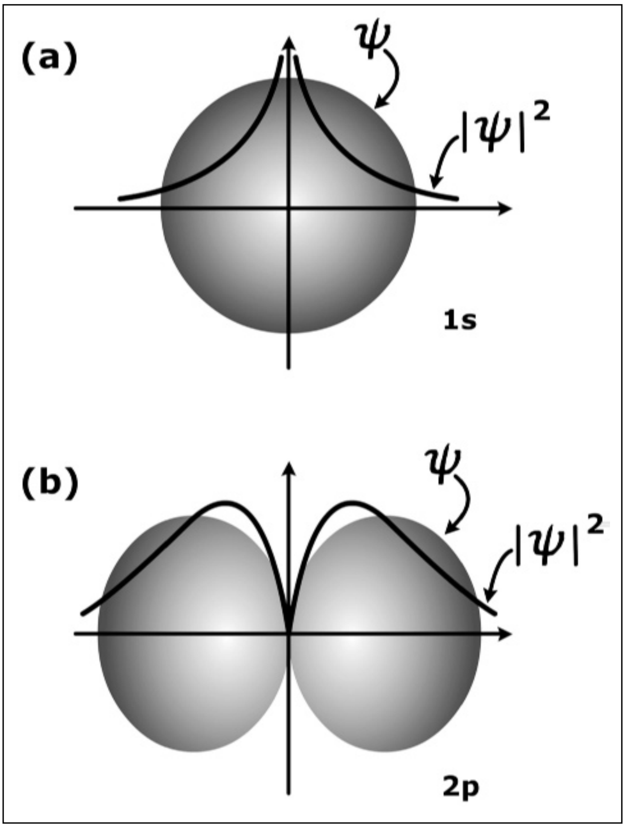

Figure 4.3: Typical Electron Orbitals. The wave function Ψ is a solution of Schrödinger's Equation and the heavy curve shows the probability function $|\Psi|^2$ for finding an electron. Illustration by Kevin deLaplante.

all; in practice this is limited by computational complexity. Large organic molecules such as proteins have very complicated orbital structures, and so solving Schrödinger's Equation to find their wave functions has to be done with computers (not available to Schrödinger in 1926!) and various kinds of approximations have to be made that require skill and judgment. However, Schrödinger's wave mechanics was *useful* in a way that matrix mechanics was not, and it immediately spawned a host of applications in chemistry and many areas of physics.

Schrödinger hoped that the success of his continuum methods would restore the classical picture of physics. One problem with this interpretation of the wave function is that wave functions tend to spread out in space, sometimes very rapidly, while electrons themselves are always found in discrete locations. Another clue that the wave function would not easily admit of a classical interpretation was that the wave function is given by a *complex*-valued function. Complex numbers are numbers of the form $a + ib$, where a and b are ordinary real numbers, and i is the square root of -1. While such functions have a definite mathematical meaning and very important applications throughout mathematics, they are hard to visualize—and in particular they cannot in fact represent the density of anything physical at all. But it would be a few months longer before this would become painfully clear even to Schrödinger.

EQUIVALENCE—NEARLY!

The pioneers of matrix mechanics were at first horrified by Schrödinger's wave mechanics, while it delighted those (including Einstein and Planck) who preferred a return to something that at least resembled the old classical certainties.

On the face of it, no two approaches to the same sets of physical problems could seem more different than matrix and wave mechanics. The former was expressed in abstract algebra and made little attempt at picturing atomic phenomena, and it treated matter and energy as essentially discontinuous. It claimed to be able to do little more than calculate probabilities of transitions. Wave mechanics, on the other hand, *almost* seemed like a visualizable continuum theory, similar to those that might be used to represent vibrating objects, waves, or fluids—the sort of theory that classically trained physicists such as Einstein and Planck were used to and comfortable with. A great apparent advantage of the Schrödinger Equation to the older guard was that it was entirely deterministic; a given wave function would evolve in a unique and perfectly definite way, and there was—or so it was hoped by many for a few months at least—no more need for talk of random jumping, or particles shooting off in any direction at all with no good reason for doing so.

A clue that all was not all that it seemed was provided by Schrödinger himself, for later in 1926 he published a paper showing that matrix mechanics and wave mechanics were mathematically equivalent; they were really just different ways of saying the same thing, or nearly the same thing. The choice of which to use would be largely a question of practicality or taste.

Dirac: Elegant Brackets

Paul Adrien Maurice Dirac (1902–1984) was born in Bristol, England, of Swiss and English parentage. His first degree was in electrical engineering, because his father wanted him to study something practical. However, he switched to mathematical physics after becoming entranced by the elegance of the Minkowski metric (the formula for the interval in Einsteinian space time), which he heard about in a lecture on the theory of relativity given by the philosopher Charles D. Broad (1887–1971). He did much of his early work under the supervision and encouragement of the British physicist and astronomer Ralph Fowler (1889–1944). Probably more than any of the founders of quantum mechanics, Dirac was a creative mathematician of very great ability. While Bohr and Einstein were very competent applied mathematicians guided by physical intuition, Dirac made his great discoveries largely out of an exquisite feeling for mathematical simplicity. What matters, he said (Cropper 1970), was to get beauty into your equations—not always easy even for Dirac.

Fowler gave Dirac a copy of Heisenberg's paper. Dirac realized that the essential feature of matrix mechanics was noncommutativity, and Dirac certainly knew his matrix algebra. Relying on similarities with a structure from classical mechanics called the *Poisson bracket,* Dirac showed that the quantum mechanical behavior of quantities such as position or momentum can be defined by their *commutator.* The commutator of p (momentum) and q (position) is just $pq - qp$. The commutator is always a multiple of Planck's constant, and this gives another answer to the question of what quantum mechanics is: since the commutator is zero if the constant of action is zero, quantum mechanics is simply physics where the size of Planck's constant matters—because that is when certain quantities (which are said to be *conjugate* to each other) will fail to commute.

Dirac generalized Heisenberg's matrices to linear operators that transform the state. A linear operator is a mathematical structure that maps vectors into vectors. Linear operators can be represented by matrices, but there are many possible matrix representations of a given operator, depending on what types of observations we make on the system; Dirac called these different coordinate systems "representatives." By mid-1926, Dirac had independently created a generalized and mathematically clearer version of matrix mechanics.

Dirac also introduced a very elegant and simple system of notation for quantum mechanics. The state of a physical system is represented by a vector called a "ket," and every ket has a sort of mirror image called a "bra." Put the two together, and one has a "bra-ket" that represents the transition amplitude from the ket state to the bra state. Dirac's bras and kets can be very easily manipulated and they greatly simplify calculations in quantum mechanics.

In later papers published in 1926 Dirac further developed his algebra of commutators, and made a distinction between what he called c-numbers and q-numbers. The former are classical quantities (such as position or momentum) given by ordinary real numbers, while q-numbers are quantum mechanical linear operators that can be represented by matrices. He also showed that the quantum mechanics of continuous quantities such as position and momentum required the use of a mathematical device commonly called the Dirac delta-function (though it had in fact been introduced by Gustav Kirchhoff in 1882). This is an idealized "pulse" function whose value is zero everywhere except at one point, and whose integral over all of space is one. Up to this point matrix mechanics had been able to deal only with discrete quantities such as energy levels, but physicists still preferred to assume that some quantities such as position and linear momentum can be treated as continuous. (Whether space and time really are continuous, or whether this is just a convenient approximation, is still a current topic of investigation.) By studying the quantum mechanics of continuous quantities, Dirac arrived at Schrödinger's Equation and thus showed that Schrödinger's very useful equation, as he had written it in terms of continuous wave functions, is a special case that arises out of the formalism of the more abstract theory of linear transformations when continuous representatives are applicable. Dirac's theory thus formed a bridge between matrix and wave mechanics, and the formalism of quantum mechanics is now usually given in terms of his notation and terminology.

BORN'S MOMENTOUS FOOTNOTE: THE PROBABILITY INTERPRETATION OF THE WAVE FUNCTION

A decisive breakthrough in understanding the wave function came in 1926, when Max Born, in a paper on the scattering of electrons from atoms, observed that the most obvious interpretation of the wave function is that it represents the *probability* of finding the electron at a given location. More precisely, he added almost off-handedly in a footnote, its *square* represents probability, and this observation is now called the *Born Rule.* (One might say jokingly that Born won a Nobel Prize for a footnote.) Wave functions are also often referred to loosely as *probability waves,* but this is a misnomer. The wave function is complex-valued, and so cannot stand for a probability by itself. However, if a complex number is squared up it gives a real number (called its modulus). The wave function can be *normalized,* which means that it is multiplied by a constant that keeps the modulus between 0 and 1; this allows the modulus to be interpreted as a probability. The wave function itself is referred to more accurately as a probability *amplitude,* which is, roughly speaking, a complex *square root* of a probability.

A study of the wave mechanics of scattering seems to have led Born to his probability interpretation. He showed that the wavefronts of scattered particles would have the approximate form of expanding spheres—and yet the particles would always be detected as discrete objects traveling in certain definite directions. One never directly detects a wave; rather, the wave gives the probability of finding a particle. Suppose a headline in the newspaper says, "Crime Wave Sweeps City"; this is just a way of describing an increased frequency of discrete criminal acts. There is no wave separate from the acts themselves.

With Born's probability interpretation, it no longer seemed possible to uphold Schrödinger's realistic interpretation of the wave function, although Schrödinger himself resisted mightily for a while. Indeterminism could not be gotten rid of as easily as Schrödinger and Einstein had hoped.

HEISENBERG'S MICROSCOPE AND THE UNCERTAINTY PRINCIPLE

In 1927 Heisenberg traveled to Copenhagen and endured intense discussions with Bohr on the meaning of quantum mechanics. One of the points that especially troubled them was that they could not see how to reconcile the existence of *apparently* continuous electron trajectories with the fundamental laws of quantum mechanics. There are several experimental contexts where it *seems* as if free electrons (that is, electrons moving outside the atom) have trajectories just like bullets or baseballs, and yet quantum theory treats all detection events as discrete. Heisenberg realized that it was necessary to examine the conditions under which an electron can be observed.

In order to find where an electron is, we have to bounce some particles off it, and as de Broglie had shown, all particles have a wavelength that is shorter the higher the energy of the particle. Suppose we use photons. It is a basic law of optics that the resolving power of a lens is determined by the wavelength of the light shone through it. We could pin down the electron quite narrowly if we used high-energy photons, such as gamma rays, but these would disturb the electron's motion and thereby change the very property we are trying to observe. If we try to use lower-energy quanta (such as ordinary light) in the hope of disrupting the particle's motion less, the quanta would have a much lower resolving power, and we would have a correspondingly larger uncertainty in the position of the particle.

Heisenberg showed that if we know exactly where the electron is, it could have any velocity at all, even greater than that of light. On the other hand, if

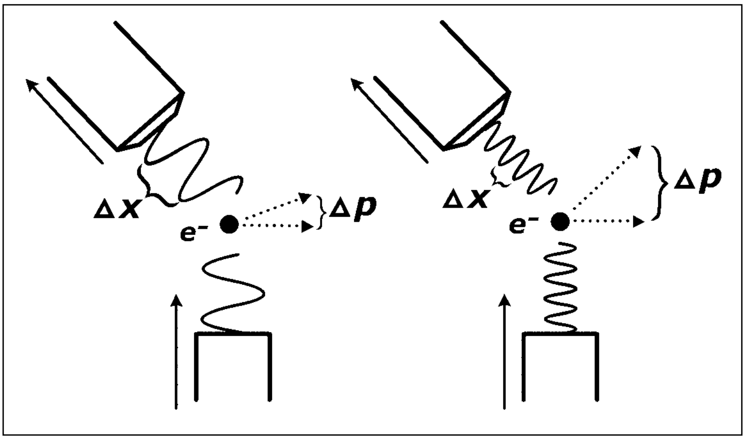

Figure 4.4: Heisenberg's Microscope. The shorter the wavelength of the gamma rays used to examine the electron, the sharper the determination of position but the larger the uncertainty in momentum. Illustration by Kevin deLaplante.

we know exactly how fast the electron is going, it could be anywhere at all. Heisenberg even went so far as to suggest that that the experimenter *creates* orbitals by attempting to observe them. In any realistic case there is always an uncertainty in both the position and the momentum, and the reciprocal relationship between these uncertainties is called Heisenberg's *Uncertainty* or *Indeterminacy Principle*. Uncertainties are usually symbolized using the Greek letter Δ (capital delta); Heisenberg's Indeterminacy Principle can then be expressed as Δ(position) times Δ(momentum) is greater than Planck's constant of action.

An interesting twist of history is that Heisenberg had almost failed his doctoral oral examination in 1922 because he did not know enough about the resolving power of lenses. Abashed, he had done some homework on the subject and later observed that this helped him in his work on uncertainty.

Bohr argued that Heisenberg's microscope is to an extent misleading. It is very easy to carelessly imagine that the electrons really do have definite positions and momenta at all times, and that the uncertainty relations are merely due to the fact that because of the finite size of the quantum of action, a certain minimum jiggling of the apparatus, and thereby a minimum experimental error, is inevitable in any experiment. The fact that this "obvious" interpretation of the uncertainty relations is wrong is one of the hardest things to grasp about quantum mechanics. However, by 1930 several authors, including Schrödinger, had shown formally that the uncertainty relations follow from the commutation relations that had been written explicitly by Dirac: any two observables that do not commute (for example, spin components in perpendicular directions) have to obey an uncertainty relation. The formalism has no way of even expressing the concept of a particle that simultaneously has sharp values of both position and momentum.

What this means, as Heisenberg and other physicists began to realize by the late 1920s, is that we do not observe continuous electron trajectories at all. What we really observe is a sequence of *discrete* snapshots of an electron. If these snapshots are very close together then we are naturally inclined to suppose that the electron is following a continuous path. But in fact we have *no warrant,* either experimentally or theoretically, to conclude that the electron exists in between observations of it.

Classical mechanics was supposed to be about particles—tiny, continuously existent chunks of matter—moving around under the influences of forces given by definite, deterministic laws. Instead, quantum theory tells us that the classical picture is only a larger-scale approximation that emerges under certain circumstances, like the apparently continuous image that is revealed if we pull back far enough from a digital image made up of thousands of pixels. Heisenberg declared that the classical picture is the picture we get when the quantum pixels, so to speak, blur together, and Schrödinger's Equation is the rule that tells us the *probability* that a pixel will appear in a particular location. In the seesaw battle between continuity and discontinuity, discontinuity had taken the lead again.

Heisenberg's Uncertainty relations can be taken as a modern quantitative version of the ancient Paradox of the Arrow stated by the Greek philosopher Zeno of Elea (ca. 450 B.C.). Zeno sought to demonstrate the unreality of motion, but his argument could easily be adapted to show the unreality of rest. Consider an arrow in flight, said Zeno; if at any moment it truly occupies a definite position in space, we cannot say that it is moving, since if it moves it is changing its position. On the other hand, Heisenberg might have added, if the arrow truly is moving, it is constantly changing its position, so that at no moment is it precisely anywhere in particular. Heisenberg's Uncertainty Relations for position and momentum express in mathematical form the paradox inherent in the very concept of motion.

The idea of an uncertainty inherent in all natural processes is also found in Plato's *Timaeus,* the work that inspired the young Heisenberg. Plato speaks of an inscrutable factor that he calls the *Errant Cause*—in effect, Plato's own Uncertainty Principle—which is a sort of zero-point energy, a tendency of everything in the natural world to be in a perpetual state of restless motion. Heisenberg does not mention Plato's Errant Cause in his memoirs (1971), but could he have been influenced by Plato's idea that everything in the physical world has an uncertainty in its very nature? Heisenberg speaks repeatedly of his search for the inner order of nature, and yet he more than any other scientist revealed that nature is founded on the tension between order and a disorder that can never be made to go away.

5

DUALITIES

THE COPENHAGEN "ORTHODOXY"

Out of the intense debates about the meaning and application of quantum mechanics from 1927 to 1935, one view quickly became dominant: the so-called *Copenhagen Interpretation* of Niels Bohr. Some have argued that the victory of the Copenhagen Interpretation was due as much to Bohr's powers of persuasion (if not intimidation) as its intellectual virtues. There are stories that Bohr would browbeat his students and colleagues in debate, sometimes reducing them to near-collapse. (It is said that Schrödinger, following a long debate with Bohr, took to his bed and regretted that he had ever had anything to do with the quantum; Kramers was hospitalized with exhaustion, and even the redoubtable Heisenberg was once reportedly reduced to tears.) However, the students and coworkers of Bohr unfailingly spoke of him in terms of greatest affection and respect and insisted that his zeal in debate arose entirely from an intense desire to find the truth. The fact remains that Bohr and his followers hammered out (sometimes painfully) a way of doing quantum mechanics that was a delicate compromise between the revolutionary and the conservative, and that, like many compromises, worked well enough to allow physicists to get on with the job of applying quantum mechanics to a host of new problems. It remains to be seen whether the Copenhagen Interpretation of quantum mechanics will stand for all time.

Complementarity

The cornerstone of the Copenhagen Interpretation is Bohr's Principle of Complementarity, which he first announced following heated discussions in 1927 with Heisenberg. Bohr thought that Heisenberg's discovery of the Uncertainty Relations was a great advance, but he also believed that Heisenberg, in his attempt to interpret his discovery, had given too much primacy to the particle picture. The Principle of Complementarity states that it is not enough to point out that pairs of canonically conjugate observables fail to be simultaneously measurable. Instead, for Bohr, the breakdown of commutativity was merely an aspect of a larger fact, which was that any part of physics where quantum effects are important requires two mutually contradictory but *complementary* modes of observation and description. Both modes are necessary in order to make all of the predictions that can be made about physical reality, and yet each mode excludes the other; that is, they cannot both be applied simultaneously. The types of experiments in which (say) position can be measured exclude the types of experiments in which momentum can be measured. In other words, for Bohr it was not enough to say that we cannot measure position precisely; rather, the issue was that there are limitations on what we can *mean* by "position" and "momentum." For Bohr, *causal* accounts of phenomena (that is, accounts in terms of dynamical quantities such as forces, momenta, and energy) are complementary to *space-time* accounts of phenomena (that is, accounts in terms of positions, times, and velocities).

An important illustration of complementarity is the wave-particle duality itself: sometimes we have to treat matter and energy as if it is composed of waves, and sometimes we have to treat matter and energy as if it is composed of particles, but it does not make sense to do both at once. In fact, it is physically impossible to observe wave properties (such as interference) with precisely the same measurements in which particle properties (such as momentum) can be observed.

One wants to ask, "But is an electron really a wave or a particle?" Bohr insisted that this question is not meaningful. He thought that we can only ask what something is when we can specify an experimental context, and the experimental contexts that allow us to observe wave-like properties exclude those that allow us to measure particle-like properties. At the same time, to do all of the physics with electrons that is *possible* requires both wave-like and particle-like measurement procedures. "But surely," the response might be, "an electron must really exist even if we can't describe it without experimentally defined terms. We don't just make it up!" Bohr would have insisted that the concept of the independent existence of the electron is not meaningful. He would have agreed that we don't just make electrons up; rather, he would have said, electrons as they are observed in various sorts of experiments are manifestations of irreducibly quantum mechanical interactions between observer and observed, which obey the probabilistic laws of quantum mechanics. And at that point it would be understandable if the questioner, like Schrödinger, took to bed in exhaustion.

The Quantum-Classical Divide

The Copenhagen Interpretation also contains an important rule about the nature of measurement: the cash value of any quantum calculation must be a prediction that can be understood in terms of unambiguous *classical* observations. No one could quarrel with this statement if all it meant was that a measurement procedure cannot make sense to humans if it cannot be expressed in procedures that humans can grasp. However, Bohr intended to make a statement about physics itself that would be true for any beings anywhere in the universe doing quantum physics, since by "classical" he apparently meant nothing other than the physics of Newton, Maxwell, and Einstein.

It is possible that Bohr had too narrow a notion of what would constitute a "classical" observation procedure. How do we know that there might not be new types of measurements or observations that are inherently quantum mechanical, but that, like Bose-Einstein condensation, can be grasped on the human scale? In other words, how do we know that our conception of what is *classical* cannot evolve in surprising ways? The fact that some quantum mechanical effects (such as Bose-Einstein condensation) can manifest themselves on the macroscopic scale suggests that this might be the case, but this remains an open question.

The Principle of Complementarity therefore contains an echo of Bohr's early 1920s attempt to defend the exact accuracy of classical electromagnetism. He had failed, with the BKS theory, to protect the absolute validity of classical electrodynamics, but now at least he thought he could show that classical physics *within its own sphere* was absolute. It could seem surprising that Bohr had to resort to such a radical position as complementarity in order to make room for his deeply conservative beliefs about classical physics.

Some remarks of Bohr's also hint at a statistical mechanics account of measurement. He said that every measurement procedure must be brought to a close by an "irreversible act of amplification" (Wheeler and Zurek 1983, p. 7). An example would be the exposure of a grain of photographic emulsion by a photon. Clearly, the kind of irreversibility Bohr had in mind here is thermodynamic or statistical, like the shattering of a wine glass on the floor, but Bohr did not develop this notion in detail.

DIRAC: TWO KINDS OF PARTICLES

While Bohr, Schrödinger, and Heisenberg wrestled with the meaning of quantum mechanics, Dirac (who had little interest in philosophical debates) kept constructing pretty equations.

By 1928, he had worked out his own version of quantum mechanics, which focused on noncommutativity as the feature that distinguished quantum from classical physics and described observations using the language of linear transformations of state vectors. There was one major deficiency, he felt, in all of the formulations of quantum mechanics up to that point including his own—they were not relativistic. This meant that they were accurate only for relative velocities that are small compared to the velocity of light and did not take into account the invariance of the speed of light, the rock upon which special relativity is founded. Dirac thought that it was time to seek an equation for the state function for the electron that could be written in *covariant* form. This means that it would be fully consistent with relativity and would treat time and space similarly, since in relativity time and space coordinates can be transformed into each other. From this point of view the Schrödinger Equation is defective because it is not *homogeneous*, which means that its derivatives are not all of the same order. It is first order in time, but second order in space coordinates.

There is a relativistic wave equation that is all second-order, usually called the Klein-Gordon Equation, although it seems to have first been discovered by Schrödinger. However, it does not describe the electron accurately. (It was eventually found to be a valid description of the state vector evolution for particles with spin-0—but that is getting ahead of the story.) Dirac then decided to see whether or not he could write a wave equation for the electron that would be first-order (first derivatives) in all four coordinates, space and time. The problem would be to find state functions with the right sort of mathematical structure to satisfy such an equation, and then try to solve the equation and see if it gave physically meaningful results. Some inspired algebraic manipulation showed Dirac that he could represent the states of electrons using *spinors*, which are four-component complex-valued vector-like objects that turned out to be constructible out of the Pauli spin matrices. (Spinors were invented by the distinguished French mathematician Élie Cartan, 1869–1951.) Using spinors, Dirac was able to write a relativistic wave equation for the electron, now called the Dirac Equation. It is essentially a covariant version of the Schrödinger Equation. It can be adapted to many other sorts of particles moving at relativistic velocities and is one of the basic tools of quantum field and particle theory. Dirac's picture of the electron also has the satisfying feature that it *predicts* the electron's intrinsic spin; it is no longer necessary to add spin into quantum theory by hand as it had been up to that point. The way this works is that in order to write a wave equation that was relativistically covariant, Dirac had to assume that the electron had two components, and

Figure 5.1: Paul Dirac. Photograph by A. Bortzells Tryckeri, courtesy AIP Emilio Segre Visual Archives.

these correspond nicely to the two possible spin states of the electron. This is the kind of result that theoretical physicists love—deriving important facts from a small number of general principles.

Particles and Antiparticles

A very odd feature of Dirac's theory was that the possible energy values turn out to be given by the square roots of a relativistic expression. Since square roots can be positive or negative, this seemed to predict the existence of negative energy states. Dirac boldly suggested that this was not a mistake. Instead, he argued, we can assume that the vacuum is an infinitely deep sea of mostly occupied negative energy states, often now called the *Dirac Sea.* By the Pauli Exclusion Principle each possible energy state could be occupied by only one electron. Normally we do not know about these energy states, precisely because they are occupied and because an energy state can only be observed if an electron can drop into it and emit a photon. However, if an electron is knocked out of a negative energy slot by a passing gamma ray it leaves a hole behind. In order to keep the electric charge balanced, the hole has to be positively charged, and the hole will move around exactly as if it were a particle. If the electron falls back into the hole then both it and the hole will disappear, and a gamma quantum will be emitted.

At first Dirac hypothesized that the hole where the electron had been could correspond to the proton, which at the time was the only other positively charged particle known. In 1930 several physicists pointed out that if the proton was Dirac's missing electron, then it would be possible for the proton and the electron in the hydrogen atom to suddenly annihilate each other, releasing two photons. All matter would disappear in a great flash of light, a process that is fortunately not observed. Furthermore, the hole corresponding to the electron would have to have the same mass as the electron, which is less than 1/1800 the mass of the proton. Finally, in 1931, Dirac took what now seems to be the obvious step and proposed that there must be a particle distinct from the proton, positively charged, but having the mass of an electron. Barely a year later American physicist Carl Anderson (1905–1991) detected positively charged particles in cosmic ray showers. Because of the amount of curvature their tracks showed in a magnetic field he deduced that they had to have very nearly the mass of the electron. However, they curved in the opposite direction to electrons, showing that they were positively charged. Anderson dubbed his new particles *positrons*—positive electrons, with precisely the same mass as the electron, but positively charged. Anderson himself did not realize at first that he had confirmed Dirac's theoretical prediction, because he was too busy getting his delicate apparatus to work to study Dirac's abstruse papers.

The discovery of the positron provided a simple interpretation for the fact that Dirac's spinors had to have four components: there are two components for the electron (one for each spin state), and two components for the positron.

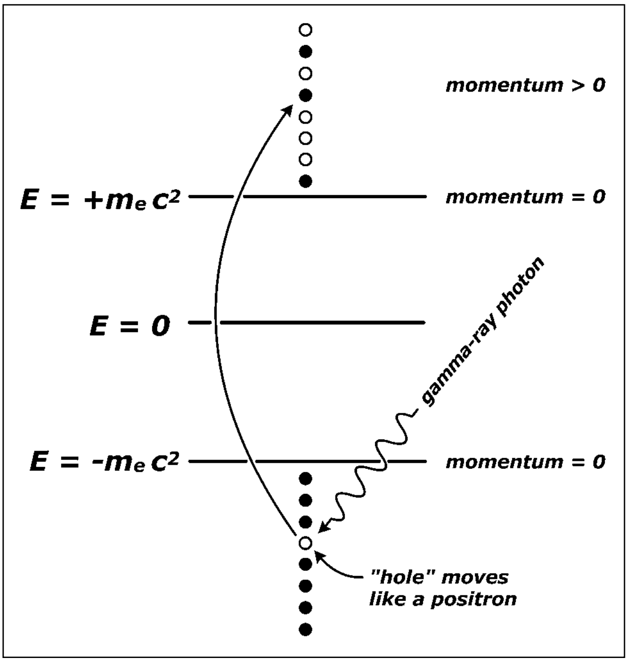

Figure 5.2: The Dirac Sea. All negative energy states are normally occupied by exactly one electron. A gamma ray can knock an electron into a positive energy state, creating a hole that looks like a positron. Illustration by Kevin deLaplante.

The confirmation of Dirac's prediction had other important implications. First, it suggested that *the vacuum is not nothing*; rather, the vacuum (which classical physics naively describes as "empty" space) could be crammed with particles that we simply aren't able to observe normally. Einstein and others noted that Dirac had, in effect, restored the ether of nineteenth-century physics, although in a strange, quantum form. Second, it showed that *particle number is not conserved.* Particles can be created and annihilated; the gamma ray bouncing (or scattering) off a negative electron will seem to split into a positron and an electron, and after these have careened around for a while they will annihilate each other, leaving a gamma photon behind. To put it another way, particle number is not (unlike mass-energy or charge) a conserved quantity. The phenomenon of particle creation and annihilation would come to have huge importance in quantum mechanics.

The discovery of the positron also suggested that other particles may have antiparticles as well, and as early as 1933 Dirac hypothesized that there had to be an antiproton and that there could even be stars or planets far away in space that were composed of antimatter.

While the discovery of the positron was a satisfying confirmation of Dirac's theory, it was also troubling, because it spoiled the nice simplicity of particle theory of the late 1920s. For a while it had seemed that it might be possible (as soon as a few more niggling technical details were worked out) to explain the whole structure of matter in terms of just two particles, the proton and the electron. How many more "elementary" particles were lurking in the mysterious nucleus of the atom, or in the vacuum itself, which was beginning to look like something with a very complex structure indeed? The answer would turn out to be—lots!

In 1936, de Broglie, with his usual gift for spotting the obvious, articulated what was becoming apparent to many physicists: for every particle there is a corresponding *antiparticle* with *opposite* quantum numbers. With one stroke de Broglie nearly doubled the number of particles, although it would take a few more years to experimentally confirm his guess.

TWO KINDS OF STATISTICS

In 1926 the precocious Italian physicist Enrico Fermi (1901–1954), together with Jordan and Dirac, made an important advance in understanding the statistics of quanta. (Fermi was to make many contributions to physics, including leading the construction of the first nuclear reactor in 1942.) They showed that there are two kinds of quanta that obey importantly different statistics.

In fact, there are three kinds of particle statistics. *Boltzmann statistics* are the classical statistics of distinguishable particles; this was described in the nineteenth century by Boltzmann and is still a useful approximation when particles can be treated as independent entities. However, Boltzmann statistics break down when quantum effects become important. When the de Broglie wavelengths of the particles begin to overlap, the particles become indistinguishable and begin to correlate in unexpected ways, and they can no longer be counted like classical objects.

Fermi-Dirac statistics are the statistics of particles that obey Pauli's Exclusion Principle. Such particles are now called *fermions.* The defining feature of fermions is that each possible quantum state can be occupied by no more than one particle. Electrons and protons, and in general, particles that can make up *matter*, are fermions. The stability of ordinary solid matter itself is a consequence of the fact that fermions obey the Exclusion Principle.

Particles obeying Bose-Einstein statistics, now called *bosons*, are particles that, in direct contrast to fermions, obey what might be called the "inclusion principle," which means that they have a higher probability of occupying a given state the more particles are already in that state. The photon was the first boson to be identified, and there would soon be many more. The laser is possible because photons will all go into precisely the same state if given a chance, making it possible to generate a beam of perfectly coherent light. Bosons typically appear as *field quanta*, which transmit interactions between fermions, the components of matter. The photon is the quantum of the electromagnetic field.

Think of fermions as rugged individualists, and bosons as conformists who prefer to disappear into a crowd. The difference in their behavior is controlled by a mere plus-or-minus sign in the distribution function (the formula stating the number of particles having a given energy). At a more formal level, the difference between fermions and bosons lies in the behavior of their wave functions. The wave function for a number of fermions is *antisymmetric* under particle exchange; this means that if two particles are interchanged in the wave function, the sign (plus-or-minus) of the wave function changes. By contrast, the wave function for bosons is *symmetric*, meaning that it stays the same if particles are interchanged.

In 1940, Pauli proved a key result called the Spin Statistics Theorem: particles with integral spin (0, +/–1, +/–2, etc.) are bosons, while particles with half-integral spin (+/–1/2, +/–3/2, etc.) are fermions. Spin is measured in multiples of Planck's reduced constant $\hbar = h/2\pi$, first introduced by Dirac in 1926. (The symbol $\hbar$ is pronounced "h-bar".) All fermions have distinct antiparticles; for instance, the antiparticle of the electron is Anderson's positron. Neutral bosons such as the photon, however, are their own antiparticles —and that is why de Broglie only *nearly* doubled the number of particles by observing that every particle has an antiparticle.

THINGS ARE MADE OF PARTICLES, BUT ARE PARTICLES THINGS?

By the time that particle statistics were clarified by Fermi and others, there could no longer be any doubt that the elementary particles of which the world is, presumably, made are unlike ordinary objects in many ways. First, it is not clear that they can persist through time the way Mt. Everest can. As Heisenberg argued, even if a particle is following a detectable trajectory, the trajectory can only be defined by a sequence of discrete detection events, and we cannot be sure that the particle even exists between detections. Second, all quanta are indistinguishable, which expresses the fact that a particle such as an electron has no other distinguishing features than its quantum numbers. Ordinary objects possess an indefinitely large amount of detail; two pennies, for instance, can always be told apart if they are examined closely enough. Third, particles obey very different sorts of statistics than do ordinary objects. For the case of bosons this difference should be observable at the macroscopic level, as Einstein first predicted, since there is no limit to how many bosons can go into one quantum state. Yet another challenge to common sense is forced on us by the fact that the structure of elementary particles cannot be described by comprehensible classical models. While by 1926 electrons had been shown beyond a doubt to possess an intrinsic angular momentum, it is impossible to model an electron like an ordinary spinning object. As Pauli showed, it would have to be spinning far faster than the speed of light. So far it has been impossible to experimentally define a radius for the electron with even the most powerful of modern particle accelerators, and so particle theory usually treats the electron as a mathematical point (even though we know, by Heisenberg, that this does not make physical sense) with a definite rest mass, intrinsic spin, and electrical charge. It can be described with great accuracy using the formalism developed in its most clear form by Dirac, but it is like nothing that we can picture or hold in our hands.

Somehow, despite these facts, ordinary matter is built up out of nothing but quantum mechanical combinations of extraordinary quantum matter. It is the job of the quantum physicist to show how this is done; and by about 1930 the tools were mostly at hand to do this, although it was less clear than ever why these tools worked. As quantum mechanics increased in predictive power, the duality between the quantum and classical pictures of the world had only sharpened.

Life in Hilbert Space

John von Neumann (1903–1957) was a Hungarian-born mathematician renowned for his phenomenal memory and powers of calculation. He made important contributions to quantum physics, mathematics, logic, computing, theoretical economics, and the development of the hydrogen bomb.

By 1932 von Neumann had distilled the work of the founders of quantum mechanics into a unified *axiomatic* version of quantum theory, which means that he set forth a set of rules or axioms from which the rest of the theory follows mathematically. There are other ways of doing quantum mechanics, but von Neumann's axiomatic formulation, expressed in Dirac's efficient notation, is probably the most widely used version of nonrelativistic quantum mechanics, and it is usually what is taught to university physics majors.

On von Neumann's view, the mathematics of quantum mechanics is a kind of linear algebra. The basic object of study is the *state vector*, which represents a possible preparation state of the physical system under study. The state vector can be written as a column vector with complex-valued components, and it is usually represented as a "ket" in Dirac notation. State vectors live in complex-valued linear spaces called *Hilbert Spaces*, after the influential Göttingen mathematician David Hilbert. There is not one Hilbert Space but many; every experimental arrangement has its own. Hilbert Space is merely a mathematical device that encodes the number of degrees of freedom of the system and its symmetries; no one thinks that there really is such a place.

The state vector for any system is a sum, or *superposition*, of components, each representing a possible state that the system could be in. For instance, the spin state of an electron is represented by a two-component vector, with one component representing spin *up* and the other representing spin *down.* Each component of a state vector is multiplied by a complex-valued constant called a *phase factor.*

State vectors are transformed into other state vectors by mathematical machines called *linear operators.* The behavior of operators is defined by their commutation relations, and, as noted, the crucial fact is that some operators do not commute; that is, one gets a different result if the measurements they represent are performed in reverse order. (The notion of noncommutativity is not so counterintuitive in itself. Toasting bread and then buttering it produces a different and more satisfactory result than buttering bread and then toasting it.) Some operators rotate a state vector, while others merely stretch it, that is, they multiply it by a constant called a *scalar.* If an operator simply stretches a vector, then the vector is called an *eigenvector* of the operator, and the factors by which the operator multiplies the eigenvector are called the *eigenvalues* of that operator. Operators can be represented by square matrices, and an operator will be represented by different matrices in different "representations" (coordinate bases that are defined by different types of observations). Recall that Heisenberg's first version of quantum mechanics was written in terms of matrices. If an operator is transformed in such a way that its matrix representation is *diagonalized* (which means that only its diagonal components are non-zero) then the diagonal values are the eigenvalues of the operator.

Quantities that can actually be observed, such as position, momentum, or energy, are represented by *Hermitian* operators (after the French mathematician Charles Hermite, 1822–1901). A Hermitian operator is simply an operator whose eigenvalues are real numbers; such operators are also called *observables.* The set of possible eigenvalues for an observable is often called the *spectrum* of that observable. A state vector is an *eigenstate* of an observable if the observable acting on that vector leads to the observation of one definite eigenvalue of the observable. The eigenstates of an observable can serve as the basis vectors for the state space of the system; basis vectors are just the unit vectors in terms of which all other vectors can be decomposed.

The connection between theory and observation is much less direct in quantum mechanics than it is in classical mechanics. Quantum mechanics gives rules for the changes that state vectors and observables undergo, but we never perceive state vectors or observables as such; they merely serve as devices for calculating eigenvalues and the probabilities of their occurrence, which are the things that can actually be measured. Another useful quantity that can be calculated is the *expectation value* of an observable, which is the average value of its eigenvalues, weighted by the probabilities of their detection. Expectation values are the long-run average values that we can expect to observe in a series of similarly prepared experiments.

When the system goes from a given initial state to a final state, via a certain measurement procedure, there is a quantity called the *amplitude* (*probability amplitude*, or *transition amplitude*) for getting to that final state. Amplitudes are complex numbers, and by themselves have no agreed-upon physical interpretation, but their moduli (squares) are the probabilities of getting various possible results. This is an abstract and generalized version of Born's Rule. Phase differences between amplitudes become crucial when probabilities are calculated using the Born Rule, since they determine the *interference* between the various components of the state vector.

State vectors and amplitudes obey the *Superposition Principle*, which states that any linear combination (superposition) of allowed state functions is an allowed state function. (There are some exceptions to this rule, called superselection rules.) The Superposition Principle allows for the possibility of interference between physical states that from the point of classical mechanics would be entirely independent (such as particles outside each other's light cones).

There is an especially important observable called the *Hamiltonian.* It represents the structure of energy in the system, but like all operators it does so in an indirect way, since its eigenvalues represent the possible energy values that the system can have. The differences in energy eigenvalues then gives, by Bohr's rules, the frequencies of the possible spectral lines an atom can emit, while the probabilities of the transitions gives the intensities of those lines.

There is more, much more, to the mathematical structure of von Neumann's quantum mechanics, but enough has been sketched here to give some familiarity with the most commonly used vocabulary of the field.

Collapse of the Wave Function

In von Neumann's picture of quantum mechanics we encounter yet another odd duality: there are *two* ways that the state vector can evolve. If the system is not under observation, the state vector evolves smoothly and deterministically according to Schrödinger's Equation. Mathematically, *Schrödinger evolution* is a rotation of the state vector in the system Hilbert Space; such transformations are reversible, like any rotation, and are said to be *unitary.* When the system is measured, however, the state vector abruptly collapses (or reduces) to an eigenstate of the observable. This process is called the *collapse of the wave function*, and it is represented mathematically by a *projection.* Since several vectors can project onto one eigenstate, a projection is in general irreversible, and information is lost. Because of the loss of information, one cannot always tell from a given experimental result what the preparation state of the apparatus was.

Hardly any physicists believe that von Neumann's collapse postulate is literally true. There are several objections to it. First of all, it seems too much like calculating the wrong answer (that the system will remain in a superposition), rubbing it out, and penciling in the observed experimental outcome. Second, it seems mathematically clumsy to have two types of system evolution. Third, there are problems with finding the right way to describe state collapse in relativistic spacetime. There is also increasing evidence that state collapse can sometimes be reversed, so it may simply be false that some components of the state vector just go *poof!* like a soap bubble.

There are several no-collapse versions of quantum mechanics, but none yet stand out as the obvious replacement for the von Neumann picture.

THE DOUBLE SLIT EXPERIMENT

There is no better way to capture the essence of the discoveries of the mid-1920s than through the double slit experiment. It expresses the mysteries of the wave-particle duality in a very clear way. Although it is usually described as a thought experiment, it is based on observations that have been confirmed innumerable times in real quantum mechanical experiments.

The purpose of the experiment is to reveal the difference between classical and quantum behavior. First, we set up a machine gun. There is a sheet of steel between the machine gun and a detector, which could simply be a wall that can absorb the shots. In the sheet of steel there are two closely spaced slits that are only slightly larger than the diameter of a bullet. When the gun is fired most of the bullets will go through either hole, but some will ricochet away and some will be deflected slightly when they glance off the edges of the holes. The result will be two humped distributions of bullet impacts. This is typical classical particle behavior.

The second setup is designed to demonstrate how classical waves behave. Set up a tank with water in it. Have a bob moving up and down and thereby generating concentric wavelets, and have a barrier with two closely spaced holes that are roughly similar in width to the wavelengths of the ripples in the tank. Dream up some way of detecting the waves that get through the barrier. (The right side of the water tank could be a gently sloping beach, and bits of driftwood in the water could show how far up the beach the waves have gone.) The waves that go through the barrier will spread out from each hole in roughly

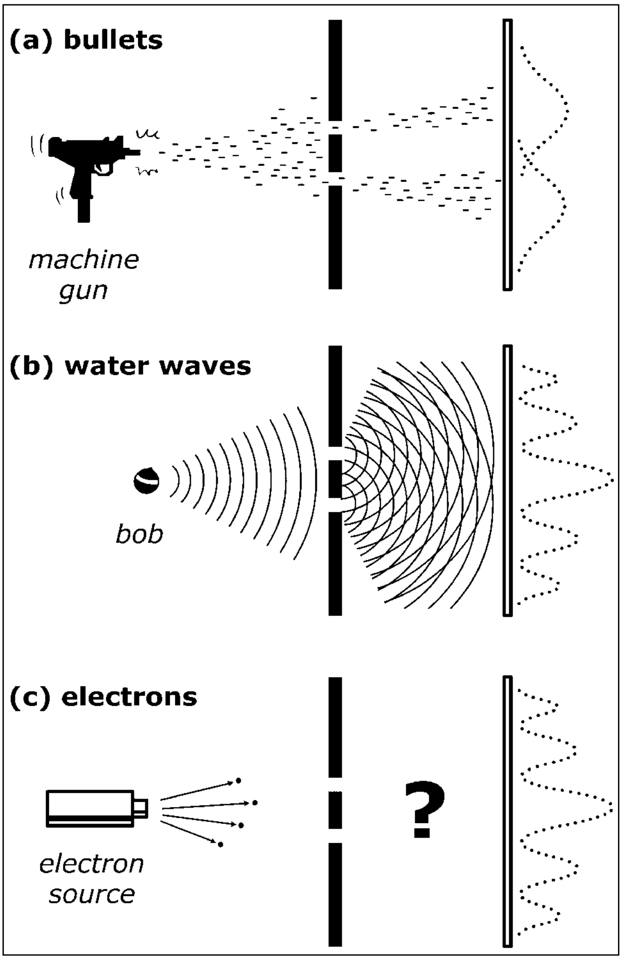

Figure 5.3: The Double Slit Experiment. The top curve shows the distribution of bullet impacts. The middle curve shows the distribution of driftwood on a beach due to water waves. The lower curve is a typical interference pattern of electron impacts on a detector screen. If we try to determine which hole the electrons go through, we destroy the interference pattern and get a bullet-like distribution. Illustration by Kevin deLaplante.

circular patterns that overlap, and when they hit the detector they will display a classic interference pattern. The peaks form where *constructive* interference between the wavelets occurs because the wavelets were in phase with each other. The troughs, or nodes, in the interference pattern form where *destructive interference* occurs because the wavelets were out of phase with each other. Interference effects such as this guided Thomas Young in the early 1800s to argue that light travels as a wave.

Now, try the same arrangement with elementary particles such as electrons or photons. Shoot them through twin slits in a barrier, with the slits roughly comparable in size to the de Broglie wavelengths of the particles. Like the bullets, each particle will hit the detector screen (which could be a photographic plate so that there will be a record of the results) as if it were a localized, discrete object, and it will leave a small spot showing where it landed. It seems that it has been shown that electrons are particles.

Not so fast! It will be seen that the pattern of particle detections forms a *wave-like* interference pattern. This is the wave-particle duality: particles are *detected* as if they are discrete, localized objects, but the *probability* of their detection is a waveform.

One is tempted to think that this could be explained by some sort of interaction between the particles as they go through the slits. To see that this is not quite right, suppose that we turn down the particle intensity so low that only one particle is going through the apparatus every week on average. As before, the particles will be detected, one by one, as discrete spots on the screen. But the truly extraordinary fact is that over a long time the detection events will, again, build up the same familiar interference pattern that appeared when there were lots of particles going through the apparatus at once. It is as if each particle interferes with other particles that were emitted long ago or that will be emitted in the future. Or perhaps it is that each particle somehow interferes with itself. (Dirac described the phenomenon this way.)

It would seem that in order for interference to be possible, the particles have to go through both slits at once. Suppose we now decide to find out which slit each particle goes through. The only way we can do this is to somehow intercept or interact with the particles as they go through the slits. We could, for instance, try shining light off them. We have to give the light a short enough wavelength for it to be able to reveal which hole the particles go through. This can be done, and it will, in fact, reveal individual particles going through one slit or the other, but not both at the same time. But when we can tell which slit each particle goes through, the interference pattern disappears, and we get a double-humped distribution just like the distribution for the machine gun bullets. The interference pattern appears only when we cannot tell which slit the particles go through. If we can tell which slit the particles go through, we get a classical "machine gun bullet" distribution.

The double slit experiment illustrates the odd connectedness of quantum phenomena. How does one electron "know" how to interfere with itself or with another electron that was emitted days earlier? It is not even clear whether this question makes sense.

The experiment also illustrates not only the inherently probabilistic nature of quantum mechanics, but an important difference between classical and quantum probabilities. What is the chance of a given major league baseball team winning the World Series? Suppose one knew nothing about the teams except how many there are; then the probability of a team winning would be simply 1 in 30, because there are 30 major league teams. However, if one knew more about the teams (such things as, for instance, their track records, or which players they have) one would be able to say that certain teams have a higher probability of winning than others. Estimates of classical probability are always conditional on the background information available, and the more background information one can get, the more accurately probabilities can be estimated. Classically, there is presumed to be no limit, other than obvious practical limits, to the amount of background information available and therefore no limit to how closely the probabilities of various events can be estimated.

A major difference—some might say *the* major difference—between quantum mechanics and classical physics is that there is in general a maximum amount of information we can get about a system before changing (usually irreversibly) the very nature of the system. A large number of quantum systems can be prepared in exactly the same way, but they will not always behave the same way; this is what is meant by saying that quantum mechanics is *indeterministic*. Quantum mechanics is very good at calculating the *probabilities* that particles will behave in various ways with high accuracy. However, as the double slit experiment illustrates, if we try to learn more than a certain amount about the particles we change the very nature of the experiment, and we get an entirely different sort of result. Quantum mechanics is probabilistic in principle, not merely by dint of practical necessity but because it appears that in many quantum mechanical experiments *there is just no more information to be had.* It was this that Einstein objected to more than anything else.

A HISTORICAL PUZZLE

The historian of science Paul Forman has argued controversially (1971) that the sudden appearance of quantum mechanics, with its emphasis on discontinuity and uncertainty, was a reflection of the breakdown of old social and political certainties in the turbulent years of the Weimar Republic in Germany following World War I. Forman's thesis is too complex to fully evaluate here. However, the history of quantum physics seems to reveal that most of the pioneers of quantum physics were not trying to justify preconceived ideas about how physics ought to be, but instead were doggedly following the leads given them by experimental results and mathematical necessity. In many cases they were surprised and even dismayed at what nature revealed to them; Heisenberg himself later spoke of his discovery of matrix mechanics as "almost frightening" (Heisenberg 1971, p. 69). Advances in physics are made by a combination (in proportions varying from scientist to scientist) of philosophical analysis, mathematical skill, experimental ingenuity, and physical intuition—and (very important) a willingness to accept unexpected results.

Perhaps youth plays a role. Most of the decisive breakthroughs in the period 1920–1930 were made by scientists younger than 30 years of age, and often younger than 25. This was noted at the time, and the phenomenon was jocularly described as *Knabenphysik*—"boy physics." Schrödinger was near 40 and Born in his early 40s when they made their major contributions to quantum mechanics, and thus they were the old men of the team. Great innovations in science, especially in mathematics and theoretical physics, are rarely made by older people. Is this merely because of aging?—or do older people get too committed to comfortable ways of thinking? Not enough is known about the nature of scientific creativity to answer these questions.

The advance of science, especially in an extraordinarily creative period such as 1925–1935, reveals yet another duality that is peculiar not just to physics but to all of science. Scientists, and academics generally, tend individually to be people of somewhat conservative character, however bold they may be in their speculations, and they usually work at universities, which have dual (Bohr might have said complementary) mandates. On the one hand, the task of a university is to preserve and transmit existing knowledge; on the other, a university exists to foster innovation and the search for new knowledge. Tensions arise from the fact that in order to seek new knowledge, researchers have to admit that their old knowledge, which they may have invested an important part of their lives in mastering, is in some respect incomplete or even *wrong*. It is sometimes very hard for academics to do this. The distinguished American physicist John Archibald Wheeler (1911–), who himself worked with Niels Bohr, once said that science advances by "daring conservatism." There seems to be no general recipe, however, for knowing when one is being too daring or too conservative, and one can see the scientists who created quantum physics, especially in the period up to 1935, struggling to find this balance.

Whether or not Forman is exactly right, there were historical forces at work in the years from 1900 to 1935 that made it possible for a group of exceptionally talented academic scientists to take the intellectual risks that made quantum mechanics possible. It would be nice to have a recipe for this historical magic so that it could be replicated whenever it is needed.

6

ELEMENTS OF PHYSICAL REALITY

This chapter covers the foundational debates about the meaning of quantum mechanics that occurred in the years 1927 to 1935. Most physicists of the time regarded these debates as largely irrelevant—at best, the sort of thing that one chatted about in the pub after a hard day at the lab—and preferred to press on and apply the powerful new techniques of quantum mechanics to the vista of unsolved problems in physics and chemistry that opened out before them. But the philosophical debates of 1935 would turn out to be the front-line research of the first years of the twenty-first century.

EARLY CAUSAL INTERPRETATIONS OF WAVE MECHANICS

The most obvious way to respond to the puzzle posed by the double slit experiment is to imagine that the wave function that determines the probabilities of particle detection really does describe some sort of actually existing wave that guides the particles to the proper spots on the detection screen. It was clear by 1927 or 1928 that electrons themselves cannot be *nothing but* waves, since they are always detected as highly localized particles even though the probabilities of their detection follows a wave-like law. But perhaps the wave function is a physically real thing that *guides* or *pilots* the particles by some mechanism to be determined, rather than merely a description of probabilities. This is called the *pilot wave* interpretation of quantum mechanics, and early versions of it were explored by several physicists in the late 1920s. All pilot wave theories have two features in common: they are continuum theories (because they attempt to explain particulate behavior in terms of wave-like structures) and they are deterministic, because the behavior of waves and particles in these theories is governed by partial differential equations of types that lead to definite, unique solutions. For the latter reason pilot wave theories

Figure 6.1: Niels Bohr and Albert Einstein. Photograph by Paul Ehrenfest, courtesy AIP Emilio Segre Visual Archives.

are examples of *causal* interpretations, and their authors hoped that if they could be made to work they would get rid of that "damned" indeterministic quantum jumping that Schrödinger had complained about. It was also hoped that causal versions of quantum mechanics would give a spacetime picture of quantum processes and thus satisfy an instinct for mechanical explanation that was frustrated by the new quantum mechanics, which was increasingly expressed in highly abstract mathematics.

An early causal interpretation of wave mechanics was offered by Erwin Madelung (1881–1972), who in 1926 outlined a hydrodynamic interpretation of wave mechanics. Hydrodynamics, the physics of fluids, is based on partial differential equations describing fluid flow. Madelung started from Schrödinger's suggestion that the wave function described a continuous distribution of charge and treated this charge distribution as an electrified fluid. Madelung had to reconcile the well-established existence of electrons as discrete particles with this model, and he proposed that electrons were in some unclear way dissolved into his hypothetical electrical fluid. His work was not found to be convincing. Still, Madelung's mathematics, differently interpreted, has appeared in other causal rewritings of quantum theory.

Louis de Broglie offered a more sophisticated causal model in 1927. At first he hoped to be able to show that the electron could be understood as a singularity in the wave field. What this means is that the electron would be a sort of knot or eddy whose structure would be determined by a dynamic law (probably nonlinear) acting on the wave function. De Broglie called this the theory of the *double solution,* because the same wave equation would have two sorts of solutions, one for waves, and one for highly localized concentrations of energy that would behave like particles. He was not able to arrive at a mathematical law that could do this ambitious job, and instead, at the 1927 Solvay Conference, he proposed a provisional pilot wave theory according to which the particles were carried along by the quantum wave field like chips of wood in a stream. The theory was inherently *nonlocal,* since the particles, in order to be able to behave in a properly quantum mechanical way, had to somehow sense the positions and momenta of all other particles in the system, instantaneously.

De Broglie's theory did not get a warm reception. Pauli argued that de Broglie had failed to give an accurate account of scattering. Wave mechanics says that when a particle scatters off a target, the wave function of the scattered particle expands outward from the target in spherical ripples, whereas particles are always detected within very narrow locations and directions. How could de Broglie explain this discrepancy? He did not have a clear answer—in 1927. Apart from Pauli's specific complaint, followers of the Copenhagen Interpretation had two kinds of objections to causal theories, one radical and one conservative. First, the Copenhagenists thought it was just mistaken to return to a picture in which positions and momenta had exact meanings independent of the experimental context. This view would soon be reinforced by a mathematical proof by John von Neumann that apparently demonstrated, to everyone's satisfaction at the time, that it is mathematically impossible for there to be a hidden variables theory that can reproduce the statistical predictions of quantum mechanics. Second, they could not accept the fact that the sort of causation contemplated in causal interpretations seemed to be inherently nonlocal, a sort of action at a distance. In the *classical* realm, Bohr insisted, relativity (the ultimate classical theory) could not be challenged.

De Broglie abandoned his theory and became, for a while, a vocal advocate of the Copenhagen Interpretation. The causal interpretation of quantum mechanics would be revived by David Bohm 25 years later, in a form that would be less easily dismissed.

SCHRÖDINGER'S CAT AND THE MEASUREMENT PROBLEM

In 1935 Schrödinger published a paper entitled "The Present Situation in Quantum Mechanics." Although it contained no new results, it was a landmark paper that raised questions that were to be debated for decades afterward.

Schrödinger invited us to consider a "fiendish device" in which an unfortunate cat is imprisoned in a box with a closed lid, so that the cat cannot be seen during the first part of the experiment. The cat's box is connected to a radioactive source that has a 50 percent probability of decaying within one hour. If the source decays, the alpha-particle it emits is detected (by a device such as a Geiger counter) and a valve is electronically triggered that releases deadly prussic acid into the box, killing the cat instantly. The quantum mechanical description of this setup says that the radioactive atom is in a superposition of states, one for it to be decayed and one for it to be not-decayed. Because the

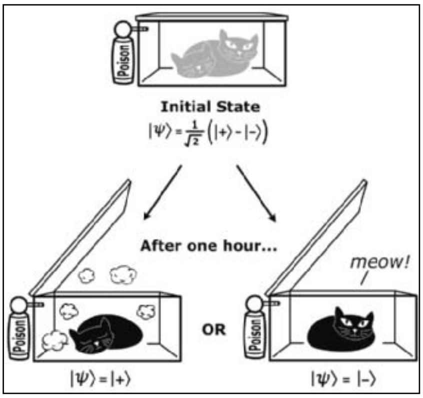

Figure 6.2: Schrödinger's Cat. The release valve for the poison is controlled by a quantum state in a superposition. Quantum theory says the cat is also in a superposition, but if the box is opened after one hour the cat will be either definitely alive or definitely dead. Illustration by Kevin deLaplante.

cat is coupled to a system that is in a superposition, its wave function gets entangled with that of the apparatus, and it is *also* in a superposition of states according to quantum formalism—one for it being dead, the other for it being alive. And yet, if the experimenter opens the lid of the box after an hour has passed, the experimenter will certainly not see the cat in a curious half-alive, half-dead state. Rather, the experimenter will either see a cat that is definitely dead (with 50% probability) or a cat that is definitely alive. In the mathematical language of quantum mechanics, the cat goes from an entangled state to a mixture (a system whose possible states can be described using classical probabilities), merely because someone opens the lid of the box.

This thought experiment illustrates Bohr's insistence that any quantum mechanical experiment will always result in a definite, classical result. It could also be taken as a confirmation of von Neumann's collapse postulate—when an observation is made, the wave function collapses into one and only one eigenstate of the system. It also shows that interesting things can happen when quantum systems are coupled to macroscopic systems; the cat in effect acts as an amplifier of an event at the subatomic scale, and there is no theoretical limit to how large an amplification factor can be achieved. (A random subatomic event could be set up to trigger a powerful nuclear bomb, for instance.) What Schrödinger really wanted to demonstrate, however, was the arbitrariness of the quantum-classical divide. The indeterminacy in the state of the radioactive sample is transferred to the cat, but it disappears when the box is opened and the cat (alive or dead) is observed by an experimenter. But at the same time the formalism of the theory says that as soon as the experimenter interacts with the system, the indeterminacy is transferred to him as well and he goes into a superposition of states—except that this is not, of course, what an actual human observer experiences.

Schrödinger had thus defined what became known as the *measurement problem.* Loosely speaking, it is simply the problem of understanding what happens during the process of measurement. More precisely, it is to explain how superpositions can turn into definite classical outcomes. Several authors have at various times shown that it is mathematically impossible for superpositions to evolve into classical states according to the Schrödinger Equation. Why is it that we always seem to get definite, very classical-looking results when quantum mechanics describes things as a blur of all possible states they could be in? Is the formalism of quantum mechanics wrong when it comes to describing measurement, or are things not as they seem to human observers?

THE MYSTERY OF ENTANGLEMENT

Einstein had suspected as early as 1905 that his light quanta were not statistically independent in the way that classical particles ought to be, and by the time the formalism of quantum mechanics was taking shape, in 1926 and 1927, it was very clear that quantum mechanical systems are interconnected in ways that defy classical intuitions. In 1935 Schrödinger introduced the term "entanglement" to describe this odd interconnectivity or mutual influence, and he described it as not "one but rather *the* characteristic trait of quantum mechanics" (1935, p. 555).

The Mathematical Basis of Entanglement

The formal basis of entanglement is fairly easy to understand in terms of the basic linear algebra of von Neumann's Hilbert Spaces. Suppose there is a system composed of two particles. (It could be any number at all.) Each individual particle can be in a number of possible eigenstates of any observable, and these eigenstates can be used as basis vectors for the state space of each particle. The state space for the *pair* of particles is called a *tensor product space* of the spaces of the individual particles, and its basis vectors are possible states of the *pairs* of particles. The most general tensor product state is a linear combination (sum) of tensor product basis states of multiparticle systems; that is, it is the space built up out of all possible linear combinations (superpositions) of the states of each individual particle. If a tensor product state is *factorizable* then the particles are statistically independent. However, it is straightforward to show that tensor product states in general cannot be factored into products of the states of the individual particles of the system; there are almost always cross-terms that translate into correlations between the particles that cannot be explained classically. From a purely mathematical point of view, therefore, entanglement is a consequence of the nonfactorizability of tensor product states, which in turn is a consequence of the superposition principle.

Entanglement can also be thought of as an interference phenomenon, like the interference of waves in the double slit experiment. Just as there are amplitudes for individual particles to go into various states, there are amplitudes for *pairs* of particles to go into various *correlated* states. Entanglement comes about when these amplitudes are out of phase and therefore interfere. From this point of view, therefore, entanglement is simply an interference phenomenon.

However it may be expressed formally, the upshot is that entangled particles cannot be treated as separate entities. Their properties are mixed up with the properties of their partners. An entangled state of two particles is not merely two separate particles connected with some sort of odd force field (although some causal interpretations would attempt to treat them this way); rather, they are more like quantum mechanically conjoined twins that do not have fully separate identities.

Historian of physics Don Howard argues that Bohr's Principle of Complementarity was a response to entanglement. Although Bohr did not use the word "entanglement," he was well aware of the phenomenon. Bohr's view seems to have been that when an experimenter makes a measurement of, say, position, his state gets entangled with the state of the apparatus. To the experimenter the apparatus is in a definite classical state (like Schrödinger's cat), but in fact the experimenter's state (because of its entanglement with the apparatus) cannot be fully distinguished from the state of the apparatus. If the experimenter should then choose to measure momentum the experimenter gets entangled in a different way, but there is no such thing as being entangled with both definite position and definite momentum states. Hence measurements of position and measurements of momentum are complementary but cannot be combined into one single process—or so Bohr seems to have thought.

QUANTUM LOGIC

An entirely new way of thinking about quantum mechanics was introduced in 1936 by von Neumann and the mathematician Garrett Birkhoff (1911–1996). They thought that it might be possible to learn something about the inner workings of quantum mechanics by trying to express it as a *logic,* that is, as a formal system of reasoning. Perhaps the fact that we live in a quantum world could be explained if it can be shown that there is deeper and more general logic to which ordinary classical logic is an approximation, just as classical mechanics is an approximation to quantum mechanics. Birkhoff and von Neumann showed that they could treat statements about possible measurement results as propositions, and represent the workings of their logic by a mathematical structure called a *lattice.* Ordinary classical logic can be represented by a so-called *orthcomplemented* lattice, while quantum logic is represented by a *non-distributive* lattice.

A distinguishing feature of quantum logic is that it fails to obey the classical distributive law of logic. Classically, saying that the master is dead and either the butler did it or the maid did it is exactly equivalent to saying that the master is dead and the butler did it or the master is dead and the maid did it. (This is called the distributive law because the "and" distributes over the "or.") Quantum mechanically, however, the distributive law fails if statements are made about noncommuting observables. For example, if spin-x is 1 and either spin-y is 1 or spin-y is –1, we cannot conclude that either spin-x is 1 and spin-y is 1 or spin-x is 1 and spin-y is –1. This is because spin-x and spin-y do not commute, so the last two statements cannot be made.

While quantum logic has contributed relatively little to practical physics so far, it has been a very important stimulus for investigations into the foundations of quantum theory, and it may yet be reborn in the new field of quantum computation.

EINSTEIN DIGS IN HIS HEELS

Albert Einstein had been one of the great pioneers of quantum mechanics. As late as his work with Bose, when he was in his mid-40s, he had been willing to forge ahead and seize at new technical results even if it was not clear how they could be reconciled with the classical worldview that he increasingly came to prefer. However, by the late 1920s Einstein ceased to participate in the development of quantum mechanics and instead became its harshest critic. He readily acknowledged that the new quantum mechanics had great empirical effectiveness. However, he felt that it could not possibly be the final form that physical theory would take, and he began to lose patience with the effort of living with contradictions in the hope that they would someday be resolved. He gradually got out of touch with recent technical developments in quantum mechanics, until he was investing almost all of his intellectual effort in his lonely search for a unified field theory that would resolve all contradictions between quantum and classical, particle and wave. Like de Broglie and Schrödinger, he thought that it should be possible to find equations that would describe elementary particles in the form of what he called "singularities" of the field. These would not be true mathematical singularities (such as what happens if one tries to divide by zero), but rather highly localized knots or concentrations of energy that would move about in a particle-like manner under the guidance of the wave field. But Einstein's vision was grander than de Broglie's, for his ultimate goal was to find a classical local field that would encompass *all* of the forces of nature, just as Maxwell had unified electricity and magnetism into a single field. Occasionally, however, Einstein found time to take stinging potshots at quantum mechanics—and especially the Copenhagen Interpretation of his friend Niels Bohr.

"God Does Not Play Dice"

One of Einstein's strongest objections to quantum mechanics was its apparently inherent indeterminism. On several occasions Einstein famously quipped that "God does not play dice with the universe." In response Bohr gently reminded Einstein that perhaps it is not for us to say what God will do. But Einstein had been troubled from the beginning by the inherently probabilistic nature of quantum physics, even as he pioneered its development by means of his own skillful use of statistical reasoning. Quantum mechanics can give extremely accurate estimates of the probabilities that an alpha particle will be emitted from a nucleus, for example, within a certain period of time and in a certain direction, but it has no way at all of telling us *exactly* when or in what direction the alpha will be emitted. Einstein was convinced that this marked an *incompleteness* in quantum mechanics: it cannot be the whole story about what is going on inside that nucleus.

Realism and the Separation Principle

Einstein was a realist in the sense that he believed that there is something about the physical world that is independent of the way humans perceive it or think of it. Like Planck, he thought of the mission of the scientist as an almost religious quest to understand this independent reality. A foundation of his conception of realism was what he called the *Separation Principle,* the statement that physical systems that are distant from each other in space at a given time are entirely physically independent from and distinguishable from each other at that time. Of course, one system can have an influence on another at a later time by such means as sound waves or light signals, but Einstein believed that any way of transmitting an influence from one system to another has to take a definite amount of time, and in no case can it go faster than light. Like his hero Newton, he thought that the very notion of action at a distance was physically absurd, and he sarcastically referred to the quantum failure of separability as "spooky action at a distance" or as a form of "telepathy."

Einstein had two reasons for his belief in separability. The first was, of course, his theory of relativity, which showed beyond a shadow of a doubt, or so he thought, that faster-than-light motion of physical influences is impossible. But there was a deeper reason for his skepticism about the apparent failure of the Separation Principle in quantum mechanics: the conception of observer-independent reality can only be maintained rigorously, he insisted, if it is physically possible to separate observers from the systems they study. In the last 25 years of his life Einstein several times stated that he regarded the Separation Principle as necessary for the very possibility of science itself. If we could not separate objects and study them in isolation, he argued, how would science be possible?

There are two answers to this. First, there is no reason to think that the world is structured for the convenience of human scientists. Second, science gets along just fine on the basis of *partial* knowledge of the parts of nature; one does not need to know *everything* about a physical system in order to make useful predictions about it. Einstein may, therefore, have simply been expecting too much of physics. Nevertheless, he was one of the first, if not *the* first, to grasp how enormous is the challenge to the classical worldview raised by quantum mechanical entanglement.

Einstein's Causal Wave Theory, and Why He Wouldn't Publish It

In 1927 Einstein, like Madelung and de Broglie, attempted his own causal version of wave mechanics and produced a mathematically sophisticated theory based on Schrödinger's Equation. Its aim was to remove indeterminism by giving a recipe for determining particle velocities uniquely in terms of the wave function. Einstein was dismayed to discover that his theory, too, violated the Separation Principle. If he tried to describe the motion of a system of particles built up out of subsystems that he assumed did not interact to begin with, he found that the wave function for the composite system could not be written merely as the product of wave functions for the individual systems, but that inevitably cross-terms appeared indicating that the particles could not be treated as separate entities within the composite system. Einstein withdrew the paper from publication, convinced that his result must be wrong.

Einstein *versus* the Uncertainty Principle

Einstein did not doubt that Heisenberg's Uncertainty Relations were accurate in practice and represented a profound insight. (In 1932 he nominated Heisenberg for the Nobel Prize.) Nor did he doubt the utility of statistical methods; after all, Einstein himself was one of the great masters of statistical mechanics. Rather, he wanted to show that Heisenberg's Δp's and Δx's were *merely* the result of (perhaps unavoidable) inaccuracies in measurement, not a sign of indeterminacies in the very *nature* of positions and momenta themselves. He wanted to show that it was contradictory to suppose that a particle did not always have a definite position and momentum, even if we might not be able to determine those quantities with arbitrary accuracy in practice.

Einstein was well aware that in the formalism of wave mechanics, as it had taken shape by the late 1920s, one cannot even express the notion of a particle having simultaneously precise position and momentum. It would be a mathematical contradiction in terms, like asking for a square circle. However, Einstein did not worry about this, since he felt that it was more important to get the physical picture right and repair the formalism afterward. Many of Einstein's greatest breakthroughs had been sparked by simple but elegant thought experiments. Throughout this period Einstein devoted a great deal of his considerable ingenuity to searching for ways to show that there was more information available in an experimental setup than allowed for by the Uncertainty Relations. Somehow, Bohr would always find an error in his arguments, and then Einstein would try again.

Embarrassment in Brussels

One of Einstein's most ingenious attempts to defeat the Uncertainty Relations was presented at the Solvay Conference of 1930 in Brussels, Belgium. It was based on another one of his deceptively simple thought experiments. Suppose there is a box containing one and only one photon. (The term "photon" was by this time in current use.) Suppose also that the box is equipped with a shutter that is controlled by a very precise timer. We design the timer so that it briefly opens the shutter for a time interval that can be set as narrowly as we want. If the photon flies out through the shutter we know the time interval within which it did so to arbitrary accuracy. We weigh the box before and after the photon leaves. The box will be slightly less massive when the photon has left, and by the relativistic equivalence of mass and energy we can determine the energy of the photon. Because Einstein took it that the box could be weighed to arbitrary accuracy, the energy of the photon could thereby be determined to arbitrary accuracy as well, and it would be possible, therefore, to violate the version of Heisenberg's Uncertainty rules that states that the product of the uncertainties in energy and time must always be greater than Planck's constant.

Bohr could not see an immediate answer to Einstein's argument, and he spent a sleepless night struggling to find the error. In the morning he appeared, triumphant, and showed that there will be inevitable uncertainties in the timer reading and measurement of photon mass that are of exactly the right amount to save Heisenberg's formula. In order to weigh the box it has to be suspended in a gravitational field, and when the photon is emitted the box recoils, introducing uncertainties in its position, momentum, and in the reading of the timer clock attached to it. The timer uncertainty comes from nothing other than Einstein's own formula for the gravitational red shift of the rate of a clock when it moves in a gravitational field.

Bohr generously emphasized how much had been learned from Einstein's clever example. (Much can sometimes be learned from making an interesting mistake.) Another important implication of Bohr's analysis of Einstein's thought experiment is that there are deeper connections between quantum mechanics and gravitation than meet the eye. The irony of the story is that 10 years earlier it had been Einstein who victoriously defended the more radical quantum reading of particles while Bohr had tried to protect the classical picture of electromagnetism. Now the tables were turned, with Bohr defeating Einstein's attempts to argue away quantum uncertainty by means of a principle from Einstein's own theory of gravitation. Bohr had decisively won round two of the Bohr-Einstein debates. But there was another round to follow.

THE EINSTEIN-PODOLSKY-ROSEN BOMBSHELL

In 1935, Einstein, in collaboration with younger colleagues Boris Podolsky (1896–1966) and Nathan Rosen (1909–1995), published his last and greatest attempt to undermine quantum uncertainty. This time his arrow struck home—although it did not hit the exact target he had been aiming for.

The title of their paper was "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" (It is usually called just the EPR paper, after the names of its authors, or perhaps after the phrase "elements of physical reality.") The EPR paper is one of the most widely cited scientific papers of the twentieth century, but its argument is subtle and impossible to describe fully without the use of mathematics. The paper was actually written by Podolsky, because Einstein was not comfortable writing in English, while the detailed calculations were carried out by Rosen, who was expert in the wave mechanics of entangled states. It is unfortunate that Einstein did not write the paper himself, because his own prose (whether in his native German or in a competent English translation) is invariably crystal clear. Later on Einstein expressed annoyance at the way Podolsky had written the paper, because he felt that the simple point of the argument was "buried in erudition."

Einstein and his coauthors began by introducing their notion of the *completeness* of a physical theory. To be complete, they declared, a theory must somehow represent every element of the physical reality it supposedly deals with, and it must do so in a way that treats these elements as existing independently of the experimenter. To test the completeness of a theory (such as quantum mechanics), one has to know what the elements of physical reality

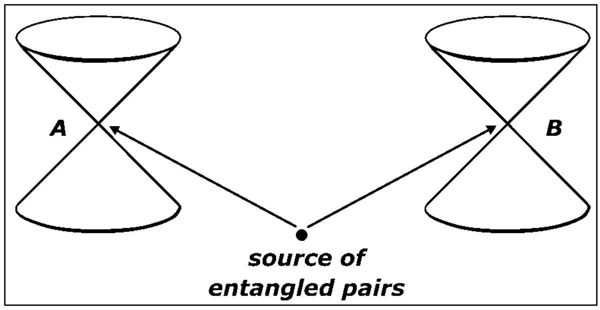

Figure 6.3: The EPR Apparatus. Particles A and B are outside each other's light cones and therefore cannot influence each other—or can they? If Alice measures the momentum of A, she knows the momentum of B. If Alice measures the position of A, she knows the position of B. Does this mean that B has definite position and momentum before Alice makes any measurements? Illustration by Kevin deLaplante.

are that it is supposed to describe. It might be very hard to come up with a complete list of such elements. But EPR declared that there was one method that would be sufficient to identify an element of reality (even if it might not give the whole list): if it is possible to predict the value of a physical quantity with certainty and without having disturbed, influenced, or disrupted the system in any way, then there must be an element of physical reality corresponding to that quantity. If a theory could *not* predict the value of that quantity with certainty, it would therefore not be complete in this sense. EPR's object was to prove that quantum mechanics fails to predict the value of a quantity that they intended to show was predictable on other reasonable grounds.

The basic structure of the apparatus in the EPR thought experiment (and many variants of this basic design have been described in the literature since 1935) begins with a source of two or more entangled elementary particles.

These particles interact dynamically with each other. There are several ways in which this could happen. They could, for example, have decayed from other particles or simply collided with each other. This dynamic interaction entangles their wave functions. The particles are then allowed to fly off in opposite directions to a considerable distance, where they interact with detectors that measure some of their physical properties.

EPR asked us to consider an entangled wave packet for two particles that is prepared in such a way that both the *total* momentum and the *difference* in position of the two particles are conserved. This is a tricky point that is often glossed over in nontechnical explanations of the EPR experiment. We know that the position and momentum for each individual particle fails to commute and therefore obeys an uncertainty relation. However, for the type of entangled wave packet they described, the *total* momentum of the system commutes with the *difference* in position; this means that both of these quantities have definite values throughout the experiment and in quantum mechanical terms therefore can be said to have simultaneous reality. The argument cannot go through without this fact in hand.

Now, we let the two particles fly off to a great distance from each other. Let the detectors be staffed by the ubiquitous quantum mechanical experimenters Bob and Alice. (We will meet them again.) To simplify matters we shall assume that Bob, Alice, and the particle source are all at rest with respect to each other. The left particle enters Bob's laboratory at precisely 12:00 noon. There is no way that he could measure both its position and its momentum at precisely the same time, for position and momentum do not commute, but he is free to choose one or the other. If he chooses to measure the position of his particle at precisely 12:00 noon then he automatically knows the position of the other particle at that time, because the difference in those positions remains constant. On the other hand, suppose Bob decides instead to measure the momentum of his particle at precisely 12:00 noon. Then he would automatically know the momentum of the other particle at 12:00 noon because the total momentum of the two particles is known.

*According to quantum mechanic*s, no one can find the position and momentum of the distant particle at 12:00 noon by a single measurement. But it has just been shown that at 12:00 noon Bob could have inferred either the position or momentum of the distant particle at that time. Here's the twist: EPR took it as obvious that because the particles are quite distant from each other, nothing done to one at precisely 12:00 noon could influence the other at exactly that time—because otherwise a causal influence would have had to travel from one to the other at infinite speed. Therefore, neither of the measurements that Bob could have carried out can influence the real physical condition of Alice's particle at 12:00 noon. Therefore, EPR concluded, Alice's particle must (at 12:00 noon) possess exact values of both position and momentum, even though the Heisenberg Uncertainty Relation says that it does not. Quantum mechanics therefore does not give a complete picture of all the elements of physical reality belonging to Alice's particle.

EPR in effect posed a dilemma (although they did not put it in exactly these words): either quantum mechanics is incomplete (in the sense that it cannot tell us about certain elements of physical reality) or else it is nonlocally causal, or both. (A nonlocally causal system is simply one that permits action at a distance.) Since nonlocal causality was, in Einstein's view, an absurdity, quantum mechanics must be incomplete. To put it another way, Einstein took the *apparent* nonlocality of quantum mechanics to be a symptom of its *incompleteness.*

The paper's publication caused a furor in Copenhagen, and Bohr labored mightily to produce a response. A few months later he published a long article with the same title as the EPR paper. It is written in Bohr's usual obscure style, and experts disagree about what he was actually trying to say. Bohr was in full agreement with Einstein on one point: there is no way that what is done to one particle has a direct, instantaneous influence on the other. That would be truly absurd, they thought. Instead, Bohr seemed to be saying that EPR made the mistake of supposing that complementary measurement procedures (measurements of position and momentum) could be taken to apply to the *same* reality—whereas in fact the notion of an independent reality has no experimental meaning, and therefore no meaning at all. Bohr could hardly disagree that entanglement violates classical expectations of separability. However, it is mistaken to try to seek a deeper "explanation" of entanglement, because that would involve trying to impose classical concepts on the quantum level of the world. Quantum mechanics is in essence just a set of recipes for calculating predictions about phenomena that can only be observed with instruments (such as Stern-Gerlach devices, mirrors, photographic plates, etc.) that can be manipulated by humans at the classical (macroscopic) scale. The very concept of an explanation of quantum phenomena just does not make sense, because anything that could count as an explanation would have to be in classical terms. Classical stories about quantum objects always come in complementary pairs (such as the wave and particle pictures); this means that there is no *consistent* picture of a particle as either *just* a wave or *just* a particle, even though there is a consistent *recipe* that tells us when to use either wave or particle concepts in making verifiable physical predictions.

To put it in simpler terms: Bohr's reply to EPR is that they had an unreasonably stringent requirement for completeness; in fact, quantum mechanics is as complete as it can be. There is no more information to be *had* than quantum mechanics can give us.

For a time, Bohr convinced most physicists that they could ignore Einstein's worries about nonseparability, but the questions raised by EPR had not truly been resolved, and a minority of especially thoughtful physicists continued to think about them. With the perspective of 70 years of hindsight, it can be seen that there are no technical errors in EPR's calculations, but their assumption that Bob's measurements could not change the state of Alice's particle is now known to be incorrect. Whether there is some mysterious faster-than-light influence (as in various causal interpretations of quantum mechanics) or whether it "just happens," if a position measurement is made on Bob's particle then the momentum measurement on Alice's particle will *probably* come out differently than it would have had Bob not made his measurement. It would be nearly 30 years before this was demonstrated by J. S. Bell.

Schrödinger (in 1935) also published a detailed study of entanglement in response to the EPR paper. Although his analysis was insightful, he made a mistake that EPR did not make: he predicted that entanglement would diminish as the correlated particles moved away from each other, just as ordinary interactions (such as gravitation or electromagnetic interactions) diminish with distance. Again, it would be 30 years or more before experiment would prove this wrong.

Eventually the EPR paper forced scientists to take seriously the fact that there is still much to be learned about entanglement; in particular, it drew attention to the nonseparability of entangled states. The EPR paper was the stimulus for work by David Bohm and J. S. Bell (to be described later) that led to the direct experimental confirmation of quantum nonseparability. Bohr was right that Einstein had failed to prove the incompleteness of quantum mechanics, because Einstein's high standard of completeness was not something that could reasonably be expected of any theory of quantum phenomena. Quantum mechanics, in a precise technical sense, is as complete as it can be, although this is another key point that would not be demonstrated formally for another 30 years. But while Bohr thus scored some points, round three of the Bohr-Einstein debate must in the end go to Einstein, because he showed that the puzzle of entanglement cannot be made to go away by soothing words. In many respects what EPR said was mistaken, although subtly and instructively so. However, as we enter the twenty-first century, there are few facts about physics more interesting, challenging, and potentially useful than quantum entanglement. Once again, Einstein was right—in the long run.

7

CREATION AND ANNIHILATION

PARTICLE PHYSICS BEFORE WORLD WAR II

By 1930 the existence of only three presumably "elementary" particles had been confirmed (if by an elementary particle is meant something out of which more complex forms of matter, atoms and molecules and radiation fields, can be built): the photon, the electron, and the proton. The existence of a neutral particle in the nucleus that would be roughly the mass of the proton had been hypothesized by Rutherford around 1920, because there had to be some way to account for the extra mass of all nuclei beyond hydrogen. It was known that electrons could be emitted from radioactive nuclei in beta decay, and so it was natural to assume that enough electrons to produce the correct nuclear charge were squeezed into the nucleus. However, Pauli and Heisenberg showed that this would not work. (The electron's de Broglie radius is too large for it to be squeezed into the nucleus; the electron can visit the nucleus, but it cannot live there.) Rutherford's suggestion of a new neutral particle remained the best contender, but it had not yet been proven. Otherwise, physicists had little inkling of the complexity that was about to burst upon them.

The Neutron

The prediction by Dirac and discovery by Anderson of the first antiparticle, the positron, has already been described. In 1932 the British physicist James Chadwick (1891–1974) confirmed the existence of a neutral particle, slightly heavier than the proton, in emissions from beryllium that had been bombarded by alpha particles from radioactive polonium, and he named it the *neutron*.

Chadwick's discovery of the neutron was the key that unlocked the door to modern nuclear physics, with all its enormous potential for both good and harm. Only a few months later, in 1932, Heisenberg used Chadwick's neutron to construct the first quantum mechanical nuclear model. The main mechanism he proposed was an *exchange force* produced by protons and neutrons passing electrons around like basketball players tossing a ball. And in September, 1933, a Hungarian-born émigré named Leo Szilard (1898–1964) was sitting in a London hotel lobby, reading a newspaper report of Lord Rutherford's recent pronouncement that any thought of releasing useable amounts of power by nuclear transformations was "moonshine" (Rhodes 1988, Ch. 1). Szilard had already made a reputation for himself with numerous inventions, including collaboration with Einstein on the invention of a new type of refrigerator, and fundamental theoretical contributions to thermodynamics. He had been an instructor in physics at the University of Berlin. However, when Hitler seized emergency powers in 1933 Szilard fled Europe, as thousands of his fellow Jewish scholars and scientists were dismissed from their posts under the Nazis' new racial laws.

Szilard thought that Rutherford had to be wrong, and it struck him immediately that because the neutron was electrically neutral it could penetrate the nucleus of the atom. If it could then cause the breakdown of the nucleus into smaller fragments *including more neutrons,* a chain reaction could be triggered in which the rate of reactions would grow exponentially, releasing huge quantities of energy. At this time no one had yet demonstrated that a neutron could split the atom as Szilard had imagined, but he knew that it was well within the realm of possibility. Szilard patented his concept for the nuclear chain reaction, and eventually turned it over to the British Admiralty. With his characteristic foresight he realized, years ahead of almost everyone else, that the first applications of nuclear power would be military.

Beta Decay and the Neutrino

Another important property of the neutron is that it when it moves freely outside the nucleus it is unstable, decaying with a half-life of around 15 minutes. Its decay products were apparently a proton and an electron. (Later it would be found that the neutron can also decay into an antiproton and a positron.) This explained the energetic electrons, or beta rays, that were produced in certain kinds of nuclear reactions, and so this process was called *beta decay.* But there was a puzzle, which was noted even before Chadwick's discovery of the neutron: the energy of beta particles follows a continuous spectrum (all values allowed over a certain range), which implied that some energy and momentum was going missing in the reaction. Bohr (as with the old BKS theory) was willing to consider that energy conservation might be violated in beta decay. However, Pauli thought that a less radical explanation was required, and in 1930 he proposed that a very light, neutrally charged particle of spin-1/2 was also emitted during beta decay, with just the right amount of energy to balance the books. Fermi in 1931 jocularly dubbed the new particle the *neutrino* ("little neutron"), and in 1934 he produced an elegant quantum mechanical description of beta decay. Part of Fermi's theory stated that the electron and the neutrino produced by beta decay shared the decay energy randomly, which accounted nicely for the continuous spectrum of the beta particle. Physicists found it hard to doubt the neutrino's existence, because they did not want to give up energy conservation. However, the neutrino is very hard to detect, because its probability of interaction with ordinary matter is so low. It was finally detected in 1956 in delicate experiments performed by Frederick Reines (1918–1998), Clyde Cowan (1919–1974), and others.

Yukawa and the Strong Force

Before the neutron was discovered, physicists supposed that protons in the nucleus were bound together electrostatically by electrons. Once it was realized that the nucleus was a combination of positively charged and neutral objects, there was no choice but to assume the existence of a powerful nuclear force that bound protons and neutrons together despite the electrostatic repulsion of the protons. This force would be almost like a glue that bonds very strongly when neutrons and protons are nearly touching, but that falls off rapidly (indeed, exponentially) with distance.

The hardworking Japanese physicist Hideki Yukawa (1907–1981) made the next step. He realized that the attempts to model the nuclear interaction on beta decay had failed, and decided that he should consider the possibility that a new kind of field, not previously studied, is responsible for the nuclear force. Just as electromagnetic forces depend on the interchange of photons, there had to be a particle that was tossed around like a ball by protons and neutrons and would account for their strong short-range interaction. In conversation, the senior Japanese physicist Yoshio Nishina (1890–1951) made the suggestion that such a particle might obey Bose-Einstein statistics. At first Yukawa thought that this was too radical, but eventually he realized that he had to let the necessary characteristics of the nuclear force field tell him what sort of particle it used. The key was that the range of a force should be inversely proportional to the mass of the particle that carries the force. Physicists were learning that the vacuum is like a bank from which the currency of energy can be borrowed temporarily. Enough energy to create the particle can be borrowed from the vacuum so long as it is paid back in a time interval allowed by the uncertainty principle for time and energy; the known range of the force thus determines the time the particle can exist, and thus its energy. Yukawa estimated the mass of the new boson to be about 200 times the mass of the electron, or around 100 to 150 MeV. (Particle mass-energies and the energies of nuclear and atomic processes are commonly measured in *electron volts,* the energy acquired by an electron when it has accelerated through a potential difference of one volt; an MeV—usually pronounced "em-ee-vee"—is a million electron volts.) Yukawa thus predicted the existence of a particle that by the late 1930s was being called the mesotron or meson, the "intermediate particle," since it was intermediate in mass between the electron and the proton.

"Who Ordered *That?*"

In 1937 Carl Anderson and Seth Neddermeyer (1907–1988) (who would play a key role in the design of the atomic bomb at Los Alamos) discovered in cosmic ray showers a particle that for a while was suspected to be the carrier of the nuclear force particle that Yukawa had predicted. Some background on cosmic rays will be helpful: *Primary cosmic rays* are extremely energetic particles (probably protons) of unclear source, entering the Earth's atmosphere from outer space. Fortunately, primary cosmic rays never reach the ground since they collide with the nuclei of oxygen or nitrogen atoms in the atmosphere. (Astronauts in orbit sometimes see bright flashes caused by primary cosmic rays blowing up nuclei inside their eyeballs.) Primary ray collisions produce showers of *secondary cosmic rays,* and thousands of these elusive particles sleet through our bodies every second.

Anderson spotted his new particle when he and Neddermeyer were studying secondary cosmic rays. It had roughly the right mass to be Yukawa's particle. The new particle came in a negative and positive version, presumably antiparticles of each other, and it was unstable, with a half-life of about two microseconds. By about 1946 it was realized that Anderson's "meson" was a very poor candidate as carrier of the strong force, because its negative version would pass through a nucleus with hardly any probability of interacting at all. It is also a spin-1/2 fermion, which (as Nishina had suspected) should rule it out from the beginning as a possible carrier of a force field, and its lifetime was about 100 times too long to be Yukawa's particle.

To make a very complicated story short, by about 1950 further studies had shown that there are *two* kinds of "meson" in cosmic ray showers. In addition to Anderson's particle of 1937, another meson (this time the quotes can be removed) was discovered by Cecil Powell (1903–1969) and collaborators by exposing photographic emulsions to cosmic rays at high altitude. Powell's meson, which became known as the pi-meson or pion, is in fact the nuclear force carrier that Yukawa had predicted, and he was duly awarded the Nobel Prize in 1949. The pion is a spin-0 boson that comes in three versions, with positive, negative, or zero electric charge. Anderson's particle of 1937 became known as the mu-meson or muon. The muon behaves exactly like an electron, only it is heavier and unstable. (The term "meson" is now reserved strictly for bosons that carry force fields.) Primary cosmic rays collide with the nuclei of atoms high in the atmosphere and release showers of pions, which in turn quickly decay into muons, and it is largely muons that are detected at ground level.

The discovery of the muon was completely unexpected, and it prompted an incredulous quip by American physicist Isidore I. Rabi (1898–1988): "Who ordered *that?*" (Kragh 1999, p. 204). Although there is now a place for the muon in the modern scheme of particles, there still is no satisfactory answer to Rabi's question.

Four Forces

By the late 1930s it was accepted that there are four forces in nature: gravitation (described by Einstein's general theory of relativity), electromagnetism, Yukawa's strong nuclear force, and the mysterious weak force or weak interaction that is responsible for beta decay. They vary greatly in strength: the strong force is 10^{38} times as strong as gravitation (the weakest force) but it falls off exponentially and is only important at distances of 10^{-13} cm. or less. Gravitation becomes the dominant force for very large masses and over large distances, because all of the mass and energy in the universe partakes in the gravitational interaction. Recent explorations in quantum gravity suggest that gravitation may again become dominant at very small scales, much smaller than any scale we can presently probe experimentally. It would become the dream of physicists (still not fully realized) to unify all these forces into one under the guidance of quantum mechanics. Before World War II there were provisional but still useful and illuminating theories of the weak force (due to Fermi) and the strong force (due to Yukawa). However, while efforts continued to find the right equations for meson fields, most quantum physicists concentrated their efforts on understanding the electromagnetic field—a task that got off to a quick start, and then ran into some surprisingly difficult challenges.

QUANTUM FIELDS

First Attempts to Quantize the Electromagnetic Field

So much had been accomplished in the few years between Heisenberg's vacation on Heligoland and the publication of the Dirac Equation that the quantum physicists (most of whom were very young) were brimming with *Hochmut* (pride) by the late 1920s. In a perhaps forgivable moment of overconfidence, even the judicious Max Born said, "Physics as we know it will be over in six months." As in the 1890s, there were just a few more loose ends to tidy up—such as the quantum mechanics of the electromagnetic field.

The first step toward quantum field theory was relativistic quantum mechanics, which began in the late 1920s almost as soon as Schrödinger's Equation appeared. The Dirac Equation has already been mentioned; it is a relativistic wave equation for the electron and other spin-1/2 particles. The Klein-Gordon Equation was also arrived at by several authors from the mid-1920s onward. (Richard Feynman was to derive it as a teenager, just for fun.) It is the wave equation for spin-0 bosons.

Dirac published the first paper on what he called *quantum electrodynamics* in 1928, and he was soon joined in developing the new theory by Heisenberg, Pauli, Jordan, and others. Quantum electrodynamics, or QED as it is often now called, is the quantum theory of the electromagnetic field, written in a way that respects the constraints of Einstein's special theory of relativity.

A distinguishing feature of quantum field theory is the nonconservation of particle number. Particles are created and destroyed—or, more precisely, particles can transform into other particles, and these transformations can involve splitting and recombination in such a way that the total number of particles is a variable quantity. This fact emerged from Dirac's Equation, which showed that positrons can be bumped out of their negative energy states in the Dirac Sea by a passing gamma ray, a process that can also be described as the splitting of a gamma photon into a positron and electron. Quantum field theory describes an unending dance of the creation of pairs of matter particles (fermions) and their antiparticles, and then the annihilation of the pair with the release of field quanta (always bosons). Mathematically this process is represented by *creation* and *annihilation* (or *destruction*) operators, which, like all quantum operators, are defined by commutation relations. A creation operator increases the number of particles by 1, and an annihilation operator decreases the number of particles by 1. It is often useful to think of a quantum system as a collection of harmonic oscillators (an idea that goes back to Planck); in this case, creation and annihilation operators are called *raising* and *lowering* operators, because the creation of a particle is equivalent to raising or lowering the energy state of a virtual oscillator. In simple terms, quantum mechanics becomes quantum field theory when the creation and annihilation of particles is taken into account. In the early years of quantum field theory, the process of turning field variables into creation and annihilation operators was called *second quantization.* The move from ordinary quantization to second quantization is a move from quantum mechanics of a single particle to that of many particles. The term "second quantization" is now largely a historical curiosity, but the procedure is not.

Physically, the creation and annihilation of particles is a consequence of Einstein's equivalence of mass and energy in combination with the laws of quantum mechanics. If a particle possesses a certain amount of energy in addition to its rest mass, it can emit a particle or particles with that amount of energy, so long as all conservation laws are respected. For instance, the electrical charges and spins have to add up. In quantum mechanics, if something *can* happen (according to conservation laws) then there is an amplitude for it to happen, and therefore a probability (sometimes vanishingly small, sometimes not) that it *will* happen. Hence, so long as there is enough energy in a system to allow for the creation of particles or their annihilation and transformation into other particles, then it will sooner or later occur.

One of the first successes of early QED in the hands of Dirac and others was that it gave methods for calculating the amplitudes for spontaneous and induced emission that Einstein had identified in 1917, and thus allowed for the most general derivation of Planck's radiation law yet found. Early QED was most successful in dealing with *free fields,* fields that are not in the presence of matter; describing how the electromagnetic field interacts with matter proved to be a much harder task.

Early in the 1930s Dirac predicted another bizarre quantum phenomenon, *vacuum polarization.* It is a consequence of pair creation. Consider an electron—a *bare electron*—sitting in the vacuum. Its charge will *induce* the creation of *virtual* pairs of electrons and positrons in the vacuum around it, for the vacuum itself is a polarizable medium. Virtual particles are those that exist for such a short period of time that the time-energy uncertainty principle forbids them to be observed; they would later be called *vacuum fluctuations.* The charge induced out of the vacuum will partially shield the electrical charge of the electron; this means that the charge that we actually observe is not the bare charge of the electron at all, but a net or effective charge.

Another early implication of quantum electrodynamics is that the energy hidden in the vacuum may be vast, many powers of 10 greater than the energy density even of nuclear matter. This has prompted unsettling speculations that the whole physical world we know is merely a higher-order perturbation, like foam on the surface of the sea, on a complex vacuum the structure of which we glimpse only dimly.

Infinitely Wrong

By the early 1930s QED ran headlong into the brick wall of the infinities. Calculations that looked perfectly valid seemed to predict that a host of electrodynamic quantities that have definite, measurable values are *divergent*, which means that they blow up to infinity. An important case is the self-energy of the electron, the energy that it possesses as a result of its electromagnetic interactions with its own field. A simple example of a divergent function is the classical formula for the electrical potential of a point charge, $V(r) = 1/r$, which becomes infinite when $r = 0$. This problem had never been entirely resolved within classical electrodynamics, but it was assumed that the electron had a finite radius and so no one worried about it very much. In quantum electrodynamics there was no getting away from the infinities, however, despite valiant and ingenious efforts by virtually everyone who worked on quantum mechanics at the time. In simple terms the problem is that there are an infinite number of ways that an electron can interact with itself. Similar infinites also appeared in the early field theories of mesons that were being written by Yukawa and several others.

Faced with the divergences, many of the more senior quantum theorists (such as Bohr, Dirac, and Heisenberg) were by the late 1930s (with World War II looming) becoming convinced that another large conceptual revolution was inevitable, one so radical that it would require the abandonment of spacetime itself as a fundamental concept. Heisenberg in 1938 suggested that it would be necessary to replace the spacetime continuum with a discrete structure based on a fundamental quantum of length, just as Planck had based quantum mechanics on a fundamental, indivisible quantum of action. But the younger generation of quantum physicists in the late 1930s—notably Hans Bethe (1906–2005), Victor Weisskopf (1908–2002), Yukawa, and Yukawa's good friend Sinichiro Tomonaga (1906–1979)—simply kept calculating, and eventually a surprisingly effective *conservative* solution began to emerge.

Renormalization: A Brilliant Stop-gap?

Progress in fundamental physics slowed during World War II, but immediately after the war physicists enjoyed unprecedented prestige (and funding) because of the success of the atomic bomb. The first item of unfinished business was the puzzle of the infinities of quantum electrodynamics.

In quantum field theory quantities have to be calculated using a *perturbation* series, because it is impossible to solve the equations exactly. Anyone who has taken first-year calculus is familiar with the idea of a Taylor series, in which a function is approximated by finding its value at a point and then writing a certain kind of infinite series. A perturbation series in quantum electrodynamics starts with a zeroeth term, which will be in terms of known quantities, and then expands around this term in a power series, usually in powers of Sommerfeld's mysterious fine structure constant (about 1/137). The first term, the free field solution, is exact but not all that interesting since it does not describe any particle interactions. The second term, although not quite exact, corresponds fairly well to a lot of observable physics. The third and higher terms diverge. For example, calculations of the self-energy of the electron come out infinite, even though it has a perfectly definite mass.

Mathematically the deep reason for the infinities is that space and time are assumed in quantum field theory to be continuous, so that at finer and finer scales there are more and more ways for the system to behave, all of which have to be taken account of. It is reminiscent of the ultraviolet catastrophe that plagued blackbody theory before Planck. This is why the notion of making space and time discrete occurred naturally to Heisenberg. But the theorists in the 1940s found that there is a less radical approach: there is a systematic way of replacing infinite quantities whenever they show up with their finite, observed values. This is called *renormalization:* erase an infinity whenever it occurs and replace it with a finite value.

Two brilliant young American physicists emerged as leaders in the battle to tame the infinities: Julian Schwinger (1918–1994), and Richard P. Feynman (1918–1988). The Japanese physicist Sinichiro Tomonaga had also developed a renormalized version of quantum electrodynamics in 1943, but because of the war this did not become known outside Japan until much later. Thus Feynman and Schwinger were repeating Tomonaga's work, although from different mathematical perspectives, and all three shared the Nobel Prize in 1965.

With renormalization, QED emerged as a tool that could predict virtually all measurable electrodynamics quantities to precisions of 10 decimal places or better. An important example was the slight shift in the spectral lines of hydrogen discovered by the American Willis Lamb (1913–). Electromagnetic attraction and repulsion is explained in terms of the exchange of photons between charged particles. QED is sometimes said to be the most accurate physical theory ever written, and it became the model for quantum field theories that would be developed in the future. It was, however, not what anyone had expected in the mid-1930s, but a compromise between quantum and classical methods that has been found to work surprisingly well. It is a compromise because quantum fields act out their probabilistic antics against a classical Minkowski space background. If Heisenberg's advice to quantize spacetime had been adopted, then it would have been necessary to reconstruct relativity theory. No one was ready to do that in the 1940s, and few are today.

Local Quantum Field Theory

The field theory that had emerged through the work of many physicists from the late 1920s to about 1950 is often called *local* quantum field theory, because it contains special "patches" to ensure that it is consistent with relativity.

The most important of these patches is a rule called "microcausality" or "local commutativity." This was first introduced by Bohr and Leon Rosenfeld (1904–1974) in the early 1930s and used by Pauli in his axiomatic construction of quantum field theory around 1940 when he proved the spin-statistics theorem. Microcausality states that all quantum observables acting at a space-like separation must commute *even if* they are observables such as position and momentum that would normally not commute if applied to the same particle. (To say that two observations are at a space-like separation is to say that they are outside each other's light cones, so that they are acting at points that could only be directly connected by a signal moving faster than light. See the Minkowski diagram in Chapter 2, Figure 2.2.) Microcausality was introduced into quantum field theory in order to ensure that its predictions would not conflict with relativity. Otherwise, or so it seemed to physicists in the 1930s, if commutativity fails at space-like separations then it would be possible to do a series of measurements on distant but entangled particles in such a way that faster-than-light signaling would be possible. The logic of microcausality is subtle, however. Given the pervasive nature of nonlocality in quantum physics, is it really safe to assume that it must never conflict with relativity? This question remains open.

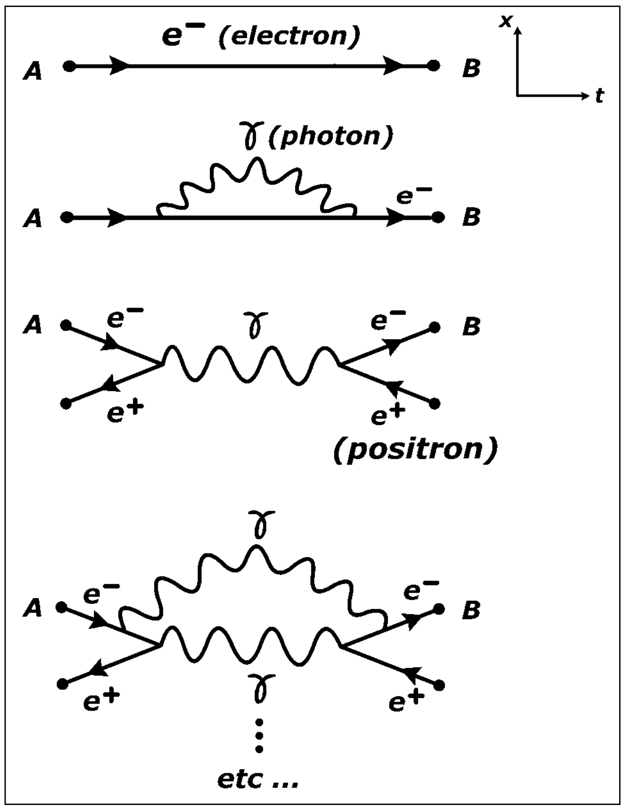

Figure 7.1: Feynman Diagrams. The probability amplitude for the electron to get from A to B has a contribution from every possible Feynman diagram for that process. Illustration by Kevin deLaplante.

Feynman's Diagrams and the Path-Integral Formulation of Quantum Mechanics

Richard Feynman had unusual mathematical skill, but his greatest virtues as a physicist were his physical intuition and his delight in finding simple and elegant approaches to problems that had baffled everyone else. While most physicists were dazzled by

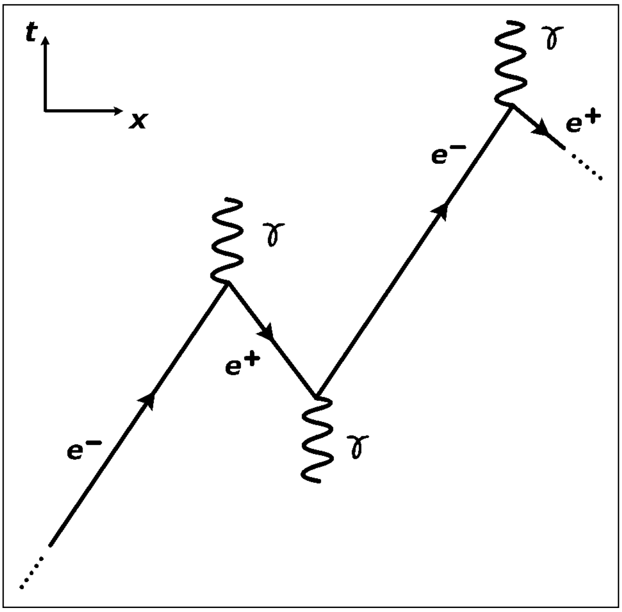

Figure 7.2: There Is Only One Electron in the Universe! A positron going forward in time is just an electron going back in time. Wheeler suggested that all electrons and positrons are the same particle, pinballing off photons backwards and forwards through time. Illustration by Kevin deLaplante.

Schwinger's mathematical virtuosity, Feynman found another way of thinking about QED that is probably as visualizable as any theory of quantum fields is ever going to be, and relatively easy to calculate with. The key idea was to represent every possible particle interaction with a spacetime diagram. Photon wordlines (their trajectories through spacetime) are represented by wavy lines, while particles such as electrons and positrons are straight lines or curves.

Each diagram represents a possible way that the system can evolve and thus represents a term in the perturbation series for calculating the transitions amplitudes for various processes. Though some ways turn out to be much more probable than others, they all have to be added in to get the right transition amplitudes. Feynman's methods are relatively easy to use compared to those of Schwinger; in fact, there was at first a suspicion that they were too simple to be correct. However, the British-born physicist Freeman Dyson (1923–) showed the mathematical equivalence of Feynman's diagrammatic approach with the operator approach of Tomonoga and Schwinger, and they were soon adopted by physicists and applied widely in particle theory.

Feynman diagrams are an application of an elegant rewriting of quantum theory by Feynman, who claimed that he could not understand standard quantum theory and had to recreate it on his own. Schrödinger's Equation plays a secondary role in Feynman's version of quantum mechanics; the principal actor is the probability amplitude. Problems are solved by adding up the amplitudes for all possible routes that a system can take. This results in a *path integral,* which gives the amplitude or *propagator* for a process (such as a particle decay or interaction) to occur.

Feynman's path integral interpretation was a natural development of work he did in the 1940s in collaboration with John A. Wheeler, his thesis advisor. In 1941 Wheeler arrived at an apparently crazy idea: there is only one electron in the universe! This is natural when thinking about what Dirac's pair creation looks like in spacetime. Picture an electron going forward in time: it scatters off a photon and goes back in time, where to ordinary human observers it looks like a positron going forward in time. Then it scatters off another photon, goes forward again for a while, then pinballs off another photon, and so forth, until its worldline has threaded itself through all of spacetime.

This perhaps fanciful vision led Wheeler and Feynman to the idea of taking seriously the apparently far-fetched notion of influences from the future. They revived an old notion due to the great German mathematician Karl Friedrich Gauss (1777–1855), who unsuccessfully tried to construct a theory of electromagnetism based on action at a distance, unmediated by any field—the very thing that Newton and Einstein had considered to be absurd. Wheeler and Feynman showed that it is mathematically possible to construct a version of electrodynamics that is entirely equivalent to ordinary classical electrodynamics, but in which the net force on each charge is a combination of *retarded* and *advanced* effects. Retarded forces are ordinary forces that propagate forward in time; advanced forces are direct influences of the future upon the past. The Feynman-Wheeler action at a distance theory was classical, but it influenced Feynman's spacetime picture of quantum mechanics, in which the amplitude for any process must include contributions from all possible routes through spacetime including the far future. The fact that in our ordinary experience influences from the past seem to dominate is purely a statistical matter.

Figure 7.3: Richard P. Feynman. AIP Emilio Segre Visual Archives, Physics Today Collection.

Summing Up Quantum Field Theory

The complex history of quantum field theory can be summarized as follows: From about 1928 to shortly after World War II, quantum mechanics grew into quantum electrodynamics (the relativistic quantum theory of the electromagnetic field). This in turn was generalized into local quantum field theory, a very powerful and effective approach to the physics of particles and fields, which began to be applied (with varying degrees of success) to the other three forces in nature.

As the predictive power of the theory increased, the process of abstraction that had begun in 1913 with Bohr's unobservable stationary states reached new heights. Even Feynman's version of quantum electrodynamics is of far greater mathematical complexity and abstraction than the wave mechanics of Schrödinger. To the extent that quantum field and particle processes can be visualized at all, it is only by means of analogies that can never be taken too seriously. This is perhaps another reason why Einstein and several of the other aging founders of quantum physics would have little to do with the new quantum field theories. Probably most particle and quantum theorists now would consider accurate visualizability (and possibly even mathematical consistency) to be a luxury or even a forlorn hope; they are happy if their theories can yield calculable and testable predictions, and even that is no longer a given in modern particle theory.

Some theorists today believe that any quantum theories of the future will be modeled on the highly successful local quantum field theories that emerged in the 1940s and 1950s. Others believe that local quantum field theory is only provisional, like the Old Quantum Theory. The problem is not only that there are jobs that today's quantum field theories cannot yet do, such as predict particle masses. There are at least two deeper theoretical worries that some authors believe cannot be swept under the carpet.

First, there is the need for renormalization. Some physicists believe that so long as renormalization can be done in a consistent way, there is nothing wrong with it at all. It has been possible to have much greater confidence in the mathematical soundness of renormalization after its mathematical basis was clarified by Leo P. Kadanoff (1937–) and Kenneth Wilson (1936–). Others believe that renormalization is merely an ingenious stopgap, and that there should eventually be a way of calculating a quantity such as the self-energy of the electron without having to, in effect, rub out the wrong answer like a schoolboy and write in the correct answer by hand. Feynman himself described renormalization as a "shell game" that he suspected is probably "not mathematically legitimate" (1985, p. 128). Should this shell game be okay just because (for certain fields) it is possible to always get away with it? Also, the fact that the gravitational field has so far proven to be nonrenormalizable (which will be discussed in more detail in Chapter 12) points to the need, say some critics, for not being satisfied with renormalization as the best solution to the infinities in any sort of field theory. Of course, another possibility, although it is crazy, is that the calculations are correct and the self-energy of the electron really *is* infinite because of its potential interactions with all other particles in the universe, except that we can only detect a finite part of its energy experimentally. If anything like this is true then there would never be a hope of writing a truly exact theory of any quantum field, although there would be many ways of finding useful approximations.

Second, the fact that quantum field theory is written on a Minkowski background worries many physicists. In his very first paper on quantum theory in 1905, Einstein warned that Maxwell's Equations might be merely an approximation that is valid only when the quantum graininess of the electromagnetic field can be ignored. But Einstein then set out to construct his theory of special relativity based on the assumption that Lorentz invariance is *exact*. As far back as the 1930s Schrödinger speculated that it may be necessary to quantize the Lorentz transformations (though he did not say how this could be done). The field theorists of the 1940s and 1950s displayed great technical brilliance in showing that it possible to write a powerful and accurate quantum field theory that is Lorentz-invariant and that respects the causal structure of the Minkowski background spacetime. It is ironic that the younger physicists in the late 1930s and 1940s were mostly the ones to perfect a highly technical *conservative* solution to the problem of the infinities—a solution that was conservative in that it respected Einstein's laws of relativistic spacetime. It was the older founding generation of quantum physicists who thought that far more radical approaches were needed to cope with the infinities of quantum field theory, such as quantizing space and time, or doing away with space and time entirely.

Many physicists today would prefer to not question the framework of local quantum field theory. Others suspect that in the long run Heisenberg will again turn out to have been right, and local quantum field theory itself will be revealed as only an approximation (an example of what is called an *effective field theory*), which breaks down at some high energy level where the metric of spacetime can no longer be treated as a classical, continuous structure. Whether local quantum field theory can continue to serve as the basic framework for quantum mechanics, or whether it has to be replaced by something else, is one of the most pressing theoretical questions that physics faces as it moves into the twenty-first century.

8

QUANTUM MECHANICS GOES TO WORK

This chapter will cover several of the important applications of quantum mechanics that grew out of the theoretical breakthroughs of the 1920s. Some of these applications were of great scientific interest in their own right, and some changed the very world we live in.

NUCLEAR PHYSICS

If the 1920s were the years in which quantum theory leaped ahead, the 1930s were the great years of nuclear physics, when, sparked by the discovery of the neutron, and using the tools of quantum theory, it went from Rutherfordian table-top experiments to the discovery of nuclear fission.

Gamow Tunnels Out of the Nucleus

The flamboyant Russian physicist George Gamow (1904–1968) engineered one of the first great successes of the new quantum mechanics when he provided a partial explanation for nuclear decay. (Gamow would later make important contributions to cosmology as well, when he was one of the first to predict, in 1948, the existence of the cosmic microwave background radiation—the faded “flash” of the Big Bang.)

Gamow visited Copenhagen in 1928, and Bohr, with characteristic generosity, made it possible for Gamow to stay in Copenhagen for a year. In 1928 Gamow devised a very simple but surprisingly accurate model that described alpha decay using the new tools of wave mechanics. Gamow knew that the alpha particle was somehow trapped within the nucleus by a very short-range force that overcomes the strong electrostatic repulsion of the protons, because otherwise the electrical repulsion of the positively charged protons would blow the nucleus apart instantly. This was six or seven years before the work of Yukawa, but Gamow did not have to know the details of how the mysterious

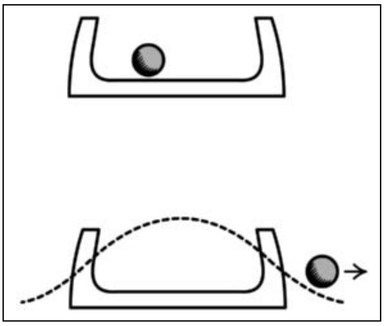

Figure 8.1: Barrier Penetration. A classical marble in a dog bowl cannot spontaneously jump out (top). However, a quantum marble can tunnel out of the bowl if the tail of its wave packet extends outside the bowl (bottom). Illustration by Kevin deLaplante.

nuclear force works in order to provide a basic quantum mechanical explanation of alpha decay.

What Gamow did would now be considered a very typical application of wave mechanics, although in 1928 he was breaking new ground. The powerful but short-range force that holds the nucleus together defines a potential barrier, a hill of energy that the alpha particle has to either get over or go through. The alpha is therefore like a marble in a dog dish. Classically, if the alpha does not have enough energy to roll out of the dish, it will stay there forever. From the viewpoint of wave mechanics, the reason that a marble cannot roll out of a dog bowl is just that its wave function does not have a tail that extends outside the bowl, and this is because the de Broglie wavelength of an ordinary marble is much smaller than the marble itself.

Gamow showed (by means of skillful approximations) that the tail of the alpha's wave packet extends outside the nucleus, so that there is a probability for the alpha to tunnel through the barrier. He thus defined the phenomenon of *barrier penetration,* which was soon found to occur in many quantum mechanical contexts. It is another example of quantum mechanics allowing something to happen that would be impossible classically. Once the alpha gets outside, electromagnetic forces take over, and the strong electrostatic repulsion between the positive alpha and the nucleus propels the alpha away from the nucleus at very high energies. Strictly speaking, therefore, the energy of radioactive alpha decay that the Curies puzzled over is electromagnetic in nature, not nuclear. The mysterious new force that held the protons together, though, would take a long time to understand.

The phenomenon of barrier penetration by quantum tunneling is very well described by wave mechanics, but it is still not entirely clear what is going on. How long, for instance, does it take the alpha particle to penetrate the potential barrier? Recently, Günter Nimtz of the University of Cologne and others have provided evidence that in some cases wave packets can tunnel at speeds faster than light, but whether this can be used for the transmission of information remains very controversial.

Splitting the Nucleus . . .

Quantum mechanics played a crucial role in understanding the nature of nuclear *fission,* which is the splitting of heavy nuclei into lighter nuclei of intermediate atomic weight. This book is not the place to tell the whole story of nuclear energy and the creation of the atomic bomb, which has been told many times before. (See Rhodes 1988 for a good introduction.) The aim here is describe certain ways in which quantum mechanics made atomic energy possible.

In 1933 German supremacy in science came to an abrupt end with the Nazi expulsion of Jewish scholars and scientists and any others, especially intellectuals, who might be critics of the regime. This coup effectively lobotomized Germany and guaranteed that Hitler would lose the world war he was planning. Only a few top-rank scientists (including Heisenberg) stayed behind, and even Heisenberg was branded by the Nazis as a "white Jew" because of his long association with Jewish scientists.

Despite the upheavals of the time, physicists continued to probe the nucleus with the new tool that Chadwick had provided. Important work was done by Irène Joliot-Curie (1897–1956) (daughter of Marie Curie) and her husband Frédéric Joliot-Curie (1900–1958), who in 1934 were the first to transmute elements artificially (using neutron bombardment). An Italian group led by Fermi in the 1930s also transmuted elements, demonstrated that slow ("thermal") neutrons could be easily captured by other nuclei, and discovered *neutron activation,* the process in which neutrons transform other elements into radioactive forms. At first most physicists working with neutrons thought that they were going to create transuranic elements, elements heavier than uranium, by reacting uranium nuclei with neutrons. The first person to realize that the uranium nucleus might be splitting into smaller fragments was Ida Noddack (1896–1978), who in 1934 published an article charging that Fermi had misinterpreted his own results and that his uranium nuclei were splitting into fragments of intermediate size. Noddack's ideas were ignored for decades, but it is clear now that she was essentially right.

One of the most active research groups was in Berlin, led by Otto Hahn (1879–1968) and Lise Meitner (1878–1968), one of the few women physicists in Europe before World War II. Meitner's particular skill seems to have been finding theoretical interpretations of Hahn's experiments. Hahn and Meitner brought into their group the chemist Fritz Strassman (1902–1980), because it was apparent that an analytical chemist were needed to identify the nuclei they were producing. Meitner, who was of Jewish heritage, lost her position and was forced to flee Germany in 1938. Hahn was unable or unwilling to protect Meitner. She ended up in Sweden, where she kept in correspondence with Hahn, who continued to seek her help in understanding the bizarre results that he and Strassman were getting.

Meitner realized that Hahn was observing the fission of the uranium nucleus, but a theoretical picture was needed. In conversations with her nephew Otto Frisch (1904–1979), who had worked in Copenhagen with Bohr on nuclear physics, she found a way to apply a model of the nucleus that had been proposed by Gamow. Not enough was known about the strong force at that time to create a fully quantum mechanical treatment of the nucleus. Gamow's model, which had been further developed by Bohr and others, was an ingenious compromise between classical and quantum concepts. It treated the atomic nucleus as if it were a drop of liquid held together by surface tension like a drop of water. The "molecules" of the liquid were the nucleons (neutrons and protons), and the surface tension was supplied by the strong force. The liquid drop model is a good example of a semiclassical model, a blend of quantum and

Figure 8.2: Lise Meitner. AIP Emilio Segre Visual Archives.

classical elements that allows the practical description of something for which a full quantum theory is not yet available or too complicated to be practical.

Meitner and Frisch calculated that a uranium nucleus, destabilized by a slow neutron, would split into two smaller nuclei of roughly the same mass, releasing the extraordinary energy of 200 MeV in the process. Most important, a few more neutrons would be released, opening the way to Szilard's nuclear chain reaction. Frisch borrowed the term "fission" from cell biology to describe this splitting process. Hahn would later receive a Nobel Prize for his part in the discovery of nuclear fission; Meitner did not.

Quantum Mechanics Goes to War

In 1939 Leo Szilard and Eugene Wigner (1902–1995) drafted a letter for Einstein to sign warning President Roosevelt of the risk that the Nazis might develop an atomic bomb. The fact that Heisenberg had remained in Germany (for reasons that he never made entirely clear) and ended up heading the German bomb project was one of the factors that made Szilard take the threat of a German bomb seriously. In 1940 Frisch and Rudolph Peierls (1907–1995) described a method by which a workable atomic bomb could be created. They were the first to show conclusively that only a few kilograms of enriched uranium brought to a critical mass would be sufficient to destroy a city.

Roosevelt ordered the creation of a group to study the new threat, and this evolved into the Manhattan Project, an industrial and scientific effort on a completely unprecedented scale led by many of the physicists who had helped to create quantum theory, including Fermi and Bohr himself. It brought about the construction of what would now be considered fission bombs of very modest yield, and the abrupt end of World War II with the atomic bombing of two Japanese cities, Hiroshima and Nagasaki.

. . . And Putting Nuclei Together Again

Fission is the process in which heavy nuclei split; *fusion* is the process in which light nuclei such as hydrogen, helium, and lithium fuse together to form heavier nuclei. In fusion the source of the energy released is the strong nuclear force. If nucleons are pushed close enough together they attract each other, releasing several million electron volts of energy per nucleon. Just as electrons in orbit around the nucleus release photons (in the few eV range) when they undergo the atomic transitions first identified by Bohr, nucleons release photons in the gamma range, the most energetic electromagnetic radiation known. Maria Goeppert-Mayer (1906–1972) and Hans Jensen (1907–1973) won a Nobel Prize for their work in describing the energy levels within the nucleus in terms of their quantum mechanical shell model of the nucleus and thereby explaining the gamma-ray spectra of various radioactive nuclei.

Fritz Houtermans (1903–1966) and Robert Atkinson (1898–1982) in 1929 noted that the atomic weights of intermediate-weight nuclei such as carbon were slightly less than the sum of the weights of lighter nuclei. They used Einstein's equivalence between mass and energy to estimate the energy equivalent of the mass lost when light nuclei fuse together. On this basis they argued that the joining together, or fusion, of light nuclei could account for the energy production of stars, which were known to be composed mostly of light elements such as hydrogen and helium. The combination of the intense gravitation of a star and its high internal temperature would allow light nuclei to overcome their electrical repulsion and get close enough together that the nuclear forces could cause them to fuse, releasing very surprising amounts of energy. The concept of fusion eventually led to an explanation for the nucleosynthesis in stars of all elements heavier than hydrogen.

The challenge of finding safe and efficient sources of energy is becoming acute because of impending fossil fuel depletion and climate change caused by carbon dioxide emissions. Unfortunately, it has only been possible so far to release large amounts of fusion energy by means of thermonuclear bombs. Fission energy cannot be the long-term solution for humanity's energy needs because it produces radioactive waste and because there is only so much uranium and thorium to go around. However, some scientists argue that fission should allow humanity to buy time until we figure out how to control fusion, which would be an ideal source of energy because the light elements that fuel it are very abundant and it produces little if any radioactive waste.

Most approaches to nuclear fusion have treated it as a problem in applied *plasma physics*. Plasmas are gasses so hot that most or all of their molecules are ionized; they are electrically conductive and display complex collective, fluid-like behavior that is still poorly understood. Plasma is sometimes called the fourth state of matter, along with gasses, liquids, and solids. The sun is composed of plasma mostly made of hydrogen, and in fact, the larger part of matter in the universe is in the plasma state. (A candle flame is a familiar example of a plasma.) Because plasmas are electrified they can be manipulated by electromagnetic fields, although with difficulty. The main approach in controlled thermonuclear fusion research since the late 1940s has been to create a "magnetic bottle" in which a hot plasma of light elements could be trapped long enough to allow it to fuse. Since we don't know (yet!) how to produce a gravitational field as intense as that of the sun, the plasma has to be extremely hot, much hotter than the plasma in the interior of the sun, in order to provide enough kinetic energy for the nucleons to be forced together and fuse.

This has proven to be an extraordinarily difficult technical problem for several reasons, especially because of the many kinds of instability that are endemic to plasmas. It is ironic to compare the slow progress made in fusion research with the explosive growth of semiconductor electronics. In the early 1950s no one had any idea how fast and compact computers were soon to become; popular science magazines depicted the personal computer of half a century later as clanking behemoths that would fill a room. On the other hand, confident predictions were made that nuclear fission would soon provide "meterless" power (that is, electricity so cheap to produce that it could be given away) and controlled fusion would not be far behind.

Currently most of the world's fusion research resources are concentrated into ITER, the International Thermonuclear Experimental Reactor to be built in France. It will use a toroidal (donut-shaped) magnetic bottle design, called a *tokamak* due to the Russian physicist and peace activist Andrei Sakharov (1921–1989). ITER is not expected to reach breakeven before about 2015, and even then it will be a long way from providing power to the grid. (Breakeven is the point at which the reactor generates more energy than it consumes.) So far, plasma physics is one of the few areas of physics where quantum mechanics is relatively unimportant, since most plasmas are so hot. But there is a lot of room for new approaches in the search for nuclear fusion, and perhaps the quantum can someday play a role after all.

CHEMISTRY BECOMES A SCIENCE

One of the most immediate and dramatic application of quantum mechanics from the late 1920s onward was to chemistry. Pauli's Exclusion Principle, the theory of spin, and the Schrödinger Equation provided the tools to define the structure of electron orbitals and the Periodic Table. The new challenge was to understand the nature of the chemical bond itself from the viewpoint of quantum mechanics. Before quantum mechanics, chemistry was largely an empirical science, which means that it amounted to trying various combinations of compounds to see what would happen. There was little or no principled understanding of why certain atoms would bond and in what way. As soon as Schrödinger's wave mechanics had been formulated, physicists and chemists began to apply it to understanding the chemical bond.

The first quantum theory of the covalent bond was developed by Fritz London (1900–1954) and Walter Heitler (1904–1981) in 1927. Their work had a strong influence on American chemists, notably Linus Pauling (1901–1994), who became the dominant figure in theoretical chemistry from about 1930 to 1950. Pauling (who would win Nobel Prizes in both Chemistry and Peace) greatly advanced the quantum theory of the chemical bond and, with his coworkers, effectively turned chemistry into a branch of applied quantum mechanics.

By the late 1940s biochemistry was branching into an entirely new discipline, *molecular biology,* the study of biologically important compounds at the molecular level. The star accomplishment of molecular biology is the discovery of the structure of DNA (deoxyribonucleic acid) and the deciphering of the genetic code by Francis Crick (1916–2004), James Watson (1928–), Rosalind Franklin (1920–1958), and Maurice Wilkins (1916–2004). Crick had been a physicist before he got interested in biology, and his expertise in wave mechanics was very helpful. He had been fascinated by a suggestion by Schrödinger (in his book *What Is Life?* published in 1944) that genetic information was transmitted by some sort of quasi-periodic molecular structure. Rosalind Franklin was expert in X-ray diffraction techniques, another application of wave mechanics. She scattered X-rays through DNA crystals, and from the resulting diffraction patterns it was possible to infer the fact that the crystal has a helical structure. Watson and Crick showed that given the possible bonds that could be formed by the components of DNA (the nucleotides and ribose sugars), there was only one way they could fit together to define a helix. The order of the nucleotides within the helix can be varied, and this is the physical basis of the genetic code.

In principle, any aspect of chemical structure and reaction dynamics can now be described and predicted using quantum mechanics. In practice, the detailed calculations required in order to predict such things as the structure of a protein molecule or the exact steps in a chemical reaction are still a challenge, often requiring the use of clever approximation techniques and computers. But quantum mechanics says that there is an answer, although it may take some skill to find it.

THE ELECTRON MICROSCOPE

Another almost immediate application of the wave mechanics of Schrödinger and de Broglie was the electron microscope. The resolution of a microscope is a function of the wavelength of the waves it uses, and because electron wavelengths are much shorter than those of visible light, electron beams can resolve much smaller objects. The German physicist Ernst Ruska (1906–1988) and others developed the first prototype electron microscope (EM) in 1931; it was called a *Transmission* EM because the beam of electrons is fired directly through the sample. Later the Scanning Electron Microscope was developed, which allows surfaces to be imaged down to near-atomic scales, and the Scanning Tunneling Electron Microscope, which can image individual atoms.

The electron microscope opened up many doors in cell biology. By the early 1960s practical electron microscopes were available that allowed scientists to study the ultrastructure of cells, and this led to a huge leap in understanding of cell biology. An important example is the theory of *serial endosymbiosis* championed by American cell biologist Lynn Margulis (1938–). Margulis proposed that many organelles within the cell, such as the mitochondria and chloroplasts, are actually bacteria that eons ago became gridlocked into a symbiotic relationship with a cellular host. Before the electron microscope this could not be confirmed, because even with the best optical microscopes a mitochondrion appears as little more than an indistinct blur. However, with electron microscopy it can be immediately seen that mitochondria and chloroplasts look exactly like bacteria. Margulis and others have since shown convincingly that eukaryotic cells (nucleated cells, such as those the human body is composed of) can be best understood as vast symbiotic colonies of bacteria.

The one drawback of the electron microscope in biology is that the high-energy beam is lethal to any living organism. Biological samples have to be fixed and prepared in various ways before they can be viewed. Another revolution in biology might occur if it became possible to image cellular ultrastructure in a living cell.

SUPER MATTER

Another duality that manifested itself in the 1920s and 1930s is the fact that many kinds of matter can exist in two phases: ordinary matter and matter where quantum mechanical effects are dominant. The transition between these two phases is usually abrupt.

Superconductivity

The Dutch physicist Kamerlingh Onnes (1853–1926) discovered that mercury immersed in liquid helium at a temperature of only 4.2°K above absolute zero loses its electrical resistance. This phenomenon has since been discovered in many other conductive materials at cryogenic (near absolute zero) temperatures. Superconductors also exhibit the *Meissner effect*, which means that magnetic fields cannot penetrate their surfaces. (This effect is named after its discoverer, Walther Meissner, 1882–1974.) Superconductors behave like ordinary conducting materials until their temperature is lowered to a critical transition temperature, below which they abruptly switch to the superconducting state.

Superconducting is still not fully understood, but it is known to be due to the tendency of charge carriers such as electrons to form Bose-Einstein condensates under the right conditions. Pairs of electrons (which are fermions) of opposite spin can be weakly coupled in certain kinds of metals and form *Cooper pairs* (named after Leon Cooper, 1930–). These pairs have a net spin of zero and thus, as a composite object, behave as bosons. This illustrates a general quantum rule, which is that if particles get combined by a dynamic interaction, the statistics of the *combination* is the statistics of the sum of the spins of the parts.

One of the Holy Grails of modern applied quantum mechanics would be a room-temperature superconductor. All superconductors to date have to be maintained at cryogenic temperatures which makes them very expensive and awkward to handle. Room temperature superconductors would enable a revolution in electrical and electronic technology.

Superfluidity

Another manifestation of Bose-Einstein condensation is superfluidity, discovered by the Russian physicist Peter Kapitsa (1894–1984) and others in 1937. The phenomenon was first observed in liquid helium-4, but it also occurs in helium-3 where it has a slightly different explanation. (Helium-4 atoms are bosons, while helium-3 atoms form Cooper-like pairs.) A superfluid will flow without any viscosity (fluid friction) whatsoever, and it has zero entropy and thermal conductivity. If superfluids are set into rotation their motion is quantized, and they form quantized vortex filaments and vortex rings that behave remarkably like fermions and bosons respectively.

The Laser

The laser ("light amplification by stimulated emission of radiation") will be described under "super matter," since it is also a large-scale manifestation of Bose-Einstein statistics. The laser was an offshoot of the *maser* ("microwave amplification by stimulated emission of radiation") created by Charles H. Townes (1915–) in 1953. In the late 1950s Townes, Arthur Schawlow (1921–1999), Gordon Gould (1920–2005), Theodore Maiman (1927–2007), and several others applied similar techniques to visible light and produced the first lasers. (The term "laser" itself was coined by Gould in 1957.)

The function of a laser is to emit a beam of *coherent* light, which means light of a uniform frequency and phase. (By contrast, the light from an ordinary light bulb is incoherent.) It exploits the process of stimulated emission of radiation, identified by Einstein in 1917. (It is not clear whether Einstein himself had any idea that stimulated emission would have practical uses; he was mainly concerned with the theoretical problem of finding equilibrium conditions between matter and the radiation field.) The laser works by a kind of chain reaction: an optical cavity with mirrors at both ends (one partially silvered) is filled with a material (solid, liquid, or gas) that can be pumped to an excited state. The material is chosen so that it will fall back into its ground state, emitting light of a definite frequency; this light reflects back and forth within the cavity, stimulating the emission of further photons in the same state, because photons obey Bose-Einstein statistics. The beam is emitted through the half-silvered mirror.

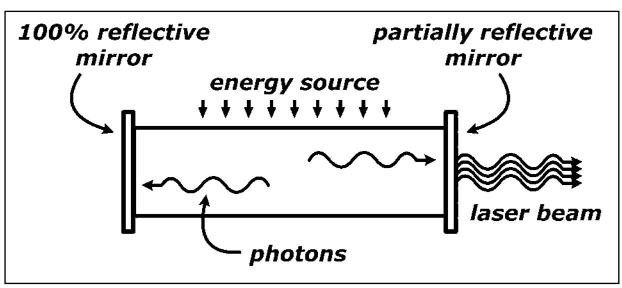

Figure 8.3: The Laser. The lasing medium (ruby crystal, CO_2, or neon gas, etc.) is pumped to an excited state by the energy source. Spontaneously emitted photons stimulate the emission of more photons of the same quantum state, producing a beam of coherent light. Illustration by Kevin deLaplante.

The applications of lasers are too numerous and well-known to detail here, except to note that exceptionally high-powered lasers are being explored as possible means of igniting nuclear fusion through a process called *inertial confinement.* If it works, this will involve compressing and heating a pellet of fusionable material by means of laser beams, and quantum mechanics will have made controlled fusion possible after all.

SPIN DOCTORING

In 1928 Heisenberg wrote the first quantum mechanical theory of ferromagnetism, the familiar sort of magnetism that makes compass needles swing to the north. There are several types of magnetism, but virtually all magnetic phenomena now can be understood in terms of the quantum mechanics of spin.

One of the most important applications of quantum mechanics is nuclear magnetic resonance (NMR), which was discovered by the American Isidore I. Rabi (1898–1988) in 1938. The nuclei of atoms that have an odd number of nucleons will have a net *magnetic moment* (a measure of magnetic field). If they are exposed to a strong external magnetic field, they can absorb a quantum of field energy in such a way that they line up in an excited state. They can then be tweaked by additional radio-frequency fields, and they will give off characteristic signals in the radio range. These signals, properly processed, give a great deal of information about the structure of the material being probed.

By 1970 NMR researchers could image a test tube of water. Since then NMR has been developed into the powerful technique of Magnetic Resonance Imaging (MRI), which can provide highly detailed images of soft tissues in the human body for diagnostic and research purposes. MRI is not only highly accurate but, unlike X-rays, safe, since there so far do not seem to be harmful effects from the exposure of the human body to the strong magnetic fields required for MRI. Recent work in neuroscience utilizes MRI for "real-time" imaging of the brain while it is in the process of performing tasks.

THE SEMICONDUCTOR REVOLUTION

Einstein's insights of 1907 on specific heats grew into the field of solid state physics, or condensed matter physics, as it is now often called. It is impossible to do justice here to everything that has been accomplished in this field. The application of quantum mechanics to semiconductors will be briefly mentioned, however, because this led to the semiconductor revolution, one of the more apparent ways in which quantum mechanics has shaped the modern world.

Semiconductors are metalloid elements such as germanium and silicon, roughly in the middle of the Periodic Table, which have electrical conductivities part-way between metals and insulators. Important contributions to the quantum theory of semiconductors were made by Rudolph Peierls in 1929. The operation of semiconductors can be understood in terms of quantum energy levels. All solids have a *band gap,* which is the energy required for the atomic electrons to jump free of the nucleus and serve as conductors of electricity. Metals have small band gaps, insulators have large band gaps, and semiconductors have band gaps that are intermediate in size. The conductivity of a semiconductor can be sensitively controlled by applied electric fields (as in an FET, or field-effect transistor) and by the addition of impurities (*doping*), which increases the number of *charge carriers,* which are either positively charged holes or electrons. A p-type semiconductor has excess holes, and an n-type has excess electrons.

The first working transistor was unveiled at Bell Labs in 1948 by John Bardeen (1908–1991), William Shockley (1910–1989), and Walter Brattain (1902–1987), although they may have drawn some inspiration from much earlier designs by Oskar Heil (1908–1994) and Julius Lilienfeld (1881–1963). (Bardeen would become the only person to win two Nobel Prizes in Physics; his other was for work on superfluids.) One of the many advantages of solid state electronic devices over their bulky vacuum tube predecessors is that they can be miniaturized. Their first crude-looking transistor has evolved into modern integrated circuit chips containing billions of microscopic transistors.

Semiconductor physics has enabled the modern electronic revolution. It is amusing to read science fiction written as late as the 1950s in which futuristic electronic devices still use vacuum tubes. What glaring failures of the imagination are we guilty of today?

9

SYMMETRIES AND RESONANCES

By the early 1960s the list of "elementary" particles had grown into the hundreds. Once renormalization had allowed physicists to tame the electromagnetic field (even if they still did not really understand it), the dominant problem was to understand the strong force (responsible for nuclear binding) and the weak force (responsible for beta decay). This long process culminated in the creation of the so-called Standard Model of elementary particles, which will be sketched below. The Standard Model is a qualified success: using it, a skilled practitioner can predict the properties (with certain important exceptions to be described) of all particles that are presently observable, and it contains within itself the unification of electromagnetism and the weak interaction into the *electroweak gauge field.* The Standard Model had settled into more or less its present form by the early 1980s. What seemed like the next natural step was the unification of the electroweak force and the strong force into a Grand Unified Theory (GUT). This was attempted during the late 1970s and early 1980s, taking the most obvious mathematical route. However, GUT failed in its most important prediction. Since the mid-1980s, particle theory has mostly gone off in another direction, pursuing a new dream of unification called *string theory.* But string theory has its own problems, as described below.

THE TOOLS OF THE PARTICLE PHYSICIST

It is not possible here to give anything more than the sketchiest presentation of detector and accelerator physics. Rutherford used natural alpha particles as probes of the nucleus. However, naturally occurring alphas have energies only up to about 7 MeV. Physicists realized that if they could hit the nucleus with a harder hammer they could get a better look at what was inside, but, as always in quantum mechanics, there is a tradeoff. Probing the nucleus with higher energy particles can reveal finer details of structure, because the wavelength of a higher-energy probe particle is shorter. However, the shorter the wavelength of the probe, the more energy it imparts to the target, and thus the more it changes the very nature of the target. Strictly speaking it is not correct to say that a 10 GeV (ten billion electron volt) proton hitting another proton detects certain particles within the target particle; rather, the incident proton in combination with whatever was inside the target particle causes certain observable products to appear. Bohr would have insisted that it is not even meaningful to ask what was inside the target before it was probed, while Heisenberg might have preferred to say that the target particle possesses certain *potentialities* (that can be expressed in the mathematical language of quantum theory) that determine its possible interactions with the incident particle. Whatever the proper interpretation, it is mistaken to suppose that a quantum mechanic looks inside a nucleon the way an auto mechanic looks under the hood of a car.

There are two main kinds of accelerators, the *linear accelerator,* which uses high voltage to accelerate charged particles in a straight line, and the *cyclotron,* which accelerates particles in a circular path. There is no known way (apart from gravity) to directly accelerate neutral particles such as the neutron, although they can be produced by various reactions and then allowed to interact with targets.

Linear accelerators evolved from the Cockroft-Walton and van de Graaf machines of the early 1930s to the two-mile-long Stanford Linear Accelerator in California (SLAC), opened in 1962. Linear accelerators remain an important tool, but the highest energies are produced by cyclotrons and their descendents.

The invention of the cyclotron is generally credited to the American physicist Ernest O. Lawrence (1901–1958) in 1929. Lawrence's idea was that charged particles could be confined to a circular path by magnets and accelerated by pulsed electromagnetic fields. As the energy gets higher, the diameter of the particle track has to go up, and cyclotrons evolved rapidly from Lawrence's first desktop device by the end of World War II into machines several feet in diameter. After World War II cyclotrons and their descendents would evolve into miles-wide monsters costing billions of dollars to construct. Particle energies climbed in the billion-electron volt (written BeV, or GeV for "giga" electron volt) range by the 1960s, and the study of elementary particles became known as *high energy physics.*

One design limitation of cyclotrons is the need for powerful magnetic fields. With the introduction of practical supercooling it became possible to use superconducting magnets, which allows for much higher beam energies. The largest operational accelerator in the world at this writing is Fermilab (near Batavia, Illinois), with a beam energy of 2 TeV (tera- or trillion electron volts).

Accelerator construction hit a financial and political wall in the early 1990s. In the United States, $2 billion had been already spent on the 54-mile diameter Superconducting Supercollider (SSC) in Texas when it was canceled by the U.S. Congress in 1993. The reasons for the cancellation were cost and doubt that it would yield scientific results of sufficient interest compared to other projects on which the money could be spent. The SSC was designed to collide protons with an energy of 20 TeV (tera-electron volts). The cancellation of the SSC meant that since the mid-1980s particle physicists have been largely unable to test their theories, which depend on predictions at energy levels beyond any accelerator currently in operation. This will change soon, however, for the Large Hadron Collider (LHC) near Geneva, Switzerland, is expected to come on line in 2008. The LHC will have beam energies of 14 TeV. This colossal machine (financed and operated by a consortium of 38 countries including the United States and Canada) is the great-grandchild of Rutherford's improvised table-top device of 1910 that demonstrated the existence of the nucleus by scattering alpha particles through a scrap of gold foil. Despite the enormous sophistication and power of modern particle accelerators, what they mostly do is just what Rutherford did: fire an energetic particle at a target and see what scatters off.

Particle detection is a complex process. Many particle detectors take advantage of the particles' electric charges and therefore cannot directly image a neutral particle. (If a pair of particles branches off apparently from nothing, however, that is a sign of the decay of a neutral particle.) An important early detector was the Wilson cloud chamber, designed in 1911 by C.T.R. Wilson (1869–1959). The chamber contains supersaturated water vapor. Charged particles such as alpha particles or electrons will ionize water, leaving a trail of mist through the chamber. If a magnetic field is applied, positively charged particles will curve one way, negative the other, and the curvature of the track is a function of the mass of the particle. Cloud chambers were used by Anderson to detect the positron and muon in cosmic ray showers.

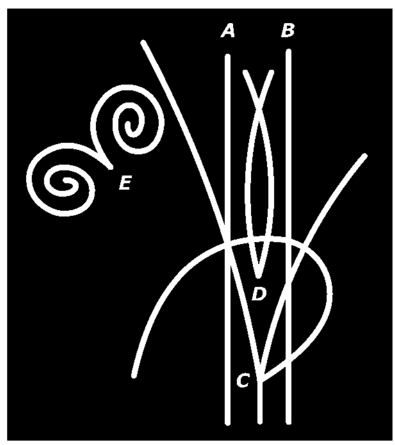

Figure 9.1: Typical Bubble Chamber Tracks. A and B are beam particles. C is a collision of a beam particle with a proton. D is the decay of a neutral particle emitted from C. E is the decay of a gamma ray into a positron-electron pair. Illustration by Kevin deLaplante.

In the 1950s hydrogen bubble chambers were introduced. A volume of liquid hydrogen is allowed to expand just as a jet of particles are fired into it from an accelerator, and myriad tracks of bubbles are produced in the hydrogen by the charged particles in the beam. Two generations of technicians have strained their eyesight recording the positions of bubble chamber tracks so that the particle trajectories could be reconstructed by computer analysis and analyzed for evidence of the creation of new particles.

Bubble chambers have been largely replaced in high energy physics by devices such as the wire chamber, drift chamber, and spark chamber, but the principle is the same: energetic charged particles can be tracked because of their ability to ionize parts of their surrounding medium.

Most particles that were being searched for have extremely short lifetimes (half-lives), and so they often are detected indirectly by their decay products. A high energy proton-proton collision, for example, can produce jets containing hundreds of by-products, the vast majority of which will be uninteresting, and the analysis of the tracks is a daunting task. The aim is to produce a three-dimensional reconstruction of the collision and its products. Sometimes high energy physicists will have a definite prediction to test, and they will be looking for particles with expected properties such as half-life, charge, and so on; sometimes they just blast away and see what comes out.

TAMING THE PARTICLE ZOO

Gauge Theory

While particles multiplied like rabbits in high energy experiments, quantum theory moved to new heights of mathematical abstraction with the publication in 1954 by C. N. Yang (1922–) and Robert Mills (1927–1999) of a new kind of field called a *gauge field*. It is impossible to describe the meaning of a gauge field adequately here. It is a generalization of the quantum theory of the electromagnetic field, and Yang and Mills intended that it could take account of symmetries that had been noted in the study of the strong nuclear force. One of the important consequences of gauge theories is that they predict the existence of new particles that mediate the forces described by the fields. But the Yang-Mills theory seemed to predict particles that did not exist, and it was so complex that it was not clear that it could be renormalized.

The Yang-Mills theory languished until the surprising proof in 1971 by Martinus Veltman (1931–) and Gerard t'Hooft (1946–) that it is renormalizable after all. It was immediately possible to apply it successfully to the strong and weak forces, as described below. The history of quantum field theory can therefore be summarized by saying that it started as the attempt to formulate a relativistic quantum theory of the electromagnetic field; it suffered a temporary setback due to the divergences; it was rewritten as a renormalizable theory, generalized as gauge theory, and is now applied with some (though not unqualified) success to all particle fields—except gravitation, which is a very special problem that will be discussed later.

The Heyday of High Energy Physics

By the early 1960s hundreds of short-lived particles, called *resonances,* had been detected in high energy collisions. Some resonances have half-lives as short as 10^{-23} seconds, and are only detectable indirectly by their decay products. The shorter the half-life, the less certainty there is in the mass of the resonance, by the uncertainty principle. Decay modes can be very complicated, with a resonance shattering into numerous fragments via the strong force, and then many of the fragments undergoing further weak force decay. Much of particle theory during the 1950s amounted to little more than desperate attempts to classify the denizens of the particle zoo. But eventually patterns would begin to emerge.

Disparity and a New Symmetry

One of the most interesting discoveries in the 1950s was *parity violation,* which was found to occur in the weak decay of certain unstable nuclei. A parity flip is a change of handedness, as in mirror reflection. It had been widely assumed that all particle interactions were invariant under change of parity; that is, if a decay were mirror-reflected it would look the same. However, it was found by C. S. Wu (1912–1997) and others that if Cobalt-60 nuclei are lined up the same way they will emit beta rays in a preferential direction, which means that the process would not look the same in a mirror.

The operator that reverses parity is symbolized as P, while the operations that flip electrical charge and time order are symbolized C and T respectively. It was eventually discovered that while some weak interactions violate parity, and some violate two of C, P, or T, all interactions in physics apparently obey CPT conservation. This means that in a particle interaction such as a decay, if one could reverse all charges, mirror-reflect the process, and run it backwards in time, one would observe precisely the same interaction. This is a bit difficult to test in practice, however. But most physicists accept CPT conservation as a fundamental property of quantum fields, because it can be shown that if CPT were violated then Lorentz invariance would be violated.

Symmetries, Made and Broken

Group theory became increasingly important in particle theory from the 1950s onward, as physicists sought clues to the dynamics of particles by trying to classify the symmetries they obeyed. Groups are a way of describing the symmetries of an object, which are the operations under which it remains unchanged. For instance, a square rotated through 90° looks just the same as it did before it was rotated. The Lorentz transformations of special relativity, as an important example, can be understood as the result of the operation of a *Poincaré group.*

The concept of *symmetry breaking* also became important. As example of symmetry breaking is the way ice might freeze in an unsymmetrical way if there is a small impurity in the water. Many current particle theories predict that all interactions are the same at some extremely high energy, but that as the universe "froze out," symmetry breaking led to the different types of interactions we observe at our energy levels. The concept of symmetry breaking gives a useful way of understanding why we do *not* observe what fundamental theory says we should observe.

Probing the Proton

When SLAC went on-line in 1962 the door was opened to a deeper understanding of particle structure through *deep inelastic scattering.* In this process high-energy particles such as electrons are scattered through heavy particles such as protons. These experiments revealed that protons and neutrons have an internal structure, just as Rutherford's scattering experiments revealed the presence of the atomic nucleus.

Various names for these new particles inside the nucleon were proposed—*partons* and *aces,* for instance—but eventually the name "quark," proposed by Murray Gell-Mann (1929–), was the one that stuck. Gell-Mann named his hypothetical elementary particles "quarks," after a line in James Joyce's *Finnegan's Wake,* in which sea-gulls flying overhead call, "Three quarks for Mr. Mark!" This started a trend of whimsical names for particles and their properties; there were so many new particles being discovered or hypothesized that bemused physicists simply did not know what else to call them.

An important characteristic of quarks is that they possess fractional electrical charges of either +/–1/3 or +/–2/3 of the charge of an electron. *Up*-quarks have a charge of +2/3, while *down*-quarks have a charge of –1/3 .

On the quark model, strongly interacting particles are classified as *hadrons,* and these can be divided into two groups, baryons (neutrons, protons, and their heavier unstable analogues), and mesons such as the pion. Baryons are fermions and are composed of three quarks, and hadrons are bosons and are composed of two quarks. The proton, on this scheme, is built up out of two up quarks and a down quark, and the neutron is an up and two downs. The positively charged version of Yukawa's pion is made of an up-quark and an anti-down quark.

Gell-Mann used the quark model and symmetry considerations to argue that groups of hadron resonances could be classified either as octets, sometimes called the *eight-fold way,* or as *decuplets,* arrays of 10 particles. One member of a decuplet, the omega-minus particle, had not yet been discovered, and when it was found in 1962 this gave an excellent verification of the quark model.

Hunting of the Quark

Evidence mounted through the 1960s and 1970s that quarks are the best explanation for the structure of hadrons and baryons. A problem, however, was that it did not seem to be possible to produce free quarks. Finally, it became clear that quarks are forever confined within baryons or hadrons. Within a particle such as a nucleon the quark can move almost as a free particle (this is called *asymptotic freedom*), but if a quark is expelled from the nucleus by a collision it polarizes the vacuum around itself, pulls other quarks out of the vacuum, and combines into a baryon or hadron too quickly to be observed as a free particle. It is as if quarks have little if any mutual attraction when they are very close to each other (inside a nucleon) but attract each other with a force that increases rapidly to a high constant value the farther apart they are.

Most particle physicists are now convinced that the quark does exist, because of the tremendous explanatory and predictive power of the quark hypothesis. But it still seems discomfiting to have particle theory dependent on an object that physicists apparently have no hope of ever observing directly.

The Standard Model

With the discovery of the quark, and the proof that gauge theories are renormalizable, it became possible to produce a unified theory of the structure of elementary particles that is now called the *Standard Model.* The main conclusions of the Standard Model will be sketched here. It is the best picture we have so far of the ultimate constituents of the physical world. It would be bewildering if not impossible to trace all of the changes in particle terminology and classification since the time of Yukawa. (For example, the term "neutron" was briefly applied to the particle that we now call the "neutrino.") Therefore, this section describes the particle zoo using terminology that has been current since the 1970s. However, beware that books and articles written before this time may not use exactly the same language.

The Standard Model says that matter and energy are described in the language of quantum gauge field theory. All particles are divided into fermions and bosons. The field quanta are bosons, and they mediate forces between particles of matter, which are fermions. This field-theoretic picture is just the wave-particle duality in a more mathematically abstract form. There is a complementarity between continuous field and discrete particle points of view. As with the double slit experiment, the cash value of a field theoretic calculation always can be expressed as a function of the probabilities of detecting particles, or other probabilistic quantities such as expectation (average) values of observables, or scattering *cross-sections* (probabilities of particle interaction).

There are two groups of fermions: quarks and leptons. Leptons include the massless neutrinos. (There is recent evidence that neutrinos may have a very small mass, but this remains controversial.) The family of quarks and leptons is divided into three generations, and in each generation there are six quarks (plus their corresponding antiquarks), a lepton, and the lepton's partner neutrino. Neutrinos move at the speed of light. As described above, all hadrons (baryons and mesons) are built up out of various combinations of quarks and antiquarks.

If quarks and leptons have a finite diameter it is less than 10^{-18} m. They are often treated as point particles—even though by Heisenberg's Uncertainty Relations the notion of a point object does not make physical sense!

There are three flavors of lepton, the electron, muon, and tau lepton, and three corresponding flavors of neutrinos, the electron neutrino (postulated by Pauli in 1930), the mu neutrino, and the tau neutrino. All particles in this scheme have now been observed in accelerators, although some of the heavier particles in the third generation were not produced until quite recently.

The three generations of particles are almost carbon copies of each other, except that the higher generations are heavier. This strongly suggests that all three generations are simply different energy states of one fundamental structure, but it remains unclear what that could be.

There are two kinds of forces in the Standard Model, the color force (a generalization of Yukawa's strong force), and the electroweak interaction. The color force is mediated by bosons called *gluons,* and the electroweak interaction is mediated by the photon and a series of heavy bosons (the so-called

Matter Quanta (Fermions)			
Generation	**Quarks**	**Leptons**	**Neutrinos**
Third	top bottom	tau	tau neutrino
Second	strange charm	muon	muon neutrino
First	up down	electron	electron neutrino

Field Quanta (Bosons)	
Field	**Quanta**
electroweak force	photon $W^{+/-}$ bosons Z^0 boson
color force	gluon
Higgs field?	Higgs particle?
gravity	graviton?
Grand Unified Theory (GUT)?	

Figure 9.2: Table of "Elementary" Particles in the Standard Model. Matter is composed of spin-1/2 quarks, which come in three generations, each with its corresponding lepton and neutrino. There are two flavors per generation. Field quanta are bosons. The "color" force between quarks is mediated by gluons, the electroweak force is mediated by photons and intermediate vector bosons, the Higgs field (if it exists) is mediated by the Higgs boson, and gravity is mediated by the so-far unobserved graviton. Simple! Illustration by Kevin deLaplante.

intermediate vector bosons). The detection of the predicted intermediate vector bosons in the early 1980s was one of the last great triumphs of the Standard Model.

The field theory of the color force is called quantum chromodynamics; it is a Yang-Mills theory. Several attempts have been made to unify quantum chromodynamics and the electroweak field, but (as discussed below under "Protons Refuse to Decay") they have not been successful.

The Standard Model is the work of literally thousands of theorists, experimental high energy physicists, and technicians over a period of nearly 40 years, with the expenditure of billions of dollars, and it is a bit difficult to assign priority to the researchers involved. Apart from those already mentioned, Nobelists Stephen Weinberg (1933–), Sheldon Glashow (1932–), and Abdus Salam (1926–1996) played an especially important role in creating electroweak theory.

In summary, the short version of the history of the Standard Model goes like this: quantum electrodynamics was generalized into gauge theory by Yang and Mills. When it was proven that gauge theories are renormalizable, they were applied to the quantum "color" fields that mediate the forces between quarks. Yukawa's early theory of the strong force follows from QCD as a low-energy approximation. The resulting field theory was called *quantum chromodynamics* (QCD). Electromagnetism and the weak force were unified into the *electroweak* field.

Protons Refuse to Decay

A number of particle theorists since the 1970s have attempted to define so-called Grand Unified Theories (GUTs) which would unify quantum chromodynamics with electroweak theory. The assumption is that at very high energies, all interactions are the same, and the differences between the three non-gravitational forces are due to symmetry-breaking. Most GUTs predict proton decay since they treat leptons and quarks as different states of the same particle. During the 1980s several experimental groups searched for proton decay in large tanks of highly purified water or various hydrocarbons (which contain many protons). Although the predicted decay modes would be very rare, with enough protons a few decays should have been observed. However, from these experiments it is now possible to say that if the proton does decay, its half-life must be substantially greater than 10^{33} years. This rules out a number of GUTs such as the model proposed in 1974 by Sheldon Glashow and Howard Georgi (1947–), but the idea remains alive.

Too Many Empirical Constants?

Despite the great success of the Standard Model, it requires the use of approximately 50 empirical parameters, which, like the old Rydberg constant of pre-Bohr spectroscopy, can be determined directly or indirectly from experiment but which have no theoretical explanation or derivation. In particular, there is no way of calculating the masses of the elementary particles. There is presently no more understanding of why the mass of the electron is .511 MeV than physicists before Bohr understood why the H-alpha line of hydrogen has a wavelength of exactly 656.3 nm.

Bohr was able to make the Rydberg constant “go away” in the sense that he found a derivation for it from the dynamics of his theory. No theory of particle structure will be truly satisfactory until it can make some of those empirical constants go away, especially the mass spectra.

FRONTIERS OF PARTICLE PHYSICS

Supersymmetry

One of the guiding hypotheses that have guided research in elementary particles since the early 1970s is supersymmetry, which says that there should be a complete symmetry between fermions and bosons: for every fermion (such as the electron) there should be a corresponding boson. This idea was introduced by Pierre Ramond (1943–) in 1970, not merely out of a love of mathematical symmetry, but because it was essential to make string theory allow for the existence of fermions. A great deal of intellectual effort has been invested in trying to predict the properties of the supersymmetric “twin” particles or *sparticles* that supersymmetry says should exist. However, so far, there is absolutely no experimental confirmation of supersymmetry, and some particle physics are beginning to doubt that the idea was viable in the first place. It is, however, essential to most versions of superstring theory, and so the idea of supersymmetry remains very attractive to many particle theorists.

Preons

Several physicists in the 1980s explored the possibility that leptons and quarks could be understood as composite particles, built up out of a hypothetical *really* elementary particle dubbed the *preon.* However, no compelling theoretical formulation and no experimental evidence has been found for preons, and the idea has largely been put on the back burner. One barrier to preon theory is that because of the Uncertainty Relations, any particle that could be small enough to fit inside the electron would have to have an enormously larger mass than the electron. Most attempts to find a unifying basis for the Standard Model therefore seek to understand particles in terms of radically different kinds of structures that would be something other than merely more particles inside particles.

Search for the God Particle

The missing link of the Standard Model is the Higgs particle or Higgs boson, proposed by British particle theorist Peter Higgs (1929–) in 1964. It is sometimes called the "God particle" because it plays such a central role in the Standard Model. According to theory, the Higgs boson is the quantum of a field that pervades all of space, and other particles such as leptons acquire mass by polarization of the Higgs vacuum field. So far, the Higgs has not been detected, and it must have a mass above 100 GeV. It is hoped that the Large Hadron Collider (LHC) coming on line at CERN will have beam energies that could be high enough to produce the Higgs particle, and some physicists are now taking bets about whether or not it will be discovered. If it is not then the Standard Theory is due for further revisions, an outcome that would surprise no one.

Heisenberg's Last Word

In 1976 the aging Heisenberg published an article in which he commented from his long experience on the state of particle theory. He argued that the most important result of particle physics over its 50-year history since Dirac predicted the existence of the positron was that there are no truly elementary particles (in the sense of an entity that cannot be changed into something else). All particles are mutable, subject only to conservation laws such as electrical charge, momentum, and mass-energy. He also pointed to the importance of symmetry principles and symmetry-breaking as guides to which particle reactions are possible. Heisenberg also noticed the similarity between the state of particle physics in 1976 and the Old Quantum Theory during its later years, with its recourse to well-educated guesswork: spotting mathematical regularities, guessing that these regularities have wider applicability, and trying to use them to make predictions. Such guesswork is very useful; a good example is Balmer's skill in spotting the mathematical structure hidden in the hydrogen spectrum. But what is needed for real understanding, Heisenberg argued, is a theory of the underlying dynamics, such as was provided by Schrödinger's Equation in 1926. The key, Heisenberg stated, is that the table of particle masses form a spectrum just as atomic energies form a spectrum, and a spectrum must imply a dynamical law whose solution, together with the right boundary conditions, would give as eigenstates the spectral values (particles) we observe. If quantum mechanics applies then there must be some kind of eigenfunction whose eigenvalues are the masses of the elementary particles, just as the spherical harmonics give the energies and structure of the electron shells in atoms. One would like to be able to understand the leptons and neutrinos, for instance, as merely different eigenstates of the same particle; the challenge is to find the operator they would be eigenfunctions of.

Heisenberg also made an interesting comparison between the atomism of Democritus (according to which the world can be broken down into atoms—*indivisibles*—and the Void) and Plato's atomism, according to whom the physical world is to be understood in terms of mathematical symmetries. Plato, argued Heisenberg, was closer to the truth. Heisenberg cautioned, however, that one must distinguish between a "phenomenological" symmetry (an approximate symmetry that is merely descriptive) and a fundamental symmetry built into the laws of physics (such as Lorentz symmetry). Heisenberg suggested that accelerator physics could be approaching the *asymptotic* regime, a region of diminishing returns in which fewer and fewer new particles will be discovered regardless of the energies applied. It was high time, Heisenberg concluded, for physicists to move beyond the gathering and classification of particles, and attend to the problem of finding the right dynamical laws.

Strings and Things

The closest thing to an attempt to answer Heisenberg's demand for a theory of the underlying dynamics of particles is string theory, which has become the most popular particle model since the 1980s. The essential idea of string theory is to replace the point-like model of quarks and leptons with a one-dimensional elastic string. The different possible particles would then be the *eigenmodes* (possible quantized vibrations) of these strings and then—in principle—it ought to be possible to calculate particle properties the way spectral energies, intensities, and selection rules can be calculated for atomic orbitals. String theory that incorporates supersymmetry is often called *superstring* theory.

String theory was sparked by the publication in 1968 by Gabriele Veneziano (1942–) of a new formula that was very successful at describing the scattering of strongly interacting particles. Over the next few years several other theorists including Leonard Susskind (1940–) realized that the Veneziano formula suggested that the force between quarks behaved rather like a quantized string, and they developed the idea in more detail. An important addition to the string picture is that one can think of both open strings and closed, loop-like strings. The two ends of an open string could be an electron-positron pair, and when they annihilate they form a closed loop, or photon. With further refinements string theory began to resemble a fundamental theory of all particles, not merely another way of describing the color force. Strings obey a very elegant law of motion: closed loop strings trace out tubes in spacetime that move in such a way as to minimize the area of the tube. Furthermore, strings seemed to automatically allow for the existence of the elusive *graviton,* the hypothetical quantum of the gravitational field.

Despite their promise, these ideas were not taken very seriously at first. In 1984 there occurred the "first superstring revolution," when John Schwarz (1941–) and Michael Green (1946–) proved that string theory is mathematically consistent so long as it is 10-dimensional. This discovery sparked a near-stampede of physicists to superstring theory, and soon most particle theorists, especially of the younger generation, were working on it.

The notion of 10 dimensions, 9 spatial and 1 for time, may seem bizarre. We only observe 3, says the string theorists, because the rest are *compactified,* meaning roughly that they are rolled up into tiny tubes too small to be observed. Unfortunately, there was no one obvious way to do this, and the door was opened to many possible string theories, which soon seemed to proliferate more rapidly than hadron resonances. It was in the 1980s that string theory began to be criticized by some senior physicists, such as Feynman and Glashow, who were unhappy that string theorists did not seem to be trying very hard to make any testable predictions. Despite these worries, string theory has continued to thrive. Almost everyone who works in particle theory these days is a string theorist, and string theorists often display a remarkable confidence that they are just a few complex calculations away from the Theory of Everything. This is despite the fact that up to now virtually the only evidence in favor of string theory is its theoretical consistency, and in particular the fact that it seems to provide a natural place for the graviton. However, one needs a very high level of mathematical training to appreciate these facts. This in itself is not necessarily a sign that something is wrong with the theory, however; theoretical physics has always been difficult.

Very recently intense controversy about string theory has flared up again. It has been heavily criticized by several prominent physicists, notably the distinguished particle and gravitational theorist Lee Smolin (1955–). He and others argue that string theory is an approach to elementary particle physics that initially had a lot of promise, but that has become an enticing dead end, the modern equivalent of the epicycles of pre-Copernican astronomy which could explain everything but predict nothing. The grounds of the criticism of string theory are simple: string theory has so far made virtually no testable quantitative predictions and there is therefore no way to check whether the theory is on the right track. All of the other successful advances in quantum physics described in this book were recognized as advances precisely because they were able to two things. First, they could explain facts that previously could not be explained, such as the structure of the hydrogen spectrum. Second, they also predicted phenomena that no one would have thought of looking for without the theory, such as Bose-Einstein condensation and various elementary particles such as the neutrino and Gell-mann's omega-minus.

There are two reasons for string theory's lack of predictive success. Both the defenders and the critics of string theory will agree that since the mid-1980s it simply has not been possible to experimentally test the predictions of any theory, string or otherwise, that attempts to predict phenomena beyond the energy limits of the Standard Model. Following the cancellation of the

Superconducting Supercollider there have not been any accelerators powerful enough to do the job. If energies in the 10 TeV range could be probed it would immediately eliminate many candidate particle models, or perhaps uncover something entirely new that would force current theory to change direction.

The other problems with string theory are theoretical. In strong contrast to quantum mechanics in its early years, string theory cannot calculate definite predictions in many cases; for instance, there is no way to calculate the mass of an electron or proton, or derive any of the numerous empirical constants needed to make the Standard Model work. The mathematics is just too hard. Second, string theory turns out to be strongly under-determined, in that there are in fact innumerable string theories (Smolin estimates 10^{500}), all equally mathematically valid. This has encouraged some string theorists such as Susskind to postulate the existence of a "landscape" of possible string theories, with the one that works for our world being essentially an accident. Smolin charges that at this point the theory has lost most of its contact with reality. String theorists say, give us more time. Smolin and other critics say, you've had time enough, and more funding should be given to alternative approaches.

The author of this book once heard a talk given by the distinguished particle theorist Howard Georgi. He joked that in the absence of experimental data, theoretical physicists get excited into higher and higher flights of theoretical fancy like so many elementary particles pumped into excited energy states, but when experimental results come along (perhaps confirming a few theories, but likely falsifying most) the theorists fall back into their humble ground states. Perhaps the Superconducting Supercollider would have been able to settle a lot of speculative dust, but it was cancelled. By the time this book is in your hands, the LHC at CERN will be up and running and may have generated some data that will cause string theorists to fall back into their ground states again, no doubt emitting many papers in the process.

In 1995, Edward Witten (1951–), today's leading string theorist, proposed that there is a yet-to-be-discovered theory behind string theory, which he called *M-theory*, an 11-dimensional picture that would unify the different versions of string theory and perhaps serve, in effect, as the long-sought Theory of Everything. (Witten's proposal is sometimes called the *second* superstring revolution.) But Witten does not yet know exactly how M-theory would work, any more than Born knew how quantum mechanics would work when he argued, in the early 1920s, that there had to be such a theory. Perhaps Witten will find the key himself or perhaps he must wait for his Heisenberg, whoever that will be, to see another sunrise on Heligoland.

10

"THE MOST PROFOUND DISCOVERY OF SCIENCE"

Throughout the 1950s and 1960s the majority of physicists who used quantum mechanics applied it to an ever-growing variety of areas in pure and applied physics—elementary particle physics, semiconductors and electronics, chemistry and biochemistry, nuclear physics, masers and lasers, superfluids, and superconductors. But a few physicists and philosophers of physics, including some of the highest ability, continued restlessly to probe the foundations of the theory. They realized that the foundational questions that had been raised by Einstein, Schrödinger, and others in the 1930s had merely been set aside, mostly because of World War II and the excitement of the new developments in particle theory and quantum electrodynamics, but had not been solved.

DAVID BOHM: THE SEARCH FOR WHOLENESS

David Bohm (1917–1992) was born in Wilkes-Barre, Pennsylvania, and did his doctoral work under the direction of J. Robert Oppenheimer (1904–1967), the "father" of the atomic bomb. Bohm possessed an exceptional combination of physical intuition and mathematical ability, and a deep fascination with the foundational problems that many other physicists preferred to ignore. Shortly after World War II Bohm wrote a paper in which he laid out the basic ideas of the renormalization theory that was soon to be developed so successfully by Feynman, Schwinger, and Tomonaga. However, when he submitted his paper to *Physical Review* it was rejected after a critical referee report by Pauli, and Bohm let the idea drop.

In the late 1940s Bohm made important contributions to *plasma physics.* This is the study of gasses that are so hot that they become a soup of ionized particles and respond collectively to electromagnetic fields in complex ways that are still poorly understood. Bohm thought deeply about the basis of

Figure 10.1: David Bohm. Library of Congress, New York World-Telegram and Sun Collection, courtesy AIP Emilio Segre Visual Archives.

quantum mechanics, and in 1951 he published a text on quantum theory in which he explained the EPR paradox in a novel way. He described the thought experiment in terms of spin measurements on particles such as electrons. Pairs of electrons are emitted at the source, and measurements on them have to obey certain global conservation laws (for example, their total spin remains constant), just as in the original EPR experiment. They travel down the arms of the apparatus and encounter Stern-Gerlach detectors, which can be set at various angles to measure their spins. Bohm's spin-measurement version of the EPR experiment paved the way to versions of the EPR experiment that could actually be performed.

Bohm's text of 1951 states the orthodox Copenhagen view in an especially clear way, but Bohm was very unsatisfied with the claim that no deeper account could be given of quantum mechanics. In 1952 he published a monumental paper in which he advanced what still is by far the most thoroughly worked-out causal interpretation of quantum mechanics. Bohm's interpretation of quantum mechanics is a "no-collapse" pilot wave theory, depending in part on mathematical steps very similar to those taken by Madelung and de Broglie years earlier. He showed that hidden in the structure of Schrödinger's Equation there exists a curious nonlocal force field that Bohm called the *quantum potential.* All particles in Bohm's theory have continuous trajectories just like classical objects, and the Heisenberg Uncertainty relations are purely statistical uncertainties that do not imply that particles do not *have* exact positions and momenta. There is a rule called the guidance condition, which sets the initial positions of the trajectories like the gates that line up racehorses before the gun is fired. The quantum potential maintains the particle correlations once the race has begun. Bohm set the guidance condition in such a way as to make the particles behave the way they ought to according to orthodox theory, because he wanted to prove that a hidden variables theory could reproduce the predictions of standard wave mechanics. However, there could conceivably be other guidance conditions; this possibility does not seem to have been explored.

The quantum potential is a function of *phase differences* in the wave functions of the correlated particles; it can therefore be distance-independent, depending on the phase structure of the system. It has physical units of energy;

therefore, part of the total energy of systems of elementary particles is tied up in the quantum potential of their wave functions. It can be shown that if the wave functions of the particles are entangled, then the quantum potential is entangled as well—which means that it cannot be broken up into separate energies linked to each of the localized particles in the system. The quantum potential, in other words, is a *nonlocal* form of energy, a property of the entangled state as a whole. It is like the energy of an atomic orbital, which is not localized until the orbital decays and emits a photon in a particular direction. A further peculiar feature of the quantum potential is that it implies the existence of a nonlocal *force* (although one of a rather complicated mathematical structure), because any potential, when differentiated with respect to distance, gives a force. And yet most physicists are very reluctant to draw this conclusion; like Einstein, they are very reluctant to take seriously any sort of "spooky action at a distance," and tend to change the subject when this topic comes up in discussion.

De Broglie realized that Bohm had found ways around most of the objections that had been raised to his own, less developed causal theory. Bohm showed that the answer to Pauli's worry about scattering was that the outgoing wave will in fact break up into packets that correspond to the motion of the outgoing particles.

De Broglie regained confidence in his old approach and developed a relativistic generalization of Bohm's theory. A defect of Bohm's theory is that (like the Schrödinger Equation on which it is based) it treats time and space classically and is valid only for nonrelativistic velocities. De Broglie found a way of writing causal quantum mechanics in a covariant, four-dimensional format. De Broglie's approach has been largely ignored, perhaps unfortunately so, and Bohm himself was uncomfortable with it because it predicts an odd kind of backwards causation (i.e., events in the future influencing events in the past).

Apart from the interest shown by de Broglie and a colleague of de Broglie's, Jean-Pierre Vigier (1920–2004), the reaction to Bohm's theory in the months and years immediately after its publication were almost entirely negative. Even Einstein, who was personally sympathetic to Bohm, and who might have been expected to welcome Bohm's apparent restoration of determinism, rejected Bohm's approach as "too cheap" (Cushing 1994, p. 146). What Einstein apparently meant by this was that one can rather easily account for distant quantum correlations if one imagines that some sort of odd action at a distance connects the particles, but that this would be unacceptable simply because any sort of action at a distance is unacceptable. The criticism directed by Pauli and other orthodox Copenhagenists toward Bohm was scathing. J. Robert Oppenheimer, who was at that time director of the Institute for Advanced Studies in Princeton, commented that "if we cannot disprove Bohm, then we must choose to ignore him" (Peat 1997, p. 133). Part of what motivated Oppenheimer's cynical remark was that Bohm had refused to testify against friends to the House Committee on Un-American Activities, and in the rabidly anti-Communist climate of the time it was politically risky to be seen supporting Bohm. Bohm lost his position in Princeton and ended up teaching in Sao Paulo, Brazil. Later he worked in Israel and ended his career at Oxford University.

In Israel, Bohm and his student Yakir Aharonov (1932–) demonstrated the existence of an odd manifestation of quantum nonlocality now called the Aharonov-Bohm Effect. They showed that a magnetic coil can influence interference fringes of electrons in a double slit experiment even if the field is zero outside the coil.

Toward the end of his life, Bohm explored a new interpretation of quantum mechanics that he called the "implicate" or "enfolded order." Suppose the three-dimensional physical space that we ordinarily experience is, in fact, merely a virtual reality like the world in a computer game. Think of what it is like to play "Space Wars," a computer game that particle physicists used to run on their PDP-12s late at night when they were supposed to be analyzing bubble-chamber data. In this game the planets and spaceships seemingly obey Newton's laws of motion and his law of gravitation, and in order to win you have to steer your rocket and launch your missiles in accordance with these laws. The little spaceships on the screen are not really moving under the influence of gravitation or inertia, however, but rather under the command of the instructions encoded in the computer program that defines the game. The program is really what is in control, not Newton's laws; the behavior of the rockets on the screen is merely a logical *implication* of the inner program. Bohm hinted that maybe something like this is happening in the real world as well. Bohm did not mean to suggest that ordinary physical reality *is* just a computer game; rather, he meant that there might be some inner logic, some program or set of laws controlling the world, that is not inherently in space and time at all, but that *produces* space and time as a byproduct of certain rules that we have yet to uncover. The inner rules would not look at all like ordinary laws of physics as we presently understand them, although perhaps we could work out those laws from the rules of quantum mechanics that we presently are familiar with.

If anything like this is true, then there might be perfectly deterministic "code" underlying the apparently random and indeterministic behavior of quantum mechanical particles. Suppose an electron is emitted from point *A,* and could be absorbed at either point *B* or point *C.* In general there is a probability that it will appear at either *B* or *C,* and if we try to assume that the electron follows a definite classical trajectory after its emission from *A,* we will in many cases get a wildly incorrect answer for the probabilities that it will arrive at *B* or *C.* It is as if the electron simply jumps from *A* to either *B* or *C* according to its own mysterious whim. (Recall Schrödinger's disgust at this "damned quantum jumping.") The electron's path is both discontinuous (because of the jump) and indeterministic (because given everything we can know about how it was emitted from *A,* we cannot tell *for sure* whether it will end up at *B* or *C*). (Don't forget, on the other hand, that the *probability amplitude* evolves deterministically, according to Schrödinger's Equation.) Now, if there is an "inner program" that controls how the electron jumps around, then the jumping is deterministic after all—there is one and only one way the electron can jump—even if we humans have no direct access to the program itself. After all, the programmer of "Space Wars" could have allowed the rockets in the game to make hyperspace jumps as well, if he had wanted to.

The implicate order might also help to explain nonlocality. Imagine a fish swimming in an aquarium. The image of the fish can be refracted (bent) through the water in such a way that if we look at a corner of the aquarium we might seem to see two strangely similar fish performing closely correlated motions. But in fact there is only one fish. Perhaps in an EPR experiment, when we seem to see two particles behaving in a way that is more correlated than local physics could allow for, there is really only one particle encoded in the "inner" order.

Bohm never worked out a detailed mathematical theory of the implicate order, and it remains a tantalizing suggestion. But if it could be made to work it might restore some semblance of the determinism sought by de Broglie and Einstein, although at the price of reducing ordinary space and time to a sort of virtual reality.

Was Bohm right that we can underpin the probabilistic predictions of quantum theory with a deterministic, realistic theory so long as we are willing to use the quantum potential? Is it really true that quantum physics is just classical physics with a quantum potential added in? It would be as if everything really commutes after all; if Bohm was right, the only reason conjugate quantities do not seem to commute is because of unavoidable statistical fuzziness caused by the quantum potential of the apparatus interfering with the quantum potential of the system under observation. Some very recent work by a number of physicists suggests that the quantum potential *itself* is a mathematical consequence of quantum uncertainty; in other words, if there were no uncertainty, there would be no quantum potential. If this is correct then we could hardly hope to explain quantum uncertainties on the basis of the quantum potential, and Bohm's theory would have to be treated as yet another very useful semiclassical approximation to the "true" quantum theory that still eludes us. But these investigations remain in a very preliminary stage.

Bohm made us aware, as no one else had, of the fact that quantum nonlocality applies to the dynamics of quantum systems as well—that is, the energy of quantum systems, especially entangled states, is nonlocal. Philosophically, perhaps the most important lesson that Bohm taught us (apart from the fact that the most respected experts in a field can sometimes be wrong) is the unbroken wholeness of the physical world. Although Einstein deplored the fact, it seems that quantum mechanics shows that nothing is ever entirely separate from everything else. This is a *physical* fact that we have yet to fully acknowledge, let alone understand. Whether it validates any particular religious or mystical view is an entirely different question.

BELL'S THEOREM TOLLS

John S. Bell (1928–1990) was an Irish-born particle physicist who made what American physicist H. P. Stapp famously called "the most profound dis-

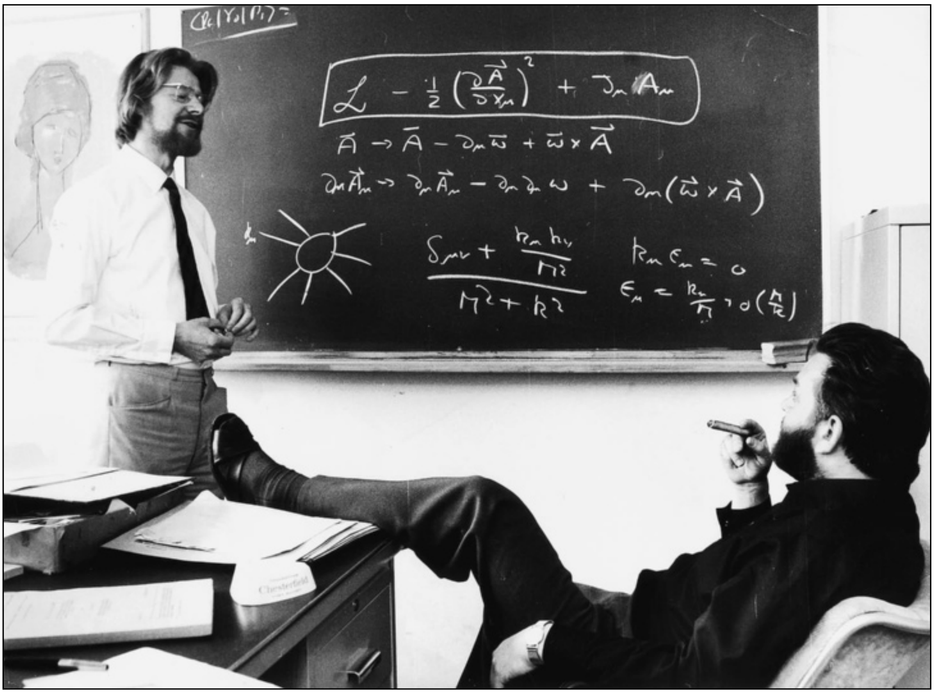

Figure 10.2: John S. Bell (on the left) with particle theorist Martinus Veltman. CERN, courtesy AIP Emilio Segre Visual Archives.

covery of science" (Bub 1997, p. 46). From Bell's writings one gets the impression that he was the sort of person who did not like to take things on authority. Although much of his work was on conventionally respectable physics, he thought deeply about the foundations of quantum mechanics. In particular, Bell wondered whether Einstein, Podolsky, and Rosen could have been right when they hinted at the completion of quantum mechanics by a theory of hidden variables.

John von Neumann had produced a proof that a hidden variables interpretation of quantum mechanics was mathematically impossible. It was hard to imagine that von Neumann could have made a mathematical mistake, and yet Bohm seemed to have done precisely what von Neumann said was impossible: he had shown that quantum mechanics (at least, nonrelativistic wave mechanics) was consistent with a picture in which particles do have definite trajectories, so long as their motions are coordinated by the quantum potential and the guidance condition. Bohm had therefore shown that a hidden variables picture of quantum mechanics is indeed mathematically possible, but in the way that Einstein himself would least have liked—namely, by invoking nonlocal (that is, faster-than-light) dynamics.

So Bell set out to solve two problems. First, where had the usually impeccable von Neumann gone wrong? Second, and much more important, did *any* hidden variables completion of quantum mechanics have to have this disturbing nonlocal character?

In 1964, Bell published a short paper in which he demonstrated the result that had so impressed Stapp: the mathematical predictions of quantum mechanics for entangled EPR states are inconsistent with *local realism,* which is the supposition that elementary particles could be programmed *at the source* with instructions that would be sufficiently complex to tell them how to respond to all possible experimental questions they could be asked in such a way as to give the predicted correlations of quantum mechanics, but *without* the particles being in any sort of communication after they leave the source. In short, if there is any explanation at all of how entangled particles stay as correlated as they do, it has to be nonlocal.

In order to demonstrate this result, Bell adapted David Bohm's 1951 version of the EPR experiment, in which spin measurements are made on pairs of electrons emitted from a source in an entangled state called the *singlet state*. Bell took the novel step of considering correlations taken with more than one combination of detector settings, and he showed how to define a *correlation function,* which expresses the relations between the measurement results at each detector in a simple way. Bell showed on very general mathematical grounds that if local realism applies—that is, if each particle can carry within itself all the information it needs in order to know how to respond to all the experimental questions it could be asked—then the correlations must obey a certain mathematical inequality, now called Bell's Inequality. Bell then calculated the correlation coefficients using standard quantum mechanics, and from his result it is easily shown that the quantum mechanical prediction violates the Bell Inequality. In other words, no *local* hidden variables theory can explain quantum correlations, but nonlocal theories (theories that countenance some sort of faster-than-light influence between the distant particles) are not ruled out.

Several physicists in the 1970s generalized Bell's Inequalities and tried to confirm his predictions directly by experiment. It is not an easy experiment to do, and the first experimental confirmation of Bell's Theorem that is considered to be decisive was performed by Alain Aspect (1947–) and coworkers in 1980. Correlated photons are sent in opposite directions along the arms of the device, where they strike polarizers. Very rapid switches change the relative angle of the polarizers while the photons are in flight, presumably ensuring that no information about one detector setting can reach the photon in the other arm of the apparatus before it hits its own detector. The Aspect experiment was thus a *delayed choice experiment,* which means that the choice of detector setting is made *after* the particles leave the source. The Aspect experiment confirmed the quantum mechanical predictions to high accuracy. Since then there have been many further tests of Bell's Theorem; in all cases quantum mechanics violates the Bell Inequality appropriate to the experimental design.

The author of this book once heard the distinguished physicist Gordon Fleming of Penn State University reminisce on the period between the publication of Bell's prediction and its confirmation by the Aspect experiments. Fleming, a field theorist with strong interests in the problems connected with nonlocal-

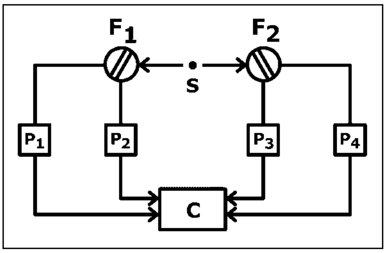

Figure 10.3: The Aspect Experiment. Pairs of entangled photons are emitted from the source S. Polarization filters F_1 and F_2 randomly either reflect or transmit the photons after they are emitted from S. Photomultipliers P_1 and P_4 detect transmitted photons, and P_2 and P_3 detect reflected photons. The coincidence counter C keeps track of the correlations between reflected and transmitted photons. (Many details are omitted!) Illustration by Kevin deLaplante.

ity, observed that during the 1960s and 1970s many physicists of his acquaintance were almost "schizophrenic" (his term) in their attitude toward Bell's Theorem: they very much wanted Bell to be proven wrong because they thought that nonlocality was crazy; on the other hand, they used quantum mechanics in their work all the time and they knew perfectly well that it, and Bell, would be proven right. Now Bell has been proven right, but the desire to have it both ways—to have quantum mechanics and to have physics be local—remains.

The title of this section is (with apologies) a double pun that appears in many variants in the foundations of quantum mechanics literature. It is based on the famous phrase "for whom the bell tolls" used by novelist Ernest Hemingway, and originally penned by the English eighteenth-century poet John Donne. One meaning of the pun is that Bell's Theorem tolls to mark the refutation of local realism; another sense of the joke is that the logic of Bell's argument is an example of a valid argument structure called *modus tollens,* which has the form *If p then q; not-p; therefore not-q.* Modus tollens is the logic behind scientific *falsification,* whose importance was emphasized (some would say over-emphasized) by philosopher of science Karl Popper (1902–1994). The way falsification works is simple: a theory makes a prediction; if experiment and observation show that the prediction is false, then the theory is false and requires repair in whole or part. Science is perpetually correcting itself by means of the feedback from experiment to theory. Bell showed that local realism predicts that quantum mechanical correlations will satisfy the inequalities that now go by his name; Aspect and many others demonstrated that the inequalities fail experimentally; therefore local realism is false. It is still unclear what, if anything, we can replace it with.

It seems unfortunate that Bell did not receive the Nobel Prize—or should we say "no-Bell" Prize?—before his sudden death in 1990. (There are in fact *two* no-Bell prizes in physics, since many people feel that British astronomer Jocelyn Bell should have won the Nobel for her part in the discovery of pulsars in 1968.)

Bell's Theorem can also be described as a failure of *common cause* explanations of quantum correlations. Suppose Alice and Bob happen to be siblings, and suppose that on 12:00 noon on a certain day they are standing next to each other. A friend notices that there are strong facial resemblances between Alice and Bob. These need not be due to any influence from Bob to Alice or vice versa *at 12:00 noon.* The similarity between their features could be due primarily to their genetic heritage, which can be traced back to a common cause in the past—in this case, their parents. Now suppose that Alice says, "Hello, Bob!" and Bob replies, "Hi to you, Alice." Bob's reply to Alice is not due to anything in his heritage *alone,* but requires for its full explanation the fact that Alice spoke to him. In the Bohm-EPR-Bell experiments, it is just as if the distant particles are "speaking" to each other, for they behave in ways that cannot be fully explained in terms of their "heritage" at their common source.

One might ask why it is not possible, no matter how unlikely, that there is some oddity in Alice and Bob's common heritage such that Alice *just happens* to say "Hello, Bob!" at noon on a certain day, and Bob *just happens* to reply as he does. What Bell proved is that while this might work for people, it is mathematically impossible to explain the apparent communication between the quantum particles this way. The fact that the two particles in the EPR experiment were emitted from a common source in a certain well-defined quantum state cannot be sufficient explanation for *all* of the details of how they behave when they encounter the detectors.

Finally, why had von Neumann made what Bell was later to characterize as a "silly" mistake? He certainly had not made any errors in his calculations. Rather, Bell showed that von Neumann had made a crucial assumption that ruled out from the beginning the very sorts of functions he needed to consider. In other words, he had implicitly assumed the proposition he had set out to prove. This error is technically known as a "circular argument," or "begging the question," and it is one of the easiest conceptual errors to fall into whenever the proposition we are supposed to be proving is something we are so convinced of that we don't quite know how to think without it.

There is still a small but active literature that seeks to find loopholes in Bell's argument or the experiments verifying his theorem, but none that have been suggested have been generally convincing. (Some critics of Bell have argued that detector inefficiencies could somehow be giving a false impression that the Inequalities are violated.) The problem for critics of Bell's argument is not only the very solid experimental evidence in support of his predictions, but the fact that the calculation of the inequality-violating correlation coefficients follows directly from the core formalism of quantum mechanics. If Bell was wrong then the quantum mechanics that has worked so well since 1926 is deeply wrong, and that just does not seem to most physicists to be a likely proposition.

It is important to see that the confirmation of Bell's Theorem is not necessarily a vindication of causal interpretations of quantum mechanics such as those proposed by de Broglie or Bohm. It is, by itself, strictly a negative result: it rules out any sort of locally realistic explanation of the correlations of entangled states, but it does not, by itself, tell us what actually makes those correlations come out the way they do. Based on what is known today, it is logically possible that entanglement could have *no explanation at all* beyond the mathematical formulas that predict its manifestations. And some physicists prefer this way of thinking about it because then they do not have to take "spooky action at a distance" seriously.

IS THERE A "BELL TELEPHONE"?

Peaceful Coexistence

Bell himself admitted that he found his theorem to be profoundly disturbing, because it seems as if the correlated particles are connected by some sort of influence moving faster than light, which Bell feared would imply that the theory of relativity might be wrong. Bell was therefore another case (like Planck, Schrödinger, and Einstein) of an innovator in quantum physics who was unhappy with what he had discovered because it shattered his conservative expectations of the way physics *should* be.

The prevailing view since the late 1970s is that despite the threat of quantum nonlocality, relativity and quantum mechanics stand in a relation of what philosopher of physics Abner Shimony (1928–) ironically called "peaceful coexistence" (Shimony 1978). The phrase "peaceful coexistence" was borrowed by Shimony from the sphere of international relations, and it suggests a state of mutual tolerance between political jurisdictions (such as the United States and the former Soviet Union) whose underlying ideologies are utterly at odds. Shimony and several other authors in the 1970s and 1980s argued that peaceful coexistence between relativity and quantum mechanics is assured because of the *no-signaling theorem,* which claims that one cannot use quantum nonlocality to signal controllably faster than the speed of light. Shimony, with tongue in cheek, suggested that quantum nonlocality should be called not action at a distance, but *passion* at a distance. Shimony (building on work by philosopher of science Jon Jarrett) made a careful distinction between what he called *Controllable Nonlocality* (sometimes also called Parameter Dependence), which would be the ability to control the nonlocal influence by means of local detector settings, and *Uncontrollable Nonlocality* (sometimes called Outcome Dependence), which is the demonstrated fact that correlations in entangled systems cannot be explained by common causes. Shimony and most other authors believe that Controllable Nonlocality is ruled out by the no-signaling theorems, and that Uncontrollable Nonlocality is sufficient to explain the violations of Bell's Inequalities.

In order to see what sort of information transmission (or "transmission") is possible with an EPR apparatus, let us arrange an EPR setup as follows: Alice and Bob will be at rest with respect to each other, but a large distance apart, and equipped with highly efficient Stern-Gerlach detectors. We will put a source of pairs of correlated electrons exactly halfway between Alice and Bob, and also at rest with respect to them, and we will have the source emit entangled pairs of electrons at regular intervals. When an electron enters the magnetic field of the Stern-Gerlach device it will be deflected either up or down. Alice and Bob will record the results as they receive particle after particle, and they may from time to time change the angles of their detectors. What each experimenter will see will be an apparently random sequence of ups and downs, like a series of coin tosses.

It will turn out that if they compare results after a long experimental run, they will find that the correlations between their results will violate a Bell

Inequality. (For electrons the correlation coefficient is given by $-\cos\theta_{AB}$ where θ_{AB} is the angle between Alice and Bob's detectors.) This means that it is mathematically impossible for their results to have been due to *preexistent* properties of each electron that they detected. It seems *as if* information is being transmitted or exchanged, faster than light speed, between the electrons in each pair of particles. This fact gives Alice and Bob an idea: is there any way that they could send messages to each other faster than the speed of light using the EPR apparatus?

Suppose that they try to test this by making the following arrangement: Bob will hold his detector at a constant angle throughout the experimental run, while Alice will turn her detector back and forth in such a way that the correlation coefficient jumps from 1 to 0 so that she can spell out a message in Morse Code. Will Bob be able to read the message? No; the most we could say is that Bob would probably detect a different random sequence of results than he would have received had Alice not manipulated her detector, but there is no way for Bob to tell this from his local measurements alone. The violations of locality appear only in the correlations between Alice's and Bob's results.

Quantum signaling with an EPR apparatus can be compared to a telephone line over which all Alice and Bob hear is static, and yet when Alice tries to speak to Bob this somehow induces correlations in the static. This means that if we made a recording of the apparently meaningless crackles received by Bob, and compared them to the apparently meaningless crackles heard by Alice, we would find that the crackles were correlated in a nonrandom way, such that it would be possible to decode what Alice was trying to say by comparison of her noise with Bob's noise. In other words, while Alice cannot send a message directly to Bob, *she can encode a message in the correlations,* and this fact is the basis of *quantum cryptography.* The key point is that it is possible for Bob to read the message, but *he has to have Alice's results in order to do so,* because the message, as noted, is built into the correlations; his own results and her own results, by themselves, still look like purely random sequences of ups and downs. And the only way that Bob can get Alice's results is by normal, slower-than-light means of transmission.

Now at this point Alice gets annoyed and decides to try something drastic. She introduces some extra magnets into the apparatus in such a way that she can force the electrons she receives always to go either up or down at her command. Surely, she reasons, if she and Bob have their detectors set at a suitable relative angle, his electrons will go *down* whenever hers go *up,* and *vice versa,* and she can send him a message. She finds, to her dismay, however, that if she tries to do this, the Bell-Inequality violating correlations *disappear,* and Bob just gets uncorrelated static no matter what Alice does with her detector. It is exactly like the double slit experiment, where if we try to determine which slit the electrons go through, the interference pattern disappears. The no-signaling theorem is the statement that this will *always* happen: the general laws of quantum mechanics guarantee that there is no arrangement of detectors that will allow Alice to utilize quantum nonlocality in order to send Bob a message faster than the speed of light.

A Quantum Fly in the Ointment

Numerous authors from the 1970s onward have published versions of the no-signaling theorem, and probably most physicists consider it to have been established beyond a reasonable doubt. However, there is a small but growing band of dissenters who question the logic of the no-signaling proofs. Briefly, the critics are worried that the standard no-signaling proofs rely upon specialized or restricted assumptions, which no doubt seemed reasonable to their authors, but which automatically rule out signaling from the beginning without giving the possibility a fair hearing. Such arguments, say the critics, simply assume what they had to prove and therefore do not rule out faster-than-light signaling in general. For example, many no-signaling proofs depend crucially on the assumption that the energy of entangled states is localized to the particles. However, as noted in the last chapter, Bohm showed that this is incorrect, although even Bohm himself did not seem to have fully grasped the implications of his own discovery. It is quite possible that the existence of the quantum potential does, in principle, allow for the possibility of a "Bell telephone," although no one at this stage has the slightest idea how it might actually be constructed.

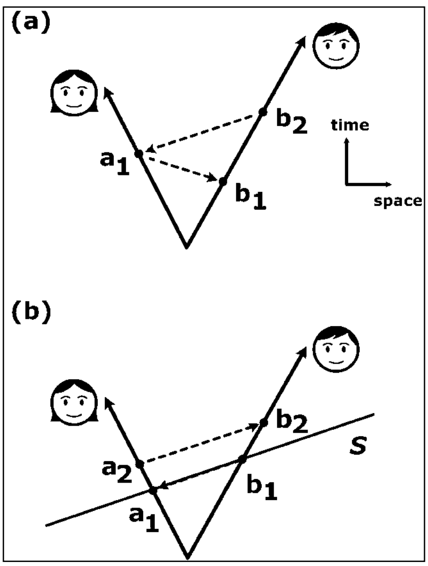

Figure 10.4: Bob Phones Alice on the Bell Telephone. In (a), Bob phones Alice faster than light. Alice's return call can arrive at Bob's worldline before he made his call, allowing for paradoxes in which Alice prevents Bob from calling her if and only if she does not. In (b), there is no risk of paradox if there is a "preferred" frame S which limits the speed of faster-than-light interactions. Illustration by Kevin deLaplante.

Causal Loopiness

A major worry about the possibility of faster-than-light signaling is that it might allow for causal paradoxes of an especially nasty sort. Suppose Alice and Bob are moving away from each other at very high speed, suppose they have quantum EPR devices that can signal at any speed faster than light, and suppose Bob wants to ask Alice out for a date. (We'll drop the assumption that they are siblings.) He sends a faster-than-light signal at point b_2 which reaches Alice at point a_1. What Bob does not know is that Alice has doctored Bob's quantum sending apparatus in such a way that she can turn it off with one of her own faster-than-light signals. Alice decides that she does not want to accept Bob's offer, and so she sends a signal to Bob that reaches his worldline at the earlier point b_1 and *turns off* his sending apparatus. It is therefore impossible for him to transmit his request at b_2. But wait!—Alice would never have sent her signal from a_1 to b_1, and thereby turned off Bob's apparatus, unless she had received Bob's signal from b_2. So Bob's apparatus gets turned off at point b_1 if and only if it is not turned off, and this is a logical contradiction.

Some physicists believe that the risk of such paradoxes, which apparently could occur whenever there is a *closed causal loop,* are sufficient to rule out the possibility of faster-than-light transmission of information, especially of a controllable kind. However, there are a few conceivable ways around the paradox.

Suppose it is not the case that Alice and Bob's faster-than-light signals could be sent at *any* velocity. Suppose there is a velocity that, although much faster than light, is still a maximum velocity for faster-than-light quantum signals. If Bob sends his message from b_1, then Alice's return message, no matter how quickly she tries to send it, will reach Bob's worldline at a point b_2, which is later than b_1 according to Bob's proper time, since neither her signals nor his can go below the line S. (See Figure 10.4 (b).) There is no risk of paradox. Many physicists are uneasy about this scenario, because the existence of S may involve a subtler violation of the principle of relativity in that it apparently defines a "preferred frame" in which the laws of physics might take a special form. On the other hand, it could be that the precise angle that S takes as it cuts through spacetime could be determined by cosmological factors, in which case there would be no violation of relativity so long as there was a proper four-dimensional description of the process. However, no one has worked out a detailed theory showing how quantum "signals" would be guided by influences from the whole universe.

A very recent model by Aharonov, Anandan, Maclay, and Suzuki (2004) seems to allow for a limited sort of nonlocal signaling, but this is still under investigation and has not yet been digested by the physics community. Aharonov's model depends on the (still controversial) possibility of "protective measurements" that do not fully collapse the wave function, and these are not covered by the standard no-signaling arguments.

BELL, BOOLE, AND PITOWSKY

A deep logical and mathematical analysis of Bell's Theorem was carried out in the late 1980s and early 1990s by the Israeli philosopher of physics Itamar Pitowsky (1950–), who showed that the Bell Inequalities are in fact special cases of mathematical inequalities first written down by the great British mathematician George Boole (1815–1864) in the 1850s. Boole was one of the founders of modern symbolic logic. The mathematics of any physical or mathematical system that has two but only two distinct "truth" values (which we can call True and False, or 1 and 0) is called Boolean algebra. Boole tried to define what could be meant by the notion of *logical consistency,* and he showed that it can be expressed mathematically by means of inequalities on correlation coefficients.

This is simpler than it sounds. Suppose Bob and Alice are examining the contents of a large urn or vat containing a large number of balls. The balls are made of several different materials and are colored differently. Bob and Alice remove the balls one by one, note their color and material, and toss them back in the urn again. Provided that nothing Alice and Bob do causes any changes to the composition or color of the balls, Bob and Alice will observe, as they build up statistics on the balls, that certain simple inequalities will hold. For instance, they will find that the frequency with which they pull out red wooden balls will be less than or equal to the sum of the frequencies with which they pull out red balls (of any composition) and wooden balls (of any color). (That's because there could be red balls made of other materials, or wooden balls of some other color.) This is an example of what Boole called a "condition of possible experience," and that's really all there is to the Bell Inequalities, although of course they can be expressed in much more general mathematical terms. Boole's inequalities could fail, however, if Alice and Bob altered the balls in various ways before they threw them back in the urn, or if the result that one person gets is somehow dependent on the result that the other gets.

Pitowsky showed that if local realism or common cause explanations hold, then observing the EPR particles is like examining balls from a Boolean urn *without changing them.* The quantum mechanical violation of the Bell Inequalities is thus a sign that quantum mechanics describes something that is inherently non-Boolean; that it, it is something that cannot have a fixed set of properties that are independent of how we investigate it. However, Pitowsky and many other contemporary authors balk at accepting the message of Bohm's causal interpretation, which is that there *literally is* a nonlocal force (almost like the "Force" of *Star Wars*) that permeates the whole universe and correlates quantum particles, no matter how far apart they may be. They prefer what might be called the "no-interpretation interpretation" of quantum mechanics: one can hope to find no deeper explanation of why particles are correlated in quantum mechanics than the mathematics of the theory itself.

The non-Booleanity of quantum mechanics was evident as early as the first work on quantum logic by Birkhoff and von Neumann in the mid-1930s, but it was proven in an especially decisive way by Simon Kochen and Ernst Specker in 1967. There is now a class of results known as Kochen-Specker Theorems; such theorems are often called *no-hidden-variables* or "no-go" theorems. Bell's Theorem of 1964 has been shown to be an example of a Kochen-Specker paradox when the quantum system under study is an entangled state spread out through space.

Schrödinger outlined the essential point of the Kochen-Specker Theorem in his cat-paradox paper of 1935, although he did not give a formal proof. The Kochen-Specker Theorem is technically complex, but the upshot can be expressed fairly simply.

Suppose we are studying a quantum-mechanical system such as a nucleus or pairs of electrons emitted from the source of an EPR device. There is a long list of experimental questions we could ask about the particles in these systems, which might include such questions as, what energies do they have? What are their spin components in various directions? What is their angular momentum? What are their electrical charges? To simplify matters, we could frame these questions in such a way that they must give either "yes" or "no" answers. For instance, we could ask questions such as, "Is the spin in the z-direction of Particle 1 *up?*" "Is the energy of Particle 6 greater than .1 electron volts?" and so forth.

Now, write a long list containing all these possible questions about the system. These will include questions about noncommuting observables. The essential content of the Kochen-Specker Theorem is that (except for a few quantum mechanical systems with an especially simple structure) it is mathematically impossible to go through this list and assign either a "yes" or a "no" answer to each question in a way that would not lead to a contradiction, where a contradiction means having to answer both "yes" *and* "no" to at least one question on the list. (Mathematically, the problem of simultaneously evaluating all of the propositions on the list is something like trying to smooth out a hemisphere onto a flat surface without a fold; it can't be done!) To put it another way, there is no consistent *valuation* (assignment of truth values) to all the questions on the list; knowing the answers to some questions on the list *precludes* knowing the answers to other questions on the list. So our uncertainty about the values of some physical parameters belonging to quantum systems does not come about merely because we don't happen to know those values; it is because they *cannot all have* yes-or-no values at one go. This is what is meant by saying that the logic of quantum propositions is *non-Boolean.*

To see how odd this is, compare it to a simple classical example. Suppose Alice wants to know how tall Bob is and what he weighs. Let's also say that Alice finds out that Bob is definitely taller than six feet. Alice naturally assumes that weighing Bob will not change the height that she just determined. But suppose that Alice measures Bob's weight and finds that he is definitely less than two hundred pounds, and then discovers that she no longer knows whether or not he is more than six feet; that is what the logic of quantum propositions is like. And it should be obvious by now that this is a consequence of noncommutativity: certain measurement operations cannot be performed independently of others. For instance, if the z component of spin of an electron is known with certainty, then the x component of spin could be either up or down with equal probability. The breakdown of Booleanity at the quantum level is due to the existence of noncommuting observables, and this in turn is due to the existence of Planck's quantum of action.

The fact that quantum mechanical systems are inherently non-Boolean tells strongly against Einstein's hope of rewriting physics in terms of an independent reality, and it also throws into question the plausibility of causal theories of quantum mechanics such as Bohm's which hope to underpin quantum statistics with an underlying Boolean mechanism. This is not to say that Bohm and de Broglie were entirely mistaken; in particular, their emphasis on the role of the nonlocal quantum potential could be quite important. But if the message of Kochen and Specker is as universal to quantum systems as it increasingly seems to be, then whatever is right about causal versions of quantum mechanics can only be *approximately* right. EPR sought to show that quantum mechanics is incomplete, by assuming that it could not be nonlocal. What is increasingly apparent is that quantum mechanics is both incomplete (because it is mathematically impossible for it to be any more complete than it is) *and* nonlocal.

11

BITS, QUBITS, AND THE ULTIMATE COMPUTER

THE ULTIMATE COMPUTER

From the mid-1970s onward physicists began to suspect that quantum mechanics could allow the construction of a radically new type of computer. The first design for a quantum computer was published by Paul Benioff in 1980. In 1982 Richard Feynman wrote an influential paper in which he approached the problem indirectly, by wondering whether an ordinary classical computer could simulate the behavior of a quantum mechanical system. He concluded that it could not, for two reasons. First, there is a barrier posed by complexity. In order to be useful, a computer has to be able to predict the behavior of systems that are more complex than it is. And yet, if we try to model the evolution of a number of particles, the complexity grows so rapidly as we increase the number of particles that any conceivable classical computer could not predict the behavior of most quantum systems as quickly as those systems can themselves evolve. Second, quantum entanglement (as in the EPR experiment) shows that quantum mechanical systems are using information to compute their own behavior that could not have been encoded within the particles themselves. And yet quantum mechanical systems quietly go on their way evolving according to the Schrödinger Equation, untroubled by Feynman's arguments. Feynman then inverted the problem, and suggested that it might be possible to construct a computer using quantum mechanical principles such as superposition to do calculations much more quickly than classical computers could, or perhaps solve some problems that classical computers cannot solve at all.

The concept of quantum computing was generalized in papers published in 1985 by the British physicist David Deutsch (1953–). Deutsch outlined a theory of a quantum mechanical version of the Turing machine, the universal computer designed by Alan M. Turing (1912–1954), one of the pioneers of modern computing theory and logic. There are several ways to model a

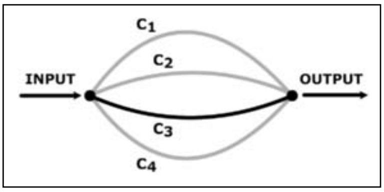

Figure 11.1: Classical Turing Machine. C1, . . . , C4 are possible computational circuits in the computing head of a Turing machine. All circuits are independent, and one is chosen at random for each computation. The probability of getting the output A is the sum of the probabilities that each possible path will be used for the computation. Illustration by Kevin deLaplante.

Turing machine as Turing originally conceived of it. The essential idea is that it is some sort of device with a memory register that records the internal state of the machine. The machine reads an infinitely long input-output device, which can be thought of as a tape. The tape is divided into discrete cells, and each cell has some data recorded in it. The information in the cells and in the machine's internal state is given in discrete, digital form. The machine is programmed with instructions aimed at performing a computation, and the instructions take the form of precise rules for replacing one bit of information in a cell with another, depending on what information is in the cell. When the machine arrives at the result it was programmed to get, it halts. (The machine might have been programmed to calculate the square root of an integer to some definite number of decimal places.) Turing showed that this seemingly mundane device is *universal* in that it can perform any algorithm whatsoever. (An *algorithm* is simply a definite set of rules or a recipe for carrying out a computation that produces a specific result.) All computers are logically equivalent to Turing's generalized computer, in the sense that whatever their differences in hardware they are doing no more computation than what the universal Turing machine can do. This implies that any algorithm that can be performed on one computer can be performed on another, although perhaps not as efficiently.

Turing also showed that his machine, although universal, has one important limitation. It cannot always tell in advance whether or not it will *halt*. That is, presented with a given computational task, it cannot determine, before it tries to complete the task, whether it will be able to do so. The only way in general to find out whether a Turing machine can do a given computation is to run it on the machine and see what happens. (There are many relatively simple problems for which the halting problem can be solved, of course; the question is whether it can be solved for *all* possible computations.) The inability of a Turing machine to solve the halting problem is closely related to the powerful incompleteness theorems of the Austrian logician Kurt Gödel (1906–1978), which (roughly speaking) say that no single Turing machine could generate all of the true theorems about the natural numbers.

The key difference between Deutsch's *quantum Turing machine* and a classical Turing machine is that in the quantum machine there is interference between possible computational pathways within the machine. Like Schrödinger's cat, the quantum computer goes into a superposition of computational states, and in each component of the superposition a version of the required computation is taking place. In simple terms the effect is massive parallelism, which allows for a huge speed-up, in principle at least. Deutsch was also able to show that his quantum Turing machine was universal in the same sense as

Turing's classical version; that is, all known types of computations can be performed on it.

The catch is that the results of a quantum computation, like all quantum processes, are inescapably *probabilistic*. Thus, while a quantum computer can come up with an answer long before an equivalent classical computer could, there would only be a certain probability that the answer would be correct. But this might be good enough for many purposes: getting an answer (like a weather forecast) that had only a 90 percent chance of being right, but getting it when you need it, might be more useful than getting an answer that is 99.9 percent likely to be right but too late to be useful. Deutsch argues that the practical design of quantum computers essentially amounts to arranging the phases of the various amplitudes of the system in such a way as to produce the desired result with the desired degree of reliability.

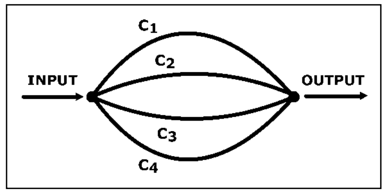

Figure 11.2: Quantum Turing Machine. C1, . . . , C4 are possible computational circuits in the computing head of a quantum Turing machine. All computational paths are used simultaneously. The amplitudes for the paths interfere whenever there is no way to tell which path was used to get the final result. The probability of getting output A is given by Born's Rule (square of the sum of the amplitudes). Illustration by Kevin deLaplante.

Just as classical information is parceled out in bits, quantum information comes in qubits and ebits. A qubit is simply a one-particle superposition, while ebits are multi-particle entangled states, and they are sometimes also called "Bell" states because of the role such states play in Bell's Theorem. Qubits and ebits are operated upon by unitary matrices, just as in the old matrix mechanics of Heisenberg and Born. (A unitary matrix is one that represents a rotation in Hilbert Space.) In quantum computation these operators are treated as quantum logic gates. They are generalizations of the classical Boolean logic gates, such as AND and OR gates, that run in computers everywhere today. From the strictly theoretical point of view, a quantum computer simply is a linear operator or series of operators designed to process quantum information in certain ways. Quantum logic gates can perform operations that do not exist in standard Boolean circuit theory or Boolean logic, such as the square root of NOT. This is a matrix that when *squared* gives the matrix that negates an input qubit. Quantum computation therefore offers a new way to think about quantum logic as a natural generalization of classical Boolean logic.

One of the most difficult problems in mathematics is factorization. This is simply the process of breaking down an integer into its prime factors; for instance, showing that $527 = 17 \times 31$. Finding the factors of numbers less than (say) 1,000 is usually pretty easy, but the difficulty of factorization mounts rapidly with the size of the number. Factorization has important applications to cryptography (code-breaking), since the security of some of the most widely used encryption systems depends upon the difficulty of factoring a large integer. In 1994 the American mathematician Peter Shor (1959–) devised a quantum algorithm that factors integers dramatically faster than any classical method yet found. So far the highest number that anyone has been able to factor using Shor's algorithm is 15, but there is little question that his method is valid.

Shor's algorithm and some other algorithms recently discovered prove that quantum computers could (if they could ever be built) enormously speed up many calculations; however, the prevailing opinion is that they cannot solve any *type* of problem that a classical Turing machine cannot solve. However, the full potential of quantum computers still remains an open and controversial question, which is the subject of intense research. It does not seem to be completely out of the question to imagine that a quantum computer might be able to solve the halting problem for itself (although again the answer would no doubt be a matter of probabilities), because there is a sense in which quantum mechanical systems have access to their own futures. It is easy to solve a problem if you already know the answer. But this possibility remains highly speculative.

There remains one major barrier to constructing practical quantum computers that could implement Shor's Algorithm for arbitrarily large integers and carry out other computational tasks that we cannot even imagine now. A quantum computer can only do what it does if it is a coherent superposition of computational states. However, to get the answer out of the computer it is necessary to make a measurement, and (at least according to the standard von Neumann account) this will collapse the wave function of the computer. The challenge, therefore, is to find a way of extracting information from the computer without destroying it in the process. Some progress has been made with quantum devices of very small size, but the general problem remains to be solved, and it may pose a challenge to the orthodox understanding of the measurement process.

Like string theory, many of the claims of quantum computation are not yet verified experimentally. Unlike string theory, however, quantum computation is a straightforward application of well-established rules of quantum mechanics, and few doubt that it will work if a few practical problems can be solved—especially finding a way to get the information out of the computer without collapsing it.

Too Many Worlds?

How could a quantum computer, which presumably is going to be instantiated on a rather small physical system of finite volume, perform calculations so quickly? Deutsch has controversially suggested that the answer to this question lies in one of the most startling interpretations of quantum mechanics, suggested by the American physicist Hugh Everett III (1930–1982).

In 1957, working under the direction of his Ph.D. supervisor John A. Wheeler, Everett produced a novel solution to the measurement problem of quantum mechanics. This problem is to explain how it is that quantum states in superpositions appear to experimenters who interact with them as if they have definite classical outcomes. Like many physicists, Everett was unhappy with the von Neumann collapse postulate and the arbitrary quantum-classical divide, and he proposed that the simplest way to resolve the measurement problem was to suppose that reality is nothing more than a quantum wave function, and that *the wave function does not collapse.* As Schrödinger emphasized, if a classical measuring device interacts with a system that is in a superposition, the wave function of the measuring device (and the experimenter who runs it) becomes correlated with each component of the superposition. (Technically, the resulting wave function is a tensor product state of the observer and the observed system.) What bothered Schrödinger is that real observers do not see superpositions, but only definite results (such as either a definitely alive cat or a definitely dead cat). Everett's answer was simplicity itself: the observer together with his or her apparatus actually does split into two components (one who perceives a dead cat, the other who perceives a living cat)—but each version of the observer is correlated with the corresponding component of the cat. That is, the new state is a superposition of two states, one with an observer perceiving a live cat, and one with an observer perceiving a dead cat. Each observer seems to perceive a definite cat-state and not a superposition. However, these two components do not interact with each other in any way. It is exactly as if the universe has split in two. Every time one system becomes correlated with another system that is in a superposition, the world splits into as many versions as there are components of the observed system, and so on, *ad infinitum.*

Everett at first called his theory the *relative state* formulation of quantum mechanics. By this he meant that every observer could be in a number of different states, each one defined relative to a state of the system being measured. This later became known as the Many-Worlds Interpretation of quantum mechanics, and Deutsch prefers to call it the *multiverse* interpretation, because on this view there literally is a colossal multiplicity of universes, multiplying exponentially or faster into more universes, with each universe playing out every possibility that is consistent with the laws of physics.

It is a dizzying vision, but Deutsch himself seems to take it literally despite its conflict with common sense. The multiverse view is preferred by some especially mathematically oriented physicists because it does away with von Neumann's arbitrary-seeming collapse postulate. Most physicists are agnostic or dismissive towards the multiverse view. However, Deutsch believes that there is an *empirical* argument for the multiverse theory. He points to the fact that a huge speed-up of calculations is possible with a quantum computer and argues that there must be somewhere that those calculations are taking place. Quantum computing, Deutsch argues, is just a kind of massively parallel computation, with all the speed-up advantages of parallelism. Any computation is something that takes place on a physical platform, and if all those bits are being crunched then there has to be some physical thing that is crunching them. Deutsch thinks it is clear that they are not taking place in the spacetime that we perceive, because Feynman and others showed that this is impossible. There just isn't enough room for it. So the calculations must be taking place in parallel universes. He challenges anyone who cannot accept the multiverse theory to come up with another explanation of quantum computing; if you don't believe in the multiverse, challenges Deutsch (1997), then where are all those calculations taking place?

QUANTUM INFORMATION THEORY: IT FROM BIT?

Recent intense interest in quantum computation has drawn attention to the nature of information in quantum theory. Some recent authors have argued that quantum mechanics is nothing other than a new form of information theory, and that the whole world, or at least the world as we can know it, is nothing but information. Although this idea has become current since the quantum computing revolution of the 1990s, it was expressed as far back as the 1960s by John Archibald Wheeler, who suggested that if we better understood the relationship between information theory, logic, and physics, we would see how to deduce "it from bit."

Does it make sense to think that the world could be made of information? There is a certain mystique surrounding information, but mathematically it is a very simple idea. Classical information theory was formulated in the late 1940s by Claude Shannon (1916–2001) of Bell Laboratories. Shannon studied the efficiency of classical communications devices such as telephones and radio, as part of a general mathematical analysis of communication. Consider a binary system (one that can be in one of two states), like a coin that can be heads or tails. There are $2^3 = 8$ possible combinations of heads or tails for three coins. One needs to know three facts to specify the state of the three coins (that is, whether each is a head or a tail). But 3 is just the logarithm to base two of the total number of combinations of the three coins. Shannon argued that the logarithm (to some convenient base, usually 2) of the number of arrangements of a system is the *information* contained in that system; classical Shannon information, therefore, is merely a logarithm. The great mathematical convenience of logarithms is that they make it easier to think about quantities of information, for logarithms are *additive:* if the number of possibilities (often called the *multiplicity*) in a system is multiplied, the increased information capacity is found by simply adding the logarithms of those numbers.

The qubit is a natural quantum generalization of a classical bit of information. The classical bit can be in two distinct states, while the quantum bit is in a superposition of states. The problem is that no one has so far found an obvious interpretation of the qubit as a logarithm, and it is therefore unclear that the interpretation of quantum states as measures of information has gone as far as it can.

Rolf Landauer (1927–1999) was a senior research scientist at International Business Machines (IBM) who made important contributions to computational theory. Landauer insisted that "all information is physical," by which he meant that if there is information present then it has to have been encoded in some form of mass or energy. There is no such thing as "pure" information except in the ideal world of mathematics. It is possible that this casts doubt on Wheeler's idea that the world could be built up out of pure information, since there is no information, according to Landauer, without a physical substrate to encode it in.

Landauer made an important contribution to a question that had dogged physicists and communications engineers for many years: what is the minimum cost in energy of a computation? The operation of electronic components such as transistors produces waste heat, and waste heat is a barrier to making circuits smaller and more powerful. Circuit manufacturers constantly strive to produce components that waste less energy. However, there had been a long debate in physics and computing theory about how far they can hope to go with this.

In 1961 Landauer proved that even if there were circuit components that were ideally efficient, any computation in which bits of information are discarded must inevitably waste a minimum amount of heat. It is the destruction of information itself that costs energy; every bit of energy lost leads to the production of a minimal amount of waste heat. Consider the logic gate known as an OR gate: this transforms any of the inputs (1,1), (1,0), or (0,1) into the output 1. Information is lost in the operation of the OR gate, since from the output 1 we cannot tell which of the three possible inputs was used. And therefore by Landauer's Principle the operation of an OR gate will inevitably result in the loss of a small amount of energy, no matter how efficient we make the components.

Surprisingly, a number of computer scientists in recent years have shown that it is theoretically possible to construct fully reversible logic circuits in which unneeded bits are shunted to one side and recirculated. This prevents the loss of energy predicted by Landauer, and it means that an ideally efficient computer could operate with no heat losses at all (except, again, one has the problem of getting useful information out of it). Surprisingly, reversible computing can be done, at least in principle, with classical circuit elements, although it would be difficult to build and unnecessarily complicated for most practical purposes. An ideal quantum computer is also fully reversible, because it is simply an example of a quantum system evolving in a unitary way. A quantum computer is therefore similar to a superfluid or superconductor. In liquid helium, for instance, it is possible to set up a frictionless circulation pattern that will flow forever so long as it is not interrupted, and so long as the temperature of the fluid is kept below the critical point. This fact again emphasizes why it is difficult in practice to build a quantum computer, because like other coherent states such as superconductors they are very sensitive to disruption.

Entanglement as a Resource

Schrödinger had speculated that entanglement would fade away with distance, but all experimental evidence to date suggests that entanglement is entirely independent of distance. This is precisely what theory indicates as well, because entanglement is purely a function of phase relationships within the wave function of the system. While phase coherence can *vary* with distance (depending on the structure of the wave packet) it does not necessarily have to decrease with distance. As far as it is known now, the entanglement in the singlet state, for instance, could persist to cosmological distances if the particles were not absorbed by something along the way.

It has been widely noted in the past 10 years or so that entanglement has properties remarkably similar to energy. Entanglement can be converted into different forms and moved around, in ways remarkably like energy. Quantum information theorists often speak of entanglement as a *resource* that can be used to store or transmit information in various ways. There is not yet a general agreement about how to define units of entanglement, however.

Very recent thinking suggests that Landauer's Principle could be used to show that entanglement does have energy associated with it. When particles are entangled they are correlated, which means that if something is known about one particle, it is possible to infer information about other particles in the system. Information that manifests itself through correlations is sometimes called *mutual* information. Another way of stating Bell's Theorem is that the distant particles in an EPR experiment can possess *more* mutual information than could have been encoded in their correlations at their common source.

If one particle is measured, then by the von Neumann rule the entanglement disappears and any nonclassical correlations disappear. This means that information is lost, and on the face of it this means that waste heat has to be released, by Landauer's Principle. Since the local energies of the particles (their kinetic and potential energies) do not necessarily change, the waste heat has to be coming from somewhere else. Because it is produced precisely when entanglement is destroyed, it seems sensible to suppose that entanglement itself has an energy associated with it. This energy will be a property of the entangled state as a whole and will not be localized to the particles, just like Bohm's quantum potential for entangled states. (In fact, entanglement energy and the quantum potential might be the same thing.) This is front-line research and it is potentially controversial, especially since there has been recent criticism of the accuracy of Landauer's argument. These questions are now in process of careful reexamination; stay tuned!

OTHER CURIOUS QUANTUM CREATURES

From the 1960s onward several other startling new applications of quantum mechanics appeared in the literature. None of these developments involved any change in the fundamental structure of quantum theory that had been laid down in the 1920s and 1930s by Dirac, von Neumann, and others, although some pose a challenge to orthodox measurement theory. They all show that we have only begun to see the possibilities inherent in quantum mechanics.

A few interesting developments are sketched here. (See Aczel 2002, Johnson 2003, McCarthy 2003, or Milburn 1997 for more detail.)

Quantum Cryptography

As described in the last chapter, Alice and Bob cannot use an EPR device to signal faster than light as far as we know (although there are some theoretical doubts about this point). Alice, however, can build a message into the correlations, because the correlation coefficient between her results and Bob's depends on the relative angle of their detectors. The catch is that the message can only be read by someone who has both sets of results. Each set of results in isolation looks like a totally random sequence of ups and downs, like the results of a toss of a fair coin repeated many times, but each result set is the key for the other. Quantum mechanical entanglement allows for the most secure method of encryption known, because it depends on quantum randomness.

Suppose Eve decides to listen in to Alice and Bob's communication by intercepting some of the particles sent out from the source. Her eavesdropping can be detected by the tendency of the correlations to obey a Bell Inequality, because eavesdropping will destroy the correlations, just as in the double slit experiment we destroy the interference pattern if we try to tell which slit the electrons go through. But Eve might decide it was worth it if she can get away with a little bit of the message, even at risk of getting caught in the process.

It is also remotely conceivable that the quantum potential of Bohm could be used to eavesdrop, but this remains an open and highly speculative question. However, quantum cryptography is one of the most active areas of current research.

The GHZ State

In 1989 Daniel Greenberger (1933–), Michael Horne (1943–), and Anton Zeilinger (1945–) (GHZ) described a theoretical spin state of three entangled particles, which permits an especially vivid illustration of Bell's Theorem without the use of inequalities. The assumption of local realism about the GHZ state produces an outright contradiction with the quantum mechanical predictions for the three-particle state using, in principle, only one measurement. The GHZ has been very recently created, and it is the most direct verification yet of the failure of local realism.

Quantum Teleportation

One of the most amazing applications of entanglement is quantum teleportation. This was theoretically predicted by IBM Research Fellow Charles Bennett (1943–) and several others in 1993, and has been demonstrated by a number of research groups since then.

The invention of teleportation was stimulated by the *No-Cloning Theorem,* which states that it is impossible to copy a quantum state if the original state is to be preserved. (The No-Cloning Theorem was arrived at in response to a faster-than-light signaling scheme proposed by physicist and author Nick Herbert in the 1970s, which involved beating the light barrier by copying quantum states. Herbert's method won't work as he designed it, but its history shows that much useful thinking can follow from a productive mistake.) Several physicists realized that a state can be copied and moved anywhere else in the universe if the original is destroyed, and *so long as the sender of the transmission does not try to read the message,* because that would collapse and thereby destroy the state of the message.

The key to quantum teleportation is the use of two channels of information, an ordinary classical channel (any normal means of communication, such as a cell phone, that sends information no faster than the speed of light) and an entangled state that acts as the carrier of the information to be teleported. Alice and Bob are each poised to observe the state of distant particles belonging to an entangled EPR pair. The object is for Alice to transmit the state of a particle to Bob. Alice takes the particle *without looking at it* and allows it to interact with her EPR particle. This entangles the target particle with Alice's EPR particle and thereby entangles it with Bob's EPR particle as well. Alice then makes a joint measurement on her EPR particle and the target particle and then phones Bob on her cell phone and gives him the (apparently random) result. Alice's measurement collapses her local state and thereby erases the information about the target particle, but if Bob makes a certain kind of measurement, and then combines his (apparently random) result with the information Alice gave him by cell phone, he can infer the state of the teleported particle. This process is therefore very similar to quantum cryptography, in which a highly nonclassical quantum effect can be exploited only with the aid of a classical channel of information. Physicists who work with quantum teleportation hasten to add that it will be a very long time, if ever, before it is possible to teleport an astronaut off of the surface of a hostile planet.

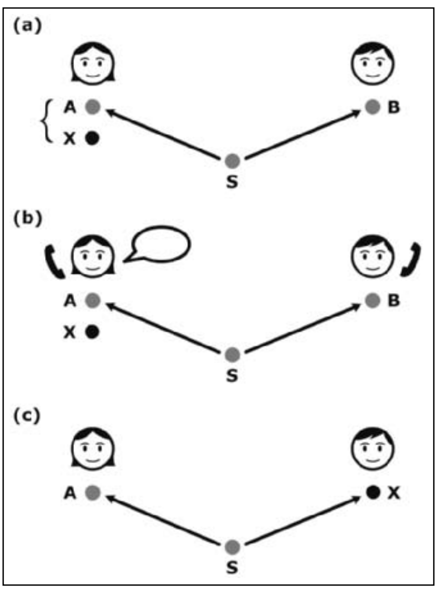

Figure 11.3: Quantum Teleportation. Particles A and B are entangled EPR pairs emitted from source S. X is the unknown particle whose state is to be teleported from Alice to Bob. In (a), Alice performs a measurement on A and X together, which entangles X with A and B. In (b), Alice sends her measurement results to Bob via a classical channel. In (c), Bob measures B and combines Alice's data with his own to reconstruct X. The original X is collapsed by Alice's measurement; hence X has not been cloned; rather, B has been transformed into X by Bob's measurement. Illustration by Kevin deLaplante.

Quantum Non-Demolition

Recently the distinguished Israeli physicist Yakir Aharonov has challenged one of the most basic dogmas of quantum mechanics, the view that any measurement on a superposition of states collapses the wave function into a single pure state. Aharonov and others are exploring the possibility that there could be *nondemolition* measurements that could extract information from a quantum state without collapsing it. A nondemolition measurement involves an *adiabatic perturbation* of the system being measured; this means that the system is interfered with very slowly by means of a very weak interaction. There is evidence that it is possible to extract some (though likely not all) information from a quantum system by means of such very gentle measurements without completely collapsing the state.

If this can be made to work reliably, it opens up the possibilities of both superluminal signaling and quantum computation. The requirements for quantum computation are remarkably similar to the requirements for signaling: in either case, one has to create a coherent entangled state that can somehow process or transmit information without being collapsed in the process. The standard arguments against faster-than-light signaling in entangled quantum states have no relevance to this process, since they do not allow for the possibility of the kinds of measurements that Aharonov and others are considering. A superluminal signaling device might therefore turn out to be nothing more than a quantum computer that extends over a large distance in space. The major difference is that superluminal signaling is regarded by most physicists as highly undesirable because it would (they think) mark the demise of special relativity, whereas quantum computing is regarded as a highly desirable outcome. If Aharonov and others who are investigating nondemolition measurements are right, it may turn out that we cannot have quantum computation without having a Bell telephone as well.

12

UNFINISHED BUSINESS

Quantum mechanics is by far the most successful physical theory ever devised, and it is also the most revolutionary, because it poses a profound challenge to conceptions of space, time, causality, and the nature of reality itself that have seemed beyond question since the beginning of the modern scientific era. Many authors have observed, though, that if quantum mechanics is revolutionary, it is an unfinished revolution. This story concludes by describing some of the unfinished business facing today's young physicists.

QUANTUM MECHANICS AND THE MIND

One of the most intriguing frontiers is the possible interactions between quantum mechanics and neuroscience. This line of investigation was stimulated by Eugene Wigner (1902–1995), who, along with Hermann Weyl (1885–1955), pioneered the use of group theory in quantum mechanics and field theory, and whose many contributions to quantum physics earned him a Nobel Prize.

In 1961 Wigner published an intriguing essay, "Remarks on the Mind-Body Question," in which he explored the possibility that the collapse of the wave function is brought about by consciousness. Schrödinger's cat paradox illustrates the puzzling fact that the dividing point between the quantum and the classical description seems to be entirely arbitrary: when the box is opened the experimenter (let us say it is Bob) sees the cat in a definite state, and yet the theory says that Bob and the cat both go into a superposition (a tensor product, technically) of the cat's state and the experimenter's. As Everett pointed out, there is no inconsistency in these two descriptions as far as they go: Bob might be in a superposition, but in each component of it he seems to perceive a definite cat state. But now, argued Wigner, suppose that Bob's friend Alice enters the room. If Bob asks her what she saw, Wigner argues that she is not going to report that she perceived a superposition; rather, she either definitely saw Bob with a dead cat or definitely saw him with a living cat. Human beings never have conscious experience of quantum superpositions, Wigner insisted. He thought that this showed that the quantum buck stops at the point at which the wave function interacts with a conscious mind.

Wigner believed that he had presented an argument for *dualism,* which is the claim that mind and body are of essentially different natures, so that mind is not subject to ordinary (quantum) physical law. Dualism has a long history in philosophy and religion, but the working hypothesis of most modern neuroscientists is that mind is purely a manifestation of physical activity within the brain and sensory system of a living being. Other contemporary scientists who have explored the possibility that quantum mechanics could be important for understanding the mind, such as H. P. Stapp, Stuart Hameroff, and Roger Penrose, work within the materialistic camp; that is, they do not advocate dualism, but instead argue for the importance of quantum mechanics in understanding the physics of the human neurosystem.

Roger Penrose (1931–) is a multitalented British mathematician, best known for his work with Stephen Hawking on general relativity. Penrose believes that quantum mechanics is needed to explain not only consciousness but the ability of the human mind to solve problems creatively. He suggests that microtubules, tiny strand-like objects with a very regular, period structure, which occur within neurons and other cells, could be the site of macroscopic-scale quantum coherence; in effect, Penrose proposes that the brain may be in part a quantum computer.

The majority of physicists and neuroscientists doubt that quantum mechanical coherence could play a role in the operations of the brain, for the simple reason that the brain is too hot. Quantum-coherent states of matter, such as Bose-Einstein condensates, superfluids, and superconductors, are typically very cold, whereas the human brain operates at temperatures around 37°C. However, the nonlocal correlations observed by Alain Aspect were in systems at normal room temperature, and there is no reason to think that quantum mechanical correlations in general are temperature-dependent. The question of whether quantum mechanics could have anything to do with whatever it is that allows brains to generate the conscious mind remains open.

QUANTUM COSMOLOGY

Quantum cosmology is the application of quantum mechanics (which arose out of the study of the smallest possible physical entities) to the largest object we know, the universe itself. This story begins with the remarkable discovery of cosmic microwave background radiation in 1965 by Arno Penzias (1933–) and Robert Wilson (1936–). Penzias and Wilson were telecommunications engineers with Bell Labs, and they were trying to eliminate an annoying hiss that was being picked up by their microwave antennas. They discovered that the hiss was due to a microwave signal reaching the Earth from all directions in space, and found that the spectral distribution (the curve of its energy as a function of frequency or wavelength) of this background radiation follows Planck's curve for the spectral distribution of a blackbody. The observed universe is therefore a blackbody cavity; in other words, the universe as a whole is a quantum mechanical object. There is therefore a deep connection between the physics of the small and the physics of the very large.

As an important example, it is highly likely that quantum mechanics may play a role in explaining *cosmic acceleration*. Probably the most surprising scientific discovery in the past ten or twenty years was the finding in 1998 by several teams of astronomers that the Hubble expansion of the universe is actually *accelerating*. It would be as if you threw a baseball straight up in the air and saw it accelerate upwards, to your surprise, rather than fall back down. Unless energy conservation is being violated on a massive scale, there has to be some presently-unseen source of *dark* energy that is causing the universe to accelerate. There is still no convincing explanation of the nature of dark energy, except that it almost certainly has something to do with the quantum mechanics of the vacuum itself. See Kirshner 2002 for an introduction to dark energy and its impact on modern cosmology.

Cosmologist Andrei Linde (1948–) has offered a startling speculation that shows how deep the connection could be between the laws of quantum mechanics and the history of the universe. The second law of thermodynamics tells us that the entropy of the universe must always be increasing, as the universe interacts with itself over and over and steadily randomizes itself. There are two linked puzzles faced by any version of the Big Bang cosmology. First, according to the Big Bang theory the universe must have started from a very low entropy state, but on the face of it this seems to be a violation of the Second Law of Thermodynamics. What physical mechanism could have gotten the entropy of the universe so low to begin with? Second, and more basic, how could something, namely a whole universe, come from nothing? Linde's clever but disturbing suggestion is that the whole universe itself might be merely a quantum fluctuation in the vacuum. As Einstein showed long ago with his theory of Brownian motion, fluctuations can and do occur, and they amount to localized pockets of temporarily lowered entropy. It is a purely probabilistic process; even in a totally undisturbed quantum vacuum there is a probability (no matter how small it might be) that an entire universe could pop out of pure nothingness if one waits long enough. It only had to happen once!

Whether or not Linde's ingenious speculation is right, it is clear that the nature and origin of the universe itself has become a problem in quantum mechanics.

THE QUEST FOR QUANTUM GRAVITY

Most physicists today agree that the central problem facing modern physics is to clarify the relation between quantum mechanics and relativity and in particular to construct a quantum theory of gravity. But the quest for quantum gravity poses technical and conceptual challenges that may be among the toughest faced by physics so far.

Understanding gravity better than we do now is not merely of theoretical interest. It is conceivable that quantum gravity might some day lead to the ability to control gravitation (perhaps making controlled fusion possible), or to other effects that we cannot presently imagine or that still belong only to the realm of science fiction. Most physicists prefer not to go so far out on the limb of speculation, but such possibilities have to be kept in mind. One thing that the history of quantum mechanics has demonstrated is that purely theoretical attempts to resolve contradictions or fill gaps in understanding can lead to unexpected practical consequences. There are, in the long run (although sometimes only the very long run), fewer more practically important enterprises than theoretical physics.

Early Efforts

One of the first approaches to quantum gravity was to treat it as a problem in quantum field theory. This meant writing a perturbation series starting with the flat-space metric (the function that determines the geometry of spacetime) as the first term and trying to find first and higher-order corrections. By the mid-1930s several theorists were able to show that if there is a particle that mediates gravitation, it has to be a spin-2 boson, massless and therefore moving at the speed of light. This hypothetical quantum of the gravitational field was dubbed the *graviton.* No such particle has ever been detected, and it would be very difficult to do so because its interactions with matter would be so weak.

Standard quantum field-theoretic methods in quantum gravity are recognized as provisional, since they are *background-dependent,* meaning that like most kinds of quantum field theory they assume a fixed Minkowski spacetime as a backdrop. This is inconsistent with the message of Einstein's general relativity, which teaches that mass-energy and spacetime geometry are inextricably entwined. One of the earliest to realize this was the brilliant Russian scientist Matvei Bronstein (1906–1938), who in 1936 outlined an early quantum theory of gravitation and argued that it may be necessary to go beyond spatiotemporal concepts in physics. Bronstein, tragically, was murdered by the Soviet secret police at the age of 32 during one of Stalin's purges.

It is impossible to give here a comprehensive picture of the many ways in which quantum gravity has been explored since the 1930s until now. Prominent names in this field include Bryce DeWitt (1923–2004), John Archibald Wheeler (1911–), Abhay Ashtekar (1949–), and numerous workers in string theory including Edward Witten (1951–). The fundamental problem with any quantum theory of gravitation that has been attempted so far is that such theories are all nonrenormalizable. Unlike the electromagnetic and Yang-Mills gauge fields, it seems to be impossible to juggle the infinities in quantum gravity so that they either cancel out or can be ignored. The physical basis for this mathematical problem is the nonlinearity of gravitation. The simple fact is that gravitation itself gravitates. The gravitational field has energy and thus has a gravitational effect, while the electromagnetic field, although it transmits electromagnetic interactions, is not itself electrically charged. (To put it another way, the photon itself is not electrically charged, while the graviton must itself gravitate.) Electromagnetic and Yang-Mills fields, if they are written the right way, can be made to add up linearly; gravitational fields add up nonlinearly. This introduces a whole new level of mathematical complexity beyond anything dealt with in quantum field theory, and so most approaches to quantum gravity so far have been linearized approximations to a physics that is profoundly nonlinear.

A Historical Perspective on Background Dependence

In 1905 Einstein erected the special theory of relativity on the assumption that the speed of light is an invariant, a quantity that must be the same for all observers in all possible states of uniform motion. Einstein did this because he believed that Maxwell's equations were more fundamental than Newtonian dynamics, so that the Newtonian picture should be modified to be consistent with electromagnetism. For many years physicists had been trying without success to explain electrodynamics in terms of classical mechanics; in 1905 Einstein turned the problem around and, instead of trying to make electrodynamics fit classical mechanics (a round peg in a square hole if there ever was one), modified classical mechanics in order to fit electrodynamics. The speed of light in vacuum should be a universal constant because it appears as such in Maxwell's Equations.

Einstein's approach was brilliantly successful, and up to the present time it has been assumed by most (though not all) physicists that quantum mechanics has to be kept consistent with the theory of relativity. However, as in 1905, it may be necessary to turn the problem around and, just as Einstein rewrote Newtonian theory to make it consistent with electrodynamics, rewrite our spacetime theories to make them consistent with quantum mechanics. There is more and more evidence that the world is *quantum all the way down.* However, there is a contradiction in twentieth-century physics that was apparent in Einstein's pioneering papers of 1905, but never resolved. In 1905 Einstein also suggested that the wave-like behavior of light is only a statistical phenomenon. If this is right, then Maxwell's theory itself, and (a crucial point) the symmetries that it obeys, could well be only statistical averages. If this is the case, then there might not be any reason at all to suppose that detailed quantum interactions are exactly Lorentz invariant. There is a parallel to the challenge faced by Boltzmann and Planck in the late nineteenth century: the rules of thermodynamics were originally formulated as exact differential laws applying to definite mathematical functions, but it became apparent they had to be understood statistically and were thus not exact (except for the First Law of Thermodynamics, to which no exceptions have been found, a fact that certainly would have pleased Planck).

As described earlier, physicists put off the problem of quantizing spacetime until the 1930s, when the infinities of quantum electrodynamics made it impossible to ignore the possibility that the smoothness of the background metric might break down at small enough distances or high enough energies. During the 1930s both Heisenberg and Schrödinger explored the possibility that space itself might be quantized, meaning that there would be a fundamental quantum of length just as there is a quantum of action. This would automatically cut off the divergences, at least in quantum electrodynamics. However, it surprisingly became possible to again put off the problem, because of the success of renormalized quantum field theory.

Many physicists now argue that there is no way to avoid the ultimate breakdown of background-dependent theories, since there is a distance range, 10^{20} times smaller than the nucleus, at which gravitation has to equal or exceed all other known forces in strength. This distance is called the *Planck length,* because it is based on Planck's proposal in 1899 that physics could be expressed in combinations of fundamental "natural" units that (unlike the meter, inch, or second) would be independent of the accidents of human history. Corresponding to the Planck length (or time) is a conjugate *Planck energy,* around 10^{16} TeV. This is vastly beyond the range of any conceivable Earth-bound particle accelerator, so anything we can say about processes at the Planck scale would have to be tested by their indirect effects—at least, for the time being!

Very recently a number of physicists have been exploring the possibility that Lorentz invariance might break down at very high energies, perhaps near the Planck energy. This implies that the vacuum would be *dispersive* at such energies, meaning that the speed of light would vary slightly at very high frequencies. (Lorentz invariance implies that the vacuum is a nondispersive medium, which means that the speed of light is the same for all frequencies.) Attempts are now being made to write versions of Einstein's special relativity that could take high-energy dispersivity of the vacuum into account, and these new versions of relativity—called Doubly Special Relativity—may disagree with the predictions of standard special relativity at high enough energy. (For an accessible introduction to recent work on Doubly Special Relativity, see Smolin 2006.)

In some respects the quantum mechanics that grew up from 1925 to 1932 represents a retreat from Heisenberg's bold vision on Heligoland. In his great paper "A Quantum-Theoretical Reinterpretation of Kinematic and Mechanical Relations," Heisenberg rewrote position and momentum as linear operators built up out of transition amplitudes between observable energy states. A particle only has a (discrete!) spectrum of possible positions when it is observed in an experimental context in which its position matrix is diagonal. Heisenberg thereby demoted position from its privileged position as the unchanging background of physics, Newton's absolute space, and made it just one of many quantum observables, any one of which can be written as functions of the others. However, shortly thereafter, by finding ways to treat continuous observables quantum mechanically, Dirac made it possible for position and momentum to be treated more like classical variables than perhaps they really are. By 1927 Heisenberg himself had retreated to a more conservative position in which space and time are continuous quantities.

Carlo Rovelli has recently argued that in order to satisfy Einstein's principle of general covariance, the foundation of general relativity, we have to construct a picture in which no observable (time, space, energy, or momentum) is privileged in the sense of being an independent c-number (classical) parameter. Instead, Rovelli insists, there should be a complete democracy of coordinates, in which all observables would be intertranslatable, and which ones are most useful would depend simply on the experimental context.

Another barrier to quantum gravity is that in the mid-1920s Pauli showed that there are severe technical barriers to constructing a quantum mechanical *time* operator, and no universally acceptable way has yet been found of getting around these difficulties so far. The result is that time is still treated like a c-number while virtually all other measurable quantities are q-numbers (quantum operator). Penrose has argued that there is no hope of constructing true quantum gravity until a way is found to construct a genuine quantum mechanical time operator, because in any relativistic theory space and time must be comparable and interchangeable. This problem also remains unsolved.

Loops or Strings?

Physics today is experiencing a tension between continuity and discontinuity that is very similar to the situation in which Planck and his contemporaries found themselves around 1900. The dominant approach to quantum gravity in the past 15 or 20 years has been string theory, based on continuous background spacetime, since it seems to predict the graviton in a natural way. However, another theory, called loop quantum gravity, is being worked on by an increasing number of theorists led by Rovelli and Smolin. This approach divides space and time up into discrete cells (Smolin calls them "atoms of space") with a quantized spectrum of possible volumes. There is a smallest possible non-zero volume, just as any discrete spectrum of eigenvalues has a smallest possible value, and the attempt to probe this volume with a high energy probe would simply create more cells. The cells can be combined into *spin networks,* based on ideas due to Roger Penrose, and each possible combination of spin networks is a possible quantum state of space itself. The key point is that Smolin's spin networks are not structures *within* space, like the particles of conventional quantum field theory; instead, space itself is built up out of them. The fact that space is discretized eliminates, in principle at least, the need for renormalization.

Despite its great conceptual attractiveness, loop quantum gravity still has not produced much more in the way of testable predictions than has string theory, and both approaches (as well as some others that cannot be described here) are being pursued vigorously.

Gravitation and Thermodynamics

Work in the past 35 years has shown that there are profound connections between gravitation and thermodynamics.

The Israeli physicist Jakob Bekenstein (1947–) noticed in the 1970s that there is a very odd parallel between the area of a black hole and the behavior of entropy. A black hole is a mass that has collapsed to within its own *event horizon,* a mathematical surface surrounding the hole over which the escape speed from the object is the speed of light. The area of a black hole is the area of its event horizon. If two black holes coalesce, their total area is always equal to or greater than the areas of each separately, and this is just like the entropy of two volumes of gas, which if the volumes of gasses are mixed must always be greater than or equal to the sum of the entropies of the separate volumes of gas. It seems as if black hole area, like entropy, can never decrease.

Bekenstein proposed a formula for the entropy of a black hole as a function of its area. The physical meaning of this is that the entropy of a black hole is a measure of the amount of information that has disappeared inside it. A black hole of a given mass is the highest possible entropy state for that mass. Many scientists have speculated that as the universe ages its entropy must gradually increase until all temperature differences in the universe are smoothed out and there is nothing left but a uniform diffuse gas, everywhere. This is called the "heat death" of the universe. Black hole thermodynamics implies that if there ever is a heat death of the universe, it will be much more dramatic. Since the highest entropy state of matter is a black hole, the "heat death" of the universe would have to amount to collapse into a vast black hole. In the 1960s Hawking and Penrose had proven, using the austere mathematics of general relativity, that gravitational collapse is inevitable: any mass or energy if left to itself must eventually become a black hole. There is therefore a deep consistency between Einstein's general relativity and thermodynamics, and it is possible (though not yet proven) that gravitation itself is nothing other than a manifestation of the tendency of entropy to increase.

Bekenstein's insights paved the way to the quantum mechanical treatment of black holes. Following Bekenstein, it is known that a black hole has a definite entropy. Furthermore, a black hole is an object that (because of its intense gravitational field) absorbs all radiation that falls upon it and therefore fulfills Kirchhoff's old definition of a blackbody. In order to be in thermal equilibrium with its surroundings it must, as Kirchhoff showed long ago, emit radiation as well as absorb it. It is just a short step to conclude that a black hole has to have a temperature and has to radiate energy with a Planck spectral distribution, and this is precisely what was shown mathematically by Stephen Hawking in 1974. Physics has thus returned to its roots in the work of Kirchhoff and Planck, with another one of those predictions that seem obvious in retrospect but that were surprising at the time.

Hawking further showed that the temperature of a black hole is inversely proportional to its mass. A black hole with the mass of the sun will have an undetectably low temperature, while black holes with the mass of a proton would radiate energy with a peak in the gamma range and would in effect detonate in a flash of lethal gamma radiation in a fraction of a second. Hawking arrived at this startling conclusion following speculation that there could be mini-black holes left over from the Big Bang, still floating about in the universe. He showed that proton-sized mini-black holes would evaporate in a flash of gamma radiation almost as soon as they were created.

The mechanism of black hole radiation is called the *Hawking effect*. It is based on the quantum field-theoretical fact that virtual particle pairs are constantly being created and destroyed in the vacuum. If this happens near the event horizon, one particle of the pair can fall into the hole, while the other carries away some of the hole's energy. The reason a black hole is black is because its escape speed is equal to the speed of light, and therefore from a classical point of view no matter or energy can get out of the event horizon of a black hole once it has fallen in. Although physicists do not usually like to put it this way, the Hawking effect is a mechanism whereby mass-energy can escape a black hole and thus a mechanism whereby quantum mechanics allows mass-energy to exceed the speed of light, at least briefly. However, no way has yet been found of getting close enough to an event horizon to check if Hawking was right.

Another startling gravitational-thermodynamic prediction is the *Unruh effect*, named after Canadian physicist William Unruh (1945–). Building on suggestions by Paul Davies (1946–), Unruh showed that an observer accelerating through an apparent vacuum will detect electromagnetic radiation having a Planck blackbody distribution, with a temperature proportional to the acceleration. The effect is too small to detect for Earthly acceleration rates, and so, like the Hawking effect, Unruh's prediction has not yet been directly verified.

Is information that falls into a black hole lost forever? Suppose one particle of an EPR pair falls into a black hole: recent theoretical work suggests that the quantum mechanical correlations between the particles will persist even when one of them has disappeared beyond the event horizon, when presumably any information that it could share with its partner particle would have to be transmitted faster than the speed of light. This could be further indication that quantum mechanics allows faster-than-light transmission of information, or it could simply indicate that all our classical intuitions about the nature of information are wrong.

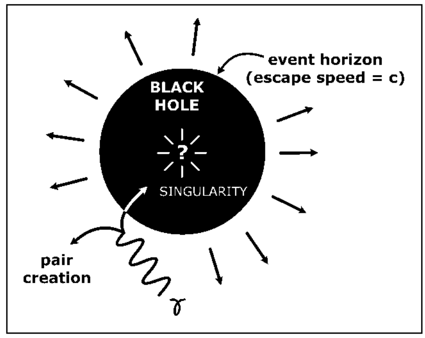

Figure 12.1: The Hawking Effect. Vacuum polarization due to the intense gravitational field near the event horizon causes pair creation. One particle falls in while its antiparticle escapes to infinity, causing the black hole to radiate like a blackbody with a temperature inversely proportional to its mass. Illustration by Kevin deLaplante.

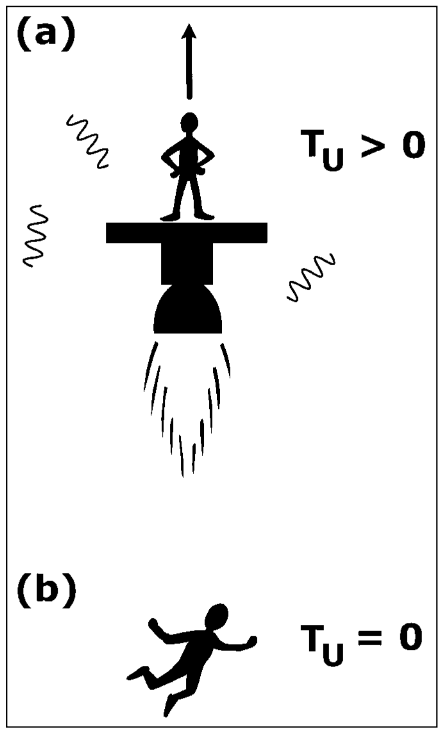

Figure 12.2: The Unruh Effect. An accelerated observer detects radiation in the vacuum with a Planck spectrum and a temperature proportional to the observer's acceleration. For an observer in free fall the temperature of the Unruh radiation is zero. Illustration by Kevin deLaplante.

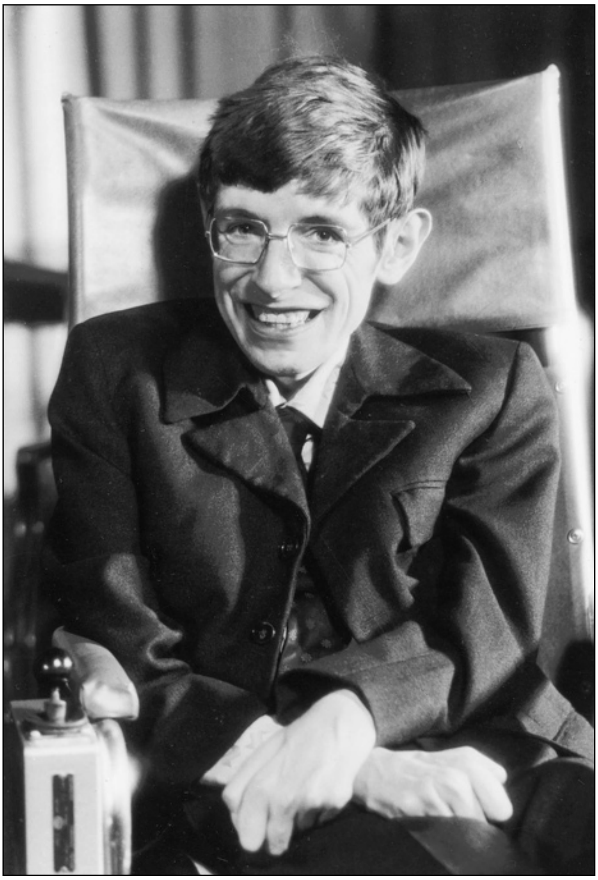

Figure 12.3: Stephen Hawking. AIP Emilio Segre Visual Archives, Physics Today Collection.

The work of Hawking and others shows that there are very deep connections between quantum mechanics, thermodynamics, and the structure of spacetime itself, and it may well be that the ultimate theory of quantum gravity could be a quantum statistical mechanics of granular spacetime.

The Topology of Spacetime

Topology is the branch of mathematics that deals with the ways in which geometric structures connect. The *connectivity* of a geometric object is, roughly speaking, the number of holes in it, and a *multiply connected* structure is one that has holes. From the topological point of view, a coffee cup with a handle and a donut are equivalent, even though *metrically* (in terms of measurable shape) they are quite distinct.

Led by John Archibald Wheeler, a number of theorists, before the advent of string theory, explored the idea that at a deep level spacetime itself is multiply connected. Wheeler suggested that the seething play of vacuum fluctuations can be described as *quantum foam,* and he also proposed the idea that a charged particle such as the electron could be understood as lines of electrical force trapped in the topology of spacetime (Misner, Wheeler, and Thorne 1973). This means that the electron would be a sort of vortex or wormhole in space, with some of Faraday's field lines threaded through it like thread through the holes of a button. The wormhole would not be able to pinch off like a classical wormhole because it would be quantized. Wheeler's elegant idea continues to intrigue physicists, but no one has yet found a way to make it work in mathematical detail.

Nonlocality of Energy in General Relativity

Another challenge for quantum gravity is the nonlocality of gravitational energy. Einstein constructed general relativity in the hope of setting up a local field theory of gravitation that would replace Newton's action at a distance picture of gravitation. However, nonlocality crept into general relativity through the back door. Any spacetime geometry has a certain energy, but (except for special cases) this energy is partially or wholly nonlocal; that is, it cannot be localized at spacetime points throughout the geometry. This fact follows from the Equivalence Principle, the basic insight on which general relativity was founded. Locally a gravitational field can be made to disappear, in that an observer falling freely in a gravitational field cannot detect the field. This means that the energy of the field must not be detectable locally either, and yet the energy of the field does not go away just because it can be transformed away locally. Is there a connection between the nonlocality of energy in general relativity, and quantum nonlocality? This remains an open question.

WHO YA GONNA CALL?

Lee Smolin's controversial critique of string theory is only part of a larger set of worries he has about the state of physics. Very recently (2006), Smolin has leveled the radical charge that physics has made less progress in the last 30 years than in any comparable period since the eighteenth century. He blames this in part on what he considers to be an obsession with string theory, but he argues that there are other systematic barriers to progress in the way modern theoretical physics is done. Above all else, Smolin (and a few other senior quantum gravity researchers, such as Carlo Rovelli) feel that innovation is hobbled by the failure of most modern physicists to think philosophically about their work. Many of the great pioneers of modern physics (notably Einstein, Bohr, Heisenberg, and Schrödinger) were not only very technically skilled but possessed broad humanistic educations and a strong interest in philosophy. Many of their key advances were stimulated by thinking that could only be described as philosophical in the sense that it involved a willingness to challenge deep assumptions about the meaning of such concepts as time, space, measurement, or causation. And like all good philosophers, Einstein and Bohr were not above taking intellectual risks, some of which (as this history shows) turned out to be wrong but instructive. Perhaps we simply need to let our young physicists make some interesting mistakes.

THE FEYNMAN PROBLEM

Faced with a bewildering variety of nonclassical and often bizarre quantum effects—quantum teleportation, superfluidity, nonlocalilty, and so on—it is hard to tell what really is the deepest puzzle about quantum mechanics. Richard Feynman, who understood quantum mechanics about as well as anyone ever has, argued that the most insistent mystery about quantum mechanics is simply this: how can anything that is both so *simple* and so utterly *basic* be so completely lacking in an explanation?

It may seem odd to describe quantum mechanics as "simple," because the mathematical applications of the theory can be dauntingly complicated. What Feynman meant is that the basic rules of quantum mechanics can be stated in a few lines using only high school mathematics. A rudimentary grasp of complex numbers and probability theory are all that is really needed.

Here is a nontechnical rendition of Feynman's statement of the basics of quantum mechanics (see Feynman, Leighton, and Sands 1965 for more detail):

1. Every physical process can be thought of as a transition from a preparation state (call this the *input,* to use more modern jargon) to a number of possible outcome states, or *outputs.* That is, we set up the system in a certain condition, something happens to it, and then we observe what result we got.
2. There can be many ways in which a physical system can undergo a transition from its input state to a given possible output state.
3. For every possible route the system can take from input to a possible output there is a complex number, called the *transition amplitude, probability amplitude,* or simply *amplitude,* for that route.
4. If there is a way of telling which route the system took to a particular output, then the probability of getting that output is found by taking the amplitude for each possible route, squaring it up to get the modulus (which will be a real number), and then adding the resulting probabilities together.
5. If it is impossible to tell which route the system took in order to get to a particular output without disturbing the system in such a way that it changes the possible outputs or their probabilities, then we find the probability of getting a particular output by adding the complex amplitudes together and then taking the modulus to get the probability.

As Feynman said, that's all there is to it, and no one has any deeper explanation of how or why this works. All the rest of quantum mechanics is merely an elaboration of these rules, using the rich mathematics of complex numbers. Rule 4 is just the classical way of adding up probabilities for statistically independent possible events: if there is a .2 probability that a certain bird will fly from its nest to its feeding ground via the forest, and a .3 probability that this bird will fly from its nest to its feeding ground via the river, then the probability that it will fly from its nest to the feeding ground via either the river or the forest is just .5. In classical probability theory, there is no such thing as a probability amplitude; we just add the probabilities directly, and probabilities have a simple interpretation in terms of frequencies of events. If, however, we are talking about an electron that has been fired through a double slit apparatus, we use rule 5 if we do not know which slit it went through. This means that we will get interference terms if the amplitudes are not perfectly in phase, because we add the amplitudes *before* we square up to get the probabilities. If, on the other hand, we slip another detector in the apparatus that tells us which slit the electron goes through, we can use Rule 4, since we have destroyed the interference.

The *Feynman problem* is simply to explain Rules 1 through 5. Where do probability amplitudes come from, and why do they superpose that way? Feynman of course knew that it has a lot to do with noncommutativity. As Dirac and Heisenberg showed, quantum nonclassicality manifests itself in the noncommutativity of certain possible measurement operations. However, there is still not a clear explanation of why noncommutativity should lead to the mathematics of probability amplitudes that was discovered by Schrödinger, Dirac, von Neumann, and others. And asking this question only leads us to the further question of why there is noncommutativity in the first place. All observables would commute if Planck's constant was zero, but it is not. To borrow Rabi's phrase and apply it to Planck's constant, "Who ordered *that*?" We are not much further ahead on this question than Planck was.

It is possible that quantum information theory might lead to a solution of the Feynman problem: if quantum information is logarithmic the way classical information is, then Rule 5 could be the simple consequence of the fact that when we multiply complexities, we add their logarithms. But what sort of complexity would a probability amplitude be a logarithm *of*? That question remains unanswered.

Some philosophers of science have speculated that there could be no explanation for quantum statistics, because there is nothing more basic in terms of which it could be explained. Others respond that it is hard to imagine that we have found the final formulation of quantum mechanics when there are still so many gaps in the theory, so many unsolved problems, so many temporary props holding up the structure. Another challenging view, explored by philosopher Colin McGinn, is that a genuine explanation for the basis of quantum mechanics could well be beyond human cognitive capacity, just as the differential calculus is beyond the grasp of any dog. This concern has to be taken seriously. We certainly would be foolishly arrogant if we did not concede the possibility that there are things that will forever be beyond the ability of any human to understand. At the same time, however, it should be obvious that we have no principled way of telling what those things are, since we would have to be smarter than we are in order to define the limits of our own understanding. We can see that dogs cannot understand certain things because we are generally smarter than dogs, but we cannot be smarter than ourselves. Maybe some day we will create a quantum computer that is smarter than we are, and it might be able to tell us what subjects to not bother trying to understand. But in the meantime, we might as well keep on trying!—especially since, as Feynman suggested, we really ought to be able to figure out the basis for a set of rules that can be expressed in such a simple way.

Despite everything that has been learned since 1875, the present situation in physics is remarkably like the way it was when an idealistic young scientist named Max Planck dedicated his life to understanding the nature of light.

TIMELINE

ca. 450 B.C.	Zeno of Elea sets forth a series of paradoxes attempting to show that the concept of motion is inconsistent.
	Democritus of Abdera argues that the world is made of atoms (tiny indivisible particles of matter) and the Void.
ca. 385 B.C.	The Athenian philosopher Plato, in his *Timaeus*, speculates that the properties of matter could be explained in terms of the symmetries of the five regular ("Platonic") solids, but states that the perfection of the physical world is inevitably marred by the Errant Cause, an early Indeterminacy Principle.
ca. 330 B.C.	Aristotle, a former pupil of Plato's, describes a qualitative theory of change, motion, and time that was to dominate physical thought for over 1,500 years. Aristotle states that time is nothing more than a "measure of motion."
1660–1680	IsaacNewton and G. W. Leibniz invent the calculus, which was to become the most important mathematical tool of physics.
1686	Newtonpublishes his *Mathematical Principles of Natural Philosophy*, setting out basic laws of mechanics and a theory of gravitation that were to become the backbone of physics for centuries to come.
1700–1850	Newtonianmechanics is developed by several mathematicians and physicists (notably Lagrange, Laplace, and Hamilton) into a powerful analytical tool, which until the end of the nineteenth century is assumed to be universally applicable.
1704	Newtonpublishes his *Opticks*, in which he describes his experiments that establish many of the laws of refraction and dispersion of light. Newton speculates that both light and matter are composed of "corpuscles," tiny discrete particles.

1801	Thomas Young demonstrates the interference of light and argues that light is best understood as a wave. This view of light becomes dominant in the nineteenth century.
1814	Spectroscopy (the study of light spectra) begins with the invention of the spectroscope by Joseph von Fraunhofer; throughout the nineteenth century Robert Bunsen, Kirchhoff, and others discover emission and absorption spectra allowing the identification of many elements; the first empirical laws governing spectra are defined.
1820–1850	Many of the basic laws of electromagnetism are developed experimentally by several researchers, including Ampère, Henry, Øersted, and Faraday. Faraday outlines the concept of the electromagnetic field.
1820–1880	The laws of classical thermodynamics are defined by several scientists, notably Carnot, Mayer, Clausius, Kirchhoff, and Helmholtz. These laws include the First Law (conservation of energy) and the Second Law (entropy must always increase).
1859	G. Kirchhoff defines the concept of the blackbody (an object that absorbs all electromagnetic radiation incident upon it) and proves that the emission spectrum of a blackbody is a function only of its temperature. However, he is not able to predict the shape of the curve.
1860–1900	Physicists, notably Maxwell and Ludwig Boltzmann, begin to understand thermodynamics in statistical terms. Boltzmann argues that entropy is a measure of disorder, which implies that the Second Law is not exact.
1861–1865	James Clerk Maxwell presents his equations describing the electromagnetic field as a unified structure. Maxwell argues that light is nothing other than transverse electromagnetic waves of a certain frequency range.
1885	Johann Balmer writes a formula expressing the wavelengths of the visible lines of the hydrogen spectrum in terms of the squares of integers; the formula is generalized by Johannes Rydberg in 1888; the spectrum depends in part on an empirical constant, which became known as the Rydberg constant.
1888	Heinrich Hertz demonstrates experimentally the existence of electromagnetic waves, thus verifying Maxwell's mathematical theory of electromagnetism.
1895	Wilhelm Roentgen discovers X-rays, high-energy electromagnetic radiation that can penetrate most solid matter.
1896	Henri Bequerel discovers that salts of uranium will fog a photographic plate, thus demonstrating the existence of spontaneous radioactivity.

1897 J. J. Thomson discovers the electron, the first elementary particle to be identified.

1898–1902 Marie and Pierre Curie isolate the radioactive elements polonium and radium.

1898–1907 Rutherford and coworkers discover the fact that radioactive elements transmute into other elements, emitting alpha radiation (which Rutherford showed was the ion of helium) and beta radiation (later shown to be comprised of electrons and positrons). Rutherford also announced his law of radioactive decay, according to which elements have a half-life and decay at an exponential rate.

1900 Paul Villard detects gamma rays emitted by uranium; these are shown by Rutherford and A. E. Andrade to be electromagnetic radiation that is more energetic than X-rays.

Max Planck discovers a formula for the spectral distribution of the radiation emitted by a blackbody at a given temperature; he then shows that this formula can be explained on the assumption that the radiation field emits and absorbs radiation only in discrete "quanta" of energy given by $E = h\nu$, where ν (nu) is light frequency, and h is a new physical constant, a fundamental "quantum" of action.

1905 Einstein's "year of miracles": he pioneers special relativity and establishes the equivalence of mass and energy, shows that Brownian motion is a statistical effect demonstrating the existence of molecules, and describes a theory of the photoelectric effect based on the assumption that light is transmitted in particulate form. He also speculates that Maxwell's theory may hold only as a statistical limit, and is the first to realize that light quanta may be correlated in ways that throws doubt on their separability.

1906–1909 Rutherford, together with Ernest Marsden and Hans Geiger, discovers the atomic nucleus by means of scattering experiments.

1907 Einstein publishes his first papers on the quantum theory of specific heats (thus founding solid state physics) and finds an explanation for the breakdown of the nineteenth-century Dulong-Petit law of specific heats.

1908 Hermann Minkowski generalizes Einstein's special relativity into a coherent geometric picture of four-dimensional spacetime (often called Minkowski space).

1909 Einstein argues that a complete theory of light must involve both wave and particle concepts.

1911 First Solvay Conference; the quantum becomes much more widely known to physicists.

Kamerlingh Onnes discovers superconductivity.

1913 Niels Bohr publishes the first version of his quantum theory of the atom. He assumes that spectral lines are due to quantum jumps between stationary states of the electrons orbiting Rutherford's positively charged nucleus, and derives the Rydberg constant.

H. G. Moseley demonstrates that atomic number is simply the positive charge of the nucleus, and predicts the existence of several new elements.

1914 James Franck and Hertz perform an experiment showing that light is absorbed by atoms in discrete energy steps; this is a further confirmation of the quantum principle.

1914–1924 The Bohr theory is elaborated under the impetus of Arnold Sommerfeld and with the collaboration of many other physicists into the Old Quantum Theory. This approach enjoys some success in calculating spectral properties of simpler atoms, but by 1924 it is clear that it has outlived its usefulness.

1916 Einstein publishes his general theory of relativity, which describes gravitation as a manifestation of the curvature of space and time.

1916–1917 Einstein develops the quantum statistical mechanics of light quanta, arguing for the quantization of light momentum as well as energy and introducing the concepts of spontaneous and induced emission of light, the latter of which would become the basis of laser physics.

1920 Bohr announces the Correspondence Principle, which states that quantum systems can be expected to approximate classical systems in certain limits, such as the limit of large (orbital) quantum numbers. Although the Correspondence Principle is not rigorous, it is a useful guide to model construction.

1922 Discovery of the electron's intrinsic magnetic moment by O. Stern and Walther Gerlach.

The "Bohr Festival" (an informal physics conference) in Göttingen, at which Bohr and Heisenberg meet and begin their momentous interactions.

1923 Discovery by Arthur H. Compton of the Compton effect, which is the scattering of gamma ray quanta off of electrons. Compton showed that the resulting shift in the wavelength can be explained neatly using relativistic rules for the conservation of momentum and energy, so long as it is assumed that light quanta interact as if they are discrete particles with momentum and energy. The Compton effect is a decisive confirmation of Einstein's view that light quanta behave as particles in their interactions with other forms of matter.

1923–1924 Louis de Broglie generalizes the wave-particle duality by suggesting that particles have wave-like properties just as light waves have particle-like properties. He derives laws relating the energy, momentum, wavelength, and wavenumber of "matter waves," and predicts that particles such as electrons should exhibit wave-like interference and diffraction effects. These predictions were confirmed in the late 1920s.

1924 Bohr, together with H. Kramers and J. Slater, publishes an abortive but influential theory in which the authors argued (incorrectly) that energy is conserved only on average in emission and absorption events; this theory also includes a qualitative notion of the field of virtual oscillators, which is a precursor of quantum electrodynamics.

Wolfgang Pauli announces his Exclusion Principle, according to which no two electrons can have precisely the same set of quantum numbers. This gives an immediate explanation for many facts about the structure of the Periodic Table, so long as it is allowed that the electron has an extra quantum number (later identified as spin).

1924–1925 S. N. Bose shows that Planck's Law can be derived from a new statistical law that assumes that light quanta are indistinguishable and inclined probabilistically to aggregate in the same energy state; Einstein generalizes Bose's methods to gasses, and predicts the existence of Bose-Einstein condensates.

1925 George Uhlenbeck and Samuel Goudsmit present the first theory of electron spin.

In June Heisenberg discovers matrix mechanics, although he does not yet realize that he is working with matrices.

In December Paul Dirac grasps that noncommutativity is the most novel feature of Heisenberg's approach, and independently discovers most of the features of matrix mechanics that would be worked out by the Göttingen school in the next few months.

1926 Appearance of modern nonrelativistic quantum mechanics: the matrix mechanics of Heisenberg and the Göttingen school is developed, culminating in the "three-man work" of Heisenberg, Pascual Jordan, and Max Born; Schrödinger elaborates de Broglie's wave theory into a complete theory of wave mechanics; Dirac develops his own version of quantum mechanics based on the noncommutative algebra of linear operators.

Pauli solves the hydrogen atom using matrix mechanics.

	Max Born argues that the wave function (or more precisely its square) is most naturally interpreted as a measure of probability.
	Schrödinger demonstrates the mathematical equivalence of matrix and wave mechanics.
1926–1932	VonNeumann develops his Hilbert Space version of non-relativistic quantum mechanics.
1927	The Uncertainty Principle is stated by Heisenberg.
	Bohr announces his Principle of Complementarity, according to which causal and spacetime accounts of quantum phenomena are complementary, meaning that they are inconsistent with each other but both are required in certain experimental contexts.
	Einstein creates a causal theory of the quantum wavefield, but refuses to publish it because it is not separable, which he is convinced is a mistake.
	Enrico Fermi and Dirac clarify the distinction between Bose-Einstein statistics (according to which particles tend to occupy the same energy states) and Fermi-Dirac statistics (according to which particles obey the Exclusion Principle). Dirac shows that photons obey Planck's Law because they are Bose-Einstein particles (bosons), while electrons are Fermi-Dirac particles (fermions).
	In October the Fifth Solvay Conference is held; de Broglie presents his first causal theory of quantum mechanics, but the Copenhagen Interpretation holds sway; foundational debates continue between Bohr and Einstein.
1928	Diracpresents his relativistic wave equation for the electron.
	First papers on quantum electrodynamics (QED) by Dirac, Jordan, and others.
	Gamow describes alpha decay in terms of barrier penetration by quantum tunneling.
	Heisenberg explains ferromagnetism by means of quantum mechanics.
1929	Houtermansand Atkinson propose nuclear fusion as the means by which stars release energy.
1930	Paulipredicts the existence of the neutrino, although the name is due to Fermi.
	Dirac reluctantly predicts the existence of a positive electron, based on his hole theory.
1931	Ruska creates a prototype electron microscope.
1932	Discoveryof the positron in cosmic ray showers by Carl Anderson.
	The existence of the long-suspected neutron is confirmed by James Chadwick.

Heisenberg creates the first theory of nuclear structure that includes the neutron.

1933 Szilard conceives of the nuclear chain reaction.

1934 Hideki Yukawa discovers a theory of the strong nuclear force that binds protons and neutrons, and predicts the existence of a meson, a new particle that would transmit the strong force.

Ida Noddack argues that nuclear fragments found in neutron-nuclei studies done by Fermi and others are due to the splitting of the uranium nucleus.

Fermi publishes a theory of beta decay, in which he introduces the concept of the weak force.

1935 Schrödinger describes his "cat" paradox and coins the term "entanglement."

The paper of Einstein, Podolsky, and Rosen (EPR) questions the completeness of quantum mechanics and inadvertently highlights the importance of entanglement.

Bohr replies to EPR, arguing that they had unreasonable expectations regarding the completeness of quantum mechanics.

1936 Quantum logic created by Garrett Birkhoff and John von Neumann.

1937 Carl Anderson discovers a particle eventually called the muon, identical to the electron except heavier and unstable, in cosmic rays; it is nearly 10 years before it is clear that the muon is not Yukawa's meson but a new and totally unexpected particle.

Kapitsa and others discover superfluidity in Helium-4.

1938 Rabi discovers nuclear magnetic resonance.

1939 Discovery of nuclear fission by Hahn, Meitner, Strassman, and Frisch.

1942 First nuclear chain reaction at the University of Chicago.

1943 Tomonaga finds a way to renormalize quantum electrodynamics, but his work is not communicated to the West until after the end of World War II.

1945 In July, the first atomic bomb is tested successfully at Alamogordo, New Mexico.

On August 6 and 9 the Japanese cities of Hiroshima and Nagasaki are obliterated by atomic bombs; World War II comes to an end.

1947 The pion is discovered by Cecil Powell. It turns out to be the quantum of the nuclear force field predicted by Yukawa in 1935.

1948 Renormalization theory is created by Schwinger and Feynman; experimental confirmation of quantum electrodynamics to high accuracy.

	The first transistor is invented at Bell Labs by Bardeen, Shockley, and Brattain.
1951	Bohm describes a performable version of the EPR thought experiment in terms of spin measurements.
1952	Bohm publishes his causal version of quantum theory.
1953	The creation of the first masers (1953) and lasers (1957) confirms Bose and Einstein's statistics for light quanta.
1954	Gauge theory published by Yang and Mills.
1956–1957	C.S. Wu and others demonstrate the violation of parity (invariance of physical law under mirror reflection) in certain beta decays; recognition of CPT as a fundamental symmetry.
1957	Hugh Everett publishes the "many worlds" interpretation of quantum mechanics.
1961	Landauer argues that the erasure of a bit of information produces a minimum amount of waste heat (Landauer's Principle).
1962	SLAC comes on line, and deep inelastic scattering experiments show that nucleons have internal structure.
	Quark model developed by Gell-Mann and others; experimental confirmation with observation of omega-minus hadron.
1964	Higgs predicts the existence of a massive boson, which should account for particle masses in the Standard Model; it so far remains undetected.
	Publication of Bell's Theorem, which shows that predicted quantum correlations are inconsistent with locality.
1968	The Veneziano scattering formula, which would lead eventually to string theory.
1970s	Unification of electromagnetism and the weak force by Weinberg and Salam.
1980	The first published design of a quantum computer, by Paul Benioff.
1980–1981	Confirmation of Bell's Theorem by Alain Aspect and others.
1981, 1984	Influential papers by Richard Feynman on quantum computation. He shows that no classical computer could simulate a quantum mechanical system in the time it takes the quantum system to evolve naturally according to quantum dynamics.
1983	Discovery of neutral vector mesons confirms the prediction of their existence by the Standard Theory.
1983–1986	The Grand Unified Theory predicts the decay of the proton, but highly sensitive experiments fail to detect proton decay.

1984 The first string theory revolution, as Schwarz and Green show the mathematical consistency of string theory; this leads to an explosion of interest in the theory.

1985 First papers on quantum computation by Deutsch, showing the possibility of a universal quantum Turing machine.

1993 Quantum teleportation predicted by Bennett, and observed in 1998 by Zeilinger and others.

1994 Peter Shor discovers a theoretical quantum computational algorithm for factoring large numbers.

1998 Astronomers discover that the expansion of the universe is accelerating. This is still poorly understood, except that it is almost certainly caused by some sort of *dark energy* that is quantum mechanical in nature.

2008+ The Large Hadron Collider at CERN comes on line and either confirms or does not confirm the existence of the Higgs "God particle."

GLOSSARY

action: A fundamental quantity in physics that has units of energy times time, or (equivalently) angular momentum.

amplitude: A complex number (often a waveform) that is associated (for unclear reasons) with a transition from an initial to a final state.

angular momentum: Momentum due to rotation; in quantum mechanics angular momentum is quantized (meaning that it is observed to have a discrete spectrum).

barrier penetration: The ability of quantum objects such as the alpha particle to tunnel through a potential barrier in a way that would be classically forbidden.

Bell Inequality: A mathematical inequality between correlations coefficients that relate measurements taken on distant particles belonging to an entangled state. Such inequalities express the (generally false) postulate of local realism, which says that all of the properties of each particle in an entangled state are local to the particle.

blackbody: Any object that absorbs without reflection any electromagnetic energy that falls upon it.

blackbody radiation: In order to be in thermal equilibrium with its surroundings, a blackbody must radiate energy with a spectral distribution that is a function only of its temperature. Planck found the correct shape of the curve, and showed that it could be explained if radiant energy is emitted or absorbed only in discrete quanta by the "oscillators" in the walls of the cavity.

black hole: A gravitationally collapsed object such as a star that has fallen inside its gravitational radius; classically, no light, matter, or information can escape from a black hole.

Born Rule: The fundamental rule of quantum mechanics (first stated explicitly by Max Born in 1927) that the probability of a transition (such as an electron jump in an atomic orbital) is given by the square (modulus squared) of the amplitude (a complex number) for the transition.

Bose-Einstein condensate: A gas of bosons that at low enough temperature undergoes a phase transition into a pure quantum state; more generally, any system of bosons that settles into a pure or nearly pure state due to the Bose-Einstein "inclusion" principle.

boson: Any particle such (as the photon) that obeys Bose-Einstein statistics. Such particles are opposite to fermions in that they tend to congregate in the same quantum state.

causal interpretations of quantum mechanics: Interpretations of quantum mechanics that treats both particles and the wavefield as actually existent objects; many versions of the causal interpretation account for quantum correlations by means of nonlocal dynamics, often mediated by the quantum potential.

cavity radiation: Another term for blackbody radiation.

commutation relations: A formula expressing whether or not two observables (such as spin components) commute.

conjugate variables or canonically conjugate variables: Observables such as position and momentum that fail to commute; all observables come in conjugate pairs.

Copenhagen Interpretation of Quantum Mechanics: The view of quantum mechanics pioneered by Niels Bohr, based on the Principle of Complementarity. It says that it is not meaningful to speak of what is going on at the quantum level independently of a definite experimental context.

Correspondence Principle: A heuristic guideline stated by Niels Bohr for the construction of quantum mechanical models; it states that quantum systems will approximate to classical systems in suitable limits, usually the limits of large quantum numbers or ignorable Planck's constant.

cross section: The probability that particles will interact when they collide.

de Broglie waves: Just as light waves have particles (photons) associated with them, particles of matter such as electrons have waves, whose properties were first described by de Broglie.

Dirac's Equation: A relativistic (covariant) version of Schrödinger's Equation, first written by Dirac in 1928. It is valid for electrons and any other spin-1/2 fermion (such as quarks), and it represents the states of the particles it describes in terms of spinors.

divergence: A mathematical function diverges at a point when its value blows up to infinity at that point; the divergences in the calculated values of many electromagnetic quantities were a major challenge to quantum electrodynamics.

ebit: An entangled qubit (sometimes in the quantum computing literature also called a *Bell state,* after J. S. Bell) representing the state of more than one entangled particle.

eigenstate or eigenvector: A state vector in which a quantum system gives a single result when a certain observable is measured on that system.

eigenvalue: A possible result of the measurement of a quantum mechanical observable.
electron: The lightest lepton, having a rest mass of about .5 MeV, electric charge of –1, and spin +/– 1/2. The electron was the first elementary particle to be identified.
elementary particle: The smallest identifiable units of matter, which must be described using the rules of quantum theory. For detailed particle terminology, see Chapter 9.
entanglement: The tendency of particles that have interacted to remain statistically correlated after they have separated to a degree that would be impossible if the particles were physically independent.
entropy: The change in entropy was defined by Clausius as the ratio between the change in energy and temperature. In statistical mechanics, entropy is a measure of disorder masked by the apparent order of a macrostate, and is given by the logarithm of the number of microstates compatible with a given macrostate.
Exclusion Principle: In the form originally prescribed by Pauli, this stated that no two electrons in the atom can have the same quantum numbers; more generally, it states that no two particles that obey Fermi-Dirac statistics can be in the same quantum state.
expectation value: The quantum mechanical long-run average value of an observable.
fermion: Any particle obeying Fermi-Dirac statistics, which implies that no two such particles can be in exactly the same quantum state (Pauli Exclusion Principle).
general covariance: A fundamental building block of Einstein's general theory of relativity, according to which the laws of physics are the same under any mathematically possible coordinate transformation.
Hamiltonian: In classical physics, the Hamiltonian is the energy of a physical system (usually the sum of its kinetic and potential energies); in quantum physics the Hamiltonian is an operator whose eigenvalues are the possible energies of a system. A lot of the skill in applying quantum mechanics to concrete problems lies in finding the correct form of the Hamiltonian; Schrödinger's Equation can then be solved to calculate observable quantities such as energies, probabilities, or scattering cross-sections.
harmonic oscillator: A system in which a mass on a spring experiences a restoring force that is proportional to its distance from the equilibrium position. There are both classical and quantum mechanical harmonic oscillators, and many types of systems in quantum physics can be modeled as collections of oscillators.
Hawking Effect: A process predicted by Stephen Hawking, in which quantum effects near the event horizon of a black hole will cause the radiation of energy at a thermal temperature inversely proportional to the mass of the black hole.
high energy physics: The experimental and theoretical study of elementary particles.

Hilbert space: The mathematical space in which state vectors live; technically, it is a vector space of complex numbers.
interference: The overlapping of wave functions that occurs because of phase differences; this leads to non-zero probabilities for classically forbidden processes such as violations of the Bell Inequalities.
Lagrangian: Classically, the Lagrangian is any function that represents the difference between kinetic and potential energy of a system; in statistical mechanics, it is the *free energy,* the energy available in the system to do work. The equations of motion (generalizations of Newton's Laws) for a system can be derived from the Lagrangian for the system. In quantum field theory, the dynamics of various kinds of fields can be derived from their Lagrangians.
laser: Acronym for Light Amplification by Stimulated Emission. Stimulated emission is a phenomenon predicted by Einstein in 1917 and is a manifestation of Bose-Einstein statistics, according to which photons tend to crowd into states that are already occupied by other photons. Lasers are devices that produce beams of coherent light, meaning light that is in phase and at the same frequency throughout.
Lorentz covariance: A system of physical laws is Lorentz covariant if it takes the same form under Lorentz transformations of its coordinates; these are the transformations used in Einstein's special relativity, and that are based on the assumption that the speed of light in vacuum is the same (invariant) for all inertial observers.
M-Theory: The hypothetical, yet-to-be-described metatheory proposed by Edward Witten, which he hopes will unify and explain string theory and thus, in effect, be a Theory of Everything.
Many-Worlds, Multiverse, *or* **Relative State interpretation of quantum mechanics:** An attempted solution of the measurement problem by Hugh Everett III according to which the wave function never collapses, but keeps on branching *ad infinitum* as systems interact with each other.
measurement problem: Loosely, this is the problem of understanding what happens when a quantum system is measured by a macroscopic measuring device; more precisely, it is the problem of understanding how superpositions are translated into definite outcomes by a measurement.
moment, magnetic: A measure of magnetic field strength. Particles or combinations of particles with net spin will have a magnetic moment, meaning that they will act like tiny magnets.
neutron: An electrically neutral, spin-1/2 fermion discovered by James Chadwick in 1932; protons and neutrons comprise the nucleus.
noncommutativity: Quantum mechanical observables fail to commute if the value of their product depends on the order in which the observables act on the wave function of the system. Noncommuting observables (such as position and momentum) come in pairs, called canonical conjugates, and their behavior is defined by a *commutation relation* that gives the value of their *commutator,* the operator that is the difference between their two possible products.

nonlocality (controllable): Controllable locality (also called Parameter Dependence) would be the ability to instantly control remote events by local manipulations of entangled states—superluminal signaling, in other words. It is generally (although not without controversy) believed to be impossible because of the no-signaling theorems.

nonlocality (uncontrollable): Uncontrollable nonlocality (also called outcome dependence) is the dependence of distant measurement outcomes in entangled states that leads to the violation of Bell's Inequalities. It is a sign of the failure of common cause or local hidden variable explanations of quantum correlations.

nucleon: Protons and neutrons are collectively called nucleons, since they make up the atomic nucleus.

nucleus: The central dense, positively charged core of the atom, discovered by Rutherford and coworkers.

observable: A linear operator with real eigenvalues acting on state vectors in a Hilbert Space. Any observable is presumed to represent a possible measurement operation, with its eigenvalues being possible measurement results.

Old Quantum Theory: The early quantum theory of the period 1913–1924, based on Bohr's atomic theory, and characterized by the opportunistic mixture of classical and quantum methods.

path integral: A key part of Feynman's interpretation of quantum mechanics, in which the total amplitude for any quantum mechanical process is found by taking an integral over all possible spacetime trajectories for that process.

peaceful coexistence: Abner Shimony's ironic term describing the relation between special relativity and quantum mechanics. Shimony argued that peaceful coexistence is guaranteed by the no-signaling theorems, but that the underlying "ideologies" of quantum mechanics and relativity are different, since the former is nonlocal and the latter is local.

phase: Loosely speaking, the phase (expressed as an angle) of a wave or other periodic process is where the system is in its cycle; the phase of a wave function (which generally takes the form of a complex exponential) is the exponent of the exponential.

photon: The massless, spin-1 particle that is the quantum of the electromagnetic field—the light particle of Einstein's 1905 theory.

Planck's constant: The size of the fundamental quantum (indivisible unit) of action, discovered by Max Planck in 1900. Its modern value is 6.626×10^{-27} erg.sec.

principle of relativity: The statement that the laws of physics are the same for all frames of reference regardless of their states of motion.

probability amplitude: In the formalism of quantum theory, a probability amplitude is a scalar product of a ket (representing a preparation state) and a bra (dual to a ket and representing an outcome state). The modulus (real-valued square) of a probability amplitude is a probability.

quantization: The representation of a physical quantity in the form of a series of discrete (often integral or half-integral) quantities.

quantum chromodynamics: The quantum field theory in the Standard Model describing the interactions between quarks and gluons.

quantum computer: A computer that would utilize quantum interference and entanglement to greatly speed up calculations.

quantum cryptography: The use of quantum correlations in entangled states to encode information; theoretically it is the most secure form of encryption known.

quantum electrodynamics: The quantum theory of the electromagnetic field.

quantum field theory: A generalization of quantum electrodynamics that can be applied to other sorts of fields such as the "color" interactions between quarks. A distinguishing feature of quantum fields is that they allow for the creation and annihilation of particles.

quantum gravity: The "Holy Grail" of modern theoretical physics, a quantum theory of space and time that would be a natural quantum mechanical extension of Einstein's general relativity, and (it is hoped!) will explain all properties of matter in the bargain.

quantum information theory: A quantum mechanical generalization of classical information theory that allows for interference between information states.

quantum jump: First postulated by Bohr, this is the transition of an electron from one orbital to another, coupled with the emission or absorption of a quantum of radiant energy; more generally, a discontinuous transition from one quantum state to another.

quantum logic: An attempt to rewrite quantum theory as a logic of propositions about possible measurement operations. It can be distinguished from classical (Boolean) logic by the failure of distributivity for propositions about noncommuting operations.

quantum potential: Mysterious nonlocal energy studied by Bohm and de Broglie as a way of explaining quantum correlations. It can be derived from Schrödinger's equation, and is a function of phase differences.

quantum teleportation: A process in which an entangled pair of particles is used as a channel or carrier of a quantum state from one particle to its distant entangled partner. It is similar to quantum cryptography in that the teleported state cannot be "read" without further classical information sent by classical means.

qubit: The state vector for a single particle, thought of as a quantum bit of information.

relativity, general theory: A generalization of special relativity in which the laws of physics are the same for all reference frames regardless of their state of relative motion; as shown by Einstein, general relativity contains within itself a theory of gravitation that supersedes Newton's theory for strong fields.

relativity, special theory: A comprehensive framework for physical laws based on the assumption that the speed of light is invariant, and that the laws of physics are the same for all frames of reference moving at constant relative velocities.

renormalization: The process in which infinities are removed in a systematic way from a quantum field theory.
scattering: The process in which one or more particles are made to collide with other particles. Sometimes they scatter *elastically,* meaning that they just bounce off, while other times they scatter *inelastically,* meaning that either the target particles or the probe particles or both break down into fragments. From the time of Rutherford, scattering has been one of the most useful tools of particle physics.
Schrödinger's cat: The unwilling protagonist of a thought experiment, outlined by Schrödinger, that demonstrates the apparent contradictions between the quantum and classical views of physics.
Schrödinger's equation: The fundamental dynamical equation of quantum physics. It describes the evolution in time of the wave function of a system as determined by the system Hamiltonian. The equation as first written by Schrödinger is nonrelativistic, though still a very useful approximation for many purposes in physics and chemistry. It can be generalized in various ways.
separability: Two physical systems are separable if they can be fully described in isolation, and if the evolution of each can be considered in isolation from the evolution of the other. Separability is a more general concept than locality. *Nonseparability* is the tendency of elementary particles to interact or to be correlated in ways that violate classical expectations of statistical independence.
spectrum: The set of possible eigenvalues of a quantum mechanical observable; this is a generalization of the concept of the spectrum of light emitted by electronic transitions in atoms.
spinor: Vector-like object in a complex space used to represent the states of certain particles, especially spin-1/2 fermions such as the electron. In 1928 Dirac used four-component spinors to describe electrons.
spin-statistics theorem: The rule that particles with half-integral spin (such as electrons and protons) are fermions, and particles with integral spin (such as pions and photons) are bosons.
"spooky action at a distance": An ironic phrase of Einstein's, which refers to the mysterious way in which quantum particles seem to be able to influence each other at arbitrary distances.
Standard Model: The model of particles and their fields based on quantum chromodynamics, which settled into its present form in the early 1980s; its predictions are well confirmed and despite the need for numerous empirical parameters it affords a fairly good explanation of the properties of all elementary particles observed so far.
state vector: A vector in Hilbert Space with complex-valued components, which represents the structure of the preparation of a physical system. State vectors can be manipulated in various ways to calculate probabilities, expectation values, and scattering cross-sections. Why the state vector can so successfully encode information about physical systems remains a matter of debate.

stationary state: As postulated by Bohr, an electron is in a stationary state when it is orbiting the nucleus but not emitting energy.

statistical mechanics: The branch of physics pioneered by Boltzmann and Maxwell, and further developed into a quantum mechanical form by Einstein and many others, which explains thermodynamics and other large-scale behavior of matter in terms of the statistics of extremely large numbers of particles.

superconductor: A material that conducts electricity with zero resistance, due to the formation of *Cooper pairs* of electrons that obey Bose-Einstein statistics.

superfluid: A fluid in which most or all of the particles all have the same quantum wavefunction due to Bose-Einstein statistics; such fluids exhibit remarkable nonclassical properties such as zero viscosity. All superfluids known can form only at cryogenic temperatures.

superposition: A linear combination of state vectors or wave functions, leading to constructive or destructive interference such as happens to waves in a ripple tank.

superstring *or* **string:** A quantized one-dimensional vibrating object like an elastic string that is hypothesized to account for the structure of elementary particles.

tachyon: A hypothetical particle traveling faster than light. Some physicists argue that tachyons are permitted by relativity theory, but their existence has not been confirmed observationally.

Thermodynamics: The theory of the transformations of heat, largely developed in the nineteenth century.

Thermodynamics, First Law: Energy conservation: energy may be transformed in many ways, but never created from nothing or destroyed.

Thermodynamics, Second Law: For closed systems, the statement that entropy always increases to a maximum (except for localized probabilistic fluctuations); for open systems, the statement that gradients exist that tend to maximize entropy locally. It can also be stated as follows: it is impossible (strictly speaking, highly improbable) to transform waste heat into useful work.

Thermodynamics, Third Law: As a system's temperature approaches absolute zero, its entropy approaches a constant value. Because of zero point energy the Third Law is not strictly correct.

Tunneling: See **barrier penetration.**

Turing machine (classical): A hypothetical digital computer first described by Alan M. Turing in 1936. It operates by changing the state of a memory register on the basis of inputs and outputs of information according to definite rules. Turing and others showed that his machine is the most general form of digital computer.

Turing machine (quantum): A quantum mechanical version of Turing's universal computer, first described by David Deutsch in 1985. The quantum Turing Machine differs from the classical version in that there can be interference between the differing computational paths it can take; this makes possible a large speed-up in calculation power, although most authors believe that the quantum Turing Machine cannot do any types of calculations that the classical Turing Machine cannot do.

Uncertainty Principle: First set forth by Heisenberg, this rule states that canonically conjugate observables (such as position and momentum, or spin components in different directions) have uncertainties (sometimes called *dispersions*) whose product must always be greater than or equal to Planck's constant. The Uncertainty Relations imply that if one observable could be measured with infinite precision, its conjugate would be completely uncertain.

Unruh Effect: A process predicted by William Unruh in which a body accelerating in the vacuum will detect radiation with a Planck spectrum and at a temperature proportional to its acceleration.

vacuum polarization: The tendency of a charged particle such as an electron to attract oppositely charged particles from the vacuum surrounding the particle; this implies that the "bare" charge of the electron is not directly observed, but rather its net or physical charge due to partial charge cancellation by the virtual particles surrounding it.

virtual oscillator: From the time of Planck onward it has been found that the behavior of quantum systems can often be modeled as if they were collections of quantized harmonic oscillators (see **harmonic oscillators**).

wavefunction: A wavefunction is a probability amplitude for the state vector to project into what Dirac called a *continuous representative*—a basis of continuous observables (normally either position or momentum).

wavefunction, collapse of: The process in the von Neumann formulation of quantum mechanics in which the state vector reduces, or projects, into a subspace (often just a pure state) when the system is measured. There remains considerable controversy as to whether collapse of the wavefunction represents a real physical process or is merely a mathematical approximation to a superluminal statistical reweighting of outcomes.

wavepacket: A wavefunction in which the phase factors of the various components interfere in such a way that the wavefunction takes the form of a fairly compact "lump" or bundle; wavepackets can represent the motion of particles.

zero point energy: A minimum energy possessed by all quantum systems; because of zero point energy, it is strictly speaking impossible to reach absolute zero.

FURTHER READING

PRIMARY SOURCES

Primary sources are papers in professional journals that set forth original results. Of the thousands of scientific papers that have been published on quantum theory and related problems, the few listed here are either are among the major turning points in the history of modern physics, or represent recent work that seems (to this author) to be especially interesting or potentially important.

Historically Important Papers

Most of the papers mentioned here are available in English in the anthologies listed below.

Aspect, A., P. Grangier, and G. Roger. "Experimental Tests of Realistic Local Theories via Bell's Theorem." *Physical Review Letters* 47 (1981): 460–67.

The first confirmation of Bell's Theorem that is generally felt to be decisive.

Bell, John S. "On the Einstein-Podolsky-Rosen Paradox." In *Quantum Theory and Measurement*, ed. J. A. Wheeler and W. H. Zurek, pp. 403–8. Princeton, NJ: Princeton University Press, 1983. First publication in *Physics* 1 (1964): 195–200.

Bell's first statement of his theorem that local realism conflicts with the predictions of quantum mechanics.

Birkhoff, G., and J. von Neumann. "The Logic of Quantum Mechanics." *Annals of Mathematics* 37 (1936): 823–43.

The first presentation of quantum logic.

Bohm, David. "A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables." In *Quantum Theory and Measurement,* ed. J. A. Wheeler and W. H. Zurek. pp. 369–96. Princeton, NJ: Princeton University Press, 1983. First publication in *Physical Review* 85, no. 2 (1952): 166–93.

The first presentation of Bohm's causal interpretation of nonrelativistic quantum mechanics, based on the quantum potential and guidance condition.

Bohr, Niels. "On the Constitution of Atoms and Molecules." *Philosophical Magazine* 26 (1913): 1–15.

The first presentation of Bohr's atomic hypothesis.

Bohr, Niels, H. A. Kramers, and J. C. Slater. "On the Quantum Theory of Radiation." In *Sources of Quantum Mechanics,* ed. B. L. van der Waerden, pp. 159–76. New York: Dover Publications, 1967. First publication in *Philosophical Magazine* 47 (1924): 785–802.

Bohr and his collaborators demonstrate that a scientific hypothesis does not have to be right in order to be instructive.

Born, Max. "On the Quantum Mechanics of Collisions." In *Quantum Theory and Measurement,* ed. J. A. Wheeler and W. H. Zurek, pp. 52–55. Princeton, NJ: Princeton University Press, 1983. First publication as "Zur Quantenmechanik der Stossvorgänge." *Zeitschrift für Physik* 37 (1926): 863–67.

Born states in a footnote what is now called the Born Rule, according to which probabilities are the squared modulus of the wave function.

Born, M., W. Heisenberg, and P. Jordan. "On Quantum Mechanics II." In *Sources of Quantum Mechanics,* ed. B. L. van der Waerden, pp. 321–85. New York: Dover Publications, 1967. First publication as "Zur Quantenmechanik II." *Zeitschrift für Physik* 35 (1926): 557–615.

The "three-man-work" which sets forth the first fully worked out version of matrix mechanics.

Dirac, P.A.M. "The Fundamental Equations of Quantum Mechanics." In *Sources of Quantum Mechanics,* ed. B. L. van der Waerden, pp. 307–20. New York: Dover Publications, 1967. First publication in *Proceedings of the Royal Society A* 109 (1926): 642–53.

Dirac's first paper on quantum mechanics, in which he generalizes Heisenberg's approach.

———. "The Quantum Theory of the Electron." *Proceedings of the Royal Society of London* 117 (1928): 610–24.

Dirac presents his relativistic wave equation for the electron, and shows that it admits of both negatively and positively charged particles.

Einstein, Albert. "Does the Inertia of a Body Depend upon Its Energy Content?" In *Einstein's Miraculous Year: Five Papers That Changed the Face of*

Physics, ed. John Stachel, pp. 161–64. Princeton, NJ and Oxford: Princeton University Press, 2005. First publication as "Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" *Annalen der Physik* 18 (1905): 639–41.

The short paper in which Einstein announces the equivalence of mass and energy.

———. "On a Heuristic Point of View Concerning the Production and Transformation of Light." In *Einstein's Miraculous Year: Five Papers That Changed the Face of Physics,* ed. John Stachel, pp. 177–98. Princeton, NJ and Oxford: Princeton University Press, 2005. First publication as "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt." *Annalen der Physik* 17: 132–45.

Einstein's first statement of his light quantum hypothesis.

——— 1905. "On the Electrodynamics of Moving Bodies." In *The Principle of Relativity,* ed. A. Einstein, H. A. Lorentz, H. Minkowski, and H. Weyl, trans. W. Perrett and G. B. Jeffery, pp. 37–65. New York: Dover Books (reprint). First publication as "Zur Elektrodynamik bewegter Körper." *Annalen der Physik* 17 (1905): 891–921.

Einstein's foundational paper on the special theory of relativity.

———. "On the Motion of Small Particles Suspended in Liquids at Rest Required by the Molecular-Kinetic Theory of Heat." In *Einstein's Miraculous Year: Five Papers That Changed the Face of Physics,* ed. John Stachel, pp. 85–98. Princeton, NJ and Oxford: Princeton University Press, 2005. First publication as "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen." *Annalen der Physik* 17 (1905): 549–60.

Einstein demonstrates that if liquids are composed of discrete molecules then they should experience statistical fluctuations that could account for Brownian motion.

———. "On the Quantum Theory of Radiation." In *Sources of Quantum Mechanics,* ed. B. L. van der Waerden, pp. 63–77. New York: Dover Publications, 1967. First publication as "Zur Quantentheorie der Strahlung." *Physikalische Gesellschaft Zürich* 18 (1917): 47–62.

Einstein further develops the light quantum hypothesis, showing that such quanta must have particle-like momenta as well as energy.

Einstein, Albert, Boris Podolsky, and Nathan Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In *Quantum Theory and Measurement,* ed. J. A. Wheeler and W. H. Zurek, pp. 138–41. Princeton, NJ: Princeton University Press, 1983. First publication in *Physical Review* 47 (1935): 777–80.

The famous EPR argument for the incompleteness of quantum mechanics.

Everett, Hugh. "Relative State Formulation of Quantum Mechanics." *Reviews of Modern Physics* 29 (1957): 454–62.

The first statement of Everett's "Relative State" or "Many-Worlds" interpretation of quantum physics.

Feynman, R. P. "Spacetime Approach to Quantum Electrodynamics." In *Selected Papers on Quantum Electrodynamics,* ed. Julian Schwinger, pp. 236–56. New York: Dover, 1958. First publication in *Physical Review* 76 (1949): 769–89.

A professional-level review of Feynman's diagrammatic method of doing quantum electrodynamics.

Heisenberg, Werner. "Quantum-Theoretical Re-Interpretation of Kinematic and Mechanical Relations." In *Sources of Quantum Mechanics,* ed. B. L. van der Waerden, pp. 261–76. New York: Dover Publications, 1967. First publication as "Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen." *Zeitschrift für Physik* 33 (1925): 879–93.

The first paper in which modern quantum mechanics appears.

Kochen, S., and E. P. Specker. "The Problem of Hidden Variables in Quantum Mechanics." *Journal of Mathematics and Mechanics* 17 (1967): 59–87.

A powerful no-go theorem that restricts the possibility of Boolean accounts of quantum predictions.

Landauer, Rolf. "Irreversibility and Heat Generation in the Computing Process." *IBM Journal of Research and Development* 32 (1961): 183–191.

The influential article in which Landauer showed that the erasure of information in a computation must produce a minimum amount of waste heat.

Planck, Max. "On an Improvement of Wien's Equation for the Spectrum." In *The Old Quantum Theory,* ed. D. ter Haar, pp. 79–81. Oxford: Pergamon, 1967. Trans. by D. ter Haar of "Über eine Verbesserung der Wien'schen Spektralgleichung." *Verhandlungen der Deutsche Physikalische Gesellschaft* 2 (1900): 202–4.

Planck's first statement of the equation that he had guessed for the blackbody radiation distribution law.

———. "On the Theory of the Energy Distribution Law of the Normal Spectrum." In *The Old Quantum Theory,* ed. D. ter Haar, pp. 82–90. Oxford: Pergamon 1967. Trans. by D. ter Haar of "Zur Theorie des Gesetzes der Energieverteilung im Normalspectrum." *Verhandlungen der Deutsche Physikalische Gesellschaft* 2 (1900): 237–45.

First publication of Planck's derivation of the blackbody radiation law using Boltzmann's statistical concept of entropy.

Schrödinger, E. "Quantization as an Eigenvalue Problem (Part I)." In *Collected Papers on Wave Mechanics,* ed. E. Schrödinger, pp. 1–12. New York: Chelsea, 1978. First publication as "Quantisierung als Eigenwertproblem." *Annalen der Physik* 79 (1926): 361–76.

The first of the series of papers in which Schrödinger develops his wave mechanical version of quantum mechanics.

Schrödinger, E. "Discussion of Probability Relations between Separated Systems." *Proceedings of the Cambridge Philosophical Society* 31 (1935): 555–63; 32 (1936): 446–51.
A searching analysis of correlated quantum systems, in which Schrödinger coined the term "entanglement."

Wigner, Eugene P. "Remarks on the Mind-Body Question." In *Quantum Theory and Measurement*, ed. J. A. Wheeler and W. H. Zurek, pp. 168–81. Princeton, NJ: Princeton University Press, 1983. First published in *The Scientist Speculates,* ed. I. J. Good, pp. 284–302. London: Heinemann, 1961.
Wigner speculates that the collapse of the wave function is somehow caused by the conscious mind of a human observer.

Yukawa, H. "On the Interaction of Elementary Particles." In *Tabibito,* ed. Hideki Yukawa, pp. 209–18. Singapore: World Scientific, 1982. First publication in *Proceedings of the Physico-Mathematical Society of Japan* 17 (1935): 27–36.
Yukawa presents his theory of the nuclear interaction between protons and neutrons, and predicts the existence of a new particle (later shown to be the pion) as a carrier of the strong force.

Some Recent Interesting Work

A good way to follow recent developments in quantum physics, particle theory, and other mathematically oriented sciences is to visit http://arXiv.org/, a preprint exchange currently operated out of Cornell University. Preprints are prepublication versions of research papers; with the advent of Web-based preprint servers, in particular arXiv.org, preprints (and the new ideas they should contain) can be distributed very rapidly.

Aharonov, Y., J. Anandan, J. Maclay, and J. Suzuki. "Model for Entangled States with Spin-Spin Interactions." *Physical Review A* 70 (2004): 052114.
Possible (though inefficient) nonlocal communications using protective measurements.

Bub, Jeffrey. "Quantum Mechanics Is about Quantum Information." http://arXiv.org/quant-ph/0408020.
An exploration of the possibility that quantum information is a "new physical primitive."

Deutsch, David. "Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer." *Proceedings of the Royal Society of London* A 400 (1985): 97–117.
Probably the first paper to explicitly introduce the concept of a universal quantum computer.

Feynman, Richard P. "Simulating Physics with Computers." *International Journal of Theoretical Physics* 21, no. 6/7 (1982): 467–88.

Feynman observes that computers as we know them could not predict the behavior of quantum systems as quickly as quantum systems can "compute" what they are supposed to do on their own. From this it is a short step to seeing that quantum mechanical computers might be able to do computations much faster than any "classical" computer.

Heisenberg, Werner. "The Nature of Elementary Particles." *Physics Today* 29, no. 3 (1976): 32–39.

Heisenberg's last word on particle physics, in which he calls for a renewed search for the dynamical laws governing particle structure.

Jordan, Thomas F. "Quantum Correlations Do Not Transmit Signals." *Physics Letters* 94A, no. 6/7 (1983): 264.

A succinct, orthodox explanation of why there is no "Bell telephone."

Mermin, N. David. "Hidden Variables and the Two Theorems of John Bell." *Reviews of Modern Physics* 65, no. 3 (1993): 803–15.

A detailed, professional-level, but exceptionally clear review of the relationships between Bell's work, the Kochen-Specker Theorem, and "hidden" variable (i.e., Boolean) interpretations of quantum mechanics.

Peacock, Kent A., and Brian S. Hepburn. "Begging the Signalling Question: Quantum Signalling and the Dynamics of Multiparticle Systems." (1999). http://arXiv.org/quant-ph/9906036.

A muck-raking critique of the conventional "no-signaling" arguments.

Pitowsky, Itamar. "George Boole's 'Conditions of Possible Experience' and the Quantum Puzzle." *British Journal for the Philosophy of Science* 45 (1994): 95–125.

Pitowsky shows that the Bell Inequalities are examples of logical consistency conditions enunciated by George Boole in 1854.

SECONDARY SOURCES

Anthologies

The following anthologies listed here contain many of the decisive papers in the history of quantum theory and related areas of physics, and often much valuable secondary material.

Einstein, A., H. A. Lorentz, H. Minkowski, and H. Weyl. *The Principle of Relativity,* trans. W. Perrett and G. B. Jeffery. New York: Dover Books (reprint).

This contains most of the foundational papers in the theory of relativity, and is a "must" on every physicist's bookshelf.

Schrödinger, E. *Collected Papers on Wave Mechanics.* New York: Chelsea, 1978.

This book collects the papers from 1926 and 1927 in which Schrödinger founded wave mechanics.

Schwinger, Julian, ed. *Selected Papers on Quantum Electrodynamics.* New York: Dover, 1958.

The main papers in the "heroic" period of quantum electrodynamics, from 1927–1949; most are in English.

Stachel, John, ed. *Einstein's Miraculous Year: Five Papers That Changed the Face of Physics.* Princeton and Oxford: Princeton University Press, 2005.

This book contains English translations of Einstein's five great papers of 1905, together with exceptionally clear and insightful commentary by the editor.

ter Haar, D., ed. *The Old Quantum Theory.* Oxford: Pergamon, 1967.

This contains key papers running from the two major publications in 1900 by Planck, through to Pauli's paper on spin, together with a very useful introductory survey by the editor.

van der Waerden, B. L., ed. *Sources of Quantum Mechanics.* New York: Dover Publications, 1967.

This book reproduces, in English translation, many of the key papers in the period 1917–1926 (not including papers on wave mechanics), together with a valuable review by the editor.

Wheeler, J. A., and W. H. Zurek, eds. *Quantum Theory and Measurement.* Princeton, NJ: Princeton University Press, 1983.

This fascinating collection contains many of the most influential papers on the foundations and interpretation of quantum theory, from the 1920s to 1981, together with learned commentary by the editors.

Historical Studies

Here are a few academic papers on the history of physics that are especially influential or insightful.

Beller, Mara. "Matrix Theory before Schrödinger: Philosophy, Problems, Consequences." *Isis* 74, no. 4 (1983): 469–91.

A close study of the critical period from the first papers on matrix mechanics to Schrödinger; the author (who was one of our most distinguished historians of physics) explores the idea that Heisenberg's unusual ability to cope with contradictions may have been a "source of his astounding creativity."

Forman, Paul. "Weimar Culture, Causality, and Quantum Theory: Adaptation by German Mathematicians and Physicists to a Hostile Environment." *Historical Studies in the Physical Sciences* 3 (1971): 1–115.

This states Forman's controversial thesis that the emphasis on uncertainty and acausality in the quantum physics of the 1920s was a reflection of post-World War I anomie in Europe.

Howard, Don. "Einstein on Locality and Separability." *Studies in History and Philosophy of Science* 16 (1985): 171–201.

This is a clear and detailed study of the historical context in which Einstein evolved his statement of the "Separation Principle."

———. "Revisiting the Einstein-Bohr Dialogue." (2005). http://www.nd.edu/~dhoward1.

A recent study of the importance of entanglement in the debates between Bohr and Einstein.

Rovelli, Carlo. "Notes for a brief history of quantum gravity." (2001). http://arXiv.org/gr-qc/0006061.

A very accessible overview of main trends in quantum gravity research from the 1930s to the present, by a current key player in the field.

Biographies and Autobiographies

There are, by now, substantial biographies of most of the influential physicists of the twentieth century. I list a few here that I have found to be especially useful or interesting.

Cassidy, David C. *Uncertainty: The Life and Science of Werner Heisenberg.* New York: W. H. Freeman, 1992.

An excellent, detailed biography of the creator of matrix mechanics.

Gleick, James. *Genius: The Life and Science of Richard Feynman.* New York: Pantheon, 1992.

This is a very good telling of the colorful and sometimes tragic life of the man who many would rate the most brilliant physicist since World War II.

Heisenberg, Werner. *Physics and Beyond: Encounters and Conversations.* New York: Harper and Row, 1971.

Heisenberg's autobiography, centered around reconstructed conversations with Einstein, Bohr, Pauli, and other founders. Heisenberg describes his preoccupation with physics as an almost religious search for the "central order."

Isaacson, Walter. *Einstein: His Life and Universe.* New York: Simon and Schuster, 2007.

There are innumerable popular biographies of Einstein; this very detailed and clear book makes good use of the most recent scholarship on Einstein's life and thought.

James, Ioan. *Remarkable Physicists from Galileo to Yukawa.* Cambridge: Cambridge University Press, 2004.

This book is "for those who would like to read something, but not too much" (as the author puts it) about the lives of the major physicists; it is very good on the period from the mid-1900s onward.

Moore, Walter John. *Schrödinger: Life and Thought.* Cambridge: Cambridge University Press, 1989.

If Hollywood were to make a movie of the history of quantum mechanics, Schrödinger would have to be played by Johnny Depp. This is an authoritative biography of the complex, creative founder of wave mechanics.

Pais, Abraham. *'Subtle is the Lord . . . ': The Science and the Life of Albert Einstein.* Oxford: Oxford University Press, 1982.

This is by far the best biography of Einstein for the reader who wishes to understand the detailed development of his scientific ideas, including his huge contributions to quantum theory. Pais alternates between non-technical biographical chapters and technical expositions of Einstein's work.

Peat, F. David. *Infinite Potential: The Life and Times of David Bohm.* Reading, MA: Addison-Wesley Publishing, 1997.

This is an engrossing story of the sometimes-tragic life of the brilliant American physicist who founded the "causal" interpretation of quantum mechanics and paved the way for the work of J. S. Bell.

Powers, Thomas. *Heisenberg's War: The Secret History of the German Bomb.* New York: Knopf, 1993.

During World War II Heisenberg headed the (fortunately) abortive German atomic bomb project. This is a very clear and interesting history of that troubled period.

Yukawa, Hideki. *Tabibito.* Singapore: World Scientific, 1982.

The engaging autobiography of the Japanese physicist who predicted the existence of the pion, carrier of the strong force.

Histories of Quantum and Atomic Physics

Cushing, James T. *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony.* Chicago and London: University of Chicago Press, 1994.

The late James Cushing was an exceptionally knowledgeable historian of physics. Cushing argues that the Copenhagen Interpretation has overshadowed the causal interpretations of Bohm and de Broglie mainly for historical reasons having little to do with the scientific merits of each approach. Professionals and novices together can learn a great deal about quantum physics and its history from this exceptionally clear book, even if they remain skeptical of Cushing's controversial thesis.

Gamow, George. *Thirty Years That Shook Physics: The Story of Quantum Theory.* New York: Doubleday, 1966.

This is a delightful survey of the heroic years of quantum mechanics from 1900 to 1935, accessible to anyone who can tolerate a little algebra. It is illustrated by Gamow's sketches and caricatures of physicists of the time, many of whom Gamow knew personally.

Jammer, Max. *The Conceptual Development of Quantum Mechanics.* 2nd ed. Woodbury, NY: Tomash Publishers/American Institute of Physics, 1989.

This is an authoritative and detailed survey of the history of nonrelativistic quantum mechanics from Planck to EPR; it is quite good on foundational questions.

Kragh, Helge. *Quantum Generations: A History of Physics in the Twentieth Century.* Princeton, NJ: Princeton University Press, 1999.

This is an excellent and nontechnical survey of the history of physics from the late nineteenth century onward. It has good coverage of both theory and applications.

Kuhn, Thomas S. *Black-Body Theory and the Quantum Discontinuity 1894–1912.* Chicago and London: University of Chicago Press, 1978.

This is a detailed and somewhat controversial historical study of the early years of quantum radiation theory, when classical physicists (Planck in particular) had to accustom themselves to the concept of discontinuity.

Mehra, Jagdish, and Helmut Rechenberg. 1982. *The Historical Development of Quantum Theory.* 6 vols. New York: Springer-Verlag, 1982–2001.

This series of books is the most complete and detailed history of quantum mechanics. The authors display a deep and accurate understanding of the physics and an exhaustive knowledge of its history.

Pais, Abraham. *Inward Bound: Of Matter and Forces in the Physical World.* Oxford: Clarendon Press, 1986.

This is a detailed and authoritative history of particle physics from Bequerel to the early 1980s and the triumph of the Standard Model, written by a participant in that history.

Rhodes, Richard. *The Making of the Atomic Bomb.* New York: Simon and Schuster, 1988.

A very useful nontechnical survey of the growth of twentieth-century physics, culminating in the development and use of atomic weapons in World War II.

Schweber, Silvan S. *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga.* Princeton, NJ: Princeton University Press, 1994.

A detailed, technical, and authoritative history of quantum electrodynamics, centering on the development of renormalization theory in the late 1940s.

For the General Reader

Listed here are a few books, most recent, that present aspects of the quantum story in an especially engaging and helpful way. Most of these books contain little or no mathematics, but they will still challenge you to think!

Aczel, Amir D. *Entanglement: The Greatest Mystery in Physics.* Vancouver, BC: Raincoast Books, 2002.

This is a very engaging and up-to-date introduction to the mysteries of quantum entanglement, with nice profiles of many of the physicists, from J. S. Bell to Nicholas Gisin, who have done recent and important foundational work.

Brown, Julian. *The Quest for the Quantum Computer.* New York: Simon and Schuster, 2000.

This is a detailed but very clear and thorough introduction to the fascinating fields of quantum information theory and computation; it is accessible to anyone with high school mathematics but meaty enough to be useful to professionals as well.

Cropper, William H. *The Quantum Physicists and an Introduction to Their Physics.* New York: Oxford University Press, 1970.

This book requires some calculus and linear algebra, but it explains in an exceptionally clear way the main mathematical ideas in the development of quantum theory from Planck to Dirac.

Deutsch, David. *The Fabric of Reality.* London: Penguin Books, 1997.

This is a very clear and nontechnical exposition of many big questions in science from quantum computing to time travel, informed by Deutsch's enthusiastic advocacy of the controversial multiverse interpretation of quantum mechanics.

Feynman, Richard P. *QED: The Strange Theory of Light and Matter.* Princeton, NJ: Princeton University Press, 1985.

A lucid nontechnical exposition by the master of quantum electrodynamics.

Greene, Brian. *The Fabric of the Cosmos: Space, Time, and the Texture of Reality.* New York: Knopf, 2004.

An exceptionally clear and authoritative account of modern physics and cosmology, from an expert in string theory.

Herbert, Nick. *Quantum Reality: Beyond the New Physics.* New York: Anchor Press/Doubleday, 1985.

Although a little out of date now, this remains one of the very best nontechnical introductions to quantum mechanics and its interpretations.

Johnson, George. *A Shortcut through Time: The Path to the Quantum Computer.* New York: Random House, 2003.

This is a very clear review of quantum computing at about the level of a *Scientific American* article; highly recommended for a quick but thought-provoking introduction to the subject.

Kirshner, Robert P. *The Extravagant Universe: Exploding Stars, Dark Energy, and the Accelerating Cosmos.* Princeton, NJ: Princeton University Press, 2002.

An accessible introduction to the accelerating universe, which in the past ten years has changed all our thinking about cosmology.

Lindley, David. *Uncertainty: Einstein, Heisenberg, Bohr, and the Struggle for the Soul of Science.* New York: Doubleday, 2007.

This book is a very readable, up-to-date, and nontechnical survey of the intense debates about the meaning of quantum mechanics that took place in the 1920s. Some would say that Lindley is a bit too hard on Bohr.

McCarthy, Will. *Hacking Matter: Levitating Chairs, Quantum Mirages, and the Infinite Weirdness of Programmable Atoms.* New York: Basic Books, 2003.

This is a very clear and fascinating introduction to the possibilities for programmable matter, which involves the use of quantum mechanics to tailor-make electronic orbitals to suit almost any purpose.

Milburn, Gerald J. *Schrödinger's Machines: The Quantum Technology Reshaping Everyday Life.* New York: W. H. Freeman, 1997.

This is a very clear and accessible introduction to the basics of quantum mechanics and many of its recent applications in quantum computing, nanocircuits, quantum dots, and quantum cryptography.

Penrose, Roger. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford: Oxford University Press, 1989.

Fascinating and controversial speculations on mathematics, physics, computers, and the mind. This book contains an exceptionally clear introduction to quantum mechanics, at a level accessible to anyone with a tiny bit of high school mathematics.

Smolin, Lee. *The Trouble with Physics: The Rise of String Theory, the Fall of a Science, and What Comes Next.* New York: Houghton Mifflin, 2006.

Smolin's highly controversial argument that for the past 30 years theoretical physics has made virtually no progress because it has become diverted into an unproductive obsession with string theory.

Smolin, Lee. *Three Roads to Quantum Gravity.* New York: Basic Books, 2001.

An exceptionally clear nonmathematical account of cutting-edge work on the challenge of unifying quantum mechanics and Einstein's theory of gravitation, by one of the leading researchers in the field.

Philosophy and Interpretation of Quantum Mechanics

I list only a few titles here, to whet the reader's appetite.

Bell, J. S.. *Speakable and Unspeakable in Quantum Mechanics.* Cambridge: Cambridge University Press, 1987.

This contains most of Bell's major papers on the foundations of quantum mechanics, and much else besides; pure gold.

Bub, Jeffrey. *Interpreting the Quantum World.* Cambridge, New York: Cambridge University Press, 1997.

This is an exceptionally detailed and thorough explanation of the "no-go" results in the foundations of quantum mechanics, such as Bell's Theorem and the Kochen-Specker Theorem.

Feyerabend, Paul. *Against Method: Outline of an Anarchistic Theory of Knowledge.* London: Verso, 1978.

A scrappy and controversial attempt to debunk mythology about the history of science. Footnote 19, p. 61, contains Feyerabend's insightful remarks about history and philosophy as scientific research tools.

Maudlin, Tim. *Quantum Non-Locality and Relativity.* 2nd ed. Oxford: Blackwell Publishers, 2002.

Maudlin, a philosopher at Rutgers University, argues controversially that the violation of Bell's Inequality implies that information in entangled quantum systems is transmitted faster than light. This book contains an exceptionally clear but elementary version of Bell's Theorem.

Shimony, Abner. "Metaphysical Problems in the Foundations of Quantum Mechanics." *International Philosophical Quarterly* 18 (1978): 3–17.

The paper in which Shimony introduces the concept of "peaceful coexistence" between quantum mechanics and relativity, founded on the no-signaling theorems.

Wilbur, Ken, ed. *Quantum Questions: Mystical Writings of the World's Great Physicists.* Boston: Shambhala, 2001.

A number of books attempt to explore the alleged parallels between quantum physics and Eastern mysticism; this is one of the more responsible.

Texts

I list here a few especially sound university-level texts that would be good places to start for the determined beginner who is willing to "drill in very hard wood" as Heisenberg put it (Powers, 1993).

Bohm, David. *Quantum Mechanics.* Englewood Cliffs, NJ: Prentice-Hall, 1951.

In this book (now one of the classic texts in quantum theory) Bohm set forth the version of the EPR experiment that would later be used by J. S. Bell to refute locality. Bohm also presents a very clear and thorough exposition of wave mechanics and the Copenhagen Interpretation (which Bohm was soon to abandon).

Brand, Siegmund, and Hans Dieter Dahmen. *The Picture Book of Quantum Mechanics.* 3rd ed. New York: Springer-Verlag, 1995.

This book is especially useful for its graphical presentation of the structure of wave-functions and wave-packets.

Cohen-Tannoudji, Claude, Bernard Diu, and Franck Laloë. *Quantum Mechanics.* Vol. I. Trans. S. R. Hemley, N. Ostrowsky, and D. Ostrowsky. New York: John Wiley and Sons, 1977.

A sound, detailed, although somewhat ponderous introduction to nonrelativistic quantum mechanics, with an especially thorough treatment of the mathematics of entanglement.

Dirac, P.A.M. *The Principles of Quantum Mechanics.* 4th ed. (revised). Oxford: Clarendon Press, 1958.

A terse but profound and thorough presentation of quantum theory by one of its creators.

Feynman, Richard P., Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics.* Vol. III: *Quantum Mechanics.* Reading, MA: Addison-Wesley Publishing, 1965.

The legendary Feynman lectures on quantum mechanics are, by now, slightly out of date, but serious students and experienced professionals alike can still benefit from Feynman's profound understanding of the quantum.

Misner, Charles, John A. Wheeler, and Kip Thorne. *Gravitation.* San Francisco: W. H. Freeman, 1973.

This tome (so massive that it bends spacetime) is one of the most complete and authoritative introductions to general relativity. A central theme is the fact that classical relativity must be replaced by a quantum theory of spacetime.

Nielsen, Michael A., and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge: Cambridge University Press, 2000.

A very detailed and well-written text in the exploding new field of quantum computation. Although it is aimed at the professional, it contains a very clear introduction to the basics of quantum theory.

Rovelli, Carlo. *Quantum Gravity.* Cambridge: Cambridge University Press, 2004.

An up-to-date and opinionated review of the search for a quantum theory of space and time. This book makes few technical concessions but is essential reading for anyone professionally interested in current work on quantum gravity.

Taylor, Edwin, and John A. Wheeler. *Spacetime Physics.* San Francisco: W. H. Freeman, 1966.

A superbly clear introduction to special relativity, requiring only high school mathematics and a certain amount of *Sitzfleisch* ("sitting muscles," as the old-time German mathematicians would have it).

REFERENCES

Aczel, Amir D. *Entanglement: The Greatest Mystery in Physics.* Vancouver, BC: Raincoast Books, 2002.

Aharonov, Y., J. Anandan, J. Maclay, and J. Suzuki. "Model for Entangled States with Spin-Spin Interactions." *Physical Review A* 70 (2004): 052114.

Bub, Jeffrey. *Interpreting the Quantum World.* Cambridge and New York: Cambridge University Press, 1997.

Cropper, William H. *The Quantum Physicists and an Introduction to Their Physics.* New York: Oxford University Press, 1970.

Cushing, James T. *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony.* Chicago and London: University of Chicago Press, 1994.

Feyerabend, Paul. *Against Method: Outline of an Anarchistic Theory of Knowledge.* London: Verso, 1978.

Feynman, Richard P. *QED: The Strange Theory of Light and Matter.* Princeton, NJ: Princeton University Press, 1985.

Feynman, Richard P., Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics.* Vol. III: *Quantum Mechanics.* Reading, MA: Addison-Wesley Publishing, 1965.

Forman, Paul. "Weimar Culture, Causality, and Quantum Theory: Adaptation by German Mathematicians and Physicists to a Hostile Environment." *Historical Studies in the Physical Sciences* 3 (1971): 1–115.

Heisenberg, Werner. *Physics and Beyond: Encounters and Conversations.* New York: Harper and Row, 1971.

Isaacson, Walter. *Einstein: His Life and Universe.* New York: Simon & Schuster, 2007.

Jammer, Max. *The Conceptual Development of Quantum Mechanics.* 2nd ed. Woodbury, NY: Tomash Publishers/American Institute of Physics, 1989.

Johnson, George. *A Shortcut through Time: The Path to the Quantum Computer.* New York: Random House, 2003.

Kirshner, Robert P. *The Extravagant Universe: Exploding Stars, Dark Energy, and the Accelerating Cosmos.* Princeton, NJ: Princeton University Press, 2002.

Kragh, Helge. *Quantum Generations: A History of Physics in the Twentieth Century.* Princeton, NJ: Princeton University Press, 1999.

Landauer, Rolf. "Irreversibility and Heat Generation in the Computing Process." *IBM Journal of Research and Development* 32 (1961): 183–191.

McCarthy, Will. *Hacking Matter: Levitating Chairs, Quantum Mirages, and the Infinite Weirdness of Programmable Atoms.* New York: Basic Books, 2003.

Milburn, Gerald J. *Schrödinger's Machines: The Quantum Technology Reshaping Everyday Life.* New York: W. H. Freeman, 1997.

Misner, Charles, John A. Wheeler, and Kip Thorne. *Gravitation.* San Francisco: W. H. Freeman, 1973.

Pais, Abraham. *'Subtle is the Lord . . .': The Science and the Life of Albert Einstein.* Oxford: Oxford University Press, 1982.

Peat, F. David. *Infinite Potential: The Life and Times of David Bohm.* Reading, MA: Addison-Wesley Publishing, 1997.

Rhodes, Richard. *The Making of the Atomic Bomb.* New York: Simon and Schuster, 1988.

Schrödinger, E. "Discussion of Probability Relations between Separated Systems." *Proceedings of the Cambridge Philosophical Society* 31 (1935): 555–63; 32 (1936): 446–51.

Schrödinger, Erwin. *What Is Life? The Physical Aspect of the Living Cell.* Cambridge: Cambridge University Press, 1944.

Shimony, Abner. "Metaphysical Problems in the Foundations of Quantum Mechanics." *International Philosophical Quarterly* 18 (1978): 3–17.

Smolin, Lee. *The Trouble with Physics: The Rise of String Theory, the Fall of a Science, and What Comes Next.* New York: Houghton Mifflin, 2006.

van der Waerden, B. L., ed. *Sources of Quantum Mechanics.* New York: Dover Publications, 1967.

Wheeler, J. A., and W. H. Zurek, eds. *Quantum Theory and Measurement.* Princeton, NJ: Princeton University Press, 1983.

INDEX

About the Author

KENT A. PEACOCK is professor of philosophy at the University of Lethbridge, in Alberta, Canada. Peacock received his PhD from the University of Toronto and has also taught at the University of Western Ontario. He has published in philosophy of science, metaphysics of time, and ecological philosophy, and he spends much of his time trying to understand why it is still not obvious that quantum mechanics should be true.