HOW SCIENTISTS EXPLAIN DISEASE

HOW SCIENTISTS EXPLAIN DISEASE

Paul Thagard

PRINCETON UNIVERSITY PRESS PRINCETON, NEW JERSEY

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
3 Market Place, Woodstock, Oxfordshire OX20 1SY

Second printing, and first paperback printing, 2000
Paperback ISBN 0-691-05083-X

The Library of Congress has cataloged the cloth edition of this book as follows

Thagard, Paul.
How scientists explain disease / Paul Thagard.
p. cm.
Includes bibliographical references and index.
ISBN 0-691-00261-4 (cloth : alk. paper)
1. Diseases—Causes and theories of causation.
2. Medicine—Research—Methodology.
3. Medicine—Philosophy. 4. Peptic ulcer—Etiology.
5. Helicobacter pylori infections. I. Title.
RB151.T47 1999
616.07'1—dc21 98-34717

This book has been composed in Times Roman

The paper used in this publication meets the minimum requirements
of ANSI/NISO Z39.48-1992 (R1997) (*Permanence of Paper*)

www.pup.princeton.edu

Printed in the United States of America

3 5 7 9 10 8 6 4 2

For Adam, Daniel, and Ziva

SANTÉ!

Contents

Figures

Tables

Preface

THIS BOOK is about the causes of disease and the causes of science. It is an attempt to answer the question: How do scientists learn about why people get sick? Explaining advances in medical science is similar to explaining diseases, in that both kinds of explanations require the assembly of complexes of interacting causes. Just as most diseases arise from the interaction of environmental and genetic factors, so medical theories arise from the interplay of psychological, physical, and social processes. I have written the book for two main audiences. The first consists of readers with a general interest in the development of medical knowledge about diseases such as peptic ulcer. The second consists of people studying the history, philosophy, psychology, or sociology of science, who will find here an investigation that combines and integrates all these approaches.

The case study at the core of this book is the development and acceptance of the theory that the primary cause of most peptic ulcers is infection by a recently discovered bacterium, *Helicobacter pylori*. I first encountered this case in 1993, when Dr. David Graham invited me to visit him at the Baylor College of Medicine in Houston. He had read my book *Conceptual Revolutions* and saw it as relevant to the rise of the bacterial theory of ulcers, which was first proposed in 1983 by two Australian physicians, Barry Marshall and Robin Warren. Initially, this theory was greeted with intense skepticism by medical experts, but by 1995 it had widespread support. Chapters 3 to 6 of this text provide an integrated explanation of these developments, discussing psychological processes of discovery and acceptance, physical processes involving instruments and experiments, and social processes of collaboration, communication, and consensus.

Chapters 1 and 2 set the stage for the ulcers case study by discussing the nature of explanations of scientific developments and of diseases, both of which are best described in terms of complex schemas that assemble multiple interacting causes. Chapter 1 presents explanation schemas that capture the main current approaches to the study of science, ranging from logical schemas favored by many philosophers to social schemas employed by sociologists. Chapter 2 reviews the most important medical explanation schemas in the history of medicine, from the Hippocratic theory of humors to very recent explanations based on molecular genetics. I argue that an integrated cognitive-social schema provides the most promising approach to explaining the growth of scientific knowledge, and chapters 3 to 6 fill out this schema in the case of the bacterial theory of ulcers.

Chapters 7 to 10 delve more deeply into cognitive mechanisms involving causality, analogy, and conceptual change. Chapter 7 discusses the meaning of the claim that bacteria cause ulcers and provides a general account of medical causal reasoning. Chapter 8 uses this account to explain why discovering the causes of diseases encounters many difficulties, which are illustrated by the development of ideas about scurvy, spongiform encephalopathies (e.g., mad cow disease), AIDS, and chronic fatigue syndrome. Analogical thinking has been important in many cases in the history of medicine that are described in chapter 9. Chapter 10 shows how the development of new medical theories can involve major kinds of conceptual change concerning diseases and their causes.

Chapters 11 to 13 investigate social processes that contribute to the growth of scientific knowledge. Collaboration was a major factor in the development and acceptance of the bacterial theory of ulcers, as it is in most current scientific work; chapter 11 provides a description and evaluation of this role. Chapter 12 describes a social process unique to medicine, the use of consensus conferences to reach authoritative conclusions that provide recommendations for medical practitioners. Increasingly, social interactions in science are being facilitated electronically by the various technologies available on the Internet, and chapter 13 discusses the contributions of these technologies to the development of scientific knowledge. In all three of these chapters, my concern is not only to describe social processes but also to evaluate their potential positive and negative effects on medical progress. Finally, chapter 14 uses ideas about distributed computing to portray science as a complex system of cognitive, social, and physical interactions. The book concludes with a defense of scientific rationality and realism.

Especially in chapter 13 but also in other chapters, I have referred to World Wide Web resources using universal resource locators beginning with "http." Web users can find live links for these references via my Web site at http://cogsci.uwaterloo.ca.

Acknowledgments

I OWE a great debt to medical researchers for very helpful conversations, and thank Drs. David Graham, Richard Hunt, Barry Marshall, J. Robin Warren, and Tadataka Yamada. Special thanks to David Graham for introducing me to the ulcers case and for providing feedback on my first attempts to understand it.

For research support I am very grateful to the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Killam Fellowship program of the Canada Council. Thanks to Kim Honeyford and Kathleen Gorman for research assistance.

For various chapters of this book, I have adapted parts of previous articles, and I am grateful to the respective publishers for permission to reprint some of their contents:

Thagard, P. (1993). Societies of minds: Science as distributed computing. *Studies in History and Philosophy of Science, 24*, 49–67. [Used in chapter 14.] Reprinted with permission of Elsevier Science.

Thagard, P. (1994). Mind, society, and the growth of knowledge. *Philosophy of Science, 61*, 629–645. [Used in chapter 1.] Reprinted with permission of the University of Chicago Press.

Thagard, P. (1995). Explaining scientific change: Integrating the cognitive and the social. In D. Hull, M. Forbes, and R. Burian (Eds.), *PSA 1994*. Vol. 2 (pp. 298–303). East Lansing, MI: Philosophy of Science Association. [Used in chapter 1.] Reprinted with permission of the Philosophy of Science Association.

Thagard, P. (1996). The concept of disease: Structure and change. *Communication and Cognition, 29*, 445–478. [Used in chapters 2 and 10.] Reprinted with permission of *Communication and Cognition.*

Thagard, P. (1997). Collaborative knowledge. *Noûs, 31*, 242–261. [Used in chapter 11.] Reprinted with permission of Blackwell Publishers.

Thagard, P. (1997). Medical analogies: Why and how. In P. Langley and M. Shafto (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* pp. 739–744. Mahway, N.J.: Erlbaum. [Used in chapter 9.] Reprinted with permission of the Cognitive Science Society.

Thagard, P. (1998). Explaining disease: Causes, correlations, and mechanisms. *Minds and Machines, 8*, 61–78. [Used in chapter 7.] Reprinted with permission of Kluwer Academic Publisher.

Thagard, P. (1998). Ulcers and bacteria I: Discovery and acceptance. *Studies in History and Philosophy of Science. Part C. Studies in History and Philosophy of Biological and Biomedical Sciences*, *29*, 107–136. [Used in chapters 3 and 4.] Reprinted with permission of Elsevier Science.

Thagard, P. (1998). Ulcers and bacteria II: Instruments, experiments, and social interactions. *Studies in History and Philosophy of Science. Part C. Studies in History and Philosophy of Biological and Biomedical Sciences*, *29*, 317–342. [Used in chapters 5 and 6.] Reprinted with permission of Elsevier Science.

For assistance with particular articles, I am grateful to Kevin Dunbar, David Graham, Daniel Hausman, Ed Hutchins, Barry Marshall, Robert McCauley, Nancy Nersessian, Gary Olson, Paul Rusnock, Cameron Shelley, Herbert Simon, Miriam Solomon, and James van Evra. Thanks to Miriam Solomon and two anonymous referees for comments on a previous draft of the whole book, and to Sam Elworthy, Ziva Kunda, and Allison Aydelotte for useful suggestions.

Part One

EXPLANATIONS

CHAPTER 1

Explaining Science

In the 1950s, a doctor whose patient was diagnosed with a stomach ulcer would typically recommend that the patient relax and drink lots of milk. By the late 1970s, however, treatment had changed, and the doctor would probably prescribe Tagamet or one of the other acid-blocking drugs that had been developed. Today, in contrast, a well-informed doctor will prescribe a combination of antibiotics for an ulcer patient to kill the bacteria that are now thought to cause most stomach ulcers.

The change in medical practice is due to general adoption in the 1990s of the theory that most peptic (gastric and duodenal) ulcers are caused by *Helicobacter pylori*, a species of bacteria that was discovered only in the early 1980s. When Barry Marshall and Robin Warren suggested that these bacteria might be responsible for peptic ulcers, their proposal was widely viewed as implausible, particularly by the specialists in gastroenterology who usually treat ulcers. But by the mid-1990s, medical consensus panels in many countries had endorsed the bacterial theory of peptic ulcers and their treatment by antibiotics.

How did this change take place? Contrast the following two pictures of scientific development. In a traditional view held by many scientists and philosophers, scientists conduct careful experiments and use the resulting observations to confirm or refute explanatory hypotheses that can provide objective knowledge about the world. In a postmodern view held by some sociologists and culture theorists, scientists conduct experiments to support the hypotheses that best suit their personal and social interests, and they negotiate with other scientists to accumulate sufficient power to ensure that their theories prevail over those of their rivals. Whereas on the traditional view science is largely a matter of logic, in the postmodern view it is largely a matter of politics. The traditional view is exemplified by such philosophers as Hempel (1965), Popper (1959), and Howson and Urbach (1989), whereas the postmodern view is found among sociologists and cultural theorists (e.g., Aronowitz 1988; Latour 1987; Ross 1996).

Neither the traditional nor the postmodern account provides much of an explanation of the discovery and acceptance of the bacterial theory of ulcers. The logical view neglects the diverse psychological and social processes that contribute to scientific development, whereas the political view ignores the

extent to which the growth of science is affected by experimental interactions with the world and by rational assessment of alternative hypotheses. By discussing the ulcers case and other important events in the history of science and medicine, this book develops a much richer view of science as an integrated psychological, social, and physical system.

Many philosophers, historians, psychologists, and sociologists of science are concerned about explaining the development of scientific knowledge, but the kinds of explanations they propose are very diverse. Some philosophers of science prefer *logical* explanations, in which new scientific knowledge derives logically (inductively or deductively) from previous knowledge. Researchers in cognitive science, including psychologists, computer scientists, and some philosophers, propose *cognitive* explanations, in which the growth of knowledge derives from the mental structures and procedures of scientists. Sociologists of science offer *social* explanations, in which factors such as the organization and social interests of scientists are used to explain scientific change.

Are these explanations competitive or complementary? During the 1980s and 1990s, since sociologists of knowledge staked claims to what had been the traditional philosophical territory of explaining the growth of scientific knowledge, there has been conflict between proponents of logical and social explanations (see, for example, Barnes 1985; Bloor 1991; Brown 1984, 1989; Collins 1985). In the meantime, cognitive approaches have emerged with explanatory resources much richer than those available within the logical tradition, but the relation between cognitive and social accounts is rarely specified. Some sociologists are intensely antagonistic toward psychological and computational explanations, even going so far as to propose a ten-year moratorium on cognitive explanations of science (Latour and Woolgar 1986, p. 280). In a similar vein, Downes (1993) attacks what he calls "cognitive individualism" and defends the claim that scientific knowledge is socially produced.

But we can appreciate science as a product of individual minds *and* as a product of complex social organizations. Not only can we see cognitive and social explanations as providing complementary accounts of different aspects of science, but we can also look for ways of integrating those explanations, bringing them together in a common approach. This chapter compares cognitive and social explanation schemas and shows how they can be brought together to form integrated explanations of scientific change. To illustrate the unification of approaches, I show how a cognitive account of the chemical revolution can be socially enriched, and how a social account of the early development of science and mathematics can be cognitively enriched. The social categories of Downes (1993) require similar enrichment. Finally, I sketch how a cognitive/social approach offers new perspectives on the question of scientific rationality.

EXPLANATION SCHEMAS

An explanation schema consists of an explanation target, which is a question to be answered, and an explanatory pattern, which provides a general way of answering the question. For example, when you want to explain why a person is doing an action such as working long hours, you may employ the following rough explanation schema:

Action Explanation Schema

Explanation target:

Why does a **person** with a set of **beliefs** and **desires** perform a particular **action**?

Explanatory pattern:

The **person** has the **belief** that the **action** will help fulfill the **desires.**
This **belief** causes the **person** to pursue the **action**.

To apply this schema to a particular case, we replace the terms in boldface with specific examples, as in explaining Mary's action of working long hours in terms of her belief that this will help her to fulfill her desire to finish her PhD thesis. Many writers in the philosophy of science and cognitive science have described explanations and theories in terms of schemas, patterns, or other abstractions (Darden and Cain 1989; Giere 1994; Kelley 1972; Kitcher 1981, 1989, 1993; Leake 1992; Schaffner 1993; Schank 1986; Thagard 1988, 1992b).

What are the explanation targets in science studies? The most straightforward is belief change, as when we ask why eighteenth-century chemists adopted Antoine Lavoisier's oxygen theory or why nineteenth-century physicians adopted the germ theory of disease. The focus of the general explanation target is why scientists abandoned their previously held belief in favor of a new theory. But there is much more to the development of science than belief change, for we can ask why conceptual changes took place involving the introduction and reorganization of whole conceptual systems (see chapter 10).

Another legitimate explanation target in science studies involves discovery. Why did Lavoisier discover the oxygen theory in the 1770s? Why did Louis Pasteur discover the germ theory of disease in the 1860s? Although such questions are not open to logical explanations, they are grist for the mills of cognitive and social theorists (see chapter 3). Similarly, cognitive and social explanations can be given for why scientists pursue particular scientific research programs. Pursuit is an intermediate stage between the initial discovery or proposal of concepts and beliefs and their eventual acceptance. Within that stage, there are many interesting questions to be answered, such as why scientists conducted particular experiments in particular ways. The remainder of

this chapter focuses on schemas for explaining belief change. But we should not forget that understanding science requires attention to other important explanation targets, such as conceptual change, discovery, and pursuit.

EXPLAINING BELIEF CHANGE

Why do scientists acquire new beliefs, sometimes abandoning old ones? My goal in this section is not to answer this question but rather to characterize the kinds of answers it has been given by means of logical, cognitive, and social explanation schemas. For all these schemas, the explanation target is as follows:

Why did a group of **scientists** adopt a particular set of **beliefs**?

But very different kinds of explanatory patterns can be used to answer this question.

For philosophers and others operating within the tradition of Frege and Russell, formal logic provides the central model for understanding knowledge, in a way roughly captured by the following schema.

Logical Explanation Schema

Explanation target:

Why did a group of **scientists** adopt a particular set of **beliefs**?

Explanatory pattern:

The **scientists** had a set of **previous beliefs**.

The **scientists** employed a **logical method**.

When applied to the **previous beliefs**, the **logical method** implies a set of **acquired beliefs**.

The **scientists** therefore adopted the **acquired beliefs**.

This schema can be made more specific by filling in the account of logical method, which might include deduction, confirmation theory, or—the currently most sophisticated candidate—Bayesian probability theory. Recent proponents of logical approaches to scientific change include Gärdenfors (1988), Howson and Urbach (1989), and Levi (1991). The logical positivists who originated this approach to understanding science were not so much concerned with explaining the growth of scientific knowledge as with providing a foundation for knowledge, but logical schemas have more recently been aimed at understanding scientific change.

Cognitive science offers a mentalistic explanatory approach that differs strongly from the antipsychologistic tradition of the logical positivists. It postulates that the human mind contains representational structures and computational processes that operate on these structures to produce new structures (Thagard 1996). These new structures include sentence-like beliefs as well as

visual images and various kinds of concepts and schemas. Oversimplifying again, we can roughly capture cognitive explanations of belief change in a group of scientists as follows:

Cognitive Explanation Schema

Explanation target:

Why did a group of **scientists** adopt a particular set of **beliefs**?

Explanatory pattern:

The **scientists** had a set of **mental representations** that included a set of **previous beliefs**.

The **scientists'** cognitive mechanisms included a set of **mental procedures**.

When applied to the **mental representations** and **previous beliefs**, the **procedures** produce a set of **acquired beliefs**.

So the **scientists** adopted the **acquired beliefs**.

This cognitive schema is more general than the logical one, since the representations and procedures that it invokes need not be those found in formal logic. Nonsentential representations such as diagrams, maps, and other visual images may be included among the scientists' mental representations in addition to sentential beliefs. Mental procedures may differ completely from methods in deductive and inductive logic and probability theory. For example, in my theory of explanatory coherence, beliefs are accepted on the basis of their coherence with other beliefs, and coherence is modeled computationally by means of connectionist algorithms that perform parallel satisfaction of multiple constraints (Thagard 1992b; see also chapter 4). The cognitive schema thus has a constraint that the antipsychologistic logical schema lacks: that the representations and procedures postulated must be plausible parts of human psychology. This constraint rules out both computationally intractable logical methods such as deductive closure and psychologically implausible methods such as Bayesian updating. Different cognitive explanations of scientific development have been offered by Churchland (1989), Darden (1991), Giere (1988), and Langley et al. (1987); for a collection of relevant papers, see Giere (1992).

Unlike logical methods, mental procedures can also explain the discovery of new concepts and hypotheses and decisions about the pursuit of research programs. Mental procedures can include those that we would not want to count as rational, such as motivated inference in which conclusions are affected by thinkers' personal goals (Kunda 1990). Thus, the cognitive schema competes with the logical schema for providing an understanding of science, since the procedures it postulates are by and large very different from logical methods. In principle, however cognitive and logical schemas could be compatible, if human belief change were fundamentally driven by logical mechanisms, but there is abundant evidence that human psychology involves

a much broader range of structures and processes than logic describes (for deduction, see Johnson-Laird and Byrne 1991; for induction, see Holland et al. 1986).

Sociologists of science tend to focus on different features of science than on logical methods and mental procedures. They note that because of their social situations scientists have various interests, ranging from personal ambition to national sentiment. The also note that the development of science depends in part on the social connections that control information flow among scientists and the power relations that make some scientists much more influential than others in determining what science is done. Amalgamating ideas from various sociologists, we can roughly summarize various social explanations for belief change with the following:

Social Explanation Schema

Explanation target:

Why did a group of **scientists** adopt a particular set of **beliefs**?

Explanatory pattern:

The **scientists** had **previous beliefs** and **interests**.

The **scientists** had **social connections** and **power relations**.

Previous beliefs and **interests** and **social connections** and **power relations** lead to **acquired beliefs**.

The **scientists** adopted the **acquired beliefs**.

This schema is incompatible with the logical schema, which assumes that epistemic matters must be kept isolated from psychological and sociological ones. However, it competes with the cognitive schema only if one assumes that the best explanation of the development of science must be either purely cognitive or purely social. But open-minded cognitivists can easily grant that scientists have the interests, social connections, and power relations postulated by sociologists, and that these qualities play some role in the development of science. Similarly, open-minded sociologists can grant that psychological structures and processes can mediate socially affected belief changes. The cognitive schema is incomplete because it fails to note how social relations can affect the spread of beliefs through the group of scientists. The social schema is incomplete because it fails to show how individual scientists came to acquire their beliefs.

A full account of the growth of scientific knowledge must therefore integrate the features of cognitive and social schemas, as is roughly illustrated by the following schema:

Integrated Cognitive-Social Explanation Schema

Explanation target:

Why did a group of **scientists** adopt a particular set of **beliefs**?

Explanatory pattern:

The **scientists** had a set of **mental representations** that included a set of **previous beliefs** and a set of **interests**.

The **scientists'** cognitive mechanisms included a set of **mental procedures**.

The **scientists** had **social connections** and **power relations**.

When applied to the **mental representations** and **previous beliefs** in the context of **social connections** and **power relations,** the **procedures** produce a set of **acquired beliefs**.

The **scientists** adopted the **acquired beliefs**.

As with the previous schemas I presented, considerable detail must be added to put this explanation schema to work. To fill in the cognitive side, we must specify the mental representations and procedures that operate on them, including logical methods. To fill in the social side, we must specify the relevant social interests, connections, and power relations. As chapter 5 shows, it is also crucial to take into account the instruments and experiments through which scientists interact with the physical world.

To make the integrated cognitive-social explanation succeed, we must provide a much fuller account of how the cognitive and social features of scientists together determine their belief changes. For example, sociological explanations that appeal to the interests of scientists should be able to draw on Kunda's account (1990) of the cognitive mechanisms by which goals affect the selection of evidence. Her experiments show that, in general, people do not simply believe what they want to believe, but rather, that what they want to believe can influence their recall and use of evidence in more subtle ways that influence but do not fully determine their conclusions.

The question of how to make such integrated explanations work cannot be pursued abstractly, since the balance of cognitive and social factors is different in different historical cases. If the explanation target is why T. H. Huxley accepted Charles Darwin's theory of evolution by natural selection, cognitive factors such as the explanatory coherence of the theory should predominate, although the social relations of the two friends should not be ignored. On the other hand, if the explanation target is why some nineteenth-century U.S. industrialists embraced Social Darwinism, social factors such as the mesh between their economic interests and the idea of survival of the fittest should predominate, although the cognitive mechanisms of motivated inference must not be ignored. Similarly, the explanation for acceptance of hormonal or sociobiological explanations of behavioral sex differences may have to weight social values more heavily than evidence evaluation (Longino 1990). I now look in more detail at two important cases of the development of scientific knowledge: the chemical revolution and the development of the mathematical-

mechanistic world view. These cases illustrate the interactions of cognitive and social factors whose contribution to medical knowledge are discussed at greater length in later chapters.

LAVOISIER AND THE CHEMICAL REVOLUTION

In previous work, I offered a cognitive account of the chemical revolution in which Lavoisier's oxygen theory of combustion overthrew the phlogiston theory of Georg Stahl (Thagard 1992b). This account has two parts; a description of the conceptual changes that took place when Lavoisier developed an alternative to the phlogiston scheme, and an explanation, in terms of explanatory coherence, of why he viewed the oxygen theory as superior to the phlogiston theory. Both parts are cognitive, in that conceptual schemes are taken to be organized systems of mental representations, and judgments of explanatory coherence are specified as psychologically plausible computational procedures. My account of the chemical revolution thus instantiates the cognitive schema presented earlier.

I remarked, however, that my account omitted the social side of the chemical revolution and did not presume to tell the whole story (Thagard 1992b, p. 113). What would a social explanation of the chemical revolution look like? My aim in what follows is not to provide a full social account of the acceptance of the oxygen theory but merely to sketch enough that the compatibility and integrability of social and cognitive explanations become evident. From a social perspective, we can look at the developments of Lavoisier's own beliefs and also at how these beliefs spread to the larger scientific community. Social treatments of the chemical revolution include those of Levin (1984), McCann (1978), and Perrin (1987, 1988); other useful sources include Conant (1964), Donovan (1988), Guerlac (1961), and Holmes (1985).

No scientist is an island. Lavoisier had numerous teachers, friends, and associates who contributed to the development of his ideas. We can mention, for example, Guyton de Morveau, who demonstrated to Lavoisier in 1772 that metals gain weight when calcined; Joseph Priestley, who showed Lavoisier in 1774 his experiments that mercury when heated forms a red "calx"; and his wife, Marie, who translated English articles for him, made entries in his notebooks, and drew figures for his publications. Lavoisier was elected at a young age (25 years) to the French Academy and participated in its meetings. He also had a smaller circle of chemists with whom he could perform experiments and discuss the defects of the phlogiston theory uninhibitedly at a time when senior chemists such as Philippe Macquer would not have approved of the aggressive proposal of an alternative theory. Although he alone wrote his most important publications on the oxygen theory, he had various other joint publications,

including the influential *Method of Chemical Nomenclature* (1787), written with Guyton de Morveau, Berthollet, and Fourcroy.

Lavoisier's broader social situation also contributed to his work. His substantial income as a tax farmer meant that he had ample resources and time to conduct his experiments (although this position ultimately led to his execution during the French Revolution). According to an early biographer, "His great wealth, his excellent education, his mathematical precision, his general views, and his persevering industry, all contributed to ensure his success" (Thomson 1813, p. 82). Understanding how the spread of oxygen theory differed between France and England requires an appreciation of the institutional differences between the two countries, which McEvoy summarizes:

> The difference between Lavoisier's corporate view of knowledge and Priestley's individualistic epistemology highlights the difference between the institutional organization of French and British science in the late eighteenth century. In the highly organized and centralized community of France, the pressures of formal education, centralized learned societies, employment opportunities, and a competitive system of reward and recognition meant that aspiring French chemists had little choice but to follow the intellectual lead of the academicians in Paris. In contrast, the organization of English science was much weaker, comprising fewer educational institutions, decentralized societies, little employment opportunity, and a looser congregation of amateurs with closer ties to entrepreneurial industry than their French contemporaries. Thus, whereas the highly integrated community of state-subsidized French theoreticians provided fertile ground for the flowering of paradigmatic conformity during the Chemical Revolution, the dissemination of Lavoisier's theory in England met with a more varied resistance. (McEvoy 1988, pp. 210–211)

Thus, a full explanation of the development of the oxygen theory should not be limited to conceptual development and belief revision, as in my cognitive account. Nevertheless, there is no incompatibility between that account and the relevant social information. No matter how much is said about how Lavoisier gained information from his associates or about how his social situation inclined him to act in certain ways, there remains the problem of describing how his conceptual system developed and changed as he formed and adopted the oxygen theory of combustion, rejecting the phlogiston theory that he had held as a young chemist. As is displayed in the Integrated Cognitive-Social Explanation Schema, cognitive and social explanations of conceptual change can coexist.

Both mind and society contributed to the development of the oxygen theory, but they do not tell the whole story either. The experiments of de Morveau, Lavoisier, Priestley, and others were an important part of the development of eighteenth-century chemistry: Neither mental nor social construction can

fully explain why experiments on combustion and calcination gave the results they did. The growth of scientific knowledge is a function of mind, society, *and* the world. The difficult task for science studies is to create a synthetic account of how mind, society, and the world interactively contribute to scientific development.

The social side of the chemical revolution becomes even more prominent if one addresses the question of how scientists other than Lavoisier came to adopt the oxygen theory. Contrary to the common view that adoption of a revolutionary theory comes only when the proponents of the previous theory die off, the oxygen theory was almost universally adopted in France and (more slowly) in England by scientists who had to abandon their previous phlogiston beliefs. A cognitive explanation of this switch goes roughly like this. Through personal contact with Lavoisier or his disciples, or through reading his argumentative publications, scientists began mentally to acquire the new scientific conceptual scheme. The new mental representations enabled them to understand Lavoisier's claims and to appreciate that the oxygen theory has greater explanatory coherence than the phlogiston theory. This appreciation was part of a cognitive process that led them to accept the oxygen theory, abandoning the phlogiston theory and its conceptual scheme.

From a social perspective, we want to know more about how information spread from scientist to scientist. Diffusion of the oxygen theory was slow, even in France (Perrin 1988). Members of Lavoisier's immediate circle, such as Pierre Laplace, were fairly quick to adopt his views, but the majority of French chemists came around only in the late 1780s and early 1790s. According to Perrin, nearly all converts initially resisted Lavoisier's theory but underwent a conversion that lasted several years. The duration of conversion has both a cognitive and a social explanation. The cognitive explanation is that developing a new conceptual system and appreciating its superiority to the old one is a difficult mental operation; the social explanation is that information flow in social networks is far from instantaneous. Lavoisier and his fellow antiphlogistinians worked to improve the flow—by giving lectures and demonstrations, by publishing articles and books, and by starting a new journal, *Annales de Chimie.* It is also possible that different scientists had different interests that made them resistant to the new theories, although I know of no documentation of this. It is certainly true that different scientists had different initial beliefs and cognitive resources. My cognitive account of Lavoisier cannot be automatically transferred over to all the other scientists, since they had different starting points and associated beliefs. In principle, we would need a different cognitive account for each scientist; but these accounts would have a great deal in common, since the scientists shared many concepts and beliefs, not to mention similar underlying cognitive processes.

Thus, there is much more to a social account of the chemical revolution than

was present in my cognitive explanation of Lavoisier. However, the expanded social account must coalesce with cognitive descriptions of Lavoisier and all the other scientists whose beliefs and conceptual systems changed.

HADDEN ON THE MATHEMATICAL-MECHANISTIC WORLD VIEW

Despite the antagonism that some sociologists display toward psychology, many sociological explanations of scientific developments can be usefully supplemented by cognitive explanations. As an illustration, consider the sociological account of some essential features of early modern mechanistic thought given by Richard Hadden. His abstract provides a summary (Hadden 1988, p. 255): "A sociological explanation is offered for certain features of the mathematical-mechanistic world view. Relations of commodity production and exchange are seen as providing an analogy of 'abstraction' for such a world view. The mediation between social relations and content of science is provided by commercial reckoners who contributed a new meaning to ancient mathematical concepts and thus paved the way for the notion that all sensually intuitable events are explicable in terms of the motion of qualitatively similar bodies." The explanation target here is the emergence in the fifteenth and sixteenth centuries of the view that nature can be understood mechanically and mathematically.

Hadden argues that social relations involving commercial arithmetic provided an analogy for how nature could be understood. "The crux of my argument is that a view of the conditions of the period gets projected onto all of nature and eventually human society as well" (Hadden 1988, p. 257). Just as in the early modern European economy the sensible properties of commodities such as bread and shoes could be abstracted into exchange values, so the sensible properties of all physical objects could be ignored in favor of their mechanical and mathematical properties. Hadden provides evidence that such developments as the replacement of ancient concepts of number were influenced by commercial concerns. Simon Stevin, for example, who was among the first to introduce the notion of decimal fractions, was very much concerned with practical mathematical problems.

Without evaluating the plausibility of Hadden's Marxian account, we can readily see that it presupposes cognitive processes. His explanation of the emergence of new mathematical ideas assumes that "social relations provided analogies and metaphors which were refined technically by thinkers whose concerns involved, at first, the reckoning up of calculable aspects of those relations" (Hadden 1988, p. 271). Thinkers such as Stevin, Hadden conjectures, used commercial social relations as analogs to develop ideas about mathematics and science. Although Hadden's documentation of Stevin's

use of analogy is sparse, later uses of social analogies in science have been well established. Darwin, for example, came up with the idea for natural selection by reading Malthus on political economy (Darwin 1958). It has also been conjectured that Lavoisier's innovative concern with conservation of matter may have been influenced by his tax farmer's familiarity with the balance sheet.

Although Hadden says nothing about how analogical thinking actually works, this is where cognitive science has much to offer, since the topic has been thoroughly investigated using psychological experiments and computational models. The process most relevant to Hadden's account is *analogical mapping*, in which some of the content of a source analog is transferred to a target analog. In Hadden's case study, the target analog involves the mathematics and physics of objects, and the antecedently understood source analog involves commercial and social objects. According to Holyoak and Thagard's theory of mapping (1989, 1995), people's cognitive processes in mapping from one domain to the other require simultaneous satisfaction of semantic, structural, and pragmatic constraints. Some kind of cognitive theory is presupposed by sociological explanations such as Hadden's, which see analogy as the mediating factor between social relations and the development of science. Cognitive theories of analogy are not alternatives to Hadden's account—the social and economic relations he discusses are an important, ineliminable part of the story. But cognitive explanations supplement the social ones by describing the mental processes of the thinkers who made the transition to new ideas. For more on analogy, see chapter 9.

Latour and Woolgar (1986) pursue their extreme anticognitive stance by ignoring the content of scientific papers and speaking only of how scientists use "inscriptions" to produce other inscriptions, as if all that mattered to the process of scientific development were the social relations of scientists and the papers they shuffle around. Latour and Woolgar clearly miss an important part of what is going on when the cognitive representations and processes of scientists enable them to read what has been written, develop and test new hypotheses, and produce new writings. A sociologist or historian who ponders scientific development without paying attention to the intellectual goals and cognitive processes of the scientists involved is like an anthropologist who does fieldwork in an alien tribe without knowing the language. Like Hadden, Latour and Woolgar can only gain from cognitive models that provide a crucial supplement to their social accounts of what laboratory scientists are doing. As Bloor pointed out in the second edition of one of the books that spawned the sociology of scientific knowledge (1991, p. 168), sociologists would be "foolish" to deny the need for a background theory about individual cognitive processes. Similarly, Barnes et al. (1996) present a sociological approach to science that is also open to philosophical and psychological approaches.

ALTERNATIVES TO COGNITIVE INDIVIDUALISM

Downes (1993, p. 452) accuses me and others of *cognitive individualism*, "the thesis that a sufficient explanation for all cognitive activity will be provided by an account of autonomous individual cognitive agents." Obviously, I do not hold this position and in fact have given a battery of arguments for why psychological reductionism in science studies is bound to fail (see chapter 14). But the kind of anticognitive view that Downes seems to prefer in alliance with Latour, Woolgar, and Collins is also bound to fail. Downes distinguishes three levels of social aspects of science, each of which can be shown to have an essential cognitive component.

The first level is the "public embodiment of scientific theories," which includes the textbooks, research papers, instruments, and other shared property of the scientific community. These things clearly exist outside the mental representations of individual scientists, and naturalistic science studies cannot ignore their significance. But part of this significance is cognitive: The use of textbooks, papers and instruments by scientists presupposes scientists' mental capacities to read, write, plan, design, and in other ways produce and use such tools. The public embodiment of scientific knowledge would be pointless if scientists lacked the cognitive processes to understand and produce the embodied objects. Use of external representations such as books and diagrams means that the thought of each scientist does not have to rely entirely on his or her own internal mental representations; but internal representations are needed to comprehend the external ones.

Downes's second level is social interaction, such as is found in complex laboratory work in which no one researcher is entirely responsible for the ultimate result. This level is indeed of great importance, as is clear from research in fields such as psychology, in which most research is collaborative, and experimental physics, in which almost all work is collaborative. But the importance of collaboration and social interaction speaks only against the most implausible forms of psychological reductionism and provides no support for purely social accounts (see chapter 11). Understanding how scientists work with each other in part requires understanding how they communicate with each other, which in turn requires cognitive theories of how they represent information and use language and other means, such as diagrams to convey information to each other. Level 2 is undeniably social, but it is also undeniably cognitive.

Downes's third social level depends on the claim that the activities of scientists make sense only when taken in the context of a broader scientific community. The difference between someone performing an experiment and someone else doing the same physical motions in a play lies in the fact that the former is part of a community of experimenters. We can grant this social distinction,

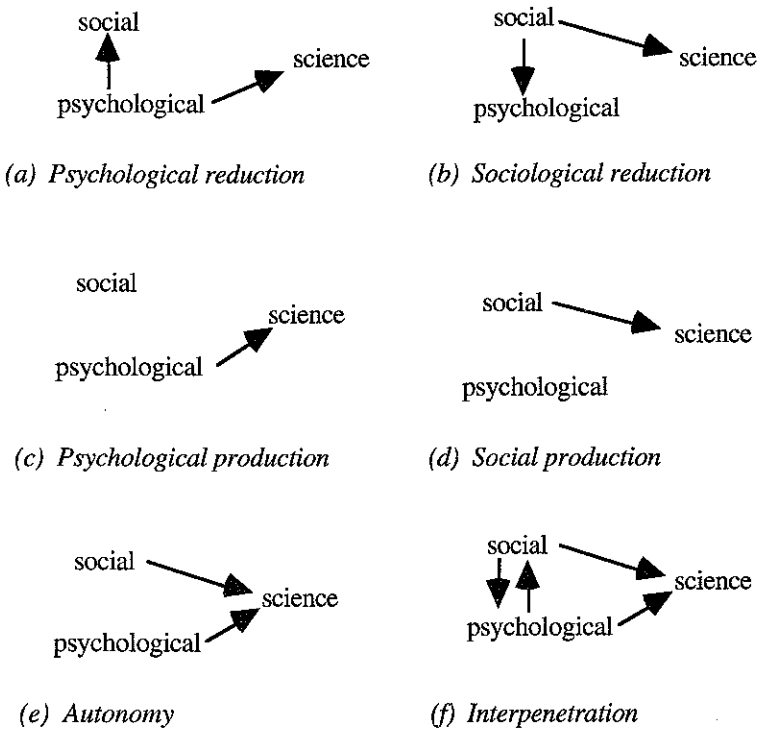

Figure 1.1. Six models of the relation of social and psychological explanations of science. The arrows signify "explains."

but we cannot help but notice that there are also obvious cognitive distinctions. The mental representations of the trained scientist are drastically different from those of the actor who is merely mouthing lines, since the scientists have absorbed an enormous amount of both declarative and procedural knowledge in the course of training. The ability of the experimenter to plan experiments and interpret the results cannot be explained purely in terms of social context but must also make reference to mental structures and procedures.

My arguments that Downes's three social levels each have a crucial cognitive aspect are in no way an attempt to explain them psychologically. We can appreciate social aspects of science at each of these levels while simultaneously appreciating relevant cognitive aspects. Figure 1.1 illustrates six possible relations between psychological and social explanations of science. Schemas *a* and *b* express extreme views about the dominance of a particular style of explanation. Psychological reductionism (*a*) is the view that everything about science, including social aspects, can be understood in terms of the

psychology of the individuals involved. An analog of this view may survive in the economic doctrine of methodological individualism, which proclaims the reduction of macroeconomics to microeconomics, but I know of no one in science studies who holds this view. Sociological reductionism (*b*) is the view that everything about science, including its psychological aspects, can be understood in terms of social factors. In their most rhetorical moments, some Marxists and social constructivists approximate to this view. A slightly more modest view (*d*) advocates social explanations of science but does not purport to explain the psychological. Similarly, schema *c* proposes to simply ignore the social explanations while providing psychological explanations of science. The last two schemas (*e* and *f*) present less dogmatic views of the relation of mind and society. Schema *e* eclectically proposes that social and psychological explanations of science can proceed in relative autonomy of science, perhaps explaining different aspects of science, whereas schema *f* presents a potentially richer and more dynamic view of science studies, in which the social and the psychological are mutually informed. The task before us is to specify these interactions in much more detail, as chapters 3 to 6 do for the development and acceptance of the bacterial theory of ulcers.

The best strategy for naturalistic studies of science is neither psychological reductionism nor sociological reductionism but an integrated approach that takes both the cognitive and the social seriously. To conclude this chapter, I argue that such an approach can be normative—prescritive of how science should be done—as well as descriptive of how it is done.

MIND, SOCIETY, AND RATIONALITY

When the sociology of scientific knowledge arose in the 1970s with its implication of supplanting logical explanation schemas with social ones, philosophers were aghast. Philosophers in the analytic tradition have viewed incursions of psychology into epistemology as assaults on rationality. Incursions of sociology seemed even worse, especially given the rampant relativism of sociologists such as Woolgar (1988), who think that scientific objectivity is an illusion. However, as epistemology and philosophy of science have come to take psychology more seriously, it has become obvious that psychologism requires new theories of rationality but need not embrace irrationalism or relativism. For example, Giere (1988), Goldman (1986), Harman (1986), and Thagard (1988, 1992b) all use psychology to challenge traditional logic-based conceptions of rationality while opening up new territory for rational appraisal.

Similarly, taking the social context of science seriously does not entail relativism. Goldman (1992, p. 194), Kitcher (1993), and Solomon (1994) have

outlined how social practices, like cognitive processes, can be subject to rational appraisal, for example, concerning the extent to which they promote reliable beliefs. Logical explanation schemas carry rationality with them for free, since any beliefs that are inferred logically are presumably warranted. With cognitive and social explanations, the matter is more complicated. We have to ask first what is the best cognitive and social account of a scientific development and only then raise the question of whether the cognitive and social processes invoked are ones that promote the ends of science. In pursuit of the first question, philosophers of science can ally themselves with psychologists, sociologists, and historians of science who, lacking an appetite for the second question, may choose to leave concern for rationality in philosophy, its traditional home. But rational appraisal of social practices and organizations has barely begun (see Goldman 1992, and chapters 11 to 13 of this text).

Solomon (1994) has made the audacious proposal that the scientific community, rather than the individual scientist, should be taken as the important unit of cognitive processing. She contends that a scientific community may reach a consensus that can be judged to be normatively correct from an empirical perspective, even though not one individual scientist in the community made an unbiased judgment. Although the view that she calls "social empiricism" is a useful antidote to past neglect of social aspects of rationality, it swings too far in that direction. My Integrated Cognitive-Social Explanation Schema allows various cognitive and motivational biases to influence the judgments of scientists. But if these biases are as dominant as Solomon suggests, it becomes mysterious how the community collectively reaches a consensus based on empirical success rather than on communal delusion. On the other hand, if scientists share cognitive processes such as those postulated by my theory of explanatory coherence (Thagard 1992b), then their convergence on the empirically successful theory despite their disparate individual biases becomes intelligible. Individual evaluations of the merits of competing theories are not all there is to rationality, but they are an indispensable part of it.

A key conclusion to draw from the interdependence of cognitive and social explanations of scientific change is that the appraisal of cognitive and social strategies must also be linked. Cognitive appraisal should consider the fact that much scientific knowledge is collaborative, and we should therefore evaluate particular cognitive strategies in part on the basis of how well they promote collaboration (see chapter 11). Conversely, social appraisal should take into account the cognitive capacities and limitations of the individuals whose interaction produces knowledge. Determining how to facilitate the growth of scientific knowledge, like the more descriptive task of explaining this development, depends on appreciating the complex interdependencies of mind and society. The next five chapters, however, are primarily descriptive and attempt to explain the development and acceptance of the bacterial theory of ulcers. I return to the question of social rationality in chapters 11 to 14.

SUMMARY

Philosophers, psychologists, and sociologists have offered alternative explanations of the development of scientific knowledge. Cognitive and social explanations can, however, be complementary rather than competitive, and can be combined to fit an Integrated Cognitive-Social Explanation Schema that incorporates both mental processes and social relations. Cognitive accounts of scientific change need to be supplemented with social explanations, just as social accounts need to be supplemented with cognitive explanations. Like cognitive processes, social processes can be evaluated according to how well they contribute to the growth of knowledge.

CHAPTER 2

Explaining Disease

TWO KINDS of explanation are important in medicine. When a patient goes to a physician with a set of complaints and symptoms, the physician's first task is to make a diagnosis of a disease that explains the symptoms. For example, if the patient has a fever, muscle aches, and a runny nose, the physician may explain these symptoms by saying that the patient has influenza. The second kind of explanation, which belongs to medical research rather than clinical practice, requires an answer to the question of why the patient became sick with influenza, which we now know is caused by a virus. Over the past one hundred and fifty years, medical science has identified and generated explanations for numerous human diseases.

This chapter shows that the explanations furnished by medical research fall under a set of basic patterns or schemas that specify the causes of various kinds of disease. After describing the humoral theory that was central to medicine up to the middle of the nineteenth century, I outline the germ theory of disease as a system of explanation schemas. Nutritional and autoimmune diseases are characterized by patterns of explanation that specify nongerm causes for diseases. During the 1980s and 1990s, advances in molecular genetics have generated new explanation patterns for diseases such as cancer. Theoretical knowledge in medicine is not like physics, in which a small number of mathematical equations can provide unified explanations of many observed phenomena. Medicine provides unifications of a different kind, by means of an organized collection of explanation schemas that characterize the causes of numerous diseases.

EXPLANATION SCHEMAS IN THE HISTORY OF MEDICINE

At the most general level, a medical explanation schema has the following form:

Disease Explanation Schema

Explanation target:

Why does a **patient** have a **disease** with associated **symptoms**?

Explanatory pattern:

The **patient** is or has been subject to **causal factors.**

The **causal factors** produce the **disease** and **symptoms.**

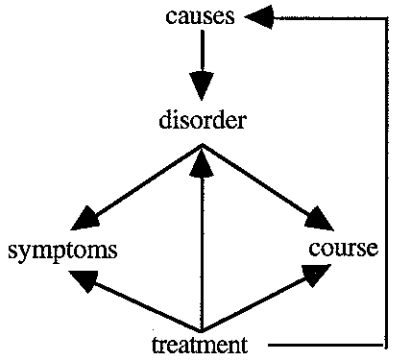

Figure 2.1. Causal structure of disease concepts. The arrows in this and subsequent figures indicate causal relations.

In this schema, the terms in boldface are to be replaced with particular patients, diseases, and so on. At this level of generality, the disease explanation is not useful, but schemas for infectious, nutritional, and other kinds of diseases have provided powerful means of medical explanation.

Disease explanation schemas can alternatively be represented by diagrams that display the causal relations that characterize a disease, as shown by Figure 2.1. Symptoms are the observable manifestations of a disease, which can develop over time in particular ways that constitute the expected course of the disease. The symptoms arise from the cause or causes (etiology) of the disease. Treatment of the disease should affect the symptoms and course of the disease, often by affecting the causal factors that produce the symptoms. For example, tuberculosis has a set of typical symptoms such as coughing and the growth of tubercles (nodules) in the lungs and elsewhere, along with a course that before the twentieth century often included wasting and death. The disorder most commonly affects the lungs, but tuberculosis can also infect many other parts of the body. In 1882, Robert Koch discovered that the cause of tuberculosis is a bacterium, now called *Mycobacterium tuberculosis*, and in 1932, Gerhard Domagk discovered that this microbe can be killed by the drug Prontosil. The drug streptomycin was discovered in 1944 and proved effective in treating the disease. Hence, today tuberculosis has a well-understood cause and a kind of treatment that is effective except for the emergence of bacterial strains resistant to antibiotics.

Hippocrates and the Humoral Theory

The first scientific disease explanation schema is due to Hippocrates, who was born on the Greek island of Cos around 460 B.C. We know little concerning what he himself wrote, but between 430 and 330 B.C., he and his disciples

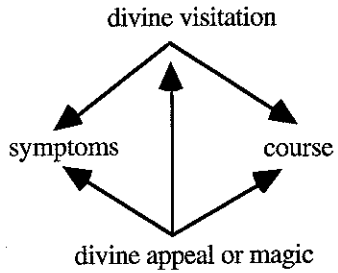

Figure 2.2. Causal structure of religious disease concepts.

produced a body of medical writing. The Hippocratic approach to medicine, as interpreted by Galen and others, dominated European medical thought well into the nineteenth century.

Hippocrates developed a naturalistic approach to medicine that contrasted sharply with the religious views that preceded him. Figure 2.2 shows the causal network that the Hippocratics rejected, for example, in their discussion of the "sacred disease," epilepsy. In the traditional view, epilepsy was caused by divine visitation and hence could only be cured by using an appeal to the gods or other magic. Little was said of the existence of a physical disorder responsible for the observable symptoms. The Hippocratics argued that epilepsy is no more sacred than any other disease and contended that it is caused by an excess of phlegm, one of the four humors (fluids) that constitute the human body.

The following quotes from Hippocratic treatises concisely summarize the humoral theory:

> The human body contains blood, phlegm, yellow bile, and black bile. These are the things that make up its constitution and cause its pains and health. Health is primarily that state in which these constituent substances are in the correct proportion to each other, both in strength and quantity, and are well mixed. (Lloyd 1978, p. 262)

> All human diseases arise from bile and phlegm; the bile and phlegm produce diseases when, inside the body, one of them becomes too moist, too dry, too hot, or too cold; they become this way from foods and drinks, from exertions and wounds, from smell, sound, sight, and venery, and from heat and cold. (Hippocrates 1988, p. 7)

To modern ears, the humoral theory sounds bizarre, but in its time it possessed considerable conceptual and explanatory coherence. Many of Hippocrates's contemporaries believed in four fundamental elements: earth, air, fire, and water. These elements possess various combinations of the four qualities of moist, dry, hot, and cold; for example, fire is hot and dry. The four humors also possess these qualities in different degrees, so that bile tends to be hot and phlegm tends to be cold.

According to the Hippocratics, diseases arise because of humoral imbalances. Too much bile, for example, can produce various fevers, and too much phlegm can cause epilepsy or angina. Imbalances arise from natural causes such as heredity (e.g., phlegmatic parents have phlegmatic children), regimen (e.g., diet and other behavior), and climate (e.g., temperature, wind, and moisture conditions). Different kinds of imbalance produce different diseases with symptoms and development that were acutely observed by the Hippocratics. The Hippocratics described in detail not only the symptoms of patients with a particular disease but also the ways that the patients tended to develop toward recovery or death. The course of a disease was affected by the development of a particular humor, producing crises that signaled basic changes in patient outcome. Fevers were classified as tertian, quartan, and so on based on the number of days before a crisis occurred.

Hippocratic treatment of a disease attempted to address either the causes of the humoral imbalance, by changing diet and environment, or the humoral balance itself. To rid the body of excess bile or phlegm, methods were used to induce vomiting or evacuation of the bowels, and veins were opened to let blood. The use of emetics, purgatives, and phlebotomy remained standard medical practice well into the nineteenth century. These techniques make sense within the Hippocratic framework because they are means of changing fluid balances. Figure 2.3 displays the structure of the causal network underlying the Hippocratic concept of disease.

The Hippocratic theory of disease causation translates into the following explanation schema:

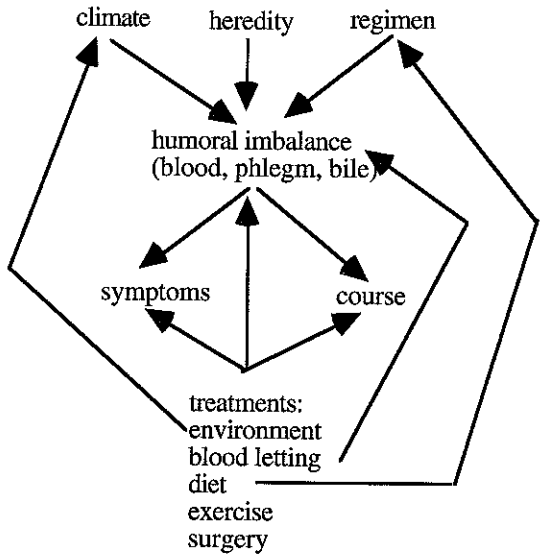

Figure 2.3. Causal structure of Hippocratic disease concepts.

Humoral Theory Explanation Schema

Explanation target:

Why does a **patient** have a **disease** with associated **symptoms**?

Explanatory pattern:

The body of the **patient** contains four humors: blood, phlegm, yellow bile, and black bile.

Nutritional and environmental **factors** produce a **humoral imbalance**.

The **humoral imbalance** produces the **disease** and **symptoms**.

In the Hippocratic view, different diseases arise from different humoral imbalances. For example, various fevers arise from too much bile, and epilepsy and angina are the result of too much phlegm. Thus, the humoral explanation schema can be instantiated for particular diseases, as in this explanation pattern for epilepsy:

Epilepsy Explanation Schema:

Explanation target:

Why does a **patient** have epilepsy characterized by seizures?

Explanatory pattern:

The body of the **patient** contains an excess of phlegm.

The excess of phlegm in the **patient** produces the epileptic seizures.

This schema can in turn be instantiated to explain why a particular patient is sick. Thus, medical explanation in the humoral theory was provided by a hierarchy of schemas that applied general beliefs about disease causation to particular cases.

Pasteur and the Germ Theory

The major blow to the humoral theory came in the 1860s, when Louis Pasteur and others developed the germ theory of disease. Pasteur was a French chemist who in the 1850s turned his attention to the process of fermentation, including the production of lactic acid in sour milk and the production of alcohol in wine and beer. Many scientists at the time believed that fermentation and putrefaction were the result of spontaneous generation. Justus von Liebig, for example, contended in 1839 that fermentation in beer is not caused by yeast but by the internal development of the beer. Pasteur was able to show that the yeast increased in weight, nitrogen, and carbon content during fermentation, and he inferred that yeast is a living organism that is the cause of fermentation in beer and wine. Pasteur proceeded in the early 1860s to identify other organisms—bacteria—that produce lactic acid fermentation. To challenge directly the theory of spontaneous generation, he conducted ingenious experiments to show that fermentation does not take place in the absence of contamination by air. Pasteur's work greatly improved the manufacture of vinegar and wine, and he

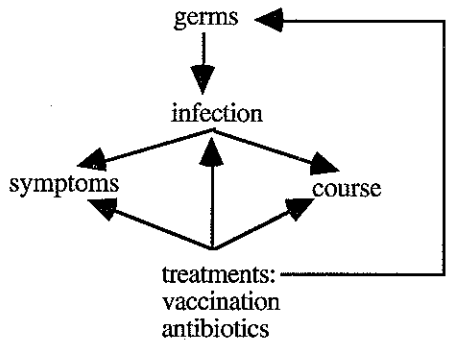

Figure 2.4. Causal structure of the germ theory of disease.

was invited in 1865 to investigate an epidemic of silkworm disease in the south of France. He also took time to study cholera, which had spread to France from Egypt. Naturally, Pasteur applied to silkworms some of the same microscopic techniques that had proved so fertile in his studies of fermentation.

Pasteur (and, independently, the British surgeon Joseph Lister) made the most important mental leap in the history of medicine, pursuing an analogy between fermentation and disease. They realized that just as fermentation is caused by yeast and bacteria, so diseases may also be caused by microorganisms. (See chapter 9 for a discussion of this and other analogies.) In the second half of the nineteenth century, bacteria were shown to be the cause of many important human diseases, including tuberculosis, cholera, and gonorrhea. The germ theory employed a concept of disease (figure 2.4) and an explanation schema that differed dramatically from that of the humoral theory:

Germ Theory Explanation Schema:

Explanation target:

Why does a **patient** have a **disease** with **symptoms** such as fever?

Explanatory pattern:

The **patient** has been infected by a **microbe.**

The **microbe** produces the **disease** and **symptoms**.

Different kinds of microbes provide variants of the Germ Theory Explanation Schema, which was originally based on bacteria. By the 1890s, it was known that some disease-causing microbes were too small to be observed with the microscope; but what we now call viruses were observed in 1939 using electron microscopes (see chapter 10). Other infectious microbes include protozoa and fungi. In the 1980s, Prusiner hypothesized that spongiform encephalopathies such as scrapie, kuru, and Creutzfeldt-Jakob disease were caused by a novel kind of infectious agent called a *prion* (see chapter 8). We can describe current knowledge about infectious diseases in terms of the hierarchy of explanation schemas shown in Figure 2.5. Falling under the Germ Theory Explana-

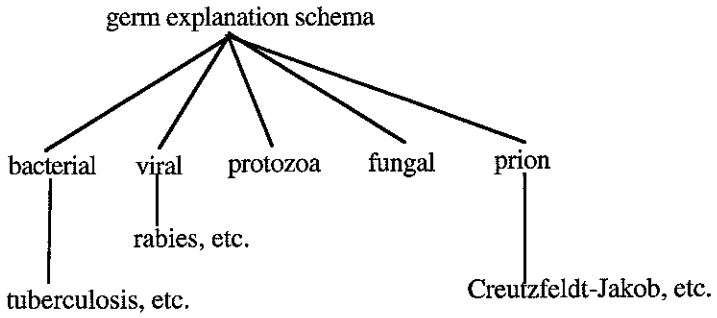

Figure 2.5. Hierarchy of infectious disease schemas.

tion Schema are at least five more specific schemas for bacterial, viral, and other kinds of infectious diseases. Particular diseases then fall under the schemas for different kinds of infectious microbes; for numerous examples, see Murray et al. (1994).

The germ theory of disease has been frequently mentioned since the 1860s, but it is difficult to state in terms of universal laws or general models. Instead, we can think of the germ theory in terms of the hierarchy of schemas (figure 2.5), which includes the general Germ Theory Explanation Schema, explanation schemas for classes of infectious diseases caused by different agents, and schemas for particular diseases. This collection of schemas provides an excellent, unifying fit with hundreds of human maladies.

Nutritional Diseases

Scientific advances in the first half of the twentieth century identified an entirely different class of noninfectious diseases caused by vitamin deficiencies (Funk 1912; see chapter 8). In 1928, for example, Albert Szent-Györgyi isolated vitamin C, deprivation of which causes bleeding gums and other symptoms of scurvy. The explanation schema for nutritional diseases is as follows:

Nutritional Disease Explanation Schema:

Explanation target:

Why does a **patient** have a **disease** with associated **symptoms**?

Explanatory pattern:

The **patient** has a deficiency of a needed **nutrient.**

Absence of the **nutrient** produced the **disease** and the **symptoms**.

In addition to being applicable to scurvy and vitamin C (see chapter 8), this schema fits such diseases as beriberi, which is due to vitamin B_1 deficiency,

and rickets, which is due to vitamin D deficiency. As with infectious diseases, the clinical importance of explanation schemas is that they suggest therapies, such as treating scurvy with vitamin C supplements and treating bacterial infections with antibiotics.

Autoimmune Diseases

During the 1950s, medical researchers led by Frank Macfarlane Burnet developed an understanding of how the body's immune system helps protect it against infectious agents (Silverstein 1989). This understanding generated a new class of diseases that occur when the immune system becomes overactive and attacks the body it is supposed to protect. For example, Grave's disease appears to originate when the immune system damages the thyroid, and lupus erythematosus is the result of an immune attack on the connective tissue. Other diseases that may have autoimmune origins include multiple sclerosis (in which there is damage to myelin in the central nervous system), rheumatic fever (in which there is damage to joint cartilage), and juvenile diabetes (in which there is damage to the pancreas) (Wyngaarden et al. 1992). Here is the general explanation pattern:

Autoimmune Disease Explanation Schema:

Explanation target:

Why does a **patient** have a **disease** with associated **symptoms**?

Explanatory pattern:

The **patient's immune system** attacks an infectious agent.

The **immune system** becomes overactive and attacks **bodily tissues**.

Damage to the **bodily tissues** produces the **symptoms**.

Advances in the understanding of infectious, nutritional, and autoimmune diseases have been monumental, but they leave many of the most important medical problems unexplained. Atherosclerosis, cancer, adult-onset diabetes, and osteoarthritis are just some of the widespread diseases whose primary causes do not appear to be infectious agents, nutritional deficiencies, or autoimmune reactions.

EXPLANATION SCHEMAS FROM MOLECULAR GENETICS

During the 1980s and 1990s, the explanation of disease has undergone major transformations owing to developments in molecular genetics. According to Edward Rubenstein (1994, p. vii):

> We are in the midst of revolutionary changes in basic science that will allow us to identify and to correct or circumvent molecular defects that give rise to some of

> the most prevalent afflictions of humanity, including many forms of atherosclerosis, hypertension, diabetes, neoplasia, autoimmune diseases, and disorders of mendelian inheritance. Henceforth, clinicians will increasingly employ diagnostic methods and therapeutic interventions made possible by the manipulation of the genes of microorganisms, plants, animals, and humans. In short, we have entered the era of molecular medicine.

Medical explanations based on molecular genetics are very different from the kinds of germ-based and nutrition-based explanations of diseases that became available after the mid-nineteenth century.

Molecular genetics has a general explanation schema with specialized versions that apply to diseases of various kinds, including those caused by defects in single genes, multifactorial diseases, and cancer. The following schema is abstracted from Strachan and Read (1996):

Molecular Genetics Disease Explanation Schema:

Explanation target:

Why does a **patient** get a **disease** with associated **symptoms**?

Explanatory pattern:

Genes in the **patient**'s body are encoded in **DNA**.

DNA specifies the synthesis of **RNA**.

RNA specifies the synthesis of polypeptides, which form **proteins**.

Normal function of a **patient**'s body requires the production of **proteins**.

Mutations produce changes in **DNA**.

Mutated **DNA** may alter the production of **proteins** needed for normal functioning of the **patient**'s body.

Abnormal functioning in the **patient** produces the **disease** and its **symptoms**.

This schema leaves open whether mutations are inherited or, as in most cancers, occur during a patient's lifetime. It also leaves open whether the alteration of protein production involves a loss of function, as in most inherited diseases, or a gain of function found in cancer growth. This style of explanation, obviously very different from the infectious and other disease explanations presented in the last section, is too general to apply to particular diseases, which fall into several different classes. The diseases most easily understood in terms of molecular genetics are those produced by defects in single genes.

Mendelian Diseases

A *Mendelian* genetic character is one whose presence or absence depends on the genotype (types of alleles) at a single chromosomal locus. In humans, more than five thousand Mendelian characters have been identified, including hundreds of inherited diseases (McKusick and Francomano 1994). A Mendelian disease is one caused by an inherited mutation in a single gene, yielding the following kind of explanation schema:

Mendelian Disease Explanation Schema

Explanation target:

Why does a **patient** get a **disease** with associated **symptoms**?

Explanatory pattern:

The **patient** has inherited a mutated **gene**.

The mutated **gene** is defective and produces the **disease** and its **symptoms**.

This schema is a specialization of the more abstract Molecular Genetics Disease Explanation Schema, in that it states that a single inherited gene is responsible for the disease. The first disease to be identified as genetic in origin, in 1902, was alkaptonuria, a rare disorder characterized by large quantities of dark-colored urine.

Further specifications of the Mendelian Disease Explanation Schema are possible because of the five different patterns of Mendelian inheritance: autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, and Y-linked. The following schema, specifies an autosonal recessive disease:

Autosomal Recessive Disease Explanation Schema

Explanation target:

Why does a **patient** get a **disease** with associated **symptoms**?

Explanatory pattern:

The **patient** has inherited a recessive mutated **gene** from both parents.

The mutated **gene** is defective and produces the **disease** and its **symptoms**.

This schema is now specific enough that it applies to particular diseases such as cystic fibrosis:

Cystic Fibrosis Explanation Schema

Explanation target:

Why does a patient get **cystic fibrosis** with **symptoms** such as excessive mucous and pulmonary failure?

Explanatory pattern:

The patient has inherited a mutated gene ΔF508 from both parents.

The mutated gene ΔF508 produces anomalous mucous **secretions**.

These **secretions** produce **symptoms** such as excessive mucous and pulmonary failure.

At this level, it is now possible to explain, in a manner that is virtually deductive, why a particular patient became sick: With few exceptions, every human

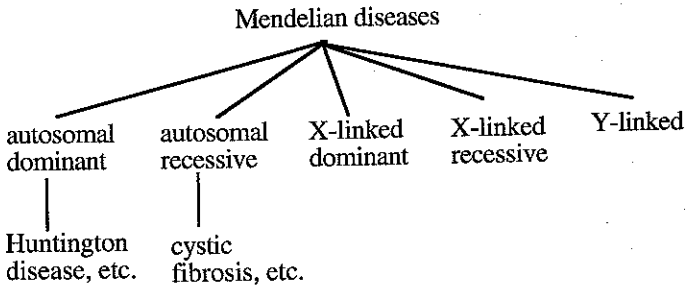

Figure 2.6. Hierarchy of Mendelian disease schemas.

who has inherited the mutated gene ΔF508 from both parents eventually gets cystic fibrosis. Medical explanations are rarely so simple, however, as the discussion of more complex diseases will show. The human genome contains about seventy thousand genes, but only about five thousand Mendelian phenotypes are known (Strachan and Read 1996). Most diseases involve the action of more than one gene, and some genes are responsible for more than one disease. For example, the gene PRNP is implicated in both Creutzfeldt-Jakob syndrome and familial fatal insomnia. The schemas for Mendelian diseases form the hierarchy shown in Figure 2.6.

Multifactorial Diseases

Non-Mendelian characters are polygenic, meaning that they depend on more than one genetic locus, and they may be multifactorial, with a substantial contribution from environmental as well as genetic factors. Modern medicine recognizes that the diseases that most commonly afflict humans—such as atherosclerosis, hypertension, cancer, diabetes, and arthritis—are multifactorial. The tendency to atherosclerosis, for example, seems to depend on hereditary factors and also on environmental factors such as diet and exercise. There is a great diversity of multifactorial diseases, but a very general explanation schema covers them. Because much research is now taking place into the genetic causes of these diseases, the diseases are increasingly falling into the realm of molecular genetics. In 1996 alone, genetic correlates were identified for such multifactorial diseases as diabetes, pancreatic cancer, and basal cell carcinoma.

Multifactorial Disease Explanation Schema

Explanation target:

Why does a **patient** get a **disease** with associated **symptoms**?

Explanatory pattern:

The **patient** has inherited various **genes**.

The **patient** is subject to various **environmental factors**.

The **genes** and the **environmental factors** interact to produce the disease.

Although identification of genes relevant to many diseases is proceeding apace, the complex processes by which genes interact with environmental factors are often hard to identify. (The next section on cancer does describe some recent successes, however.) Explanation schemas for particular diseases such as atherosclerosis do not fully describe causal processes but can describe the general causality of disease, as in the following schema:

Atherosclerosis Explanation Schema:

Explanation target:

Why does a **patient** get **atherosclerosis** with associated **symptoms** such as chest pain?

Explanatory pattern:

The **patient** has inherited various **genes** that encourage the development of risk factors such as hyperlipidemia, hypertension, and diabetes.

The **patient** is subject to various **environmental factors**, such as a high-fat diet.

The **genes** and the **environmental factors** interact to produce the disease.

The number and diversity of multifactorial diseases is too great to diagram a hierarchy of multifactorial disease explanation schemas. Medical textbooks usually classify diseases according to the organ system affected, as in diseases of the lungs or of the stomach. Diseases are also classified in terms of type of disease, as in infectious or nutritional diseases. As more is learned about the genetic influences of multifactorial diseases and about the interaction of genetic and environmental factors, new classifications of these diseases should become possible.

Cancer

There are more than one hundred kinds of cancer, and medical professionals including pathologists, epidemiologists, and oncologists have traditionally treated them as diverse diseases. But molecular genetics has made possible a theory of cancer causation that ties these diseases together, as Bishop and Weinberg report (1996, p. 1): "There is now good reason to believe that a unifying explanation for cancer has been found. No matter what form cancer takes, it remains a malady of genes, and most, if not all, causes of cancer act by damaging genes directly or indirectly." Thus, cancer, the second leading cause of death in advanced countries after heart disease, falls under the Molecular Genetics Disease Explanation Schema.

Since the early 1980s, medical research has discovered that cancer is fundamentally a disease of individual cells and that the behavior of cells can be

understood in terms of the genes operating within them. Our bodies contain approximately 10^{14} cells, whose frequent divisions offer abundant opportunities for harmful genetic mutations to occur. But it is estimated that six or seven successive mutations are needed to convert a normal cell into an invasive carcinoma (Strachan and Read 1996). The genesis of tumors is a multistep process in which successive damage to various genes leads to different kinds of cancer. Recent research has identified three kinds of genes that are frequently mutated in cancer: oncogenes, tumor suppressor genes, and mutator genes. An oncogene is a gene involved in cell proliferation that can help transform a normal cell into a tumor cell. More than one hundred oncogenes have been identified, such as the E6 and E7 oncogenes found in the human papilloma virus HPV16, which can lead to cervical cancer. Some oncogenes are inserted into cells by viruses, but others are mutated versions of genes that are involved in a variety of normal cellular functions. These normal genes, called *proto-oncogenes*, can be transformed by mutations that produce a gain of function, such as increased production of a protein or production of a modified protein, which leads to the stimulation of cell growth. Causes of such mutations can include environmental factors such as smoking and chemical exposure. But oncogenes alone are not sufficient to produce cancer, because cells contain numerous ways of repairing DNA damage. Tumor suppressor genes produce proteins that constrain cell proliferation and help control the unceasing cell growth that oncogenes can cause. Cancers generally arise when the operation of an oncogene, produced by a virus or a mutation in a proto-oncogene, is followed by mutation in a tumor suppressor gene, which then fails to perform its function of controlling growth. This is called the *two-hit* theory of carcinogenesis. More than a dozen tumor suppressor genes have been discovered, such as BRCA1 and BRCA2 which, when rendered ineffective by mutation, often contribute to breast cancer. Other genes implicated in cancer are mutator genes whose loss of function makes a cell prone to errors in information transfer. The role of various oncogenes and tumor suppressor genes in many human cancers is now well established.

The result of these developments is the following disease schema that provides the unifying explanation advocated by Bishop, Weinberg (1996), and other researchers:

Cancer Explanation Schema

Explanation target:

Why does a **patient** get a **cancer?**

Explanatory pattern:

The **patient** has **cells** with active **oncogenes** resulting from a viral infection or a mutation of proto-oncogenes.

These **cells** also contain mutated **tumor suppressor genes.**

The **tumor suppressor genes** have failed to stop the stimulation of growth in the **cells** produced by the **oncogenes**, generating the **patient's cancer.**

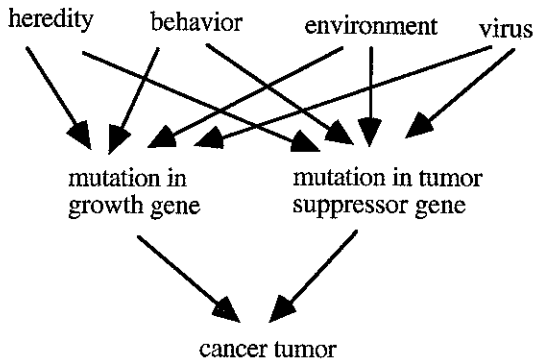

Figure 2.7. Mechanisms of cancer production. Arrows indicate possible causal relations.

This schema applies to many different cancers, involving different kinds of cells (e.g., those of the lung, breast, or prostate) containing different kinds of oncogenes and tumor suppressor genes. For example, a lung cancer explanation schema would specify the typical process of genetic damage caused by smoking. Cancer is clearly a disease that is both polygenic and multifactorial, involving a number of different genes and various inherited and environmental factors that can contribute to mutations.

Explanation schemas for multifactorial diseases can be vividly depicted using causal network diagrams such as the one in Figure 2.7, which displays how mutations in oncogenes and tumor suppressor genes can arise from various causes and can together produce cancer. Verbal schemas such as the Cancer Explanation Schema and pictorial displays of causal networks such as Figure 2.7 both generate explanations as a matter of fit to a particular situation. To explain why a patient became sick is a matter of finding an explanation schema that fits well with the patient's disease and relevant causal factors such as heredity and environment. To make the fit complete, we need to instantiate the terms in the explanation schema (or, equivalently, the factors in the causal network) to provide an answer to the explanatory target, which concerns why the patient became sick. Instantiation can be based on factors known to apply to a patient, such as smoking, or on factors hypothesized to apply, such as genetic disposition based on the frequent familial occurrence of a particular kind of cancer. See chapter 7 for further discussion of causal explanation.

For example, to explain why Fred, a patient with lung cancer, got his disease, we may be able to construct a causal network like the one in Figure 2.8.

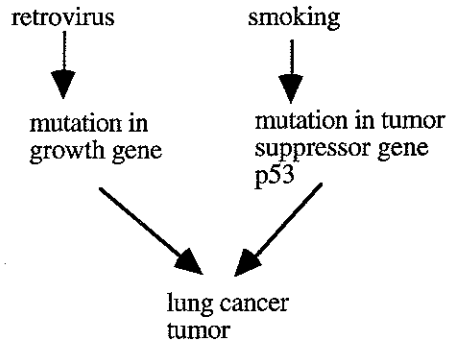

Figure 2.8. Causal explanation of a particular cancer.

This instantiated causal network provides a good fit to Fred's personal situation, although some of the connections in the causal network may be highly conjectural. The explanation of why Fred got cancer does not conform to a deductive or statistical pattern, because there are no universal laws to produce deductions, nor are their sufficient data to provide precise statistical connections between causal factors and carcinogenesis. As with the germ theory, it is difficult to state the new "unifying explanation" extolled by Bishop and Weinberg in terms of universal laws or general models. But the preceding Cancer Explanation Schema captures the kind of causal mechanism that is now believed to be responsible for many different kinds of cancer.

EXPLANATORY AND CONCEPTUAL UNIFICATION

Science is of course much more than a collection of observed facts. In physics, theories such as general relativity and quantum mechanics provide general principles that apply to many phenomena. Evolutionary theory and genetics provide similar unification to biology, as does the theory of plate tectonics to earth science. In medicine, however, unified understanding does not come from the availability of a general overarching theory but from the availability of a system of explanation schemas (these are partly shown in Figure 2.9). Maximum simplicity would result from the applicability of a single explanation schema that accounted for all diseases, as in the eighteenth-century claim by Benjamin Rush that there is only one disease and only one cause: "irregular or convulsive action in the system affected" (quoted in Shryock 1969, p. 3). But as Albert Einstein is reputed to have said, everything should be as simple as possible but not simpler.

My analysis of medical explanation schemas supports the view of Schaffner

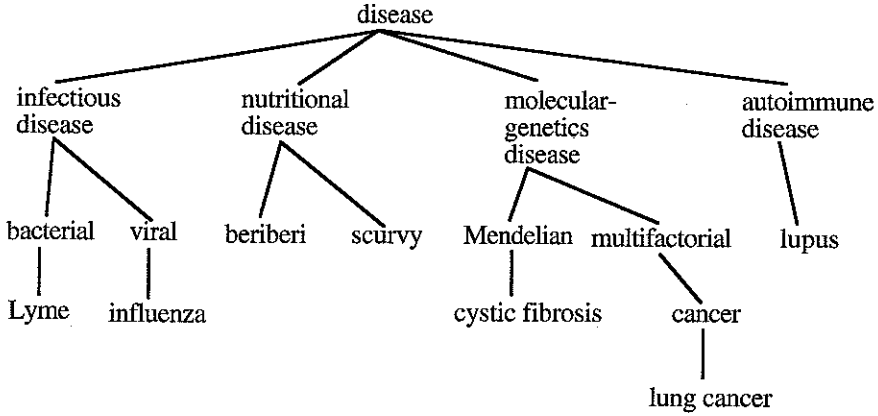

Figure 2.9. Hierarchical organization of disease explanations, with examples of particular diseases. See chapter 10 for further discussion.

(1993) that middle-range biomedical theories are best characterized in terms of hierarchical cognitive structures. These structures provide explanations by fitting particular diseases into general concepts based on common mechanisms. Medical explanation is a matter of fit with causal schemas at different levels of generality, ranging from particular patients to particular diseases to levels of kinds of diseases. Unification in medicine is simultaneously explanatory and conceptual, because what ties explanations together is an organized system of disease concepts.

In recent decades, molecular genetics has made possible new explanation schemas that are having rich applications to Mendelian diseases, cancers, and a wide range of multifactorial diseases that afflict many humans. Chapter 1 argued that the origins of scientific knowledge are also multifactorial and involve a complex of explanation schemas. The next four chapters provide a multifaceted explanation of an important recent development in the theory and treatment of a common disease.

The history of medicine in the nineteenth and twentieth centuries is much richer than my schematic account indicates. Before Pasteur, there were variants of a germ theory of disease as well as a theory that some diseases are caused by miasmas (atmospheric vapors). Much fuller historical treatments include those of Have et al. (1990), Heidel (1941), Hudson (1983), King (1982), Kiple (1993), Magner (1992), Nuland (1988), and Temkin (1973). Philosophical discussions of the nature of disease include the work of Caplan et al. (1981) and Reznek (1987). Also relevant to the germ theory of disease are works on the history of microbiology, such as those by Brock (1961), Collard (1976), Grafe (1991), and Lechevalier and Solotorovsky (1974).

SUMMARY

Disease explanation schemas provide patterns of causal relations responsible for diseases and their symptoms. In the nineteenth century, the humoral theory of disease gave way to the germ theory, which employed different explanation schemas involving infectious agents. Since then, medical research has added explanation schemas for diseases due to nutritional deficiencies, autoimmune reactions, and processes of molecular genetics. Diseases such as cancers can be explained by being fit into a general pattern of genetic and environmental factors. Unified knowledge in medicine comes not from a general set of principles but from the broad applicability of an organized system of explanation schemas.

Part Two

THE BACTERIAL THEORY OF PEPTIC ULCERS

CHAPTER 7

Causes, Correlations, and Mechanisms

I NOW examine in more detail some of the cognitive processes involved in the development of the bacterial theory of ulcers and in other cases of the growth of medical knowledge. This chapter concerns how causes are inferred from correlations and other information about mechanisms and alternative causes. It first discusses the inference from correlation to causation, integrating recent psychological discussions of causal reasoning with epidemiological approaches to understanding disease causation. In addition to the bacterial theory of ulcers, this chapter considers the evolution over the past several decades of ideas about the causes of cancer, particularly lung cancer. Both of these developments involved progression from observed correlations to accepted causal hypotheses (e.g., bacteria cause ulcers, smoking causes cancer), followed by an increased understanding of the mechanisms by which the causes produce the diseases. There is much more to causal reasoning than simply noticing that two factors are associated with each other. I describe how causal mechanisms represented by causal networks can contribute to reasoning that involves correlation and causation. An understanding of causation and causal mechanisms provides the basis for presentation of a model of medical explanation as causal network instantiation.

CORRELATION AND CAUSES

Explanation of why people get a particular disease usually begins by noticing associations between the disease and possible causal factors. For example, the bacterial theory of ulcers originated when Barry Marshall and Robin Warren noticed an association between duodenal ulcer and infection with *Helicobacter pylori* (see chapter 3). They were aware that their study did not establish a cause-and-effect relation between bacteria and ulcers, but they took it as evidence that the bacteria were etiologically related to the ulcers and undertook studies to determine whether eradicating the ulcers would cure the bacteria (see chapters 4 and 5).

A similar progression from correlation to causation has taken place with various kinds of cancer. Over two thousand years ago, Hippocrates described cancers of the skin, stomach, breast, and other body locations and held that cancer is caused, like all diseases, by an imbalance of bodily humors, particu-

larly an excess of black bile. In the eighteenth century, rough correlations were noticed between cancers and various practices: using snuff and nose cancer, pipe smoking and lip cancer, chimney sweeping and scrotum cancer, and being a nun and breast cancer (Proctor 1995, pp. 27–28). The perils of causal reasoning are shown by the inferences of the Italian physician Bernardino Ramazzini, who concluded in 1713 that the increased incidence of breast cancer in nuns was caused by their sexual abstinence, rather than by their not having children. Early in the twentieth century, it was shown that cancers can be induced in laboratory animals by radiation and coal tar.

Lung cancer rates increased significantly in Great Britain and the United States during the first half of the twentieth century, correlating with increase in smoking. Carefully controlled studies, however, began to appear only in the 1950s (Hennekens and Buring 1987, p. 44). In one classic study conducted in England, 649 male and sixty female patients with lung cancer were matched to an equal number of control patients of the same age and sex. For both men and women, there was a strong correlation between lung cancer and smoking, particularly heavy smoking. By 1964, when the U.S. Surgeon General's Report asserted a causal link between lung cancer and smoking, there had been twenty-nine controlled studies performed in numerous countries that showed a high statistical association between lung cancer and smoking. Although the exact mechanism by which smoking causes cancer was not known, more than two hundred different compounds had been identified in cigarette smoke that were known carcinogens.

To grasp how disease explanations work, we need to understand what correlations are, what causes are, and how correlations can provide evidence for causes. Patricia Cheng's (1997) Power PC theory of how people infer causal powers from probabilistic information provides a useful starting point. She proposes that when scientists and other people infer the causes of events, they use an intuitive notion of causal power to explain observed correlations. She characterizes correlation (covariation) in terms of probabilistic contrasts: how much more probable an effect is with a cause than without a cause. The association between an effect e and a possible cause c can be measured by the following equation:

$$\Delta P_c = P(e/c) - P(e/\sim c)$$

The probability of e given c (e/c) is calculated minus the probability of e without c ($e/\sim$c). However, in contrast to many philosophers who try to give a purely probabilistic account of causality, she introduces an additional notion of the *power* of a cause c to produce an effect e, p_c, which is the probability with which c produces e when c is present. Whereas $P(e/c)$ is an observable frequency, p_c is a theoretical entity that is hypothesized to explain frequencies, just as theoretical entities such as electrons and molecules are hypothesized to

explain observations in physics. In Cheng's account, causes are used to provide theoretical explanations of correlations, just as theories such as the kinetic theory of gases are used to explain laws such as those linking observed properties of gases (i.e., pressure, volume, temperature).

Terminologically, I take *correlation* to be interchangeable with *covariation* and *statistical association*. Correlations are not always measured by the statistical formula for coefficient of correlation, which applies only to linear relationships. As with Cheng's theory, the work of Peng and Reggia (1990, p. 101f) involves "probabilistic causal models" that rely not on conditional probabilities of the form P(effect/disease) but on "conditional causal probabilities" of the form P(disease causes effect/disease). Both probabilistic and causal power ideas have a long history in philosophy. On probabilistic causality, see, for example, Eells (1991), Shafer (1996), and Suppes (1970). On causal powers, see, for example, Cartwright (1989) and Harré and Madden (1975).

According to Cheng (1997), a causal power p_c is a probability, but what kind of probability? Philosophers have debated whether probabilities are frequencies, logical relations, or subjective states, but the interpretation of probability that seems to fit best with Cheng's view is that a probability is a propensity, that is, a dispositional property of part of the world to produce a frequency of events in the long run. The causal power p_c cannot be immediately inferred from the observed frequency $P(e/c)$ or the contrast ΔP_c, because the effect e may be due to alternative causes. Celibate nuns get breast cancer more than non-nuns, but it is nonpregnancy rather than celibacy that is causally related to breast cancer. To estimate the causal power of c to produce e, we need to take into account alternative possible causes of e, designated collectively as a. If there are no alternative causes of e besides c, then $P(e/c) = p_c$, but they will normally not be equal if a is present and produces e in the presence of c, that is, if $P(a/c)*p_a > 0$, where p_a is the causal power of a to produce c and * indicates multiplication. In the simple case which a occurs independently of c, Cheng shows that p_c can be estimated using the following equation:

$$p_c = \Delta P_c / 1 - P(a)*p_a$$

The causal relation between e and c can thus be assessed by considering positively the correlation between e and c and negatively the operation of other causes a. When these alternative causes do not occur independently of c, then ΔP_c may not reflect the causal status of c.

Cheng's characterization of the relation between correlations and causal powers fits well with epidemiologists' discussions of the problem of determining the causes of diseases. Her account also fits with the view of Chinn and Brewer (1996) that data interpretation is a matter of building mental models that include alternative explanations. According to Hennekens and Buring

TABLE 7.1
Framework for the Interpretation of an Epidemiological Study

A. Is there a valid statistical association?
   1. Is the association likely to be due to chance?
   2. Is the association likely to be due to bias?
   3. Is the association likely to be due to confounding?

B. Can this valid statistical association be judged as cause and effect?
   1. Is there a strong association?
   2. Is there biological credibility to the hypothesis?
   3. Is there consistency with other studies?
   4. Is the time sequence compatible?
   5. Is there evidence of a dose-response relationship?

Source: From Hennekens and Buring (1987, p. 45).

(1987, p. 30), a causal association is one in which a "change in the frequency or quality of an exposure or characteristic results in a corresponding change in the frequency of the disease or outcome of interest." Elwood (1988, p. 6) says that "a factor is a cause of an event if its operation increases the frequency of the event." These statements incorporate both ΔP_c, captured by the change in frequency, and the idea that the change in frequency is the result of the operation of the cause (i.e., a causal power). Further, epidemiologists stress that assessing whether the results of a study reveal a causal relation requires one to consider alternative explanations of the observed association, such as chance, bias in the design of the study, and confounding alternative causes (Table 4.1; see also Evans 1993; Susser 1973). Thus, the inference from correlation to cause must consider possible alternative causes, p_a.

Hennekens and Buring (1997) summarize their extensive discussion of epidemiological studies in the framework reproduced in Table 7.1. Questions A1 to A3 reflect the need to rule out alternative causes, and questions B1 and B3 reflect the desirability of high correlations, ΔP_c. Cheng's account of causal reasoning captures five of the eight questions relevant to assessing causal power, but the remaining three questions are beyond the scope of her model, which is restricted to induction from observable input. Hennekens and Buring (p. 40) state that "the belief in the existence of a cause and effect relationship is enhanced if there is a known or postulated biologic mechanism by which the exposure might reasonably alter the risk of the disease." Moreover (p. 42), "for a judgment of causality to be reasonable, it should be clear that the exposure of interest preceded the outcome by a period of time consistent with any proposed biologic mechanism." Thus, according to Hennekens and Buring, epidemiologists do and should ask mechanism-related questions about biological credibility and time sequence; this issue is discussed in the next section. Hennekens and Buring's last question concerns the existence of a dose-response

relationship, that is, the observation of a gradient of risk associated with the degree of exposure. This relation is not just ΔP_c, the increased probability of having the disease given the cause, but rather the relation that being subjected to more of the cause produces more of the disease, as when heavy smokers get lung cancer more than light smokers.

Hennekens and Buring (1987) show how answers to the questions in Table 7.1 provide a strong case for a causal connection between smoking and lung cancer. Many studies have found a strong association between smoking and cancer, with a nine- to ten-fold increase in lung cancer among smokers (B1, B3), and the high statistical significance of the results makes it unlikely that the association is due to chance (A1). The conduct of the studies ruled out various sources of observation bias (A2), and researchers controlled for four potential confounding factors: age, sex, social class, and place of residence (A3). By 1959, cigarette smoke was known to contain more than two hundred different compounds that were known carcinogens, providing possible mechanisms to support the biological credibility of the hypothesis that smoking causes cancer (B2). Moreover, there was evidence of a temporal relationship between smoking and cancer, because people are more likely to get lung cancer if they have been smoking for a long time, whereas people who stop smoking dramatically drop their chances of getting cancer (B4). Finally, there is a significant dose-response relationship between smoking and lung cancer, in that the risk of developing lung cancer increases substantially with the number of cigarettes smoked per day and the duration of the habit.

The development of the bacterial theory of ulcers can be interpreted in terms of Cheng's theory of causality (1997) and Hennekens and Buring's framework for epidemiological investigation (1987). As described in chapter 4, when Marshall and Warren first proposed that peptic ulcers are caused by bacteria, most gastroenterologists were highly skeptical. They attributed the presence of bacteria in Warren's gastric biopsies to contamination, and they discounted the correlation between ulcers and bacterial infection as a likely result of chance or incorrect study design. Moreover, an alternative explanation that ulcers are caused by excess acidity was widely accepted because of the success of antacids in alleviating ulcer symptoms. But attitudes toward the ulcer hypothesis changed dramatically when numerous other researchers observed the bacteria in stomach samples, and especially when other research teams replicated Marshall and Warren's finding that eradicating *H. pylori* usually cures ulcers.

The key question is whether bacteria cause ulcers, which requires attributing to *H. pylori* the causal power to increase the occurrence of ulcers. Initial evidence for this attribution was the finding that people with the bacteria have ulcers more frequently than do those without the bacteria:

$$P(ulcers/bacteria) > P(ulcers/no\ bacteria)$$

The early studies, however, could not establish causality because they did not address the question of the existence of possible alternative causes for the ulcers. Whereas lung cancer investigators had to use case-control methods to rule out alternative causes by pairing up patients with lung cancers with similar patients without the disease, ulcer investigators could use the fact that *H. pylori* can be eradicated by antibiotics to perform a highly controlled experiment with one set of patients, comparing them before eradication and after. The eradication experiments described in chapters 3 and 4 show a high value for ΔP, $P(ulcers/bacteria) - P(ulcers/no\ bacteria)$, under circumstances in which no alternative causal factors such as stress, diet, and stomach acidity were varied.

Dose-response relationship has not been a factor in the conclusion that bacteria cause ulcers, since it is not easy to quantify how many bacteria inhabit a given patient's stomach. Time sequence is not much of an issue, since the common presence of the bacteria in children implies that people have the bacteria long before they develop ulcers. But biological credibility, concerning the mechanism by which bacterial infection might produce ulcers, has been the subject of much investigation, as I discuss in the next section. The correlation between ulcers and bacteria might be taken to suggest that ulcers cause bacterial infections, rather than the other way around. But the presence of bacteria is too widespread for this to be plausible: $P(bacteria/ulcers) - P(bacteria/\ no\ ulcers)$ is not high, since the bacteria are quite common, infecting as much as 50% of the population. Moreover, *H. pylori* bacteria were not found to be prominent on gastric ulcer borders, suggesting that the ulcers were not responsible for bacterial growth (see chapter 4).

In sum, much of the practice of physicians and epidemiologists in identifying the causes of diseases can be understood in terms of Cheng's theory, which states that causal powers are theoretical entities that are inferred on the basis of finding correlations and eliminating alternative causes. But mechanism considerations are also often relevant to assessing medical causality.

CAUSES AND MECHANISMS

What are mechanisms, and how does reasoning about them affect the inference of causes from correlations? A mechanism is a system of parts that operate or interact like those of a machine, transmitting forces, motion, and energy to one another. For millennia, humans have used simple machines such as levers, pulleys, inclined planes, screws, and wheels. More complicated machines can be built out of these simple ones, all of which transmit motion from one part to another by direct contact. In the sixteenth and seventeenth centuries, natural philosophers increasingly understood the world in terms of mechanisms, culminating with Newton's unified explanation of the motion of earthly and heavenly bodies. His concept of force, however, went beyond the operation of simple machines by direct contact to include the gravitational interaction of objects at a distance from each other. In the history of science, progress has been made in many sciences by the discovery of new mechanisms, each with interacting parts affecting each other's motion and other properties. Table 7.2 displays some of the most important of these mechanisms. The sciences employ different kinds of mechanisms in their explanations, but each involves a system of parts that change as the result of interactions among them that transmit force, motion, and energy. Mechanical systems are organized hierarchically, in that mechanisms at lower levels (e.g., molecules) produce changes that take place at higher levels (e.g., cells).

TABLE 7.2
Sketch of Some Important Mechanisms in Science

Science	*Parts*	*Changes*	*Interactions*
Physics	Objects such as sun and planets	Motion	Forces such as gravitation
Chemistry	Elements, molecules	Mass, energy	Reactions
Evolutionary biology	Organisms	New species	Natural selection
Genetics	Genes	Genetic transmission and alteration	Heredity, mutation, recombination,
Geology	Geological formations such as mountains	Creation and elimination of formations	Volcanoes, erosion
Plate tectonics	Continents	Motion such as continental drift	Floating, collision
Neuroscience	Neurons	Activation, synaptic connections	Electrochemical transmissions
Cell biology	Cells	Growth	Cell division
Cognitive science	Mental representations	Creation and alteration of representations	Computational procedures

Medical researchers are similarly concerned with finding mechanisms that explain the occurrence of diseases, for therapeutic as well as theoretical purposes: Understanding the mechanism that produces a disease can lead to new ideas about how the disease can be treated. In cancer research, for example, major advances were made in the 1970s and 1980s in understanding the complex of causes that lead to cancer (see chapter 1; Weinberg 1996). There are more than one hundred different kinds of cancer, but all are now thought to result from uncontrolled cell growth arising from a series of genetic mutations, first in genes for promoting growth (oncogenes) and then in genes for suppressing the tumors that are produced by uncontrolled cell growth. The mechanism of cancer production then consists of parts at two levels—cells and the

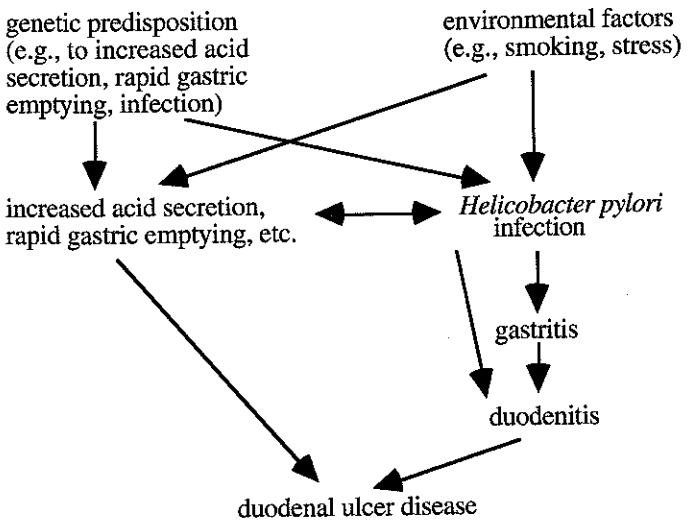

Figure 7.1. Possible mechanism of duodenal ulcer production, providing a richer causal network than that in Figure 4.1. Gastric ulcer causation is similar. Modified from Graham (1989, p. 51).

genes they contain, along with changes in cell growth produced by a series of genetic mutations. Mutations in an individual can occur for a number of causes, including heredity, viruses, and behavioral and environmental factors such as smoking, diet, and exposure to chemicals. Figure 2.7 summed up the current understanding of the mechanisms underlying cancer. This understanding is currently generating new experimental treatments based on genetic manipulations such as restoring the function of tumor suppresser genes (Bishop and Weinberg 1996).

Ulcer researchers have been concerned with the mechanism by which *H. pylori* infection produces ulcers. Figure 7.1 displays a mechanism similar to one proposed by David Graham (1989) that shows some of the interactions of heredity, environment, infection, and ulceration. Research is underway to fill in the gaps about these processes; it is looking, for example, at interactions between particular strains of *H. pylori* and the immune defenses of particular hosts.

Recent psychological research by Woo-kyoung Ahn and her colleagues has found that when ordinary people are asked to provide causes for events, they seek out information about underlying causal mechanisms as well as information about correlations (Ahn and Bailenson 1996; Ahn et al. 1995). For example, if people are asked to state the cause of John's car accident, they do not survey a range of possible factors that correlate with accidents but rather focus on the process underlying the relationship between cause and effect, such as

John's being drunk leading to erratic driving that led to the accident. Whereas causal attribution based on correlation (covariation) alone would ignore mechanisms connecting cause and effects, ordinary people are like medical researchers in that they seek mechanisms that connect cause and effect. Koslowski (1996) reports that causal reasoning in both children and adults makes good use of mechanism information as well as correlation information.

As Cheng (1997) points out, however, the emphasis on mechanism does not by itself provide an answer to the question of how people infer cause from correlation: Knowledge of mechanisms is itself knowledge of causally related events that must have somehow been previously acquired. Medical researchers inferred that bacteria cause ulcers and that smoking causes cancer when little was known about the relevant causal mechanisms. Reasoning about mechanisms can contribute to causal inference, but it is not necessary for such inference. In domains in which causal knowledge is rich, there is a kind of feedback loop in which more knowledge about causes leads to more knowledge about mechanisms, which leads to more knowledge about causes. But in less well-understood domains, correlations and the consideration of alternative causes can get causal knowledge started in the absence of much comprehension of mechanisms.

To understand how reasoning about mechanisms affects reasoning about causes, we need to consider four different situations that arise in science and ordinary life when we consider whether a factor *c* is a cause of an event *e:*

1. There is a known mechanism by which *c* produces *e*.
2. There is a plausible mechanism by which *c* produces *e*.
3. There is no known mechanism by which *c* produces *e*.
4. There is no plausible mechanism by which *c* produces *e*.

For there to be a known mechanism by which *c* produces *e*, *c* must be a component of or an occurrence in a system of parts that is known to interact to produce *e*. Only very recently has a precise mechanism by which smoking causes cancer become known: A component of cigarette smoke (Benzo[a]pyrene) was identified that produces mutations in the tumor suppresser gene *p53* (Denissenko et al. 1996). As we just saw, however, there has long been a *plausible* mechanism by which smoking causes lung cancer.

When there is a known mechanism connecting *c* and *e*, the inference that *c* causes *e* is strongly encouraged, although careful causal inference still needs to take into account information about correlations and alternative causes: A different mechanism may have produced *e* by an alternative cause *a*. For example, drunk driving often produces erratic driving that produces accidents, but even if John was drunk, his accident might have been caused by a mechanical malfunction rather than his drunkenness. Similarly, even though there is now a plausible mechanism connecting *H. pylori* infection and ulcers, we should not immediately conclude that a patient with an ulcer has the infection,

since approximately twenty percent of ulcers are caused by the use of nonsteroidal anti-inflammatory drugs such as aspirin. But an awareness of known and plausible mechanisms connecting *c* and *e* clearly facilitates the inference that *c* causes *e*, in a manner that is more fully spelled out later. Another way in which the plausibility of a mechanism can be judged is by analogy: If a cause and effect are similar to another cause and effect that are connected by a known mechanism, it is plausible that a similar mechanism may operate in the original case. There was a plausible mechanism by which *H. pylori* caused stomach ulcers, since other bacteria were known to produce other sores.

Sometimes causal inference from correlation can be blocked when there is no plausible mechanism connecting the event and its cause, that is, when possible mechanisms are incompatible with what is known. When Marshall and Warren first proposed that bacteria cause ulcers, the stomach was widely believed to be too acidic for bacteria to survive so there was no plausible mechanism by which bacteria could produce ulcers. Later it was found that *H. pylori* produce ammonia, which neutralizes stomach acid and thereby allows them to survive, removing the implausibility of the bacteria-ulcer mechanism. Similarly, when Alfred Wegener proposed continental drift early in this century, his theory was rejected in part because the mechanisms he proposed for continental motion were incompatible with contemporary geophysics. Only when plate tectonics was developed in the 1960s was it understood how continents can be in motion.

The two cases just mentioned are ones in which the implausibility of mechanisms was overcome, but there are many cases in which a rejection of causal relations remains appropriate. Even though there are some empirical studies that provide weak correlational evidence for extrasensory perception (ESP), it is difficult to believe that people have such powers as telepathy and telekinesis, which have properties that conflict with known physical mechanisms, such as being unaffected by spatial and temporal relations. Similarly, homeopathic medicine, which uses minute doses of drugs, violates established views concerning the amounts of substances needed to be chemically effective. A more extreme case is the theory of Velikovsky that the planet Venus once swung close to Earth and caused many historical events such as the parting of the Red Sea for Moses. Such planetary motion is totally incompatible with Newtonian mechanics, so there is no plausible mechanism by which Venus's motion could have had the claimed effect.

How can medical researchers and ordinary people combine information about mechanisms with information about correlations and alternative causes to reach conclusions about cause and effect? Recall Cheng's view (1997) that causes are theoretical entities to be inferred on the basis of correlations and alternative causes. I have argued that the justification of scientific theories, including their postulation of theoretical entities, is a matter of explanatory coherence, in which a theory is accepted because it provides a better explanation of the evidence (see chapter 4; Thagard 1992b). Explanatory coherence of a hypothesis is a matter of both the evidence it explains and its being explained by higher level hypotheses. Charles Darwin, for example, justified the hypothesis of evolution in terms of both the biological evidence it explained and the supposition that evolution could be explained by the mechanism of natural selection. Moreover, he explicitly compared the explanatory power of his theory of evolution by natural selection with the explanatory limitations of the dominant creationist theory of the origin of species. These three factors—explaining evidence, being explained by mechanisms, and a consideration of alternative hypotheses—are precisely the same considerations that go into an evaluation of a causal hypothesis.

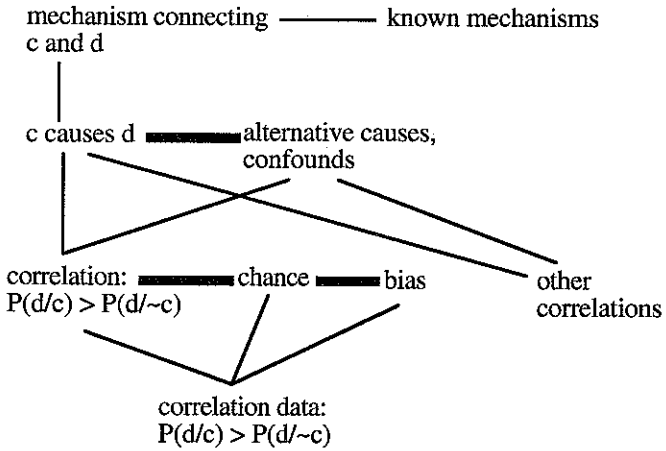

Figure 7.2. Inferring a cause *c* from correlation data about a disease *d*. That there is a correlation between *d* and *c* must be a better explanation of the observed correlation than chance or bias (or fraud). That *c* causes *d* must be a better explanation of the correlation and other correlations than alternative confounding causes. The existence of a mechanism connecting *c* and *d* provides an explanation of why *c* causes *d*. In the figure, thin lines are explanatory relations, whereas the thick lines indicate incompatibility.

Figure 7.2 shows how the inference that *c* causes a disease *d* can be understood in terms of explanatory coherence. When medical researchers collect data that find a correlation between *c* and *d*, that is, a high value for $P(d/c) - P(d/\sim c)$, there are several possible explanations for these data. That a correlation does exist in the relevant population between *d* and *c* is one explanation for the data, but experimenters must rule out other explanations, for example, that the correlation in the data arose from chance or experimental bias. Mayo (1996) provides a thorough discussion of the use of statistical tests to rule out errors derived from chance and other factors. Another possible source of error

is fraud, in which the observed correlations are based on fabricated data. Careful experimental designs involving such techniques as randomization and double blinding help rule out bias, and appropriate techniques of statistical inference tend to rule out chance, leading one to accept the hypothesis that there is a real correlation between *c* and *d*. However, before researchers can conclude that *c* causes *d*, they must have reason to believe that this hypothesis is a better explanation of the correlation than other confounding causes that might have been responsible for the correlation. Again, careful experimental design that manipulates only *c* or that otherwise controls for other potential causes is the key to concluding that *c* causes *d* is the best explanation of the correlation. In addition, the existence of a known or plausible mechanism for how *c* can produce *d* increases the explanatory coherence of the causal hypothesis. On the other hand, if all mechanisms that might connect *c* with *d* are incompatible with other scientific knowledge, then the hypothesis that *c* causes *d* becomes incoherent with the total body of knowledge. As Hennekens and Buring (1987) suggest, a major determinant of whether a causal hypothesis makes sense is whether it comes with a plausible underlying mechanism.

Figure 7.2 points to a synthesis of Cheng's ideas about causal powers, probabilities, and alternative causes with considerations of mechanism. Mechanisms are not a necessary condition for causal inference, but when they are known or plausible, they can enhance the explanatory coherence of a causal hypothesis. Moreover, causal hypotheses incompatible with known mechanisms are greatly reduced in explanatory coherence. Inference to causes, like inference to theoretical entities in general, depends on explanatory coherence as determined by evidence, alternative hypotheses, and higher level hypotheses.

Inference to medical causes is similar to legal inference concerning responsibility for crimes. In a murder case, for example, the acceptability of the hypothesis that someone is the murderer depends on how well that hypothesis explains the evidence, on the availability of other hypotheses to explain the evidence, and on the presence of a motive that would provide a higher level explanation of why the accused committed the murder. Motives in murder trials are like mechanisms in medical reasoning, in that they provide nonessential but coherence-enhancing explanations of a hypothesis.

This section has discussed how knowledge of mechanisms can affect inferences about causality, but it has passed over the question of how such knowledge is obtained. There are three possibilities. First, some knowledge about basic physical mechanisms may be innate, providing an infant with a head start for figuring out the world. It is possible, for example, that infants are innately equipped to infer a causal relation when one moving object bangs into another object that then starts moving. Second, some of the links in the causal chains that constitute a mechanism may be learned by induction from observed correlations as described in Cheng's Power PC model (1997). For example, we can

observe the relations among pressure, temperature, and volume changes in gases and infer that they are causally connected. Third, sometimes mechanisms are abduced, that is, posited as a package of hypothetical links used to explain something observed. In cognitive science, for example, we posit computational mechanisms with various representations and processes to explain intelligent behavior. Darwin abduced the following causal chain:

variation + competition → natural selection → evolution of species

The difference between abductive and inductive inference about mechanisms is that in inductive inference the parts and processes are observed, whereas in abductive inference they are hypothesized. Knowledge about mechanisms involving theoretical (nonobservable) entities must be gained abductively, by inferring that the existence of the mechanism is the best explanation of the results of observation and experimentation. Different domains vary in the extent to which knowledge about mechanisms is innate, induced from correlations, or abductive.

DISEASE EXPLANATION AS CAUSAL NETWORK INSTANTIATION

The previous description of the interrelations of correlations, causes, and mechanisms provides the basis for an account of the nature of medical explanation. First we can eliminate a number of defective alternative accounts of explanation, including accounts in which explanation is essentially deductive, statistical, or involves single causes.

1. Explanation is not deductive. The deductive-nomological model of Hempel (1965), according to which an explanation is a deduction of a fact to be explained from universal laws, clearly does not apply to the kinds of medical explanation discussed here. Deductive explanations can be found in other fields such as physics, in which mathematical laws entail observations. But there are no general laws about the origins of ulcers and cancer. As we saw, most people with *H. pylori* do not develop ulcers, and many people without *H. pylori* do develop ulcers because of nonsteroidal anti-inflammatory drugs. Similarly, most smokers do not get lung cancer, and some nonsmokers do get lung cancer. The development of ulcers, like the development of cancer, is far too complex for general laws to provide deductive explanation.

2. Explanation is not statistical. Statistics are certainly relevant to developing medical explanations, as we saw in the contribution of the equation *P(ulcers/bacteria) – P(ulcers/no bacteria)* to the conclusion that bacteria cause ulcers. But correlations themselves have no explanatory force, since they may be the result of confounding alternative causes. As we saw in Figure 7.2, the

conclusion that there is a causal and hence explanatory relation between a factor and a disease depends on numerous coherence considerations, including the full range of correlations explained, the applicability of alternative causes, and the availability of a mechanism by which the factor produces the disease. A medical explanation need not show that a disease was to be expected with high probability, since the probability of getting the disease given the main cause may well be less than 0.5, as is the case for both ulcers and bacteria and lung cancer and smoking.

3. Explanation is not in terms of single causes. Although it is legitimate to see bacteria as the major causal factor in most ulcers and to see smoking as the major causal factor in most cases of lung cancer, it is simplistic to explain someone's ulcer only in terms of bacterial infection, or someone's lung cancer only in terms of smoking. As Figures 2.7 and 7.1 displayed, ulcer causation and cancer causation are complex processes that involve multiple interacting factors. Medical researchers are increasingly stressing the multifactorial nature of disease explanations. Adult-onset diabetes, for example, is now understood as arising from a complex of factors including heredity, obesity, and inactivity, all of which contribute to glucose intolerance, possibly because of a mechanism that involves a protein that reduces glucose uptake.

I propose instead that medical explanation should be thought of as *causal network instantiation*. For each disease, epidemiological studies and biological research establish a system of causal factors involved in the production of a disease. The causal network for cancer is a more elaborate version of Figure 2.7, and the causal network for ulcers is a more elaborate version of Figure 7.3, which is an elaboration of Figure 7.1. A crucial point is that the nodes in this network are connected not merely by conditional probabilities, *P(effect/cause)*, but by causal relations inferred on the basis of multiple considerations, including correlations *P(effect/cause)* – *P(effect/~cause)*, alternative causes, and mechanisms. Given this network, we explain why a given patient has a given disease by instantiating the network, that is, by specifying which factors operate in that patient. For a patient with stomach pains, a physician can start to instantiate the network in Figure 7.3 by determining whether the patient takes large quantities of nonsteroidal anti-inflammatory drugs, (e.g., because of arthritis). Different instantiation can take place on the basis of tests (e.g., endoscopy or a breath test) to determine whether the patient's stomach is infected with *H. pylori* bacteria. Some instantiation will be abductive, making hypotheses about the operation of factors that cannot be observed or tested for. The physician might make the abduction that a patient has a hereditary inclination to excess acidity, which would explain why he or she, unlike most people with *H. pylori*, has an ulcer; the hereditary abduction would be strengthened if the patient's parents and other relatives had ulcers. Similarly, to explain patients' lung cancers, we instantiate a causal network with information about their smoking, their other behaviors, their heredity, and so on. Recent work on

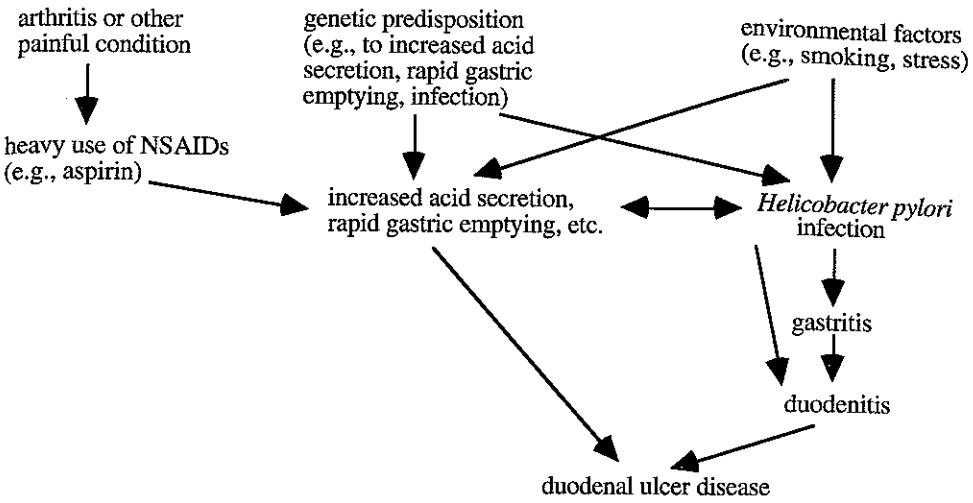

Figure 7.3. General causal network for duodenal ulcers, expanding on the network in Figure 7.1. NSAIDs are nonsteroidal anti-inflammatory drugs.

causal networks includes Glymour et al. (1987), Iwasaki and Simon (1994), Pearl (1988), and Shafer (1996).

Instantiation of a causal network such as the one in Figure 7.3 produces a kind of narrative explanation of why a person becomes sick. We can tell several possible stories about a patient, such as the following:

1. The patient became infected with *H. pylori* and developed ulcers because of a predisposition to excess acidity.

2. The patient took a lot of aspirin for arthritis and developed ulcers because of the resulting vulnerability to acidity.

But medical explanation is not just story telling, since a good medical explanation should point to all the interacting factors for which there is causal evidence and for which there is evidence of relevance to the case at hand. A narrative may be a useful device for communicating a causal network instantiation, but it is the ensemble of statistically based causal relations that is more crucial to the explanation.

Causal networks provide an explanatory schema or pattern, but they differ from the sorts of explanatory schemas and patterns proposed by others. Unlike the explanatory patterns of Kitcher (1981, 1993), causal networks are not deductive. Deductive patterns may well have applications in fields such as mathematical physics, but they are of no use in medicine, in which causal relationships are not well represented by universal laws. Unlike the explanation patterns of Schank (1986), causal networks are not simple schemas that are used to provide single causes for effects, but they instead describe complex

mechanisms of multiple interacting factors. My account of medical explanation as causal network instantiation is compatible with the emphasis on mechanistic explanations by Salmon (1984) and Humphreys (1989), but it provides a fuller specification of how causal networks are constructed and applied. As already mentioned, my account of causal network installation is not compatible with interpreting the relations between factors in a causal network purely in terms of conditional probabilities.

Like the explanation of a disease in a particular patient, the explanation of why a group of people are prone to a particular disease is also a matter of causal network instantiation. People in underdeveloped countries are more likely to have gastritis than are North Americans, because poorer sanitation makes it more likely that they will acquire *H. pylori* infections that produce ulcers. Nuns are more likely to get breast cancer than are other women, because women who do not have full-term pregnancies before the age of 30 are more likely to get breast cancers, probably because of some mechanism by which pregnancy affects breast cell division. When we want to explain why a group is more likely to get a disease, we invoke the causal network for the disease and instantiate the nodes based on observations and abductions about the disease factors possessed by members of the group. Thus causal network instantiation explanations of the occurrence of both individual and group disease are structurally identical.

CONCLUSION

This chapter has shown how correlations, causes, and mechanisms all figure in the construction of causal networks that can be instantiated to provide medical explanations. The main criterion for assessing a model of disease explanation is whether it accounts for the explanatory reasoning of medical researchers and practitioners. We have seen that the causal network instantiation model of medical explanation fits well with methodological recommendations of epidemiologists such as Hennekens and Buring as well as with the practice of medical researchers working on diseases such as ulcers and lung cancer. Additional examples of the development and application of causal networks could easily be generated for other diseases such as diabetes. My account of medical explanation as causal network instantiation gains further credibility from the fact that its assumptions about the relations of correlations, causes, and mechanisms are consistent with (and provide a synthesis of) Cheng's and Ahn's psychological findings about human causal reasoning. I have not attempted to define cause in terms of explanation or explanation in terms of cause; rather, causes, mechanisms, explanations, and explanatory coherence are intertwined notions.

For some fields such as physics, the existence of universal laws and mathematical precision often makes possible explanations that are deductive. On the other hand, in fields such as economics, the lack of causal knowledge interrelating various economic factors may restrict explanations to those based on statistical associations. I expect, however, that there are many fields, such as evolutionary biology, ecology, genetics, psychology, and sociology, in which explanatory practice fits the causal network instantiation model. The possession of a feature or behavior by members of a particular species, for example, can be explained in terms of a causal network that involves mechanisms of genetics and natural selection. Similarly, the possession of a trait or behavior by a human can be understood in terms of a causal network of hereditary, environmental, and psychological factors. In psychology, as in medicine, explanation is complex and multifactorial in ways well characterized as causal network instantiation.

SUMMARY

There is much more to inferring the cause of a disease than noticing correlations with another factor. Causal reasoning requires the abductive inference that a factor has the power to produce an effect. This inference involves noticing that the effect is more probable given the factor than otherwise, but it also requires considering alternative causal factors and the plausibility of mechanisms by which the factor produces the effect. Overall, the inference that a factor is the cause of a disease is a matter of the explanatory coherence of the causal hypothesis. Disease explanations are best characterized not as deductive or statistical inferences but as instantiations of complex causal networks.

CHAPTER 8

Discovering Causes: Scurvy, Mad Cow Disease, AIDS, and Chronic Fatigue Syndrome

HUMANS ARE subject to many hundreds of diseases. Some of the diseases, such as cancer and epilepsy, were familiar to the ancient Greeks, whereas others such as AIDS (acquired immunodeficiency syndrome) and Lyme disease have become known only in recent decades. Solid understanding of the causes of diseases is relatively recent, stemming from such sources as the investigation of infectious diseases in the second half of the nineteenth century, the explanation of nutritional diseases in the first half of the twentieth century, and the more recent understanding of many common diseases in terms of molecular genetics (see chapter 2).

This chapter uses the history of ideas about four diseases to draw some general conclusions about why it is often so difficult to determine the causes of diseases. I first describe four overlapping stages in the development of medical understanding: disease characterization, cause specification, experimentation, and mechanism elaboration. The operation of these stages is shown in the history of four diseases (or classes of disease) that differ in the extent to which they are understood: scurvy, spongiform encephalopathies (including mad cow disease), AIDS, and chronic fatigue syndrome. Understanding of these diseases ranges from that of scurvy, which has been known to result from vitamin C deficiency since early in this century, to that of chronic fatigue syndrome, whose nature and etiology are still highly controversial. The account of causal reasoning given in chapter 7 provides a framework for understanding the difficulties of discovering the causes of disease.

STAGES OF DISEASE UNDERSTANDING

The first stage of understanding a disease is its characterization, that is, its identification as a special kind of process with its own set of symptoms that differentiate it from other diseases. This stage is not as simple as it seems, since it requires first the association of a set of characteristic symptoms and second the differentiation of the newly associated symptoms from those of other diseases. The ancient Greeks had a large category of diseases called

fevers, which were considered specific diseases, not symptoms of various diseases. Many diseases have similar symptoms, so that historically there has been confusion between such diseases as smallpox and measles. Characterization of a disease, at least since the time of Hippocrates, has also involved describing the course of the disease, that is, the way its symptoms change over time.

The second stage of disease understanding is the specification of possible causes. This stage may be intermixed with the first: Sometimes two diseases are differentiated only when they are found to have different causes. Usually, however, medical practitioners have some idea about the nature of a disease before they begin to speculate about its causes. Here are three different ways in which the causes of diseases can be proposed:

Correlation: An observed factor is found to occur with a disease, so the factor is considered to be a cause of the disease.
Postulation: An unobserved factor is hypothesized to cause a disease.
Biochemical analysis: Close examination of the biochemical nature of a host identifies new factors that may be a cause of the host's disease.

Historical examples of causal reasoning may involve mixtures of these three kinds of reasoning, as when the germ theory of disease led Robert Koch to postulate that a bacterium was responsible for tuberculosis and subsequently to use microscopy to identify bacteria that correlated with the disease. More examples of cause specification are provided later.

The third stage, moving from a consideration of possible causes to a conviction that the cause of a disease has been found, can be accomplished only by experimentation. Correlation, postulation, and biochemical analysis are useless unless carefully controlled experiments show that a factor is plausibly the cause of disease. The history of medicine is littered with discarded causes. To show that a correlated or hypothesized factor is the cause of a disease, it is necessary to consider other possible causes that might be responsible for the disease. The hypothesis that one factor is the cause of the disease must have more explanatory coherence than alternative hypotheses concerning other factors. A correlation between a factor and a disease might be the result of chance or other causal relations, such as there being a common cause of the factor and the disease. The four cases described subsequently illustrate the difficulty of conducting experiments that determine causality.

The fourth stage of disease understanding is the elaboration of mechanisms by which a disease is produced by its cause or causes. As earlier chapters described, many diseases are multifactorial, involving a collection of interacting causes. Mechanism elaboration usually follows experimental determination that a factor causes a disease, but an understanding of biochemical processes has sometimes been crucial in suggesting what the causes of a disease might be. Hence, these four stages of disease understanding should not be seen

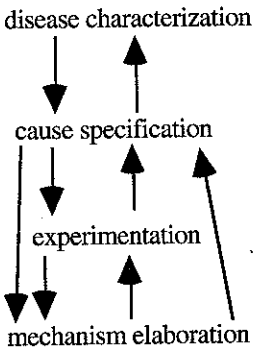

Figure 8.1. Interacting stages of disease understanding.

as discrete temporal periods but rather as four subprocesses of the process of disease understanding, as shown in Figure 8.1. For peptic ulcers, disease characterization is centuries old, but cause specification and experimentation have taken place only recently, and mechanism elaboration is still under way. To further illustrate these stages and the difficulties of discovering causes, I now briefly review the history of ideas about scurvy, spongiform encephalopathies, AIDS, and chronic fatigue syndrome.

SCURVY

In the fifteenth century, explorers from Portugal, Spain, France, and England began to make long sea voyages. By 1498, it had been noticed that sailors on these voyages often fell ill, with weakness, swollen limbs and bleeding gums (Carpenter 1986). Within a few decades, commentators had remarked that ill sailors could quickly be cured with fresh food such as oranges. But scurvy, as the disease was soon called, was not established as a nutritional deficiency until the twentieth century. Why did it take more than three hundred years to determine the cause of this disease?

Characterization of scurvy was relatively unproblematic because of the salience of the complex of symptoms that affected sea voyagers, including sores and multiple purple spots as well as the gum and limb problems just mentioned. Although it seems obvious now that the sailors suffered from a lack of vitamin C, the concept of a nutritional disease is less than a century old, dating from the discovery that beriberi was caused by a diet of polished rice (McCollum 1957). In 1622, Sir Richard Hawkins, a British sea captain who led expeditions to South America, recommended oranges and lemons as a treatment for scurvy, but he also suggested the oil of vitriol (sulfuric acid) as beneficial (Carpenter 1986, p. 15). Other recommended preventive measures included

keeping the ship clean, burning tar, wearing dry clothes, exercising, and eating breakfasts of bread and diluted wine.

Sea voyages and their associated conditions were clearly responsible for the disease, but there were many unusual features of such voyages that suggested themselves as possible causes of disease. Disease specification proceeded partly in terms of correlation, as in the suggestion that the frequently damp conditions of sea travel caused sailors to become scorbutic. Another fact about sea voyages was that sailors ate a lot of salt meat, so that this was proposed as a cause of scurvy; what sailors did eat was perhaps more salient that what they did not eat. In the sixteenth through the eighteenth centuries, the dominant medical framework was still the humoral theory of Hippocrates and Galen, according to which illness is the result of a bodily imbalance among the four main fluids (humors): blood, phlegm, yellow bile, and black bile. Accordingly, Dutch physician John Echth wrote in the mid-sixteenth century that scurvy is a disease of the spleen caused by an excess of the melancholic humor, black bile.

In 1734, another Dutch physician, John Bachstrom, proposed that the absence of fresh vegetable food was the sole cause of scurvy, but his proposal was largely ignored, in part because it did not fit with general contemporary views of disease. In 1753, British surgeon James Lind published *A Treatise of the Scurvy*. Despite having performed an experiment (possibly the first controlled experiment in medical history) that found that oranges and lemons were a much better treatment for scurvy than cider, oil of vitriol, vinegar, or sea water, Lind's theory was that scurvy was the result of a cold, wet climate producing constriction of the pores that blocked perspiration. This blockage produced a concentration of humors in the body that induced scurvy, which could be treated with improvements of ships' air, acids, and lemon juice. For Lind, moisture was a more important causal factor in scurvy than diet; he argued that abstinence from vegetables and fruits could not be the primary cause of scurvy, for the ancients had not observed the disease in besieged towns where food was severely limited (Lind 1953, p. 73). The definition of scurvy in the 1933 edition of the Oxford English Dictionary states that it is "induced by exposure and by a too liberal diet of salted foods."

British naval officers such as Captain Cook took various measures against scurvy, but their mixture made it difficult to determine which ones were actually effective. One commonly used treatment was wort, a fermenting infusion of malt. By 1800, British ships routinely carried lemon or lime juice, and scurvy became much rarer, although it later arose during special situations such as the Irish potato famine of the 1850s and the long treks of arctic explorers. Lemon and lime juice did not always succeed in preventing scurvy on long voyages, probably because of the dilution and deterioration of vitamin C. Scientific developments about substances such as acids, oxygen, protein, and

potassium suggested new possible causes of scurvy. The second half of the nineteenth century saw many medical breakthroughs based on the germ theory of disease, and bacterial contamination was accordingly considered as a possible cause of scurvy.

Aside from Lind's comparison of different treatments for scurvy, controlled experimentation played almost no role in the development of ideas about scurvy until the twentieth century. In 1907, two Norwegian researchers performed systematic experiments involving the causes of scurvy in guinea pigs, with results suggesting that diet was both the cause and cure of scurvy. However, a leading U.S. nutrition researcher, E. V. McCollum, rejected this view because diet did not produce scurvy in rats. At the time, it was as reasonable to take rats as a medical analog of humans as it was guinea pigs, although we now know that rats differ from humans in being able to synthesize vitamin C, so they do not require it in their diets. Experiments with monkeys showed that lemon juice provided full protection against scurvy, and attempts were made to identify the vitamin responsible for this protection. The concept of a vitamin—a factor necessary for good nutrition—was formed by Funk (1912).

In the 1920s, a Hungarian scientist, Albert Szent-Györgyi, isolated a substance he called hexuronic acid. In the 1930s, it was shown that this substance could be extracted from lemon juice and used to cure scurvy, so it was renamed ascorbic acid. Scurvy was firmly established as a nutritional disease caused by insufficient ascorbic acid. The key steps in this establishment were animal experiments showing that diet was responsible for scurvy and chemical identification of the specific substance. Later research showed that ascorbic acid is required for collagen metabolism, yielding the following mechanism:

$$\text{poor diet} \rightarrow \text{ascorbic acid deficiency} \rightarrow$$
$$\text{defective collagen biosynthesis} \rightarrow \text{scurvy.}$$

These twentieth-century breakthroughs should not obscure the previous centuries of laborious unsuccessful attempts to identify the cause of scurvy, which illustrate general difficulties of causal reasoning discussed later in this chapter.

SPONGIFORM ENCEPHALOPATHIES

Spongiform encephalopathies are members of a class of diseases found in humans and other animals. Bovine spongiform encephalopathy, also known as BSE and mad cow disease, has infected more than 100,000 British cattle since 1986, and by 1996 there was some evidence of the spread of the disease to humans. Spongiform encephalopathies are characterized by neurological degeneration that leads to progressively severe psychomotor dysfunction and death. Many researchers now believe that these diseases are caused by novel

infectious proteins called prions (pronounced "pree-ons"), which were hypothesized in 1982. A review of the history of ideas about this class of diseases reveals numerous interesting aspects of medical causal reasoning.

The first disease of this class to be described was scrapie, a disease of sheep and goats that has been recognized in these animals for more than two hundred years (Collinge and Palmer 1992). A similar neurodegenerative disease in humans, marked by a rapidly progressive dementia usually followed by death within a year, was identified in the 1920s and called Creutzfeldt-Jakob disease. In the 1950s, physicians who had studied the Fore people of Papua New Guinea identified a disease the Fore called *kuru*, which was characterized by a loss of coordination, dementia, and death (Gajdusek and Zigas 1957). Initially, there was much uncertainty about the cause of the disease, which did not seem to have infectious, nutritional, or toxic origins. Early suspicion that the disease was genetic gave way, however, to the conviction that ritual cannibalism was the main cause of this disease, because of its prevalence among women and children who ate the brains of deceased relatives (Mathews et al. 1968).

The symptoms of spongiform encephalopathies are so striking that characterization of the various diseases has been unproblematic. But determination of the causes of these diseases has taken many decades, and the prion hypothesis is still somewhat controversial. In 1959, the similarity between kuru and scrapie was noticed by W. J. Hadlow (1959, 1992), a scrapie researcher who saw a museum exhibit on the Fore brain disease. He systematically laid out the similarities between the two diseases:

- Both are endemic in confined populations with a low incidence of one or two percent.
- Both have onset with no fever or other signs of illness.
- Both are almost always fatal within only a few months.
- Both involve ataxia and severe behavioral changes.
- Both are accompanied by widespread neuronal degeneration.

On the basis of the fact that scrapie had been induced experimentally in sheep, Hadlow suggested an experiment involving the induction of kuru in a laboratory primate. He could not argue analogically that the cause of kuru might specifically be identified with the cause of scrapie, which was equally unknown, but he made the important conjecture that pathogenesis of the two diseases might be similar. Daniel Gajdusek and his colleagues performed experiments showing that both kuru and Creutzfeldt-Jakob disease could be transmitted to chimpanzees, and they classified these diseases together as "transmissible spongiform encephalopathies" (Brown and Gajdusek 1991; Gajdusek, et al. 1966).

At the time, the most plausible hypothesis for the causes of these diseases concerned some kind or kinds of slow-acting virus. But Stanley Prusiner

(1982) audaciously proposed that scrapie is caused by a novel proteinaceous infectious particle, or prion. The scrapie agent had been purified from sheep brains, and investigation showed that it contained a protein required for infectivity but did not contain nucleic acids characteristic of viruses. The scrapie agent was inactivated by chemical treatments that destroyed protein but not by chemical treatments that destroyed nucleic acid. Prusiner is now sufficiently confident that spongiform encephalopathies all are caused by prions that he classes them together as *prion diseases* (Prusiner 1996; Prusiner et al. 1992). He analogically hypothesizes that other neurodegenerative diseases of humans such as Alzheimer's disease and multiple sclerosis might also be causally linked to prions (Prusiner 1982; Prusiner et al. 1992). Prusiner was awarded the 1997 Nobel Prize for medicine.

The prion hypothesis is not universally accepted, however. Some researchers believe that tests have been insufficiently sensitive to detect viral nucleic acid in the scrapie agent and that an unknown small retrovirus capable of altering host protein is the primary cause (Dal Canto 1991; Rohmer 1991). The prion hypothesis is impressive, but researchers have not been able to explain how protein particles replicate or how prions produce neurological degeneration. Moreover, viruses occur in different strains, which would explain the difficulty of transferring spongiform encephalopathies across species, whereas prions do not have multiple strains. Prusiner describes genetic differences at the level of protein production that explain why transference across species is difficult but not impossible. Work is under way to determine the mechanism by which abnormal prion protein can spread, prevent normal protein reproduction, and thereby produce defective brain development.

Acceptance of the prion hypothesis has required several important kinds of conceptual change. First, prions are a new kind of infectious agent, very different from the bacteria and viruses that have been identified as the causes of many human diseases. Second, if Prusiner's terminology is correct, the prion hypothesis has generated a new class of diseases, prion diseases, and thereby expanded the tree of infectious diseases. Third, the prion hypothesis altered the normal classification of diseases that distinguishes between infectious and hereditary diseases. Creutzfeldt-Jakob disease can develop in humans as the result of inherited defects in protein production or as the result of infection by medical procedures such as brain surgery or by eating infected beef. Thus Crentzfeldt-Jakob disease is both a hereditary and an infectious disease, collapsing the usual disease classification. A similar collapse occurs with cancer (see chapter 2), because genetic abnormalities that cause cancer can arise through inheritance, viral infection, or environmental causes. Cancer is not classed as an infectious disease, but some cancers are caused by infectious agents (e.g., Kaposi's sarcoma, which is discussed in the next section). Creutzfeldt-Jakob disease occurs sporadically and has an unknown etiology.

The apparently beef-induced cases of the disease found in 1996 seem to reflect a new variant of the disease, since they affect people much younger than those who usually get Crentzfeldt-Jakob disease.

AIDS

Between 1980 and 1995, more than 300,000 people in the United States died of AIDS, a previously unknown disease that is now the leading killer of people in that country aged twenty-five to forty-four years. Most scientists now believe that AIDS is caused by the human immunodeficiency virus (HIV). The history of medical understanding of AIDS can be divided into several periods:

1980–1984 Characterization of AIDS as a disease and the discovery of HIV.
1984–1994 Complications in and challenges to the HIV theory of AIDS.
1995–1997 Deeper understanding of the mechanisms by which HIV produces AIDS, with effective treatments using protease inhibitors.

Characterizing and explaining AIDS has been a complex and difficult project that has required the expenditure of enormous financial and human resources.

In 1980, a strange new disease was identified in gay men in Los Angeles, involving symptoms such as fever, weight loss, swollen lymph nodes, diarrhea, and thrush (Grmek 1990). Around the same time, medical personnel in New York City and San Francisco were noticing unusual occurrences of a rare cancer, Kaposi's sarcoma, in gay patients. By the end of 1981, many more cases had been identified of what was initially called GRID, for gay-related immune deficiency. Within a year, however, the disease had also been found in people other than gays, including heroin addicts and hemophiliacs, and was renamed AIDS (acquired immunodeficiency syndrome) in 1982.

The incidence of the disease strongly suggested that AIDS was caused by a bloodborne infectious agent, possibly a virus such as cytomegalovirus, which was often found in patients with Kaposi's sarcoma. Analogies with animal diseases (e.g., feline leukemia) and human infectious agents (e.g., hepatitis B virus) suggested the hypothesis that a virus was responsible for AIDS (Grmek 1990). Robert Gallo, an U.S. researcher who had discovered the first human retrovirus—HTLV-I—in 1980, suspected that the cause of AIDS was a close relative of HTLV-I (Gallo 1991, Gallo and Montagnier 1989). In 1983, separate teams of researchers led by Gallo and by Luc Montagnier in Paris identified a new retrovirus that was proclaimed as the cause of AIDS. Gallo and Montagnier (1989, p. 4) summarized the evidence:

> That HIV is the cause of AIDS is by now firmly established. The evidence for causation includes the fact that HIV is a new pathogen, fulfilling the original

> postulate of "new disease, new agent." In addition, although the original tests found evidence of HIV infection in only a fraction of people with AIDS, newer and more sensitive methods make it possible to find such evidence in almost every individual with AIDS or pre-AIDS. Studies of blood-transfusion recipients indicate that people exposed to HIV who have no other risk factors develop AIDS. The epidemiological evidence shows that in every country studied so far AIDS has appeared only after HIV. What is more, HIV infects and kills the very T4 cells that are depleted in AIDS.

The last point concerns the mechanism by which HIV causes AIDS: HIV kills T4 cells, which are crucial to immune system operation, and thereby weakens the immune system to such an extent that infections and cancers can occur.

The discovery of HIV, however, was slow to lead to an effective treatment for AIDS, and some anomalies about the development of AIDS emerged. Researchers observed a wide variation in disease progression, with some people showing symptoms of AIDS within two or three years of infection but others showing no symptoms even after twelve years. Some scientists, such as the Berkeley virologist Peter Duesberg (1988), argued that the evidence that HIV causes AIDS was inadequate, and instead explained AIDS as the result of multiple factors such as drug use and AZT (azidothymidine). Root-Bernstein (1993) contended that AIDS is a "multifactorial, synergistic disease" that arises when the immune system is overcome by combinations of drugs, multiple infections, and allogeneic insults such as semen. Critics of the hypothesis that HIV causes AIDS complained that it had been rendered tautologous, since AIDS was diagnosed only in patients found to be HIV positive. Montagnier and Gallo began to discuss cofactors that might be required before HIV infection produced AIDS. AIDS researchers were able to rebut Duesberg's arguments by pointing to hemophiliacs and medical personnel who had acquired HIV from blood alone (Cohen 1994), but uncertainty remained about the course and treatment of the disease.

In 1995 and 1996, new research dramatically altered understanding of the causes and treatment of AIDS. HIV-1, one of the two types of HIV, was found to have at least ten distinct genetic subtypes that might vary in transmissibility (Anderson et al. 1996). Moreover, some people have genetic mutations that enable immune system cells to resist the virus. Other research refuted the view that HIV became dormant after initial infection and in fact showed that HIV produced about 10^9 virions daily (Ho et al. 1995). The immune system manages to keep HIV production in check for a long time, until the virus produces variants that can overwhelm the immune cells. This new understanding of the dynamics of HIV development coincided with the availability of a new class of antiviral drugs—protease inhibitors, which render HIV incapable of infecting new cells (Bartlett 1996). The combination of protease inhibitors with other antiviral agents has shown dramatic effects in curtailing the amount of

virus in the blood and in reducing the onset and symptoms of AIDS. The effectiveness of these anti-HIV treatments strongly confirms the hypothesis that HIV causes AIDS. AIDS and peptic ulcers are similar in that both diseases involve host-strain interactions that make some people more susceptible than others, and both are best treated with combinations of drugs that overcome microbial resistance.

CHRONIC FATIGUE SYNDROME

There is now complete medical consensus concerning the causes of scurvy, and there is substantial scientific agreement concerning the causes of spongiform encephalopathies and AIDS. In sharp contrast, chronic fatigue syndrome is controversial not only with respect to its possible causes but even concerning whether it is a disease. Even the name is controversial: In the past, the disorder has been given such names as *chronic Epstein-Barr virus syndrome* and *post-viral fatigue syndrome*; in Great Britain and elsewhere, it is called *myalgic encephalomyelitis*; and some researchers and patients prefer the name *chronic fatigue/immune dysfunction syndrome* (CFIDS). Even more than the other cases I have discussed, chronic fatigue syndrome illustrates the vicissitudes of causal reasoning in medicine.

The term *chronic fatigue syndrome* was introduced in 1988 in response to reports of widespread illnesses that emerged in the United States in the mid-1980s, but retrospectively there appear to have been previous outbreaks, for example, in London in 1955 (Bell 1995, Johnson 1996, Straus 1994). Wessely (1994) argues that chronic fatigue syndrome is identical to the common nineteenth-century condition of neurasthenia. Typical symptoms of chronic fatigue syndrome include severe disabling fatigue, headache, malaise, short-term memory loss, muscle pain, trouble concentrating, joint pain, depression, abdominal pain, and many others. This multiplicity of symptoms causes great problems in characterizing and diagnosing the disorder. Bell (1995, pp. 17f) draws an analogy with AIDS:

> The parallels in the history of the recognition of AIDS as a specific disease and the recognition of CFIDS are remarkable. For years physician and health care administrators said that no illness could explain fatigue, weight loss, lymph node cancer, unusual parasitic pneumonias, and the purple spots of Kaposi's sarcoma. Because patients with AIDS were dying, it was finally and somewhat reluctantly agreed that this constellation of unusual symptoms and events was not psychosomatic. And with the discovery of the HIV virus, a theory could be put forward that explained these findings.

No similar theory has emerged to provide a unified account of why people get chronic fatigue syndrome.

In contrast to HIV tests used in diagnosing AIDS, there are no directs tests for chronic fatigue syndrome, which is diagnosed only after alternative medical and psychiatric causes of chronic fatiguing illness have been excluded (Fukuda et al. 1994). If severe fatigue has lasted more than six months and if there is no evidence for alternative causes, then chronic fatigue syndrome is diagnosed if four or more of the following eight symptoms are present: impaired memory or concentration, sore throat, tender cervical or auxiliary lymph nodes, muscle pain, multijoint pain, new headaches, unrefreshing sleep, and postexertion malaise.

Since the 1980s outbreak, researchers have looked for a viral cause of chronic fatigue syndrome. An early proposal that a retrovirus similar to HTLV-II is responsible was not confirmed, and numerous hypotheses about the nature of chronic fatigue syndrome are still under debate:

1. Chronic fatigue syndrome is not a disease or even a syndrome (i.e., a recurring pattern of symptoms) but an ill-formed category that covers fatigue resulting from many other medical and psychiatric conditions, such as multiple sclerosis.

2. Chronic fatigue syndrome is a psychiatric illness primarily due to depression or neurosis.

3. Chronic fatigue syndrome is caused by an undiscovered virus that overactivates the immune system, producing excessive amounts of cytokines such as interferon that cause multiple symptoms (Bell 1995).

4. Chronic fatigue syndrome is an immune system disorder that can be triggered by many different infectious agents, including enteroviruses, the Epstein-Barr virus, and human herpesvirus-6 (Fekety 1994).

Defenders of the reality of chronic fatigue syndrome argue against the first hypothesis by pointing to the commonality of symptoms among people with chronic fatigue in geographically identifiable outbreaks such as that in Lake Tahoe in 1985. The second hypothesis is challenged by pointing out that depression is found in only sixty percent of chronic fatigue syndrome patients and is characterized by frustration at not being able to perform normal activities rather than by despair and apathy.

It is currently impossible to choose between the third and fourth hypotheses, neither of which has much evidential support. If chronic fatigue syndrome is indeed like AIDS, a novel virus will be identified that can produce the appropriate range of symptoms, and the third hypothesis will meet with rapid acceptance. On the other hand, acceptance of the fourth hypothesis will require substantial advances in knowledge concerning the mechanisms of infection and immune system reaction, displaying a common pathway from infection to fatigue. Psychiatric aspects such as depression and stress may well turn out to be cofactors influencing this pathway.

To someone seeking medical simplicity, chronic fatigue syndrome is a condition with too many names, too many symptoms, and too many possible causes. Perhaps it will fade into medical history, as neurasthenia did in the nineteenth century. A more medically satisfying outcome will require research breakthroughs concerning the causes and mechanisms of chronic fatigue syndrome.

COMPLEXITIES OF CAUSAL INFERENCE

The diseases whose history I have sketched illustrate the difficulties of determining the causes of disease, which can be framed in terms of the model of causal inference proposed in chapter 7. The inference that a factor is a cause of a disease is based on explanatory coherence: We can infer that the factor causes the disease if this hypothesis is part of the best explanation of the full range of evidence. Collecting data that the factor and the disease are positively correlated (i.e., that the probability of the disease given the factor is greater than the probability of disease without the factor) does not suffice to show that the factor causes the disease. The correlation in the data may be due to chance or bias in data collection, and we must be able to infer that a genuine correlation is the best explanation of the observed correlation. Even if the correlation is genuine, it may not indicate a causal relation, since various alternative causes may be responsible for the correlation. That the factor causes the disease must be a better explanation of the correlation between the factor and the diseases than the assertion that some other cause is responsible for both the factor and the disease. Confidence that the factor causes the disease is increased if there is a familiar mechanism that explains why or how the factor causes the disease (see Figure 7.2).

Disease Characterization

Before causal inference can get underway, there needs to be a disease to be explained. This is problematic, however, in cases such as AIDS and chronic fatigue syndrome, in which many different symptoms are involved. Historically, it has not been easy to demarcate symptoms, syndromes, and diseases. AIDS was initially identified as a syndrome but has been recognized as a disease since the causal factor HIV was identified as common to all cases. Chronic fatigue syndrome will likely remain just that—a syndrome—until the causes and mechanisms are better understood. Many other diseases, however, such as scurvy and kuru have sufficiently distinct symptomologies that they could be characterized as diseases long before their etiologies are understood. But for some diseases, indeterminacy of symptoms is an impediment to the

development of causal understanding. Additional impediments fall into three classes: identifying possible causes, experimentally demonstrating causality, and establishing mechanisms.

Cause Identification

Identifying possible causes of a disease can be difficult for several reasons. First, as in the history of scurvy, there can be too many possible causes to sort out. Sea voyages on which sailors contracted scurvy were as strongly associated with damp air as with bad diet, and even within the diet there were factors such as salty meat that were more salient that the absence of fresh fruits and vegetables. AIDS was found to correlate with many factors, including both sexual activity and drug use, making it difficult to determine which correlations were causally significant.

A second impediment to identifying causally relevant correlates of diseases can be background causal beliefs. In the first two centuries of the investigation of scurvy, there was no natural place for dietary deficiency in the humoral theory of disease or in the germ theory of disease that successively dominated medical thinking. Similarly, recognition that beriberi is a nutritional disease was impeded by attempts to find a microbial cause. The prion hypothesis was initially suspect because of the belief that infectious agents require DNA or RNA for replication. Convictions that chronic fatigue syndrome is a psychiatric disorder discourage the search for a responsible causal agent. Similarly, when the theory that peptic ulcers are caused by bacteria was first proposed in 1983, it was greeted with skepticism in part because of the belief that the stomach's acidity produces a sterile environment (see chapter 4). The delay of almost two hundred years from the observation of bacteria by Antonie van Leeuwenhoek to the development of the germ theory of disease was partly the result of the influence of the humoral theory.

A third difficulty in identifying possible causes of diseases is that many causes are not directly observable. Bacteria became observable only with the invention of the optical microscope, and viruses became observable only with the invention of the electron microscope (see chapter 5). Even with modern technology, bacteria and viruses are not always easy to identify, as is shown by discoveries in only the past few decades of new kinds of bacteria that are responsible for peptic ulcers, Lyme disease, and Legionnaire's disease, as well as the discoveries of medically important viruses such as HIV. Correlating a disease with a factor is obviously impeded by an inability to observe the factor. The difficulty is even greater when the cause is not a microbe but rather a complex biochemical process that involves interactions of genes, proteins, and environmental conditions (see chapter 2).

Experimentation

The three difficulties just described concern the search for possible causes to correlate with diseases. Finding that a factor is correlated with a disease obviously does not show it to be a cause of the disease, for the factor may be a result of the real cause of the disease or may be only accidentally related. The best way to show that correlation indicates causation is to conduct controlled experiments that rule out other causal factors. In medicine, however, such experiments are not always possible. Sailors could not go on voyages without damp air, and researchers could not ethically inject patients with HIV to see if they develop AIDS. Hence the first difficulty in experimentally demonstrating the causes of diseases is that fully controlled experiments are often not practicable or ethical.

The second difficulty in experimentally showing the causes of diseases is that animal models may be unavailable or misleading. Animal models often provide a means of conducting controlled experiments, but they are not always available or accurate. Demonstration that scurvy is caused by nutritional deficiency benefited from experiments with guinea pigs, but experiments with rats were misleading. Gajdusek's experiments with chimpanzees were crucial in establishing that kuru and Creutzfeldt-Jakob disease are caused by a transmissible agent, but there appear to be differences between the prions involved in diseases in different animals. These differences make some inferences problematic, for example, concerning how likely it is that bovine spongiform encephalopathy will spread to humans. Animals models for HIV infection have been difficult to establish and interpret, because the virus behaves differently in other animals than it behaves in humans. Chronic fatigue syndrome is not even close to having any kind of animal model.

The third difficulty in experimentally showing disease causality arises from the complexity of many disease processes. Some diseases, such as spongiform encephalopathies and AIDS, may take years to develop, making it difficult to determine the effects of different kinds of experimental manipulations. Moreover, many diseases are multifactorial, with many contributing causes. Infectious diseases not only are the result of the invasion of the host by a microbe but also may depend on various features of the host, such as immune system status, and on interactions between the strains of the microbe and the host. When a disease has interacting causes, it can be difficult to isolate experimentally a particular factor as a major cause. Some sailors on long voyages did not get scurvy; only a few British beef eaters have so far developed the new variant of Creutzfeldt-Jakob disease; and exposure to HIV does not always produce infection and AIDS. There are some diseases (e.g., genetic conditions

such as Huntington's disease) for which we can unambiguously establish a unitary cause-disease relation, but most human diseases involve more complicated processes that have multiple causes.

Mechanism Elaboration

In addition to the three difficulties of identifying possible causes and the three difficulties of experimentally demonstrating causality, there is a remaining difficulty concerning the description of mechanisms. Our confidence that a factor really is a major cause of a disease is greatly increased if we can describe in detail the biochemical process by which the cause produces the disease and its symptoms. By and large, such understanding has become possible only in the last few decades through the rapid developments in molecular biology. We can say that vitamin C deficiency causes bleeding gums and other problems because it is needed for collagen metabolism. Prion researchers are increasingly understanding how defects in proteins can lead to brain disorders. The molecular genetics of HIV are sufficiently well understood that effective antiviral drugs such as protease inhibitors have been produced. Unfortunately, there are many conditions and diseases, ranging from chronic fatigue syndrome to atherosclerosis to arthritis, for which the causal mechanisms are poorly understood.

The difficulties of determining causes that occur at the different stages of disease understanding are summarized in Table 8.1. It is impressive that, despite these difficulties in determining the causes of diseases, modern medicine has made remarkable progress. It took more than three hundred years to identify vitamin C deficiency as the crucial factor in scurvy, and sixty years to identify prions as the cause of Creutzfeldt-Jakob disease. Strikingly, the period from the characterization of AIDS to the identification of HIV as the plausible cause of AIDS was only three years. Progress on chronic fatigue syndrome has not been so impressive, but serious investigation using the full resources of epidemiology and molecular medicine has been underway only since 1988. The scientific sophistication of medical research has expanded dramatically since the mid-nineteenth century, with improved theories, technologies, and experimental methodologies. Randomized controlled studies became the accepted norm for medical research only in the second half of the twentieth century (see chapter 12).

Initially, it seems amazing that the cause of peptic ulcers was discovered only in the 1980s, but the four diseases discussed in this chapter show that the path to uncovering disease causality is often difficult. The 1980s investigation of peptic ulcers did not have the difficulty found with AIDS and chronic fatigue syndrome of having a confusing complex of symptoms, but all the other

TABLE 8.1
Difficulties of Discovering Causes in Four Stages of Disease Understanding

Disease characterization
1. A condition may have diverse symptoms and be hard to recognize as a disease.

Cause specification
2. A disease may be correlated with many possible causes.
3. Background theories may impede the recognition of plausible causes.
4. Causes of diseases may be nonobservable.

Experimentation
5. Controlled experiments in humans may be impracticable or unethical.
6. Animal models may be unavailable or misleading.
7. Multifactorial diseases involve complex interactions.

Mechanism elaboration
8. Causal mechanisms may be difficult to discover and describe.

seven difficulties listed in Table 8.1 apply. And disease characterization *is* a problem for dyspepsia, which is sometimes caused by *H. pylori*; (see chapter 12).

Discovering and establishing causal factors for diseases is a complex cognitive task that requires great ingenuity in identifying possible causes and in performing controlled experiments to rule out alternative causes. My account in this chapter of scurvy, spongiform encephalopathies, AIDS, and chronic fatigue syndrome has been much briefer than my description of the bacterial theory of ulcers, and it has ignored physical and social processes in favor of cognitive ones. But it has served to display further the complexities of causal reasoning in medicine.

SUMMARY

Understanding a disease requires characterizing its symptoms, specifying possible causes, determining actual causes experimentally, and elaborating the mechanism by which the cause produces the disease. It took four hundred years for the causes of scurvy to be understood, because of the multiplicity of possible causes and the interference of the humoral and germ theories of disease. An understanding of the spongiform encephalopathies was delayed for decades by difficulties in identifying the highly unusual infectious agent that is responsible for the class of diseases. The cause of AIDS was found within a few years of the characterization of the disease but remained controversial until more was learned about the behavior and variants of the

HIV virus. Chronic fatigue syndrome is a difficult subject for scientific investigation, because its symptoms are variable and no causal agent has been identified. Determining the causes of diseases is a complex process that can be hindered by serious impediments to discovery and experimentation (see Table 8.1).

CHAPTER 9

Medical Analogies

CAUSAL REASONING based on explanatory coherence is a major part of medical thinking, but other cognitive processes are also important to understanding the development and application of medical knowledge. One such process is analogy, in which a previously solved problem serves as a source for solving a new target problem. I have already briefly mentioned some analogies that have been important in medical cases, such as the analogies between disease and fermentation (see chapter 2), between ulcers and other infectious diseases (see chapter 3), and between kuru and scrapie (see chapter 7). This chapter describes in more detail the purposes served by medical analogies (i.e., why they are used) and the different cognitive processes that support those purposes (i.e., how they are used). Historical and contemporary examples illustrate the theoretical, experimental, diagnostic, therapeutic, technological, and educational value of medical analogies. Four models of analogical transfer illuminate how analogies are used in these cases.

MODELS OF ANALOGICAL TRANSFER

The widespread use of analogies in cognition, including scientific reasoning, has been well documented (e.g., Biela 1991; Gentner et al. 1997; Holyoak and Thagard 1995; Leatherdale 1974). Analogical transfer, in which people use a source problem to provide a solution to a target problem, can take place in at least four different ways. The model of analogical transfer most commonly discussed in cognitive science works as follows. First, someone attempts to solve a target problem and then remembers or is given a similar source problem for which a solution is known. The target problem is then solved by adapting the solution to the source problem to provide a solution to the target. Many psychological experiments have followed this pattern (e.g., Gick and Holyoak 1980). And many computational models of analogical problem solving, including most work on case-based reasoning, also fit this pattern (e.g., Kolodner 1993). Accordingly, I use the term *standard model* for this pattern of retrieving a source to solve a target problem.

There are, however, other ways in which people use analogies to solve problems. In the standard model, the target problem serves as a direct retrieval cue for the source problem, but retrieval can also take place more indirectly using

a schema that is abstracted from the target problem. According to the *schema model*, an attempt to solve a target problem produces an abstract schema that then serves as a powerful retrieval cue for finding a source problem that provides a solution to the target problem. Although the abstraction may directly suggest a solution to the target problem, it may less directly suggest a solution by producing recall of a particular case that is sufficiently similar to the target to serve as the source of a solution. Darden (1983) discusses analogies in terms of shared abstractions.

In both the standard and schema models, the thinker starts with a target problem and retrieves a source, but there are important cases in which the act of reminding works in the opposite direction. These cases are ones in which an attempt to solve a target problem has failed, and the problem solver leaves it aside. Later, however, the problem solver serendipitously encounters a solved problem that can serve as a potential source, and this new source prompts recall of the unsolved target problem. Instead of the target providing a retrieval cue for the source, the source provides a retrieval cue for the target. The *serendipity model* refers to a pattern of analogical transfer in which a target problem is recalled and solved using a source accidentally encountered after initial solutions fail (cf. Langley and Jones 1988). Darwin's discovery of the theory of evolution by natural selection fits well the serendipity model: Darwin had long wondered about how biological evolution occurs, he found a solution only when he read Malthus and realized that Malthus's ideas about human population growth could be adapted to provide an explanation of species evolution in terms of the struggle for existence.

In all three models so far described, the source problem exists independently of the target problem. But there are rich analogies in which the source problem is constructed to provide a solution to the target problem. Nersessian (1992) describes how Maxwell generated a theoretical explanation of electromagnetism by constructing a mechanical analog. He did not understand electromagnetism in terms of any known mechanical system but instead concocted a new mechanical system that suggested the equations that he was then able to apply to electromagnetism. I use the term *generation model* for analogical transfer that takes place when a target problem is solved by analogy with a specially constructed source problem. The process of generation of a source analogy is roughly this:

1. Start with a target problem.
2. Retrieve or encounter a very approximate analog.
3. Fill out the approximate analog by looking at the target and identifying aspects of the constructed analog that need identification.
4. Transfer from the newly constructed source to the target.

The standard, schema, serendipity, and generation models are complementary accounts of analogical transfer rather than competitors (Figure 9.1). Different episodes of human analogical problem solving employ all four of the reasoning strategies that the models describe. In particular, there are important medical analogies that instantiate each of these models.

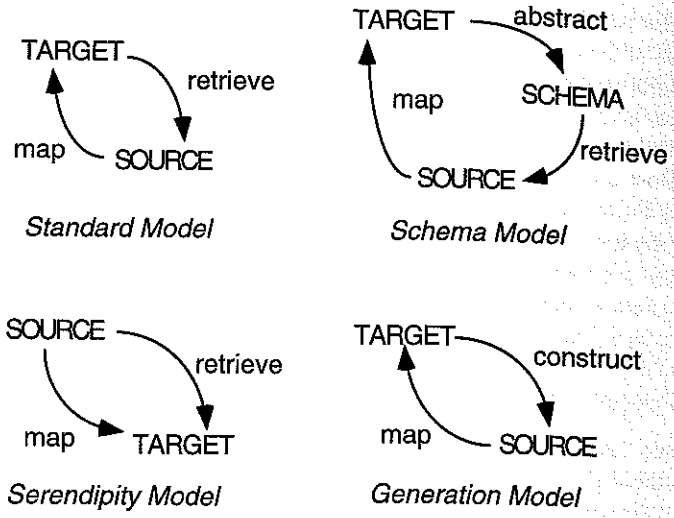

Figure 9.1. Models of analogical transfer.

THEORETICAL ANALOGIES

Theoretical analogies are those that are important in the development and justification of explanatory hypotheses. Important theoretical analogies in physics include the comparison of sound with water waves and of light waves with sound waves. Biology has also employed analogies that have contributed to theoretical development, such as Darwin's analogy between natural and artificial selection. Theoretical analogies have been equally important in the history of medicine, from the Hippocratics to the development of the germ theory of disease and beyond. The ancient Greeks explained health in terms of a balance of the various qualities that constituted the body, using a term for balance, *isonomia*, that also connoted equality of political rights (Temkin 1977, p. 272). The great seventeenth-century physician Thomas Sydenham conceived of diseases as akin to biological species, maintaining that just as characteristics of a plant species are extended to every individual so the characteristics of a disease apply to every individual who has it (Bynum 1993, p. 341).

In 1847, a physician in Vienna, Ignaz Semmelweiss, used a serendipitous analogy to form a hypothesis concerning the cause of childbed fever that was common among women who had been examined by medical students (Sinclair

1909). His colleague Kolletschka cut his finger during an autopsy and became very sick with the same symptoms as women with childbed fever. Semmelweiss hypothesized that Kolletschka had become ill because of contamination from a cadaver, and he proposed analogously that women were being made ill by medical students who had been performing autopsies.

The most important theoretical analogy in the history of medicine was used by Louis Pasteur and Joseph Lister in the development of the germ theory of disease. In the 1850s and 1860s, they realized that just as fermentation is caused by yeast and bacteria, so diseases may also be caused by microorganisms. Pasteur's ideas about infection moved from using microorganisms to explain why milk, beer, and wine ferment to proposing similar explanations of diseases in silkworms to explaining human diseases such as rabies in terms of germs. Pasteur wrote concerning his work on fermentation:

> What meditations are induced by those results! It is impossible not to observe that, the further we penetrate into the experimental study of germs, the more we perceive sudden lights and clear ideas on the knowledge of the causes of contagious diseases! Is it not worthy of attention that, in that Arbois vineyard (and it would be true of the million *hectares* of vineyards of all the countries in the world), there should not have been, at the time I made the aforesaid experiments, one single particle of earth which would not have been capable of provoking fermentation by a grape yeast, and that, on the other hand, the earth of the glass houses I have mentioned should have been powerless to fulfill that office? And why? Because, at the given moment, I covered that earth with some glass. The death, if I may so express it, of a bunch of grapes, thrown at that time on any vineyard, would infallibly have occurred through the *saccharomyces* parasites of which I speak; that kind of death would have been impossible, on the contrary, on the little space enclosed by my glass houses. Those few cubic yards of air, those few square yards of soil, were there, in the midst of a universal possible contagion, and they were safe from it. . . . Is it not permissible to believe, by analogy, that a day will come when easily applied preventive measures will arrest those scourges which suddenly desolate and terrify populations; such as the fearful disease (yellow fever) which has recently invaded Senegal and the valley of the Mississippi, or that other (bubonic plague), yet more terrible perhaps, which has ravaged the banks of the Volga? (translated in Vallery-Radot 1926, pp. 287–288; for the original, see Pasteur 1922, Vol. 2, p. 547).

Pasteur's theoretical analogy had the following structure :

Fermentation is caused by germs.
Disease is like fermentation.
Disease may therefore also be caused by germs.

As far as one can tell from the historical record, the development of Pasteur's ideas appears to fit with the standard model of analogical transfer. In working on silkworms, he was able to draw on his previous work on fermentation, and in working on human diseases, he drew on the ideas and techniques that had been useful with silkworms. The previously understood problems of fermentation and silkworm diseases provided sources for analogical solution of the subsequent target problem of human disease.

A similar theoretical analogy was also important in the development of modern surgery. Before the 1860s, many surgical patients suffered serious infections, which were not explained until the British surgeon Lister realized the significance of Pasteur's ideas about fermentation and recognized that germs in the air can cause infection of wounds, just as they cause fermentation. He wrote in 1867:

> Turning now to the question how the atmosphere produces decomposition of organic substances, we find that a flood of light has been thrown upon this most important subject by the philosophic researches of M. Pasteur, who has demonstrated by thoroughly convincing evidence that it is not to its oxygen or to any of its gaseous constituents that the air owes this property, but to the minute particles suspended in it, which are the germs of various low forms of life, long since revealed by the microscope, and regarded as merely accidental concomitants to putrescence, but now shown by Pasteur to be its essential cause, resolving the complex organic compounds into substances of simpler chemical constitution, just as the yeast plant converts sugar into alcohol and carbonic acid. . . . Applying these principles to the treatment of compound fracture, bearing in mind that it is from the vitality of the atmospheric particles that all mischief arises, it appears that all that is requisite is to dress the wound with some material capable of killing those septic germs, provided that any substance can be found reliable for this purpose, yet not too potent as a caustic. (reprinted in Brock 1961, p. 84)

The structure of Lister's reasoning was as follows:

Fermentation is caused by germs.
Putrefaction (infection) following surgery is like fermentation.
Putrefaction may therefore be caused by germs.

This analogical transfer does not fit the standard model, since Lister must have worried about the problem of wound infection for many years before reading Pasteur's work on fermentation, which reminded him of the preexisting wound target problem. In this case, the source problem (fermentation) prompted retrieval of the target problem (infection), so it best fits the serendipity model of analogical transfer.

The analogy between fermentation and infection was a remote one, since on the face of it there is little apparent similarity between grapes becoming alcoholic and wounds becoming infected. Closer analogies are ubiquitous in medical research, which relies heavily on the use of animal models to determine the causes of disease. For example, Robert Koch determined that tuberculosis is caused by a bacterium by performing experiments on guinea pigs. He showed that injecting guinea pigs with bacteria taken from other guinea pigs with

tuberculosis induced tuberculosis in them. Obviously, it would be unethical to induce tuberculosis in humans in this way and therefore impossible to do a controlled experiment of tuberculosis in humans. Koch used animals to *generate* an analog to human disease (Brock 1988). This is not a case of analogical transfer by reminding or serendipity but rather of constructing an animal analog that can then be used to make inferences about human diseases. The structure of the analogical transfer in these cases is roughly as follows:

We want to know the causes of a disease in humans.
Animals (e.g., guinea pigs) have the same (or similar) disease.
In animals, the disease is caused by X.
The human disease may therefore also be caused by something like X.

The constructive nature of animal analogies is even more evident when new animal strains are created to provide models for human diseases. Biologists, for example, have used genetic engineering to create a strain of mouse that develops Alzheimer's disease. Because the mouse develops the types of plaques on the brain that are found in humans with Alzheimer's and also suffers memory problems, it can be used in experiments that are aimed at determining the causes of and possible treatments for Alzheimer's disease. Analogies based on animal models are also important for therapeutic purposes (see later). Sometimes, animal models are arrived at serendipitously, as when researchers who set out to genetically engineer rats as a model of human arthritis discovered that they had created a model of ulcerative colitis. As I mentioned in chapter 3, early attempts to use pigs as an animal model for *H. pylori* infection failed, but later efforts with bacteria-free piglets were more successful. Animal experimentation has nevertheless played a very minor role in development of the bacterial theory of ulcers.

Animal analogies were important in the development of ideas about nutritional deficiency diseases. Understanding of beriberi was greatly advanced when a similar disease was found in chickens that had been fed polished rice, and chapter 8 described how guinea pigs served as a valuable animal model for human scurvy. Funk (1912), having isolated what he thought was the vitamin needed to prevent beriberi, analogically suggested methods for isolating the vitamin that he conjectured was similarly responsible for scurvy. Moreover, on the basis of similarities with beriberi and scurvy, he correctly speculated that pellagra and rickets are also deficiency diseases.

Critics of animal experimentation have raised doubts about the ability of such models to provide explanations of human diseases (LaFollette and Shanks 1995). Animal models often break down because of physiological differences between humans and the animals used, which also lead to differences in causality and treatment effectiveness. Treatments that are effective in animals or in the test tube often do not work on humans. Analogical reasoning is frequently a risky kind of inference, but Holyoak and Thagard (1995) describe various steps that can be taken to improve the quality of analogical reasoning. We urged analogists to use *system mappings*, ones based on deep similarities of causal relations rather than on superficial similarities. When animal experimentation uses animals whose physical processes are known to be similar to those of humans, there can be a system mapping based on the existence of similar causal mechanisms. We also urged analogists to use multiple analogies, that is, to consider the relevance of various possible source analogs for the case at hand. Well-informed medical researchers look at various possible animal models for a human disease and base their experimental conclusions on deep causal similarities between the animals and humans. Under these conditions, animal models provide generated theoretical analogies that are at least suggestive about the causes of diseases in humans.

A related issue is the value of animal models of human thinking. On the one hand, the assumption of behaviorist psychologists that the rat could serve as a full model of human learning grossly underestimated the cognitive differences between humans and rats. On the other hand, neuroscientific research has found important similarities in humans and other animals with respect to visual and emotional systems.

Medical thinking about some human diseases has also been aided by analogies with similar diseases. Researchers on tuberculosis made comparisons with similar infectious diseases such as smallpox and syphilis, and researchers on yellow fever made comparisons with malaria. These analogies are relatively close ones that generally fit the standard model of analogical transfer, as with Robert Gallo's attempt to find a virus that causes AIDS that is analogous to those viruses with which he was already familiar (see chapter 8). In turn, AIDS has served as a suggestive analogy for some investigators of chronic fatigue syndrome. Recent speculations that atherosclerotic coronary heart disease might be caused by an infection of *Chlamydia pneumoniae* are defended by comparison with the discovery of a bacterial cause for another inflammatory/degenerative disease, peptic ulcers (Muhlestein 1997).

As I mentioned in chapter 8, analogies contributed to the development of ideas about kuru. Hadlow (1959) noticed similarities between the sheep disease scrapie and the New Guinea disease kuru and suggested experiments to determine if the latter was also transmissible. Research on these brain diseases led to Stanley Prusiner's (1982) hypothesis concerning a novel infectious agent called prions, which he analogically suggested might also be responsible for other diseases, such as Alzheimer's. According to Rhodes (1997, p. 101), two of the anthropologists who first made the connection between kuru and cannibalism did so because of a strange analogy. Shirley Lindenbaum and Robert Glasse left Australia for New Guinea in 1961 and in 1962 read a *Time* magazine story that a scientist had trained flatworms to find their way through a maze, chopped them up, and fed them to other flatworms that then got through the maze. This result (which has since been discredited) provided the

anthropologists with a rough model to suggest that the brain problems of the Fore people might be caused by their cannibalism.

I have included in this section only analogies that are important in the development and justification of explanatory hypotheses. Explanatory analogies whose function is primarily expository are discussed in the later section on educational analogies. Some philosophers and scientists are skeptical that analogy can play any role in justifying hypotheses; see Thagard (1992b) for a defense of the relevance of analogy to justification as well as discovery. Analogy is a contributor to explanatory coherence, and analogical mapping and retrieval can be understood computationally as coherence problems (Holyoak and Thagard 1995; Thagard and Verbeurgt 1998).

EXPERIMENTAL ANALOGIES

To establish a medical hypothesis, controlled experiments are needed to distinguish causation from mere correlations. Epidemiologists have established numerous standards for designing experiments that address the causes of diseases. Because of the complexity of experimentation, however, it is unlikely that medical researchers design their experiments from scratch. Experiments can be designed via an application of the standard model of analogical transfer, when a researcher remembers a previous experiment that suggests how to do the desired new experiment. Dunbar (1995, 1997) describes the frequent use of analogies in the design of experiments in molecular biology, and Kettler and Darden (1993) describe a program that uses analogy to help design protein sequencing experiments.

Experimental analogies have the following structure:

We need to do an experiment to accomplish X.
A previous experiment accomplished Y, a task similar to X.
We can therefore do a modification of the previous experiment.

It is also possible that analogical transfer in experimental design could fit the serendipity model. A researcher might wonder how to design an experiment to test a hypothesis and then encounter a paper describing an experimental procedure that tests a similar hypothesis. The researcher could then design a similar experiment.

DIAGNOSTIC ANALOGIES

Medical research aims at discovering the causes of diseases, but the reasoning task facing most physicians consists of diagnosing the presence of disease in individual patients. The physician needs to decide what disease or complex of

diseases provides the best explanation of the patient's symptoms. This task often does not involve analogy. In straightforward cases, it can be almost deductive: If the patient has symptoms S1, S2, and S3, then it is almost certain that the patient has the disease D. In more complex cases, the reasoning is abductive, with the physician having to select, from a variety of diseases that would explain the patient's symptoms, a diagnosis that fits best with what is known.

Sometimes, however, a diagnosis problem does not admit a simple deductive or abductive solution, and analogies may then be useful. The general structure of diagnostic analogies is as follows:

The patient P has the unusual set of symptoms, S1, S2, and S3.
Another patient with similar symptoms had a disease D.
The patient P may therefore also have the disease D.

Koton (1988) describes a case-based–reasoning program that produces causal explanations of a heart patient's symptoms by retrieving examples of similar patients.

THERAPEUTIC ANALOGIES

In addition to performing the task of diagnosis, medical reasoners want to be able to treat patients in ways that cure their diseases or at least reduce their symptoms. Berlinger (1996) describes a dramatic case of a baby born with a cystic cygroma that made it very difficult to breathe. When the baby stopped breathing, it became crucial to insert a tube in the baby's airway, but a cluster of yellow cysts hid the airway so that it was not clear where to insert the tube. Berlinger fortunately remembered a previous case in which an emergency technician had inserted a breathing tube to save the life of a snowmobiler with a severed windpipe by sticking the tube where bloody bubbles indicated the airway. Analogously, Berlinger pushed down on the baby's chest to push air out through the cysts, generating saliva bubbles that he could use as a guide for insertion of the breathing tube. This therapeutic analogy fits the standard model of analogical transfer, with the physician retrieving a source problem (the snowmobiler's inability to breathe) to solve the target problem (the baby's inability to breathe). There are undoubtedly more prosaic cases in which physicians prescribe treatments because they worked previously with the same or similar patients.

Therapeutic analogies can also be based on similarities between diseases. Greenberg and Root (1995) describe a case in which a physician was unable to diagnose a particular disease or diseases in a patient with a complex set of symptoms. However, because the patient's symptoms were similar to those of patients with identified diseases who had been successfully treated, the

physician recommended a similar treatment. This case fits the standard model of analogical transfer.

At a more general level, therapeutic analogies can be drawn from animal models used in experiments to determine the effectiveness of treatments for diseases. The general structure of these analogies is as follows:

We want to know the medical effects of a treatment in humans.
Animals (e.g., guinea pigs) are similar to humans.
We can therefore try the treatment first in animals.
We can then transfer the conclusions (positive or negative) back to humans.

As with the animal model analogies described in the previous section on theoretical analogies, these analogies fit the generative model of transfer, and the value of the animal therapeutic analogies depends on the relational similarity of the relevant causal processes in animals and humans.

Finally, here is an analogy used to suggest early and aggressive treatment of HIV infections (Ho et al. 1995, p. 126):

> The CD4 lymphocyte depletion seen in advanced HIV-1 infection may be likened to a sink containing a low water level, with the tap and drain both equally wide open. As the regenerative capacity of the immune system is not infinite, it is not difficult to see why the sink eventually empties. It is also evident from this analogy that our primary strategy to reverse the immunodeficiency ought to be to target virally mediated destruction (plug the drain) rather than to emphasize lymphocyte reconstitution (put in a second tap).

TECHNOLOGICAL ANALOGIES

Medicine requires many technologies for the diagnosis, treatment, and prevention of disease. A technological analogy is one in which transfer produces a new medical tool. I discuss three examples: Lister's treatment of wounds, the invention of the stethoscope, and the invention of the polymerase chain reaction.

Lister's analogy between fermentation and putrefaction suggested a means of preventing infection. He recalled that carbolic acid had been used in Carlisle on sewage to prevent odor and diseases in cattle that fed on the pastures irrigated from the refuse material; he accordingly began to use carbolic acid to sterilize wounds, which dramatically decreased the infection rate. This analogical transfer fits the standard model. Having inferred from Pasteur's work that germs from the air might cause putrefaction, he generated a new solution to the target problem of how to prevent germs from infecting wounds. This problem reminded him of the use of carbolic acid in Carlisle, which he then applied successfully (if not pleasantly) to surgery.

Earlier in the nineteenth century, a French physician had used analogy in the invention of the most widely used piece of medical technology, the stethoscope. There are two different historical accounts of this discovery, alternatively fitting the schema and serendipity models of analogical transfer. Here is Théophile Laennec's (1962, pp. 284–285) own description in 1819 of how he invented the stethoscope:

> In 1816, I was consulted by a young woman labouring under general symptoms of diseased heart, and in whose case percussion and the application of the hand were of little avail on account of the great degree of fatness. The other method just mentioned [application of the ear to the chest] being rendered inadmissible by the age and sex of the patient, I happened to recollect a simple and well-known fact in acoustics, and fancied, at the same time, that it might be turned to some use on the present occasion. The fact I allude to is the augmented impression of sound when conveyed through certain solid bodies,—as when we hear the scratch of a pin at one end of a piece of wood, on applying our ear to the other. Immediately, on this suggestion, I rolled a quire of paper into a sort of cylinder and applied it to one end of the region of the heart and the other to my ear, and was not a little surprised and pleased, to find that I could thereby perceive the action of the heart in a manner much more clear and distinct than I had ever been able to do by the immediate application of the ear.

This account fits with the schema model of analogical transfer: Laennec solved the target problem of how to listen to the woman's heart by abstracting it into a general acoustic problem that reminded him of pin scratching a piece of wood. The wood then served as a source that suggested the use of a rolled-up piece of paper to listen to the woman's heart. In Laennec's account, a general acoustic fact provided the retrieval cue for finding a source problem that could be used to produce a solution to the target problem.

A different account has, however, found its way into the historical record, owing to Laennec's friend Lejumeau de Kergaradac:

> As the author told me himself, he owed to chance the great discovery that immortalized his name. We must say at once that these chances would only occur to a man of genius. One day while crossing the court of the Louvre, he noticed children with their ears held to two ends of long pieces of wood, transmitting the noise of small pin strokes on the opposite end. This everyday acoustic experience was a revelation for him. He conceived on the spot the thought of application to heart disease. The next day at his clinic at the Necker hospital, he rolled his appointment notebook, tied it compactly while keeping a central tube, then placed it on a diseased heart. This was the first stethoscope. (my translation of passage quoted by Grmek 1981, p. 113)

Whereas Laennec described himself as using acoustic principles to think of the wooden source analogy, his friend's account described Laennec as seren-

dipitously encountering children listening to a pin scratch wood. The children's game provided a fortuitous source analog that reminded him of his ongoing target problem of effectively listening to patients' chests. In accord with the serendipity model of analogical transfer, the encountered source provided a retrieval cue for the target problem rather than vice versa. The historical record is not adequate to establish which of these accounts is correct, although an authority leans toward Laennec's own story (Grmek 1981). Nevertheless, the two versions of the story are useful for distinguishing between the schema and serendipity models of analogical transfer, and Laennec's discovery under either description qualifies as a technological analogy of great medical importance.

In 1983, Kary Mullis, a biologist at Cetus Corporation in California, invented polymerase chain reaction (PCR), a technology that now has many applications in molecular medicine. PCR is a method in which an enzyme called a polymerase is used to act along a strand of DNA to produce unlimited quantities of selected genetic material for further investigation. The idea for PCR came to him on a drive to his cabin in Mendocino County. He had been looking for a general procedure to identify a single nucleotide at a given position in a DNA molecule. According to Rabinow (1996, p. 96) the discovery came about because Mullis had been experimenting with fractals and other computational procedures that involved iteration and exponential amplification:

> This was the breakthrough moment. His tinkering with fractals and other computer programs had habituated him to the idea of iterative processes. This looping, back and back again, as boring and time consuming as it might be on the level of physical practice, was nearly effortless on the computer. Mullis made the connection between the two realms and saw that the doubling process was a huge advantage because it was exponential.

This discovery appears to fit the standard model of analogical transfer. Wondering about how to solve the target problem of producing large quantities of genetic principle led Mullis to think of a kind of computational problem with which he was familiar. The iterative processes of fractals then provided a source problem that suggested a solution to the target problem. Thus, technological analogies exemplify the standard as well as the schema and serendipity models of analogical transfer.

EDUCATIONAL ANALOGIES

All the analogies I have discussed so far are highly creative ones in which new solutions were suggested for important theoretical, experimental, diagnostic, therapeutic, and technological problems. Much more common, however, are

more prosaic educational analogies that function to enable someone who understands something about the nature of disease to convey that information to someone else. Polemics in favor of the bacterial theory of ulcers drew comparisons with other infectious diseases such as smallpox, cystitis, and polio. Zamir (1996) explained why regular exercise is important for healthy hearts by using an extended financial analogy that compares coronary output to bank deposits. Strachan and Read (1996, p. 458) provided an analogy that helps distinguish the roles of different cancer-causing genes: "By analogy with a bus, one can picture the oncogenes as the accelerator and the TS [tumor suppressor] genes as the brake. Jamming the accelerator on (a dominant gain-of-function mutation in an oncogene) or having all the brakes fail (a recessive loss-of-function mutation in a TS gene) will make the bus run out of control." Medical researchers and practitioners can also use analogies to explain new ideas about disease causality to others. Analogies can also be used to give practical advice, as with the following anonymous comparison inspired by mad cow disease. Safe eating is like safe sex: You may be eating whatever it was that what you are eating ate before you ate it.

I have described how analogies are useful in medicine for theoretical, experimental, diagnostic, therapeutic, technological, and educational purposes. The processes of analogical reasoning are not, however, always the same, and different cases of medical analogizing fit different models of analogical transfer, although the standard model in which source analogs are remembered and applied to solve a target problem is probably the most common. The additional examination of historical cases and ongoing medical practice will undoubtedly provide more illustrations of different ways in which analogical transfer can contribute to medical thinking.

SUMMARY

Analogy is a cognitive process that is important in many kinds of creative thinking, so it is not surprising that it also contributes to the growth of medical knowledge. Analogical transfer often fits the standard model in which a source is remembered to solve a target problem. But sometimes it is the target problem that is retrieved, and sometimes the source problem is constructed rather than retrieved. Pasteur, Lister, and other medical researchers have used analogies in their theoretical advances. Analogies have also been useful for designing experiments, suggesting diagnoses, proposing therapeutic treatments, inventing new medical technologies, and enhancing education.