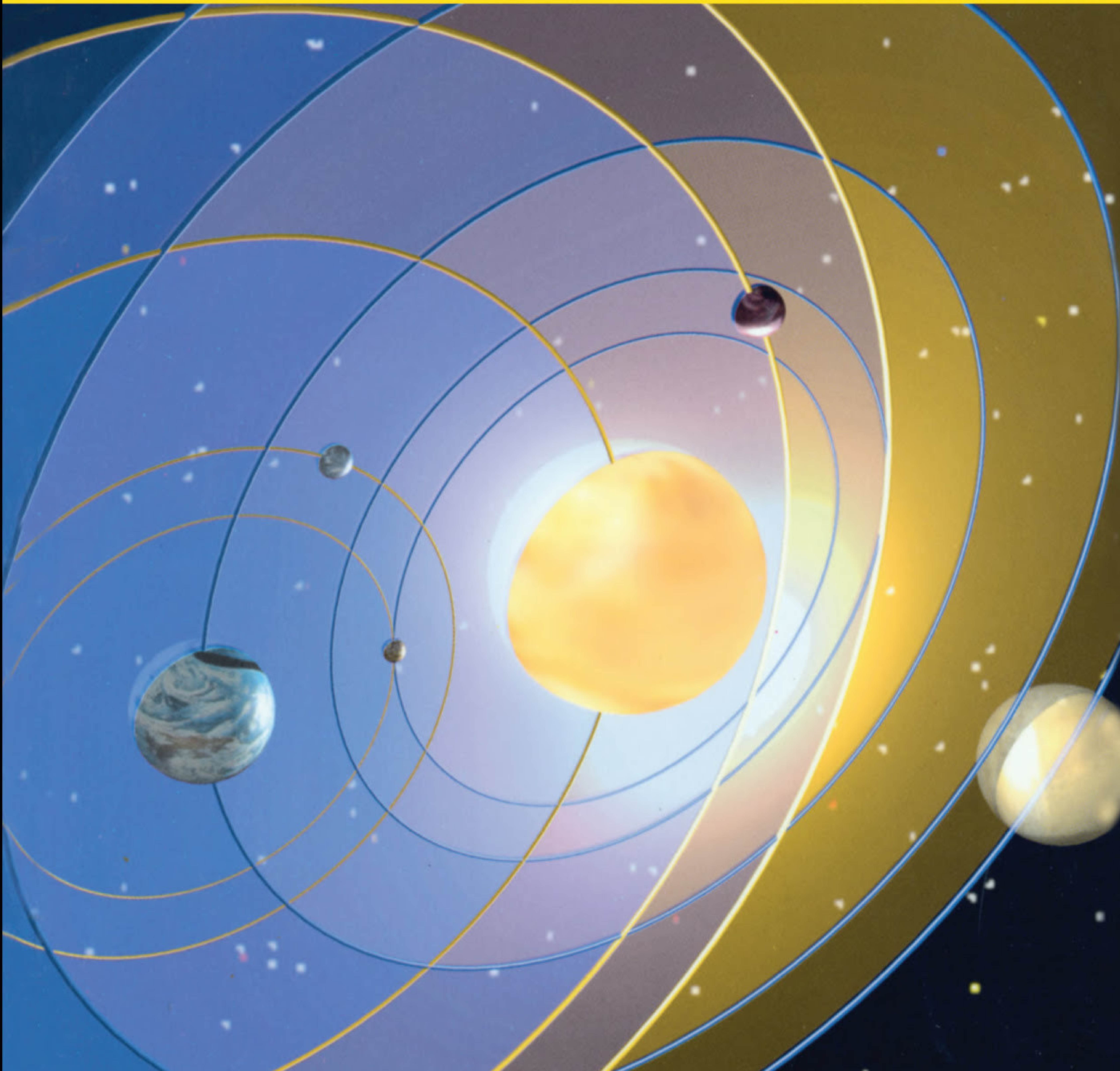
CONCEPTUAL REVOLUTIONS



PAUL THAGARD

CONCEPTUAL REVOLUTIONS

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, Chichester, West Sussex

Library of Congress Cataloging-in-Publication Data

Thagard, Paul.
Conceptual revolutions / Paul Thagard.
p. cm.
Includes bibliographical references (p.) and index.
ISBN 0-691-08745-8
ISBN 0-691-02490-1 (pbk.)
1. Science—Philosophy. 2. Science—History. 3. Concepts—History. 4. Change (Psychology) 5. Biology—Philosophy. 6. Cognition. I. Title.
Q175.T4793 1992
501—dc20 91-20210 CIP

First Princeton Paperback printing, 1993

This book has been composed in Adobe Times Roman

Princeton University Press books are printed on acid-free paper and meet the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources

Printed in the United States of America

5 7 9 10 8 6 4

Contents

Figures

Tables

Acknowledgments

I AM GRATEFUL to many people for their help in developing the ideas in this book. Here is only a partial list: William Bechtel, William Brewer, Susan Carey, Lindley Darden, Myrna Gopnik, Steve Hanson, Keith Holyoak, Gilbert Harman, Daniel Hausman, Philip Johnson-Laird, Trevor Levere, Daniel Kimberg, Pat Langley, Michael Mahoney, Peter Meyers, Robert McCauley, George Miller, Greg Nelson, Nancy Nersessian, Gregory Nowak, Paul O'Rorke, Michael Ranney, Brian Reiser, and Jeff Shrager.

For valuable comments on a previous draft, I am particularly grateful to William Brewer, Lindley Darden, Philip Johnson-Laird, Noretta Koertge, Rachel Laudan, Robert McCauley, Gregory Nowak, Michael Ranney, Robert Richardson, and two anonymous referees.

During the period in which this book was written, my research was supported by contracts from the Basic Research Office of the Army Research Institute for Behavioral and Social Sciences, and by a grant to the Princeton University Human Information Processing Group from the McDonnell Foundation. The Princeton University Cognitive Science Laboratory provided an excellent working environment.

I am grateful to my research assistant Gregory Nowak for his historical analyses of the geological, Copernican, and Newtonian revolutions. Chapter 7 is based on articles written with him, and Chapter 8 also benefited from his research. Greg Nelson wrote the graphics program used to display ECHO runs. Malcolm DeBevoise of Princeton University Press and Ron Twisdale provided valuable assistance.

To my wife, Ziva Kunda, I am indebted for the title, numerous suggestions, and much more.

This project was conceived largely as a whole, but as it developed various parts became articles, which have now been excerpted, revised, and integrated into the current text. I am grateful to the respective publishers for permission to include vestiges of previously published articles:

Nowak, G., and Thagard, P. (forthcoming). Copernicus, Ptolemy, and explanatory coherence. In R. Giere (Ed.), *Cognitive models of science, Minnesota Studies in the Philosophy of Science*, vol. 15. Minneapolis: University of Minnesota Press. (chs. 4, 8)

Thagard, P. (1989). Explanatory coherence, *Behavioral and Brain Sciences, 12*, 435–467. Cambridge University Press. (chs. 4, 6)

Thagard, P. (1989). Extending explanatory coherence (reply to commentators), *Behavioral and Brain Sciences, 12*, 490–502. Cambridge University Press. (ch. 4)

Thagard, P. (1989). Scientific cognition: Hot or cold? In S. Fuller, M. de Mey, T. Shinn, and S. Woolgar (Eds.), *The cognitive turn: sociological and psychological perspectives on science*. Dordrecht: Kluwer Academic Publishers, 71–82. (ch. 5)

Thagard, P. (1990). Concepts and conceptual change, *Synthese, 82*, 255–274. Kluwer Academic Publishers. (chs. 2, 3)

Thagard, P. (1990). The conceptual structure of the chemical revolution. *Philosophy of Science, 57*, 183–209. (chs. 1, 2, 3, 5)

Thagard, P. (1990). Modelling conceptual change. In J. Tiles, G. McKee, and G. Dean (Eds.), *Evolving knowledge in natural science and artificial intelligence*. London: Pitman Publishing, 201–218. (chs. 1, 3, 8)

Thagard, P. (1990). Philosophy and machine learning. *Canadian Journal of Philosophy, 20*, 261–276. (ch. 2)

Thagard, P. (forthcoming). Philosophical and computational models of explanation. *Philosophical Studies*. Kluwer Academic Publishers. (ch. 5)

Thagard, P., and Nowak, G. (1988). The explanatory coherence of continental drift. In A. Fine and J. Leplin (Eds.), *PSA 1988*, vol. 1. East Lansing, MI: Philosophy of Science Association, 118–126. (ch. 7)

Thagard, P., and Nowak, G. (1990). The conceptual structure of the geological revolution. In J. Shrager and P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. San Mateo: Morgan Kaufmann, 27–72. (ch. 8)

A final note: In July 1992, I become Professor of Philosophy at the University of Waterloo in Waterloo, Ontario, Canada.

CONCEPTUAL REVOLUTIONS

CHAPTER 1

The Problem of Revolutionary Conceptual Change

SCIENTIFIC knowledge often grows slowly with gradual additions of new laws and concepts. But sometimes science undergoes dramatic conceptual changes when whole systems of concepts and laws are replaced by new ones. By analogy with political upheavals, such changes are called *scientific revolutions*.

Although many historians and philosophers of science have stressed the importance of scientific revolutions, there has been little detailed explanation of such changes. How do conceptual revolutions occur? How can a new conceptual system arise and replace an old one? What are these conceptual systems whose transformation is so fundamental to scientific development? Are scientific revolutions rational?

I shall propose answers to these questions from a viewpoint that is psychological and computational. In an earlier book, I advocated the use of techniques derived from artificial intelligence (AI) to understand the structure and growth of scientific knowledge; I called the enterprise *computational philosophy of science* (Thagard 1988). Here I shall show the relevance of ideas from the cognitive sciences to the most dramatic phenomena in the history of science: scientific revolutions. The theory of revolutionary conceptual change developed is germane to central issues in cognitive psychology and artificial intelligence, as well as to disputes in the philosophy of science.

1.1 THE IMPORTANCE OF CONCEPTUAL CHANGE

In the philosophy and history of science, the question of revolutionary conceptual change became important with the 1962 publication of Kuhn's *Structure of Scientific Revolutions*. For several previous decades, philosophy of science had been dominated by the logical empiricist approach. Exemplified by such philosophers as Carnap (1950) and Hempel (1965), the logical empiricists used the techniques of modern formal logic to investigate how scientific knowledge could be tied to sense experience. Like the views of Popper (1959), the logical empiricists emphasized the logical structure of science rather than its psychological and historical development.

Kuhn, along with other historically inclined philosophers such as Hanson

(1958) and Feyerabend (1965), charged the logical empiricists with historical irrelevance. Kuhn (1970, 1) wrote: "History, if viewed as a repository for more than anecdote or chronology, could produce a decisive transformation in the image of science by which we are now possessed." Kuhn's general view of scientific revolutions and his accounts of particular scientific episodes must be questioned, but his basic claim that the development of scientific knowledge includes revolutionary changes can be sustained.

Kuhn's claims about scientific revolutions have caused great consternation among philosophers of science because of their apparent implication of irrationality. Kuhn so stressed the dramatic and noncontinuous nature of scientific change that transitions in scientific theories or "paradigms" took on the appearance of cataclysmic, nonrational events. According to Kuhn's early statements, later moderated, a scientific revolution involves a complete change in standards and methods, so that rational evaluation of competing views using external standards appears impossible. He even said that when one theory or "paradigm" replaces another, scientists work in a different world.

Although Kuhn's emphasis on revolutionary change was an antidote to the simplistic models of the logical empiricists, a finer-grained theory of revolutionary change than Kuhn presented need not succumb to irrationalism. To develop such a theory, however, we need tools different from both the formal ones of the logical empiricists and the vague historical ones of Kuhn. Artificial intelligence offers the possibility of developing such tools for describing the structure of scientific knowledge and the processes that advance it. We can begin to characterize the structure of conceptual systems before, during, and after conceptual revolutions; and we can investigate the cognitive mechanisms by which conceptual changes occur.

The importance of the problem of conceptual change is not restricted to the history and philosophy of science. Conceptual change is of general psychological interest, since people other than scientists also experience it. Children acquire much knowledge through observation and education, and developmental psychologists have recently been arguing that children's acquisition of knowledge is not simply a matter of accretion of new facts. Rather, it involves an important restructuring of their conceptual systems. Carey (1985) has suggested that children undergo a fundamental restructuring of their biological ideas between the ages of 4 and 10, and she explicitly compares this restructuring to scientific revolutions. Vosniadou and Brewer (1987, 1990) have similar speculations about children's learning of astronomy. McCloskey (1983) describes the difficulties of children and some adults in appreciating Newtonian physics. Chapter 10 shows that conceptual change in children can be understood within the same computational framework that sheds light on scientific revolutions, although scientific revolutions involve conceptual change that is more radical than what occurs in the ordinary cognitive development of children. Human learning and scientific discovery are continuous processes, but scientific revolutions are rare events that involve more dramatic changes than are experienced in everyday cognition.

My approach both to the history of science and to developmental psychology is computational. For many researchers in cognitive science, thinking can not only be modeled computationally like the weather and wind resistance: thinking *is* a form of computation. To model mental structures and processes, programs are designed with data structures corresponding to the postulated mental structures and with procedures corresponding to the postulated mental processes. Running the program and doing psychological experiments provides a way of judging whether the model corresponds to psychological reality.

Artificial intelligence offers numerous tools for constructing these kinds of cognitive models. The AI subfield of *knowledge representation* is concerned with techniques of representing information in a computer for intelligent processing. AI work on problem solving and planning is highly relevant to the problem solving activities of scientists. *Machine learning* is the AI subfield concerned with how computational systems can improve their performance by acquiring and modifying their structures and procedures. Ideas from knowledge representation and machine learning will figure prominently in the theoretical developments in later chapters.

We cannot, however, simply take over existing ideas from these subfields of AI. Although much work in machine learning has been done on topics such as concept formation, we shall see that available techniques are not adequate to account for the origins of many important scientific concepts. Moreover, AI researchers have concentrated on cases of learning by accretion of knowledge, rather than on cases of revolutionary replacement of complexes of concepts. New computational ideas are required to account for this kind of replacement. Previous work on scientific discovery (e.g., Langley, Simon, Bradshaw, and Zytkow 1987) has neglected conceptual change. Thus the theory of conceptual change developed here is an extension of research in machine learning, not just an application. The problem of conceptual change is open for the field of artificial intelligence, as well as for philosophy and history of science.

1.2 ARE THERE SCIENTIFIC REVOLUTIONS?

Many critics of Kuhn have challenged whether the concept of *revolution* is appropriately applied to the development of science (Toulmin 1972). The concept of revolution has itself undergone interesting changes, from its original application concerning objects such as celestial bodies going round and round, to modern usages involving political, social, and scientific changes (Cohen 1985; Gilbert 1973). The old view that a revolution was fundamentally a return to a previous state has been abandoned, and instead the term "revolution" is applied primarily to cases in which major transformations have occurred. Cohen (1985) argues that there have indeed been scientific revolutions, but his account is purely historical, judging developments in the history of science as revolutionary if scientists and historians have so described them.

I shall count conceptual changes as revolutionary if they involve the replacement of a whole system of concepts and rules by a new system. The two key words here are "replacement" and "system." Merely adding a new set of ideas poses no special problems, and replacement of a single concept or rule should be a simple process. What is much harder to understand is how one system can be replaced by another.

If knowledge in science were neatly accumulative, fact piling on top of fact, there would be no need for a theory of revolutionary conceptual change. But there are episodes in the history of science that strongly suggest the importance of conceptual revolutions. As principal data for my theory of revolutionary conceptual change, I shall take seven historical cases that have most universally been dubbed revolutions:

Copernicus' sun-centered system of the planets, which replaced the earth-centered theory of Ptolemy.
Newtonian mechanics, which, in addition to synthesizing celestial and earth-bound physics, replaced the cosmological views of Descartes.
Lavoisier's oxygen theory, which replaced the phlogiston theory of Stahl.
Darwin's theory of evolution by natural selection, which replaced the prevailing view of divine creation of species.
Einstein's theory of relativity, which replaced and absorbed Newtonian physics.
Quantum theory, which replaced and absorbed Newtonian physics.
The geological theory of plate tectonics that established the existence of continental drift.

Examination of these cases in the light of the new cognitive theory of conceptual change will display, from a cognitive perspective, what is revolutionary about them. I am eager not to adulterate the overused term "revolution"; the importance of conceptual revolutions is so great in part because they are so rare. Science does not make revolutionary leaps very frequently, but when it does the epistemic consequences are enormous.

1.3 THESES ON CONCEPTUAL REVOLUTIONS

To introduce the theory of conceptual revolutions developed in later chapters, I now advance six theses that sketch my major claims. First, it is necessary to characterize scientific revolutions from a cognitive perspective. The concept of revolution was originally applied to scientific developments by analogy with political and social developments. Political revolutions involve major transformations in political structures; in the American Revolution, for example, power was transferred from the British monarchs and their representatives to American citizens. Social revolutions involve major transformations in social structure, with some social classes wresting wealth and power from other social classes, for example during the Chinese revolution of 1949 (Skocpol 1979). Similarly, to understand scientific revolutions, we need to have an understanding of the kinds of structures undergoing transformation. Thesis 1 accordingly states:

1. Scientific revolutions involve major transformations in conceptual and propositional systems.

But how are concepts organized? Since the pioneering work of Quillian (1968), it has been common in AI systems to have much of the organization provided by *kind* or *is-a* hierarchies. For example, Tweety is a canary, which is a kind of bird, which is a kind of animal, which is a kind of thing. Similarly, psycholinguists have noticed the importance of kind-hierarchies for organizing the mental lexicon (Miller and Johnson-Laird 1976), and in addition have emphasized the role that *part-whole* hierarchies play. Part-hierarchies have different inferential properties from kind-hierarchies: because canaries are a kind of bird and birds have feathers, you can generally infer that canaries have feathers, but you cannot infer that beaks have feathers because beaks are parts of birds. Part-hierarchies have not received nearly as much attention in AI, although Darden and Rada (1988) show their importance in the development of the notion of a gene. Nevertheless, part-hierarchies are important for organizing concepts because they provide orderings such as: a toe is part of a foot, which is part of a leg, which is part of a body. I therefore conjecture:

2. Conceptual systems are primarily structured via kind-hierarchies and part-hierarchies.

Chapter 2 discusses the nature of concepts and the hierarchies that organize them. It also argues that there is more to conceptual change than mere revision of beliefs.

From theses 1 and 2 follows the conjecture that all scientific revolutions involve transformations in kind-relations and/or part-relations. We shall see in later chapters that this conjecture is true. A thorough analysis of the chemical revolution displays major changes in both kind-relations and part-relations (Chapter 3). Darwin not only proposed a major reorganization of kind-hierarchies by reclassifying humans from being a special kind of creature to being a kind of animal, he also transformed the meaning of *kind* by substituting a historical conception based on common descent for a notion of kind based on superficial similarity (Chapter 6). In the geological revolution, plate tectonics brought with it reorganizations of the kind-hierarchies and part-hierarchies involving continents and the seafloor (Chapter 7). Revolutions in physics discussed in Chapter 8 also display transformations in kind-hierarchies and part-hierarchies. One major ingredient in the revolution wrought by Copernicus is the reclassification of the earth as a kind of planet. Newton differentiated mass from weight, and reconceived gravity as a kind of centripetal force. Einstein's relativity theory brought with it a conceptual organization very different from that found in Newtonian mechanics, viewing mass and energy as manifestations of mass-energy. Moreover, the meaning of part-relations was changed dramatically by the substitution of an integrated space-time for commonsense concepts of space and time. Finally, quantum theory blurred the distinction between wave and particles, since light waves are quantized and particles have wavelengths. In his most recent writings, Kuhn (1987) has identified changes in taxonomic categories as characteristic of scientific revolutions.

How do new concepts and laws arise? Empirical laws are usually framed in the same vocabulary as the observational descriptions on which they are based or in terms directly derived from observational ones. In contrast, theories often invoke entities and processes that are unobservable or at least unobserved. Electrons, quarks, and mental processes are examples of theoretical entities postulated because of the explanatory power of the hypotheses that state their existence. The distinction between theoretical and observational concepts is not absolute, since better instruments can render a theoretical entity observable, as the electron microscope did for the gene reconceived as a sequence of DNA. Concepts referring to theoretical entities or processes cannot be derived from observation, so how can they arise? Theoretical concepts can be formed by *conceptual combination*, in which new concepts derive from parts of old ones (Thagard 1988). For example, the concept of sound waves, which are not observable, is the result of conjoining the concept of sound with the concept of a wave, both derived from observation. I therefore conjecture:

3. New theoretical concepts generally arise by mechanisms of conceptual combination.

This conjecture seems to fit well with additional cases from the history of science, such as natural selection and continental drift, but I have not conducted a sufficiently complete canvas of scientific theories to feel confident that it is true. Other mechanisms are probably necessary.

Theses 2 and 3 deal with the structure and origin of conceptual systems, while theses 4 and 5 address the same questions for propositional systems. Chapter 4 will describe a computational theory of explanatory coherence designed to account for revolutionary and nonrevolutionary cases of theory evaluation. On this account, the most important relations between proposi-

1.5 SUMMARY

Conceptual revolutions occur when whole systems of concepts are replaced by new ones. Scientific revolutions include the development of Copernicus' theory of the solar system, Newtonian mechanics, Lavoisier's oxygen theory, Darwin's theory of evolution, Einstein's theory of relativity, quantum theory, and the geological theory of plate tectonics. A theory of conceptual revolutions should illuminate these cases by saying what concepts are, how they are organized into systems, and how conceptual systems are formed and replaced.

1.6 APPENDIX: CHRONOLOGY OF REVOLUTIONS

4th century B.C. Aristotle.

2nd century A.D. Ptolemy.

1543	Copernicus' *On the Revolutions.*
1609	Kepler's *Astronomia Nova* proposes elliptical orbits.
1628	Harvey's *On the Motion of the Heart and Blood.*
1632	Galileo's *Dialogue Concerning the Two Chief World Systems.*
1644	Descartes' *Principles of Philosophy.*
1687	Newton's *Principia.*
1723	Stahl's *Fundamenta Chymiae.*
1772	Lavoisier, age 29, conjectures that calxes contain air.
1777	Lavoisier proposes that "pure air" combines with metal to produce metals.
1783	Lavoisier attacks the phlogiston theory.
1831–1836	Darwin travels on voyage of the Beagle
1837	Darwin concludes that species evolved.
1838	Darwin, age 29, discovers natural selection.
1859	Darwin's *On the Origin of Species.*
1871	Darwin's *Descent of Man.*
1887	Michelson-Morley experiment.
1905	Einstein discovers special relativity. Einstein uses a quantum hypothesis to explain the photoelectric effect.

1913	Bohr model of the atom. Watson introduces behaviorism.
1915	Einstein develops general relativity. Wegener's *Origins of Continents and Oceans* proposes continental drift.
1919	Bending of light is confirmed.
1923	de Broglie introduces particle-wave duality.
1926	Schrödinger's wave mechanics.
1927	Heisenberg's uncertainty principle.
1956	Cognitivist works by Miller and by Newell and Simon.
1959	Harry Hess postulates seafloor spreading.
1965	J. Tuzo Wilson develops concept of transform fault.
1968	Jason Morgan develops mathematical theory of plate tectonics.

CHAPTER 2

Concepts and Conceptual Systems

THIS chapter develops an account of the nature and organization of concepts that provides the basis for a theory of revolutionary conceptual change. It begins with a survey of the varied theories of concepts that have been proposed in philosophy. The major burden of this chapter is to contest the view that there is little need to discuss conceptual change because the development of scientific knowledge can be fully understood in terms of *belief revision.* This view is widely taken for granted in philosophy and artificial intelligence. A psychological discussion of the functions and nature of concepts, along with an account of the organization of concepts into hierarchies, shows how a theory of conceptual change can involve much more than belief revision.

2.1 PHILOSOPHICAL THEORIES OF CONCEPTS

From Plato to current journals in psychology and AI, discussions of knowledge and mind frequently deal with concepts, schemas, and ideas. A critical survey of all the different accounts of concepts that have been offered in philosophy, psychology, and AI would take a volume in itself. Fortunately, the task here is more focused: to give an account of concepts that can support a theory of conceptual revolutions.

Writers in different fields and at different times have had various goals behind their accounts of what concepts are. For philosophers, the goals have often been metaphysical and epistemological: to describe the fundamental nature of reality and to show how knowledge of reality can be possible. Psychologists have proposed theories of concepts in the service of the empirical goal of accounting for observed aspects of human thinking. AI researchers develop accounts of conceptual structure to enable computers to perform tasks that require intelligence when done by humans. I share all these general goals, but in this project they are subordinated to the particular problem of making sense of revolutionary conceptual change.

In epistemology and science as in politics, to be ignorant of the past is to be condemned to repeat it. The problem of the nature of concepts has been with us since Plato and Aristotle, and while summarizing the views of major philosophers at a one-paragraph level can be misleading, I want to give a general overview of the range of opinions concerned with what concepts might be.

The first influential theory of concepts was proposed by Plato twenty-three hundred years ago and was explicitly designed to *avoid* problems of change. Heraclitus had argued that the world was in a state of perpetual flux, and Plato concluded that valid knowledge would have to be of a heavenly realm of ideas removed from the changing, subjectively perceived world. In dialogues such as the *Euthyphro*, *Meno*, and *Republic*, Socrates served as Plato's mouthpiece in investigating questions about the nature of piety, virtue, knowledge, and justice (Plato 1961). Typically, Socrates argued that understanding of these concepts is not to be had through particular examples or through rough definitions. Plato's view was that knowledge of conceptual essences could only come from the use of education to regain the acquaintance with the heavenly forms that was lost at birth. Learning is then just recollection of the essences of things, and concepts are abstract, unworldly, and immutable.

Aristotle's *Metaphysics* (1961) systematically criticized Plato's theory of eternal ideas, arguing that there was no need to postulate an eternal substance. Aristotle proposed instead that only particular objects were substances: there are no universal concepts independent of the objects, since universals are only what is common to all things to which a predicate applies. The essence of *human* is to be found by a definition that states what all humans have in common, not by acquaintance with a heavenly idea of the human. In Aristotle we find fundamental worries about the relation of the general and the particular that have survived to this day. Should a concept be treated as something additional to all the objects that fall under that concept? Related questions arise today in cognitive science, when researchers debate whether there needs to be a mental representation of a concept over and above the representation of the objects that fall under it.

For Plato and Aristotle, the point of discussing ideas was metaphysical, not psychological: the issue was whether ideas are a separate kind of substance, not whether they are part of thought. In contrast, seventeenth-century discussions of ideas tied them closely to thinking. In his *Meditations*, Descartes said that ideas are thoughts that are like images of things, as when one thinks of a man or an angel (Descartes 1980, 68). Some of these ideas are innate, while others derive from an external source and others are constructed by the thinker. Descartes used his method of doubting everything to determine what ideas are "clear and distinct" and therefore safe sources of truth. For Descartes as for Plato, the most important knowledge was true *a priori*, independent of experience. Leibniz also treated an idea as something in the mind (Leibniz 1951, 281) and maintained that knowledge is primarily found through clear innate ideas.

Writing in the seventeenth, eighteenth, and nineteenth centuries, the British empiricists offered a very different view of the origins of ideas and knowledge. Locke, Berkeley, Hume, Hartley, Reid, and Mill all contended that ideas derive from sense experience. Locke's seventeenth-century *Essay Con-*

cerning Human Understanding claimed against Descartes that there are no innate principles in the mind. He characterized an idea as "whatsoever is the object of the understanding when a man thinks" (Locke 1961, vol. 1, p. 9). Ideas are divided into simple and complex, with simple ideas coming directly from sensation and complex ideas being formed from the mind out of simple ones. Not all thinkers of this period, however, adopted a mentalistic view of concepts. According to Weitz (1988), the seventeenth-century thinker Hobbes understood concepts as linguistic entities given through definitions.

Like Locke, Berkeley maintained in 1710 that the objects of human knowledge are either ideas imprinted on the senses or ideas formed by help of memory and imagination (Berkeley 1962, 65). He differed from Locke primarily in denying that there are any abstract ideas separate from ideas of particular things. A few decades later Hume distinguished the perceptions of the human mind into impressions and ideas: the former are forceful and lively perceptions while the latter are faint images of perceptions (Hume 1888, 1). Hume approved of Berkeley's rejection of abstract ideas, and extended Berkeley's skeptical conclusion that there is no material substance independent of ideas with an argument that the self should only be considered to be a bundle of sense impressions. Hume made a sharp division of objects of human reason into *matters of fact* that are acquired directly through sense experience and *relations of ideas* that arise by thinking with ideas acquired by sense experience.

Partly in response to Hume, Kant gave an account of the contents of mind that is often viewed as a synthesis of the rationalist account of Descartes and the empiricist account of Locke. He thought that there are pure, a priori concepts of the understanding, but that these can be applicable to sensory appearances by means of *schemata,* which are obscurely characterized as a rule of synthesis or a universal procedure of the imagination (Kant 1965, 182). Like the empiricists, Kant was concerned with thought, but he placed more emphasis on the a priori preconditions of thought. Like all his predecessors, Kant had a static view of concepts, considering in the abstract the conditions of their application.

Consideration of conceptual dynamics began only in the nineteenth century; Hegel probably deserves to be understood as the founder of the study of conceptual change. Whereas Kant tried to find a foundation for knowledge using both empiricist and rationalist ideas, Hegel stressed the importance of conceptual development. From Plato to Kant, philosophers were looking for a method to achieve direct and certain knowledge, either by reason or from the senses. Hegel for the first time rejected the need for any direct foundation for knowledge, emphasizing instead the need for consciousness to develop a grasp of truth through a dialectical process of passing through and criticizing successive stages of complexity, from sense experience to much more abstract thought (Hegel 1967). Even less than the other philosophers I have been

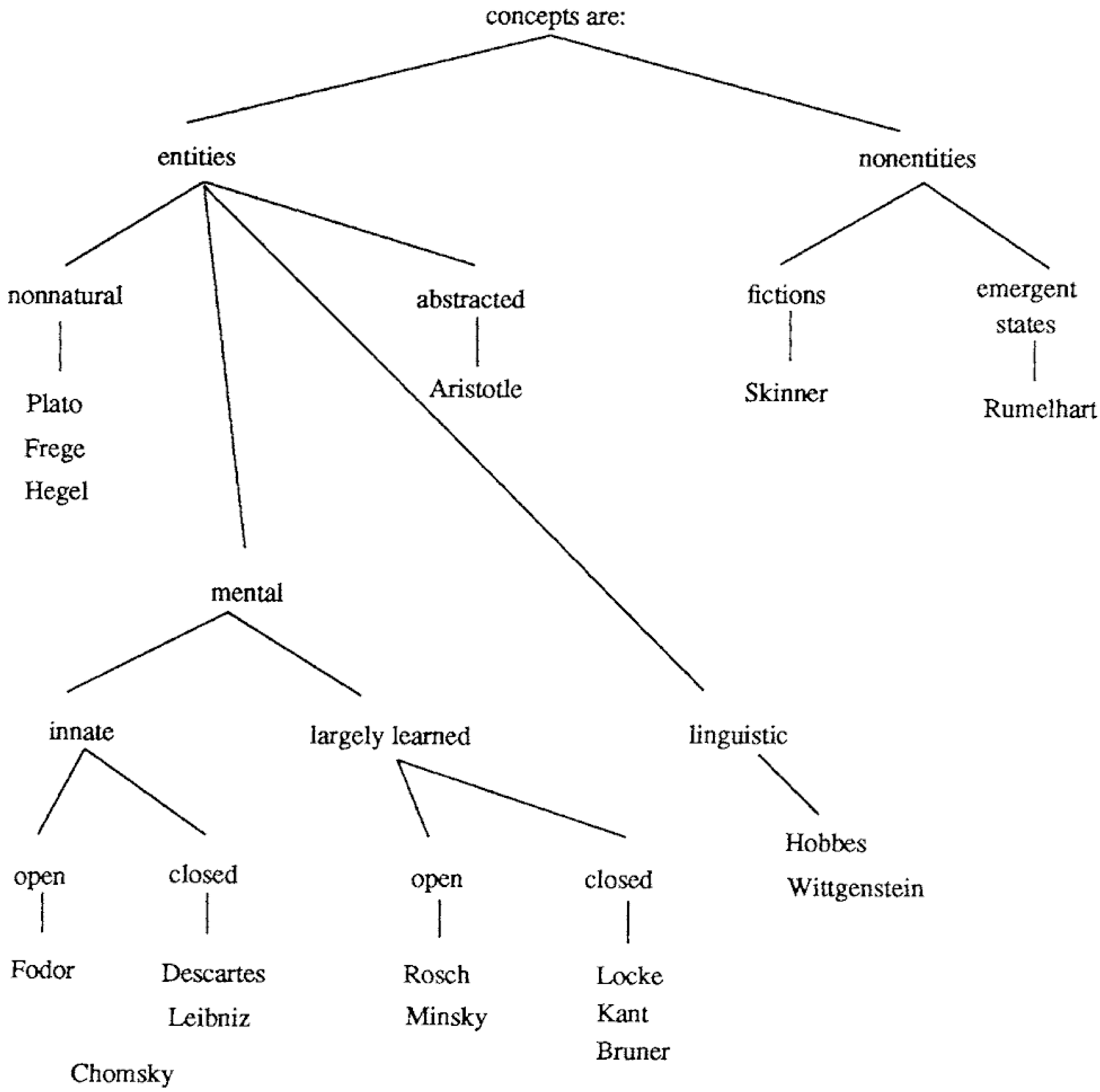


Figure 2.1. Taxonomy of theoretical views of the nature of concepts. Concepts are *closed* if they are definable in terms of necessary and sufficient conditions, and open otherwise.

terms of necessary and sufficient conditions. Among modern theorists, few besides Chomsky and Fodor hold that concepts are innate. Chomsky (1988, 134) claims that in addition to setting grammatical parameters "the language learner must discover the lexical items of the language and their properties. To a large extent this seems to be a problem of finding what labels are used for preexisting concepts." It is not clear whether Chomsky believes that concepts are open or closed. The innateness of concepts has also been defended by Fodor (1975), who maintains that concepts are open (Fodor, Garrett, Walker, and Parkes 1980).

Today, most theorists who view concepts as mental entities hold with Locke that concepts are largely learned from experience, while agreeing with

Kant that such learning presupposes an innate apparatus. Much psychological research has been conducted on concept learning, initially under the assumption that concepts are definable in terms of necessary and sufficient conditions (e.g., Bruner, Goodnow, and Austin 1956). In the 1970s, however, experiments by Rosch and her colleagues (Rosch et al. 1976) were used to support the view of concepts as *prototypes* lacking sharp defining conditions. Related critiques of the traditional view of concepts and definitions were mounted by Wittgenstein (1953), Putnam (1975), and Minsky (1975).

Theorists who deny that concepts are entities fall into two camps. The most radical camp claims that concepts are mere fictions and should not be part of a scientific analysis of language and mind. Skinner (1976), for example, rejected the postulation of mental entities such as concepts in favor of a focus on verbal behavior; see section 9.2 for further discussion of behaviorism. A novel nonentity view has emerged in recent years from connectionists such as Rumelhart and his colleagues, who assert about concept-like schemata: "Schemata are not 'things.' There is no representational object which is a schema. Rather schemata emerge at the moment they are needed from the interaction of large numbers of much simpler elements all working in concert with one another. Schemata are not explicit entities, but rather are implicit in our knowledge and are created by the very environment that they are trying to interpret—as it is interpreting them" (Rumelhart, Smolensky, McClelland, and Hinton 1986, 20). Concepts, then, are not entities but emergent states of neural networks. The connectionist view of concepts is discussed in section 2.4. My own view, developed later in this chapter, is that concepts are mental entities that are largely learned and open.

2.2 BELIEF REVISION VERSUS CONCEPTUAL CHANGE

Although discussions of concepts have been important in the history of philosophy, they have been much rarer in contemporary philosophy. According to Ian Hacking (1975), current analytic philosophy is the "heyday of sentences." Whereas seventeenth-century thinkers talked of *ideas*, contemporary philosophers take sentences to be the objects of epistemological investigation. Knowledge is something like true justified belief, so increases in knowledge are additions to what is believed. Epistemology, then, consists primarily of evaluating strategies for improving our stock of beliefs, construed as sentences or as attitudes toward sentence-like propositions. For example, Gärdenfors (1988) models the epistemic state of an individual as a consistent set of sentences that can change by expansion and contraction.

In the cognitive sciences, however, the intellectual terrain is very different.

In cognitive psychology, the question of the nature of concepts receives far more attention than the question of belief revision. Researchers in artificial intelligence often follow philosophers in discussing belief revision, but they also pay much attention to how knowledge can be organized in concept-like structures called *frames* (Minsky 1975; for reviews see Thagard 1984, 1988). Nevertheless, even a philosopher like Alvin Goldman (1986) who takes cognitive science very seriously places belief revision at the center of his epistemology, paying scant attention to the nature of concepts and the question of conceptual change. Gilbert Harman has written both on epistemic change (1986) and on the nature of concepts (1987), but has not much discussed the relevance of the latter topic to the former. Historically oriented philosophers of science such as Kuhn (1970) have suggested the importance of conceptual change but have not provided accounts of conceptual structure that are sufficiently developed for epistemological application.

The central question in current epistemology is: when are we justified in adding and deleting beliefs from the set of beliefs judged to be known? Without denigrating this question, I propose that epistemology should also address the question: what are concepts and how do they change? Concepts are relevant to epistemology if the question of conceptual change is not identical to the question of belief revision. But maybe it is; consider the following argument.

> The issue of conceptual change is a red herring. Whenever a concept changes, it does so by virtue of changes in the beliefs that employ that concept (or predicate, if you are thinking in terms of sentences). For example, if you recategorize whales as mammals rather than fish, you have made an important change in the concept *whale*. But this amounts to no more than deleting the belief that whales are fish and adding the belief that whales are mammals. Your concept of mammal may also change by adding the belief that whales produce milk, but this merely follows from the other belief addition. So as far as epistemology is concerned, conceptual change is redundant with respect to the central question of belief revision.

Thus anyone who thinks conceptual change is important has to give an account of it that goes beyond mere belief revision.

The above argument assumes that the principles according to which beliefs are added and deleted operate independently of considerations of conceptual structure. If you are a Bayesian, belief revision is just a matter of changing probability distributions over the set of propositions (see, for example, Horwich 1982). But suppose that you want to take a more psychologically realistic approach to belief revision, one that could account for why some revisions are harder to make than others and why some revisions have more global effects. Perhaps such facets of belief revision can only be understood by noticing how beliefs are organized via concepts. There may be a difference between deciding that whales are mammals and deciding that whales have fins, a difference that can only be understood in terms of the overall structure of our conceptual system, relating *whale* to *mammal* in ways more fundamental than simply having the belief that whales are mammals. For the moment, this is only a possibility, not a refutation of the argument that conceptual change is just belief revision. But it is enough to suggest that it is worth exploring the cognitive science literature on concepts for suggestions about how conceptual structure could matter to belief revision.

2.3 WHAT ARE CONCEPTS FOR?

Before proceeding further, some clarification is in order concerning concepts and predicates, and sentences and propositions. Sentences are syntactic entities, marks on paper. Among their constituents are predicates such as "whale" in the sentence "Gracy is a whale." In contrast, I shall treat concepts and propositions as mental representations, with concepts corresponding to predicates and propositions corresponding to sentences. In my usage, concepts are mental structures representing what words represent, and propositions are mental structures representing what sentences represent. This mentalistic interpretation is not the only one possible: a Platonist could treat concepts as the meaning of predicates and propositions as the meaning of sentences independent of what is in anybody's head. Instead of discussing abstract meanings, I follow many researchers in psychology and artificial intelligence in supposing that concepts are mental structures analogous to data structures in computers. Anyone who is wedded to concepts and propositions as abstractions should reinterpret my concepts as "conceptual representations" and my propositions as "propositional representations." Whereas words are parts of sentences and not vice-versa, concepts and propositions as mental representations can be parts of each other (see section 2.5).

To prevent additional terminological confusion, it is necessary to point out that researchers in cognitive psychology and artificial intelligence tend to use the terms "knowledge" and "belief" differently from philosophers, who often characterize knowledge as something like true justified belief. Their use of "knowledge" is closer to philosophers' use of "belief." Cognitive scientists have also taken to using the term "epistemology" broadly to cover anything having to do with knowledge in a diluted sense ignoring justification. In this book I generally use "knowledge" and "epistemology" in their traditional philosophical senses that presuppose questions of justification.

Psychologists have many reasons for being interested in the nature of concepts. Whereas the epistemologist's primary concern is with the question of justification, the psychologist must try to account for many different kinds of behavior. Here is a list, undoubtedly incomplete, of various roles that concepts have been deemed to play, using the concept *whale* as an example.

1. *Categorization.* Our concept *whale* enables us to recognize things as whales.
2. *Learning.* Our concept *whale* must be capable of being learned, perhaps from examples, or perhaps by combining other existing concepts.
3. *Memory.* Our concept *whale* should help us remember things about whales, either in general or from particular episodes that concern whales.
4. *Deductive inference.* Our concept *whale* should enable us to make deductive and inductive inferences about whales, for example enabling us to infer that since Gracy is a whale, she has fins.
5. *Explanation.* Our knowledge about whales should enable us to generate explanations, for example saying that Gracy swims *because* she is a whale.
6. *Problem solving.* Our knowledge about whales should enable us to solve problems, for example how to get an errant whale out of the harbor.
7. *Generalization.* Our concept *whale* should enable us to learn new facts about whales from additional examples, for example to form new general conclusions such as that whales have blubber under their skin.
8. *Analogical inference.* Our concept *whale* should help us to reason using similarities: if you know that dolphins are quite intelligent and are aquatic mammals like whales, then whales are perhaps intelligent too. Metaphor should also be supportable by the concept, as when we say that someone had a whale of an idea.
9. *Language comprehension.* Our understanding of sentences such as "Gracy is a whale" depends on our knowing something about the concept *whale.*
10. *Language production.* We need to be able to utter sentences like "Gracy is a whale" and "Whales are less friendly than dolphins."

Ignoring the last two language issues, which introduce problems not directly connected to belief revision, we can examine whether the first eight roles require that belief change pay attention to conceptual structure. Categorization might be seen as a straightforward case of belief application: you believe that any large sea-object that moves and blows water into the air is a whale, so you categorize the large blob in the ocean producing spray as a whale. You thereby add the belief "the blob is a whale" to your set of beliefs. But categorization is rarely so simple as this deduction, since unexceptionable rules are hard to come by. Submarines are also large sea-objects that move and can blow water into the air. So in categorizing the blob as a whale rather than as a submarine you will need to decide which concept fits the blob better, and fitting the concept may be more than a matter of simple belief application (see the discussion of categorization in Holland, Holyoak, Nisbett, and Thagard 1986, ch. 6).

Identifying the blob as a whale presupposes that you have already learned