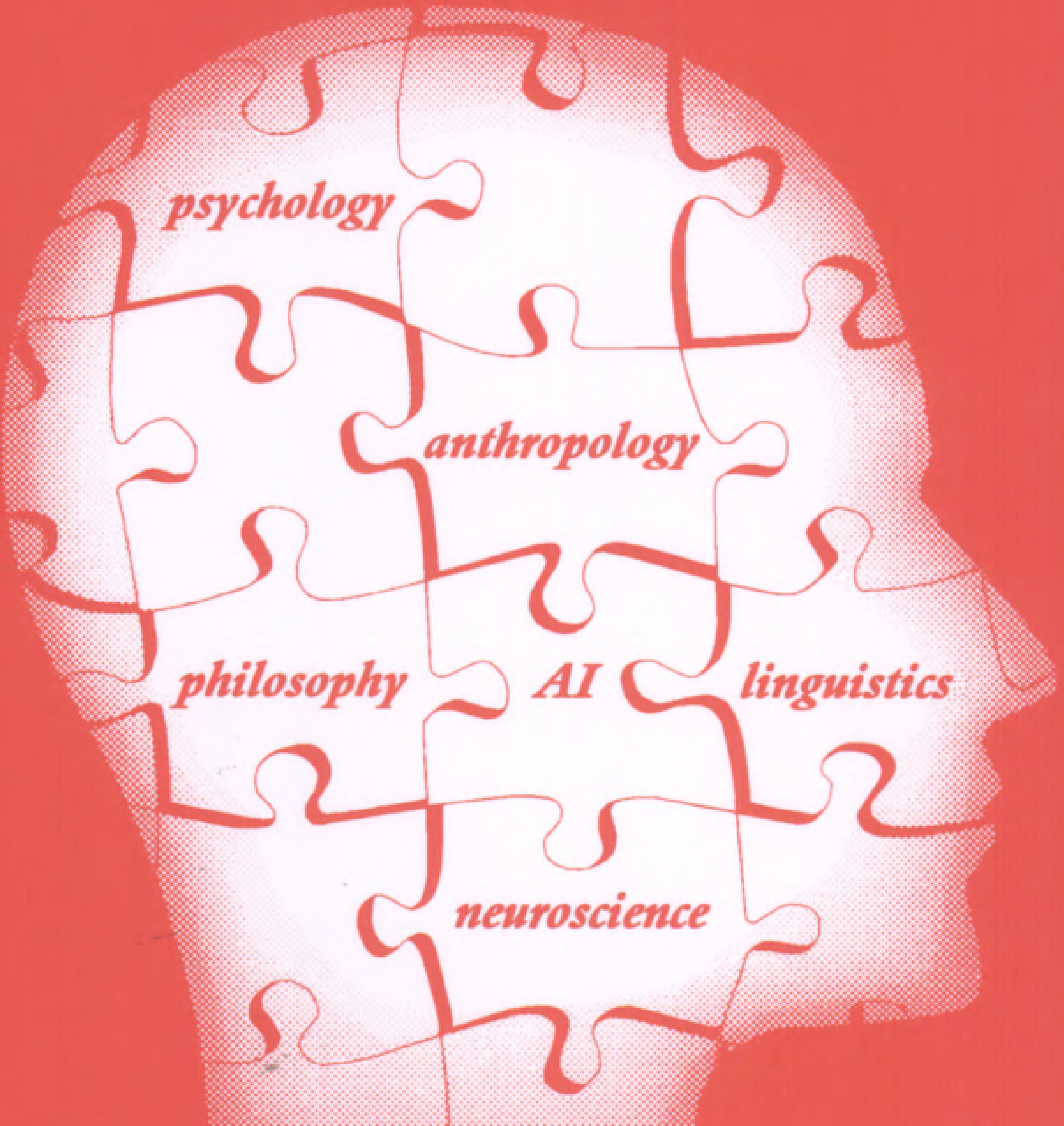
MIND

Introduction to Cognitive Science

SECOND EDITION



Paul Thagard

Mind

Introduction to Cognitive Science

second edition

Paul Thagard

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 5 Cambridge Center, Cambridge, MA 02142.

This book was set in Stone sans and Stone serif by SNP Best-set Typesetter Ltd., Hong Kong. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Thagard, Paul.
Mind: introduction to cognitive science/Paul Thagard.—2nd ed.
p. cm.
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-20154-2 (hc : alk. paper)—ISBN 0-262-70109-X (pbk. : alk. paper)
1. Cognitive science—Textbooks. I. Title.

BF311.T42 2005
153—dc22

2004053091

10 9 8 7 6 5 4 3 2 1

1 Representation and Computation

Studying the Mind

Have you ever wondered how your mind works? Every day, people accomplish a wide range of mental tasks: solving problems at their work or school, making decisions about their personal life, explaining the actions of people they know, and acquiring new concepts like *cell phone* and *Internet.* The main aim of cognitive science is to explain how people accomplish these various kinds of thinking. We want not only to describe different kinds of problem solving and learning, but also to explain how the mind carries out these operations. Moreover, cognitive science aims to explain cases where thinking works poorly—for example, when people make bad decisions.

Understanding how the mind works is important for many practical activities. Educators need to know the nature of students' thinking in order to devise better ways of teaching them. Engineers and other designers need to know what potential users of their products are likely to be thinking when they use their products effectively or ineffectively. Computers can be made more intelligent by reflecting on what makes people intelligent. Politicians and other decision makers can become more successful if they understand the mental processes of people with whom they interact.

But studying the mind is not easy, since we cannot just pop one open to see how it works. Over the centuries, philosophers and psychologists have used a variety of metaphors for the mind, comparing it, for example, to a blank sheet on which impressions are made, to a hydraulic device with various forces operating in it, and to a telephone switchboard. In the last fifty years, suggestive new metaphors for thinking have become available through the development of new kinds of computers. Many but not all cognitive scientists view thinking as a kind of computation and use computational metaphors to describe and explain how people solve problems and learn.

What Do You Know?

When students begin studying at a college or university, they have much more to learn than course material. Undergraduates in different programs will have to deal with very different subject matters, but they all need to acquire some basic knowledge about how the university works. How do you register for courses? What time do the classes begin? What courses are good and which are to be avoided? What are the requirements for a degree? What is the best route from one building to another? What are the other students on campus like? Where is the best place to have fun on Friday night?

Answers to these questions become part of the minds of most students, but what sort of part? Most cognitive scientists agree that knowledge in the mind consists of *mental representations*. Everyone is familiar with nonmental representations, such as the words on this page. I have just used the words "this page" to represent the page that you are now seeing. Students often also use pictorial representations such as maps of their campuses and buildings. To account for many kinds of knowledge, such as what students know about the university, cognitive scientists have proposed various kinds of mental representation including rules, concepts, images, and analogies. Students acquire rules such as *If I want to graduate, then I need to take ten courses in my major.* They also acquire concepts involving new terms such as "bird" or "Mickey Mouse" or "gut," all used to describe a particularly easy course. For getting from building to building, a mental image or picture of the layout of the campus might be very useful. After taking a course that they particularly like, students may try to find another similar course to take. Having interacted with numerous students from different programs on campus, students may form stereotypes of the different kinds of undergraduates, although it may be difficult for them to say exactly what constitutes those stereotypes.

The knowledge that students acquire about college life is not acquired just for the sake of accumulating information. Students face numerous problems, such as how to do well in their courses, how to have a decent social life, and how to get a job after graduation. Solving such problems requires doing things with mental representations, such as reasoning that you still need five more courses to graduate, or deciding never to take another course from Professor Tedium. Cognitive science proposes that people have mental *procedures* that operate on mental representations to produce thought and action. Different kinds of mental representations such as rules and concepts foster different kinds of mental procedures. Consider different ways of representing numbers. Most people are familiar with the Arabic numeral representation of numbers (1, 2, 3, 10, 100, etc.) and with the standard procedures for doing addition, multiplication, and so on. Roman numerals can also represent numbers (I, II, III, X, C), but they require different procedures for carrying out arithmetic operations. Try dividing CIV (104) by XXVI (26).

Part I of this book surveys the different approaches to mental representations and procedures that have developed in the last four decades of cognitive science research. There has been much controversy about the merits of different approaches, and many of the leading cognitive science theorists have argued vehemently for the primacy of the approach they prefer. My approach is more eclectic, since I believe that the different theories of mental representation now available are more complementary than competitive. The human mind is astonishingly complex, and our understanding of it can gain from considering its use of rules such as those described above as well as many other kinds of representations including some not at all familiar. The latter include "connectionist" or "neural network" representations that are discussed in chapter 7.

Beginnings

Attempts to understand the mind and its operation go back at least to the ancient Greeks, when philosophers such as Plato and Aristotle tried to explain the nature of human knowledge. Plato thought that the most important knowledge comes from concepts such as *virtue* that people know innately, independently of sense experience. Other philosophers such as Descartes and Leibniz also believed that knowledge can be gained just by thinking and reasoning, a position known as *rationalism*. In contrast, Aristotle discussed knowledge in terms of rules such as *All humans are mortal* that are learned from experience. This philosophical position, defended by Locke, Hume, and others, is known as *empiricism.* In the eighteenth century, Kant attempted to combine rationalism and empiricism by arguing that human knowledge depends on both sense experience and the innate capacities of the mind.

The study of mind remained the province of philosophy until the nineteenth century, when experimental psychology developed. Wilhelm Wundt and his students initiated laboratory methods for studying mental operations more systematically. Within a few decades, however, experimental psychology became dominated by *behaviorism,* a view that virtually denied the existence of mind. According to behaviorists such as J. B. Watson (1913), psychology should restrict itself to examining the relation between observable stimuli and observable behavioral responses. Talk of consciousness and mental representations was banished from respectable scientific discussion. Especially in North America, behaviorism dominated the psychological scene through the 1950s.

Around 1956, the intellectual landscape began to change dramatically. George Miller (1956) summarized numerous studies that showed that the capacity of human thinking is limited, with short-term memory, for example, limited to around seven items. (This is why it is hard to remember long phone or social security numbers.) He proposed that memory limitations can be overcome by recoding information into chunks, mental representations that require mental procedures for encoding and decoding the information. At this time, primitive computers had been around for only a few years, but pioneers such as John McCarthy, Marvin Minsky, Allen Newell, and Herbert Simon were founding the field of artificial intelligence. In addition, Noam Chomsky (1957, 1959) rejected behaviorist assumptions about language as a learned habit and proposed instead to explain people's ability to understand language in terms of mental grammars consisting of rules. The six thinkers mentioned in this paragraph can justly be viewed as the founders of cognitive science.

The subsequent history of cognitive science is sketched in later chapters in connection with different theories of mental representation. McCarthy became one of the leaders of the approach to artificial intelligence based on formal logic, which we will discuss in chapter 2. During the 1960s, Newell and Simon showed the power of rules for accounting for aspects of human intelligence, and chapter 3 describes considerable subsequent work in this tradition. During the 1970s, Minsky proposed that conceptlike frames are the central form of knowledge representations, and other researchers in artificial intelligence and psychology discussed similar structures called schemas and scripts (chapter 4). Also at this time, psychologists began to show increased interest in mental imagery (chapter 6). Much experimental and computational research since the 1980s has concerned analogical thinking, also known as case-based reasoning (chapter 5). The most exciting development of the 1980s was the rise of connectionist theories of mental representation and processing modeled loosely on neural networks in the brain (chapter 7). Each of these approaches has contributed to the understanding of mind, and chapter 8 provides a summary and evaluation of their advantages and disadvantages.

Many challenges and extensions have been made to the central view that the mind should be understood in terms of mental representations and procedures, and these are addressed in part II of the book (chapters 9–14). The 1990s saw a rapid increase in the use of brain scanning technologies to study how specific areas of the brain contribute to thinking, and currently there is much work on neurologically realistic computational models of mind (chapter 9). These models are suggesting new ways to understand emotions and consciousness (chapters 10 and 11). Chapters 12 and 13 address challenges to the computational-representational approach based on the role that bodies, physical environments, and social environments play in human thinking. Finally, chapter 14 discusses the future of cognitive science, including suggestions for how students can pursue further interdisciplinary work.

Methods in Cognitive Science

Cognitive science should be more than just people from different fields having lunch together to chat about the mind. But before we can begin to see the unifying ideas of cognitive science, we have to appreciate the diversity of outlooks and methods that researchers in different fields bring to the study of mind and intelligence.

Although cognitive psychologists today often engage in theorizing and computational modeling, their primary method is experimentation with human participants. People, usually undergraduates satisfying course requirements, are brought into the laboratory so that different kinds of thinking can be studied under controlled conditions. To take some

In addition to descriptive questions about how people think, philosophy concerns itself with normative questions about how they *should* think. Along with the theoretical goal of understanding human thinking, cognitive science can have the practical goal of improving it, which requires normative reflection on what we want thinking to be. Philosophy of mind does not have a distinct method, but should share with the best theoretical work in other fields a concern with empirical results.

In its weakest form, cognitive science is merely the sum of the fields just mentioned: psychology, artificial intelligence, linguistics, neuroscience, anthropology, and philosophy. Interdisciplinary work becomes much more interesting when there is theoretical and experimental convergence on conclusions about the nature of mind. Later chapters provide examples of such convergences that show cognitive science working at the intersection of various fields. For example, psychology and artificial intelligence can be combined through computational models of how people behave in experiments. The best way to grasp the complexity of human thinking is to use multiple methods, especially combining psychological and neurological experiments with computational models. Theoretically, the most fertile approach has been to understand the mind in terms of representation and computation.

The Computational-Representational Understanding of Mind

Here is the central hypothesis of cognitive science: Thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures. Although there is much disagreement about the nature of the representations and computations that constitute thinking, the central hypothesis is general enough to encompass the current range of thinking in cognitive science, including connectionist theories. For short, I call the approach to understanding the mind based on this central hypothesis *CRUM*, for *Computational-Representational Understanding of Mind.*

CRUM might be wrong. Part II of this book presents some fundamental challenges to this approach that suggest that ideas about representation and computation might be inadequate to explain fundamental facts about the mind. But in evaluating the successes of different theories of knowledge representation, we will be able to see the considerable progress in understanding the mind that CRUM has made possible. Without a doubt, CRUM has been the most theoretically and experimentally successful approach to mind ever developed. Not everyone in the cognitive science disciplines agrees with CRUM, but inspection of the leading journals in psychology and other fields reveals that CRUM is currently the dominant approach to cognitive science.

Much of CRUM's success has been due to the fact that it employs a fertile analogy derived from the development of computers. As chapter 5 describes, analogies often contribute to new scientific ideas, and comparing the mind with computers has provided a much more powerful way of approaching the mind than previous metaphors such as the telephone switchboard. Readers with a background in computer science will be familiar with the characterization of a computer program as consisting of data structures and algorithms. Modern programming languages include a variety of data structures including strings of letters such as "abc," numbers such as 3, and more complex structures such as lists (A B C) and trees. Algorithms—mechanical procedures—can be defined to operate on various kinds of structures. For example, children in elementary school learn an algorithm for operating on numbers to perform long division. Another simple algorithm can be defined to reverse a list, turning (A B C) into (C B A). This procedure is built up out of smaller procedures for taking an element from one list and adding it to the beginning of another, enabling a computer to build a reversed list by forming (A), then (B A), then (C B A). Similarly, CRUM assumes that the mind has mental representations analogous to data structures, and computational procedures similar to algorithms. Schematically:

Program	*Mind*
data structures + algorithms = running programs	mental representations + computational procedures = thinking

This has been the dominant analogy in cognitive science, although it has taken on a novel twist from the use of another analog, the brain. Connectionists have proposed novel ideas about representation and computation that use neurons and their connections as inspirations for data structures, and neuron firing and spreading activation as inspirations for algorithms. CRUM then works with a complex three-way analogy among the mind, the brain, and computers, as depicted in figure 1.1. Mind, brain,

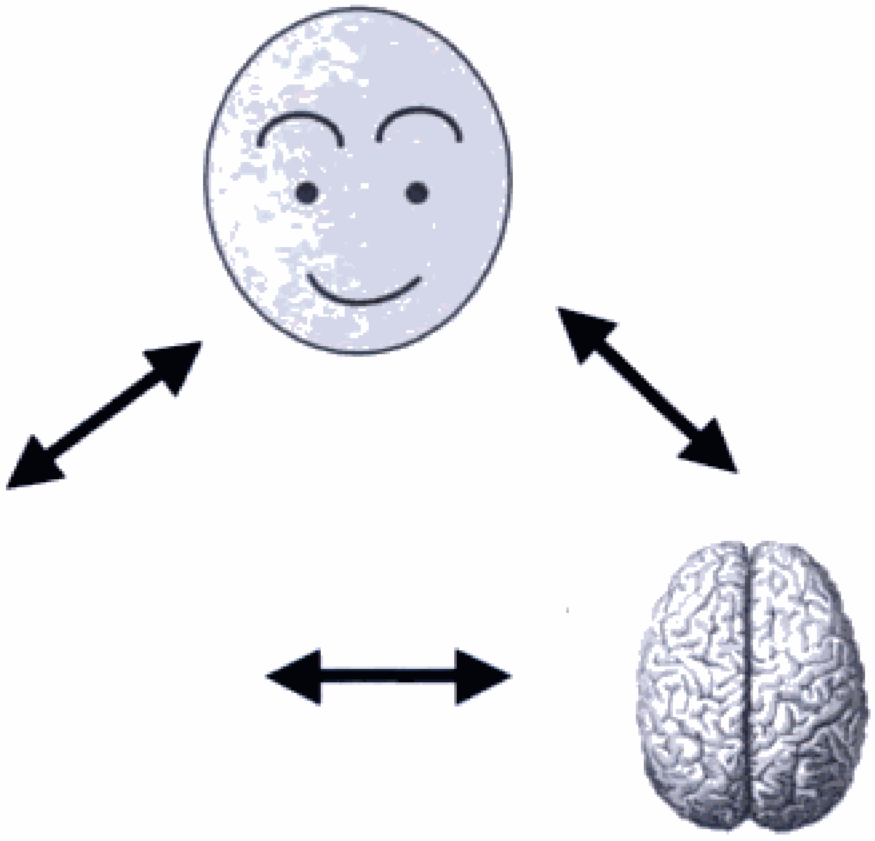


Figure 1.1
Three-way analogy between minds, computers, and brains.

and computation can each be used to suggest new ideas about the others. There is no single computational model of mind, since different kinds of computers and programming approaches suggest different ways in which the mind might work. The computers that most of us work with today are serial processors, performing one instruction at a time, but the brain and some recently developed computers are parallel processors, capable of doing many operations at once.

If you already know a lot about computers, thinking about the mind computationally should come fairly naturally, even if you do not agree that the mind is fundamentally like a computer. Readers who have never written a computer program but have used cookbooks can consider another analogy. A recipe usually has two parts: a list of ingredients and a set of instructions for what to do with them. A dish results from applying cooking instructions to the ingredients, just as a running program results from applying algorithms to data structures such as numbers and lists, and just as thinking (according to CRUM) results from applying computational procedures to mental representations. The recipe analogy for thinking is weak, since ingredients are not representations and cooking instructions require someone to interpret them. Chapters 2–7 provide simple examples of computational procedures that map much more directly onto the operations of mind.

Theories, Models, and Programs

Computer models are often very useful for theoretical investigation of mental processes. Comprehension of cognitive science models requires noting the distinctions and the connections among four crucial elements: theory, model, program, and platform. A cognitive *theory* postulates a set of representational structures and a set of processes that operate on these structures. A computational *model* makes these structures and processes more precise by interpreting them by analogy with computer programs that consist of data structures and algorithms. Vague ideas about representations can be supplemented by precise computational ideas about data structures, and mental processes can be defined algorithmically. To test the model, it must be implemented in a software *program* in a programming language such as LISP or Java. This program may run on a variety of hardware *platforms* such as Macintoshes, Sun Workstations, or IBM PCs, or it may be specially designed for a specific kind of hardware that has many processors working in parallel. Many kinds of structures and processes can be investigated in this way, from the rules and search strategies of some traditional sorts of artificial intelligence, to the distributed representations and spreading activation processes of newer connectionist views.

Suppose, for example, that you want to understand how children learn to add numbers together in problems such as 13 + 28 = ? A cognitive theory would postulate how children represent these numbers and how they process the representations to accomplish addition. The theory would propose whether 13 is to be represented by a single structure, a combined structure such as *10 plus 3*, or by a complex of neuronlike structures. The theory would also propose processes that operate on the structures to produce a result such as 41, including the carrying operation that somehow turns 30-plus-11 into 41. A computational model would specify the nature of the representations and processes more precisely by characterizing programmable structures and algorithms that are intended to be analogous to the mental representations and processes for addition. To evaluate the theory and model, we can write a computer program in a computer language such as LISP, running the program to compare its performance with human adders and checking that the program not only gets the same right answers as the humans but also makes the same kind of mistakes. Our program might run on any number of different platforms such as PCs, or it might be specially tailored to a particular kind of computer such as one that mimics the neuronal structure of the brain.

The analogy between mind and computer is useful at all three stages of the development of cognitive theories: discovery, modification, and evaluation. Computational ideas about different kinds of programs often suggest new kinds of mental structures and processes. Theory development, model development, and program development often go hand in hand, since writing the program may lead to the invention of new kinds of data structures and algorithms that become part of the model and have analogs in the theory. For example, in writing a computer program to simulate human addition, a programmer might think of a kind of data structure that suggests new ideas about how children represent numbers. Similarly, evaluation of theory, model, and program often involves all three, since our confidence in the theory depends on the model's validity as shown by the program's performance. If the computer program for doing addition cannot add, or if it adds more perfectly than humans, we have reason to believe that the corresponding cognitive theory of addition is inadequate.

The running program can contribute to evaluation of the model and theory in three ways. First, it helps to show that the postulated representations and processes are computationally realizable. This is important, since many algorithms that seem reasonable at first glance do not scale up to large problems on real computers. Second, in order to show not only the computational realizability of a theory but also its psychological plausibility, the program can be applied qualitatively to various examples of thinking. Our addition program, for example, should be able to get the same kinds of right and wrong answers as children. Third, to show a much more detailed fit between the theory and human thinking, the program can be used quantitatively to generate detailed predictions about human thinking that can be compared with the results of psychological experiments. If there are psychological experiments that show that children get a certain percentage of a class of addition problems right, then the computer program should get roughly the same percentage right. Cognitive theories by themselves are normally not precise enough to generate such quantitative predictions, but a model and program may fill the gap between theory and observation.

take into account how efficient the computation is. Imagine a procedure that takes only a second to be applied once, but twice as long the second time, and twice as long as that the third time, and so on. Then twenty applications would take 2^{20} seconds, which are more seconds than there have been in the approximately 15 billion years since the universe was formed. Both naturally and artificially intelligent systems need to have sufficient speed to work effectively in their environments.

When people solve a problem, they are usually able to learn from the experience and thereby solve it much more easily the next time. For example, the first time that students register for classes is usually very confusing since they do not know what procedures to follow or how to go about choosing good classes. Subsequently, however, registering typically gets a lot easier. Part of being intelligent involves being able to learn from experience, so a theory of mental representation must have sufficient computational power to explain how people learn. In discussing different approaches to mental representation, we will encounter diverse kinds of human learning, ranging from the acquisition of new concepts such as *registration* and rules such as *Never sign up for an 8:30 class* to more subtle kinds of adjustment in performance.

In addition to problem solving and learning, a general cognitive theory must account for human language use. Ours is the only species on Earth capable of complex use of language. General principles of problem solving and learning might account for language use, but it is also possible that language is a unique cognitive capacity that must be dealt with specially. At least three aspects of language use need to be explained: people's ability to comprehend language, their ability to produce utterances, and children's universal ability to learn language. Different approaches to knowledge representation provide very different answers to how these work.

If artificial intelligence is viewed as a branch of engineering, it can develop computational models of problem solving, learning, and language that ignore how people accomplish these tasks; the question is just how to get computers to do them. But cognitive science has the goal of understanding *human* cognition, so it is crucial that a theory of mental representation not only have a lot of representational and computational power, but also be concerned with how people think. Accordingly, the third criterion for evaluating a theory of mental representation is psychological plausibility, which requires accounting not just for the qualitative capacities of humans but also for the quantitative results of psychological experiments concerning these capacities. Relevant experiments include ones dealing with the same high-level tasks that were discussed under the heading of computational power: problem solving, learning, and language. The difference between this criterion and the last is that a cognitive theory of mental representation must not only show how a task is possible computationally, but also try to explain the particular ways that humans do it.

Similarly, since human thought is accomplished by the human brain, a theory of mental representation must at least be consistent with the results of neuroscientific experiments. Until recently, neurological techniques such as recording EEGs of brain waves seemed too crude to tell us much about high-level cognition, but the past two decades have brought new scanning techniques that can identify where and when in the brain certain cognitive tasks are performed. Cognitive neuroscience has thereby become an important part of reflection on the operations of mind, so we should try to assess each approach to knowledge representation in terms of neurological plausibility, even though information about how the brain produces cognition is still limited (see chapter 9).

The fifth and final criterion for evaluating theories of mental representation is practical applicability. Although the main goal of cognitive science is to understand the mind, there are many desirable practical results to which such understanding can lead. This book considers what each of the approaches to knowledge representation has to tell us about four important kinds of application: education, design, intelligent systems, and mental illness. For educational purposes, cognitive science should be able to increase understanding of how students learn, and also to suggest how to teach them better. Design problems, such as how to make computer interfaces that people like to use, should benefit from an understanding of how people are thinking when they perform such tasks. Developing intelligent systems to act either as stand-alone experts or as tools to support human decisions can directly benefit from computational ideas about how humans think. Different theories of mental representation have given rise to very different sorts of expert computer systems, including rule-based, case-based, and connectionist tools. Other potential practical applications of cognitive science include understanding and treatment of mental illness.

As we will see, no single approach to mental representation fully satisfies all these criteria. Moreover, there are aspects of human thinking such as perception (sight, hearing, touch, smell, taste), emotion, and motor control that are not included in these criteria (see chapters 10–12). Nevertheless, the criteria provide a framework for comparing and evaluating current theories of mental representation with respect to their accomplishments as well as their shortcomings.

Summary

Researchers in psychology, artificial intelligence, neuroscience, linguistics, anthropology, and philosophy have adopted very different methods for studying the mind, but ideally these methods can converge on a common interpretation of how the mind works. A unified view of cognitive science comes from seeing various theoretical approaches as all concerned with mental representations and procedures that are analogous to the representations and procedures familiar in computer programs. The Computational-Representational Understanding of Mind operates with the following kind of explanation schema:

Explanation target

Why do people have a particular kind of **intelligent behavior**?

Explanatory pattern

People have mental **representations**.

People have algorithmic **processes** that operate on those **representations**.

The **processes**, applied to the **representations**, produce the **behavior**.

The words in boldface are placeholders, indicating that to explain various kinds of intelligent behavior, various kinds of representations and processes can be considered. Currently, there are six main approaches to modeling the mind, involving logic, rules, concepts, analogies, images, and neural connections. These can be evaluated according to five criteria: representational power, computational power, psychological plausibility, neurological plausibility, and practical applicability.

The fundamental presuppositions that have guided the writing of this book are:

1. The study of mind is exciting and important. It is exciting for theoretical reasons, since the attempt to investigate the nature of mind is as challenging as anything attempted by science. It is also exciting for practical reasons, since knowing how the mind works is important for such diverse endeavors as improving education, improving design of computers and other artifacts, and developing intelligent computational systems that can aid or replace human experts.

2. The study of mind is interdisciplinary. It requires the insights that have been gained by philosophers, psychologists, computer scientists, linguists, neuroscientists, anthropologists, and other thinkers. Moreover, it requires the diversity of methodologies that these fields have developed.

3. The interdisciplinary study of mind (cognitive science) has a core: the Computational-Representational Understanding of Mind (CRUM). Thinking is the result of mental representations and computational processes that operate on those representations.

4. CRUM is multifarious. Many kinds of representations and computations are important to understanding human thought, and no single computational-representational account now available does justice to the full range of human thinking. This book reviews (in chapters 2–8) the six major current approaches to understanding the mind in terms of representations and computation.

5. CRUM is successful. The computational-representational approach has exceeded all previous theories of mind in its theoretical ability to account for psychological performance and its practical ability to improve that performance.

6. CRUM is incomplete. Not all aspects of human thought and intelligence can be accounted for in purely computational-representational terms. Substantial challenges have been made to CRUM that show the necessity of integrating it with biological research (neuroscience) and with research on social aspects of thought and knowledge.

Discussion Questions

1. What are additional examples of things that students learn when they go to college or university?

2. Why have researchers in different fields adopted different methods for studying the mind?

3. Can you think of any alternatives to the computational-representational understanding of mind?

4. What aspects of human thinking are most difficult for computers to perform or model? What would it take to convince you that a computer is intelligent?

5. Are theories and models in cognitive science like theories and models in physics and other fields?

6. Are there additional criteria that you would want a theory of mental representation to meet?

Further Reading

Three recent reference works contain valuable articles on many aspects of cognitive science: *The MIT Encyclopedia of the Cognitive Sciences* (Wilson and Keil 1999), *A Companion to Cognitive Science* (Bechtel and Graham 1998), and *Encyclopedia of Cognitive Science* (Nadel 2003).

On the history of cognitive science, see Gardner 1985 and Thagard 1992, chap. 9. Other introductions to cognitive science include Johnson-Laird 1988, Stillings et al. 1995, Dawson 1998, and Sobel 2001. General collections of articles include Polk and Seifert 2002 and Thagard 1998.

Textbooks on cognitive psychology include Anderson 2000, Medin, Ross, and Markman 2001, and Sternberg 2003. For introductions to artificial intelligence, see Russell and Norvig 2003 and Winston 1993. Graham 1998 and Clark 2001 provide introductions to the philosophy of mind and cognitive science. An introductory linguistics text is Akmajian et al. 2001. For accessible introductions to cognitive neuroscience, see LeDoux 2002 and Kosslyn and Koenig 1992; Churchland and Sejnowski 1992 present a more computational approach. D'Andrade 1995 provides an introduction to cognitive anthropology.

Web Sites

Note: Live links to all the sites mentioned in this book can be found at my own Web site, http://cogsci.uwaterloo.ca/courses/resources.html.

Artificial Intelligence in the news (American Association for Artificial Intelligence): http://www.aaai.org/AITopics/html/current.html

Aristotle's discovery of how to analyze syllogisms purely in terms of their form, ignoring their content, has had a major influence on logic. The discovery's usefulness, however, has been challenged from a psychological perspective, as we will see below in the section on psychological validity.

The syllogism is a form of *deductive* inference, in which the conclusion follows necessarily from the premises: if the premises are true, the conclusion is true also. *Inductive* inference is more dangerous since it introduces uncertainty. If all the students you know are overworked, you might inductively infer that all students are overworked. But your conclusion might well be erroneous—for example, if there are basket-weaving majors you do not know who take it easy.

Although the syllogism dominated discussions of formal logic for two thousand years, it is not sufficient to represent all inferences. Syllogisms are fine for simple predicates like "is a student" but they can not handle relations such as *take* in sentences like "Students who take courses get credit for them." Here *take* is a relation between a student and a course. Modern logic began in 1879 with the work of the German mathematician Gottlob Frege (1960), who devised a formal system of logic much more general than Aristotle's. Subsequently, Bertrand Russell and many other logicians have found ways of increasing the representational and deductive power of formal logic.

The early theory of computation was developed by logicians such as Alonzo Church and Alan Turing. In the 1930s, Church, Turing, and others developed mathematical schemes for specifying what could be effectively computed. These schemes turned out to be mathematically equivalent to each other, providing support for the thesis that the intuitive concept of effective computability can be identified with well-defined mathematical concepts such as Turing-machine computability. When digital computers became available in the late 1940s and 1950s, the mathematical theory of computability provided a powerful tool for understanding their operations. It is not surprising that, when artificial intelligence began in the mid-1950s, mathematically trained researchers such as John McCarthy took logic to be the most appropriate tool. We shall see, however, that other pioneers such as Allen Newell, Herbert Simon, and Marvin Minsky preferred different approaches.

Representational Power

Modern formal logic has the resources to represent many kinds of deductive inferences. The simplest system of formal logic is propositional logic, in which formulas like "*p*" and "*q*" are used to stand for sentences such as "Paula is in the library" and "Quincy is in the library." Simple formulas can be combined into more complex ones using symbols such as "&" for "and," "v" for "or," and "→" for "if-then." For example, the sentence

If Paula is in the library, then Quincy is in the library.

becomes

$p \rightarrow q.$

Such if-then sentences are called conditionals, consisting of antecedents (the "if" part) and consequents (the "then" part). To express negation, "*not-p*" can be written $\sim p$. From these building blocks we can construct formalizations for complex statements such as "If Paula or Quincy is in the library, then Debra is not," which can be formalized as

$(p \vee q) \rightarrow \sim d.$

Here, "*p*" stands for "Paula is in the library," "*q*" stands for "Quincy is in the library," and "*d*" stands for "Debra is in the library."

More complicated logics have been developed that allow different kinds of propositional operators. Modal logic adds operators for necessity and possibility, so that we can represent statements such as "It is possible that Paula is in the library." Epistemic logic adds operators for knowledge and belief, so that *Kp* represents "It is known that *p*." Deontic logic represents moral ideas such as that *p* is permissible or forbidden.

Propositional logic requires treating statements such as "Paula is a student" as an indivisible whole, but predicate logic allows us to break them down. Predicate calculus distinguishes between predicates such as "is a student" and constants referring to such individuals as Paula or Quincy. In the version of predicate calculus usually taught in philosophy courses, "Paula is a student" is formalized as "*S(p)*," where "*p*" now stands for Paula rather than a whole proposition. Computer scientists tend to express this more mnemonically as "is-student (paula)." In addition to simple properties, predicates can be used to express relations between two or more things. For example, "Paula takes Philosophy 256" becomes: "takes (Paula, Phil256)."

Predicate calculus can formalize sentences with quantifiers such as "all" and "some" by using variables such as "x" and "y." For example, "All students are overworked" becomes

(for-all x) (student(x) → overworked (x)).

Literally, this says "For any x, if x is a student, then x is overworked," which is equivalent to saying that all students are overworked. The sentence "Students who take courses get credit for them" could be formalized as

(for-all x) (for-all y) [(student (x) & course (y) & take (x, y)) → get-credit-for (x, y)]

This looks complicated, but what it is saying in English is "For any x and y, if x is a student, y is a course, and x takes y, then x gets credit for y."

Readers whose interest lies predominantly in human psychology might now be asking, why are you throwing these mathematical symbols at me? The answer is that some rudiments of formal logic are required for understanding much current work in cognitive science, including some proposals about how humans do deduction. At a minimum, we have to notice that people can comprehend such statements as "Students who pass courses get credit for them" and use them to make inferences. Predicate logic, unlike some other approaches to representation we will discuss, has sufficient representational power to handle this example.

Although predicate logic is useful for many purposes, it has limitations that become obvious as soon as we try to translate a natural language text. For example, try to put the last paragraph into logical form. Its first sentence includes the word "now," and extending predicate logic to deal with time is not an easy matter. It also contains the word "you," which the reader can figure out refers to Paul Thagard, the author of this book, but it is not obvious how to express this in logic. Moreover, the structure of this sentence includes the relation "asks," which involves both an asker and the proposition that is asked, so that we need to be able to embed a proposition within a proposition, which is not naturally done in the usual formalism for predicate logic. If translation from language to logical formalism were easier, we could have greater confidence that formal logic captures everything that is necessary for mental representation.

Propositional and predicate logic work well for making assertions that take statements to be true or false, but they provide no means to deal with uncertainty, as in "Paula is probably in the library." For such assertions, formal logic can be supplemented with probability theory, which assigns numbers between 0 and 1 to propositions. We can then write "$P(p) = 0.7$" to symbolize that the probability that Paula is in the library is 0.7.

Computational Power

Representations by themselves do nothing. To support thinking, there must be operations on the representations. To derive a conclusion in logic, we apply *rules of inference* to a set of premises. Two of the most common rules of inference make it possible to draw conclusions using conditionals (if-then sentences):

Modus ponens

$p \rightarrow q$

p

Therefore, q.

Modus tollens

$p \rightarrow q$

not-q

Therefore, *not-p*.

From the conditional "If Paula is in the library, then Quincy is in the library" and the information that Paula is in the library, modus ponens enables us to infer that Quincy is in the library. From the information that Quincy is not in the library, it follows by modus tollens that Paula is not in the library.

In predicate logic, there are rules of inference for dealing with the quantifiers "all" and "some." For example, the rule of universal instantiation allows the derivation of an instance from a general statement, licensing the inference from (for-all x)(cool (x)) to cool(Paula), that is, from "Everything is cool" to "Paula is cool." A more complicated application applies the generalization that all students are overworked: (for-all x) (student(x) → overworked (x)). Applying this to Mary, we get the conclusion that if Mary is a student, she is overworked: student (Mary) → overworked (Mary).

Abstract rules of inference such as modus ponens are not in themselves processing operations. To produce computations, they need to be part of a human or machine system that can apply them to sentences with the appropriate logical form. From a logical perspective, deductive reasoning consists of applying formal inference rules that consider only the logical form of the premises.

Problem Solving

Planning Many planning problems are open to solutions that employ logical deduction. Suppose Tiffany is a student who wants to get a degree in psychology. Her college or university catalog tells her that she needs to take ten psychology courses, including two statistics courses, Statistics 1 and Statistics 2. The first of these is a prerequisite for the other, and the second is a prerequisite for Research Methods, which is also required for the degree. From the general description in the catalog, Tiffany can infer by the inference rule universal instantiation the conditionals that apply to her, including

take (Tiffany, Stat1) → can-take (Tiffany, Stat2)

can-take (Tiffany, Stat2) & open (Stat2) → take (Tiffany, Stat2)

take (Tiffany, Stat2) → can-take (Tiffany, RM)

can-take (Tiffany, RM) & open (RM) → take (Tiffany, RM)

take (Tiffany, RM) & take (Tiffany, Stat1) & take (Tiffany, Stat2) & take (Tiffany, seven-other-courses) → graduate-with (Tiffany, psychology-degree).

The last conditional is a somewhat awkward formalization of the statement that if Tiffany takes Research Methods, the two statistics courses, and seven other courses, then she can graduate with a psychology degree. Tiffany can use these conditionals and the inference rule modus ponens to derive a plan, which in logical terms is a deduction from her initial state, where she has taken no psychology courses, to the goal state, where she graduates. Tiffany can construct the deductive plan that she can take Statistics 1, and then Statistics 2, and then Research Methods, and then seven other courses, and finally graduate with a psychology degree.

For planning to be computationally realizable, deduction must be more constrained than the general set of inference rules found in formal logic. For example, propositional logic contains the following conjunction rule:

niques have been developed for keeping probabilistic reasoning computationally tractable (Neapolitan 1990; Pearl 1988, 2000). A different issue treated below is whether people's normal decision making uses probabilities.

Explanation Whereas in a planning problem you are trying to figure out how to accomplish a goal, in an explanation you are trying to understand why something happened. Suppose that Sarah was expecting to meet Frank at the student bar, but he did not show up. She would naturally try to generate an explanation for his absence. Like plans, explanations can sometimes be viewed as logical deductions: you can try to deduce what you want to explain for what you know. Someone might tell Sarah that Frank is studying for an exam, and that whenever he studies he forgets about social engagements. From this information Sarah can deductively explain why Frank did not show up.

The view that explanations are logical deductions was developed and defended by the philosopher of science Carl Hempel (1965). Especially in mathematical areas of science such as physics, explanations can be described as logical deductions. We shall see in later chapters, however, that not all explanations are deductive. Moreover, not all deductions are explanations. For example, we can deduce the height of a flagpole from information about its shadow along with trigonometry and laws of optics, but it seems odd to say that the length of a flagpole's shadow explains the flagpole's height.

In rare cases, the reason Frank did not show up could be deduced—for example if he is a rigid person who misses appointments if and only if he is sick. Sarah could then apply modus ponens: if Frank misses an appointment, he is sick; Frank missed an appointment; therefore Frank is sick. But normally there will more than one explanation available. Just like a planner constructing multiple paths to a goal, Sarah might be able to construct several deductive explanations based on conditionals such as

If Frank is sick, then he will not arrive.

If Frank has had a car accident, then he will not arrive.

If Frank has fallen in love with someone else, then he will not arrive.

If Sarah did not actually know that Frank is sick, or that he has had a car accident, or that he has fallen in love, then she would not immediately be able to deduce that he will not arrive. But the three conditionals just given can be used to form hypotheses about what happened: maybe he's sick, or maybe he had a car accident, or maybe he has fallen in love. This kind of inference, where you form a hypothesis in order to generate an explanation, was called *abduction* by the nineteenth-century American philosopher Charles Peirce (1992). Sarah may abduce that Frank is sick because this hypothesis, in conjunction with the rule that if Frank is sick he will not arrive, allows her to deductively explain why Frank did not arrive. Abductive inference is a risky but powerful kind of learning.

Learning

Intelligent systems should be able not only to solve various kinds of problems but also to use experience to improve their performance. How can we improve planning, decision making, and explanation? Little work has been done within the logical approach on direct improvements to problem solving, but logical representations are useful for describing some kinds of learning programs.

Consider the learning problem faced by students first arriving on campus. They usually start with little knowledge about the kinds of course offerings available or the kinds of people they will meet. But they quickly accumulate information about particular examples of courses or types of people and naturally proceed to make inductive *generalizations* about them. Crude generalizations might include such statements as that philosophy classes are fun (or boring, as the case may be) and that statistics classes are demanding. These generalizations are inductive in that they involve uncertainty, a leap from what is definitely known to what is at best probable. Students who have taken two philosophy classes might be prepared to generalize from information that could be expressed in logical form as follows:

fun (Phil100)

fun (Phil200)

Therefore, (for-all x) (philosophy-course (x) $\rightarrow$ fun (x)).

The conclusion is that all philosophy courses are fun. But it is obviously possible that these two courses might be fun whereas other philosophy courses (e.g., Philosophy of Basket Weaving) are boring.

Computer programs for inductive generalization do not always use logical representations for input. One of the most widely used learning programs is Quinlan's (1983) ID3 program. It can be classified as within the logical approach because it uses probabilities to form generalizations from sets of instances. For example, it could be given a sample of students from different sections of a university along with a description of their traits. It could then start to form generalizations concerning how students from such areas as arts, sciences, and engineering differ with respect to personal, social, and intellectual characteristics.

Like inductive generalization, but unlike deduction, abduction is obviously a very risky sort of inference. There may be all sorts of reasons unknown to Sarah that explain why Frank did not show up for an appointment with her. But abduction is indispensable in science and everyday life, whether paleontologists are trying to generate explanations of why the dinosaurs became extinct or students are trying to understand their friends' behavior. Since abduction's purpose is to generate explanations, and explanations can sometimes be understood in terms of logical deduction, it is natural to treat abduction within a logical framework (e.g., Konolige 1992). Later chapters describe alternative ways of thinking about abduction.

Sarah does not want to find just *some* explanation of why Frank did not arrive, she wants to find the *best* explanation. From a logical perspective, assessing the best explanation involves probabilities. Sarah will want to be able to assess the conditional probability of Frank being sick, given that he did not arrive, as well as the conditional probabilities of all the other hypotheses. A theorem of the probability calculus, Bayes's theorem, is potentially very useful. In words, it says that the probability of a hypothesis given the evidence is equal to the result of multiplying the prior probability of the hypothesis, $P(h)$, by the probability of the evidence given the hypothesis, all divided by the probability of the evidence. For Sarah, the prior probability that Frank is sick is her estimate of how likely he is to be sick in general, without considering his failure to arrive. To apply Bayes's theorem, she also needs to consider the probability of his failure to arrive, assuming he is sick. Probabilistic approaches to the problem of how to choose explanatory hypotheses have been popular in both artificial intelligence (Pearl 1988, 2000) and philosophy (Howson and Urbach 1989; Glymour 2001). But alternative approaches are available, as we will see in chapter 7.

The term "induction" can be very confusing, since it has both a broad and a narrow sense. The broad sense covers any inference that, unlike deduction, introduces uncertainty. The narrow sense covers only inductive generalization, in which general conclusions are reached from particular examples. Abduction (forming explanatory hypotheses) is induction in the broad sense but not in the narrow one. My practice in this book is to use "learning" for the broad sense of induction and "inductive generalization" for the narrow sense. Additional computational accounts of learning will be encountered in later chapters.

Language

Linguists have sometimes taken formal logic to be a natural tool for understanding the structure of language. There are even two editions of a book called *Everything That Linguists Have Always Wanted to Know about Logic—But Were Ashamed to Ask* (McCawley 1993). The philosopher Richard Montague (1974) contended that there are no important theoretical differences between natural languages and the artificial languages of logicians. Most linguists and psychologists would disagree with this claim, however, and formal logic has played a minor role in the understanding of human language. Stabler (1992) has used logic to formalize some of Chomsky's recent ideas about language, which include the postulation of a level of "logical form" at which meaning is most explicitly represented (Chomsky 1980). Later chapters discuss how other kinds of representation, particularly rules and concepts, have been used to describe and explain human use of language.

Psychological Plausibility

Historically, logicians have disagreed about the mutual relevance of logic and psychology. Some early writers on logic, such as John Stuart Mill, saw an intimate connection between human psychology and logic, which was construed as the art and science of reasoning. In contrast, the founders of modern formal logic, Gottlob Frege and Charles Peirce, emphatically distanced their work from psychology. Today, we can distinguish at least three positions concerning the relations and relative merits of formal logic and psychology:

1. Formal logic is an important part of human reasoning.
2. Formal logic is only distantly related to human reasoning, but the distance does not matter, since the role of logic in philosophy and artificial

intelligence is to provide a mathematical analysis of what constitutes optimal reasoning.

3. Formal logic is only distantly related to human reasoning, so cognitive science should pursue other approaches.

The first position is advocated by a few psychologists who have provided experimental evidence that people use rules like modus ponens. The second position is popular among philosophers and artificial intelligence researchers who prefer formal approaches. The third position is probably now the dominant view in psychology, but is less popular in philosophy and artificial intelligence.

The psychologists who have most aggressively defended the first position are Martin Braine (1978; Braine and O'Brien 1998) and Lance Rips (1983, 1986, 1994). Rips (1986, 279) lists several kinds of psychological evidence for mental logic. Theories of mental logic successfully predict the validity judgments that subjects give for a fairly wide range of propositional arguments. For example, people recognize as valid arguments that have the same form as modus ponens, but reject arguments of the form "If A, then C; C, therefore, A." Theories of mental logic also account for reaction times and help make sense of what subjects say when they think aloud about validity decisions.

Nevertheless, other kinds of experiments have made many psychologists skeptical about mental logic. The best-known experimental technique uses Wason's (1966) selection task, in which subjects are informed that they will be shown cards that have numbers on one side and letters on the other. They are then given a rule such as *If a card has an A on one side, then it has a 4 on the other.* The subjects are then shown four cards and asked to indicate exactly which cards must be turned over to determine whether the rule holds. They can be given, for example, the four cards shown in figure 2.1. Then they must decide which of these cards should be turned over. Most people realize that it is necessary to turn the A over to check whether it has a 4 on the other side. This can be interpreted as an application of modus ponens, since the rule *If A then 4* combined with the premise A suggests checking to see if there is a 4. On the other hand, a great many people neglect to check the 7, failing to realize that if this card has an A on the other side, it refutes the rule in question. Recognition that the card with a 7 needs to be turned over requires an appreciation of modus tollens:" If A then 4; 7 means not-4; so not-A is required for the rule to hold." Some

Johnson-Laird argues that the comparative difficulties that people have with different kinds of inferences of this sort correspond exactly to the complexity of different kinds of models that have to be constructed. Rips (1994) and O'Brien, Braine, and Yang (1994) have responded with arguments that mental logic accounts for the psychological evidence about deductive inference better than mental models do. But mental model theory has been applied to many kinds of human thinking, including causal reasoning (Goldvarg and Johnson-Laird 2001).

Just as Johnson-Laird has challenged the relevance of formal logic to human deductive reasoning, psychologists have done experiments that suggest that human inductive reasoning may not have much to do with probability theory. Tversky and Kahneman (1983), for example, have shown that people sometimes violate the rule that the probability of a conjunction will also be less than or equal to the probability of one its conjuncts, $P(p \& q) \leq P(p)$. Suppose you are told that Frank likes to read a lot of serious literature, attend foreign movies, and discuss world politics. You are then asked to estimate the probability that Frank is college educated, that Frank is a carpenter, and that Frank is a college-educated carpenter. Not surprisingly, people in experiments like this one tend to judge it to be more probable that Frank is college educated than that he is a carpenter, but they often violate probability theory by judging it to be more likely that Frank is a college-educated carpenter than that he is a carpenter. When people approach such examples, they seem to employ a kind of matching process that judges the degree of fit between the description of the individual and their stereotypes such as college-educated and carpenter (see chapter 4). Numerous other instances have been found where people's inductive reasoning appears to be based on something other than formal rules of probability theory (Kahneman, Slovic, and Tversky 1982; Gilovich, Griffin, and Kahneman 2002). However, just as Rips and others have defended mental deductive logic, some psychologists have offered different interpretations of Tversky and Kahneman's results that are consistent with the view that people employ probabilistic reasoning (Gigerenzer, Hoffrage, and Kleinbölting 1991; Gigerenzer 2000).

One open possibility is that mental logic may give an appropriate account of some narrow kinds of human reasoning such as applying modus ponens, whereas more vivid representations such as mental models are needed to account for more complex kinds of human reasoning such as that involving "all" and "some." It is at least obvious that the logical approach is not the only possible way of understanding human thinking, and various alternatives are discussed in the chapters to come. Of course, philosophers and artificial intelligence researchers not interested in psychology can maintain that whether or not people use logic in their thinking is less important than developing formal logical models of how people and other intelligent systems *should* think. What they risk missing is the appreciation that human intelligence and the kind of machine intelligence we want to build may rest on representational structures and computational processes that differ markedly from those that logic affords.

Neurological Plausibility

Until recently, little was known about the neurological plausibility of formal logic. Metaphorically, every synaptic connection between neurons looks like a miniature inference using modus ponens: if neuron 1 fires, then neuron 2 fires. Neuron 1 fires, so neuron 2 fires. However, it is obvious that single neurons do not represent whole propositions, and how groups of neurons perform inferences is unknown. However, it is now possible to investigate at a larger scale how the brain performs deductive reasoning. Brain scanning experiments are being used to determine whether people perform deductions using just the left half of the their brains, as suggested by the mental logic view that deduction is formal and independent of content. The alternative hypothesis is that people perform deductions using the right half of their brains, as suggested by the mental models view that deduction requires regions in the right hemisphere of the brain that involve spatial reasoning (Wharton and Grafman 1998). (See chapter 8 for an introduction to how brain scanning is used to identify neural correlates of different kinds of thinking.)

Goel et al. (1998) used brain scans to identify regions involved in reasoning tasks such as syllogisms. They found no significant right-hemisphere activation, suggesting that deductive reasoning is purely linguistic as implied by the mental logic theory. However, Kroger, Cohen, and Johnson-Laird (forthcoming) compared brain regions involved in logical reasoning and mathematical calculation and found that parts of the right half of the brain were more active in reasoning than in calculation. They judged that their results are incompatible with a purely linguistic cognitive science

Mind
Introduction to Cognitive Science
Second Edition
Paul Thagard

Cognitive science approaches the study of mind and intelligence from an interdisciplinary perspective, working at the intersection of philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology. With *Mind,* Paul Thagard offers an introduction to this interdisciplinary field for readers who come to the subject with very different backgrounds. It is suitable for classroom use by students with interests ranging from computer science and engineering to psychology and philosophy.

Thagard's systematic descriptions and evaluations of the main theories of mental representation advanced by cognitive scientists allow students to see that there are many complementary approaches to the investigation of mind. The fundamental theoretical perspectives he describes include logic, rules, concepts, analogies, images, and connections (artificial neural networks). The discussion of these theories provides an integrated view of the different achievements of the various fields of cognitive science.

This second edition includes substantial revision and new material. Part I, which presents the different theoretical approaches, has been updated in light of recent work in the field. Part II, which treats extensions to cognitive science, has been thoroughly revised, with new chapters added on brains, emotions, and consciousness. Other additions include a list of relevant Web sites at the end of each chapter and a glossary at the end of the book. As in the first edition, each chapter concludes with a summary and suggestions for further reading.

Paul Thagard is Professor of Philosophy, with cross appointments to Psychology and Computer Science, and Director of the Cognitive Science Program at the University of Waterloo. He is the author of *Coherence in Thought and Action* (MIT Press, 2000) and the editor of *Mind Readings: Introductory Selections on Cognitive Science* (MIT Press, 1998).

A Bradford Book

"This little gem of a book has three major virtues. First, it is easy to read and easy to understand. Second, it clearly states the central thesis of cognitive science and precisely lays out the explanatory patterns underlying various theories of cognition. Third, the book is unique in its presentation of the material, arranging it along various types of knowledge representations such as rules, concepts, and images."
—Ashok Goel, College of Computing, Georgia Institute of Technology

"The second edition of *Mind* represents a significant advance for an already excellent book. My enthusiasm for continuing to use Thagard's accessible and consistently informative volume for Berkeley's large Introduction to Cognitive Science course has been fully refreshed, as the updates in the new edition have made it a superb text for undergraduates."
—Michael Ranney, Graduate School of Education, Department of Psychology, and the Institute for Cognitive and Brain Sciences, University of California, Berkeley

The MIT Press
Massachusetts Institute of Technology
Cambridge, Massachusetts 02142
http://mitpress.mit.edu

0-262-70109-X

90000

9 780262 701099