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 examples from later chapters, psychologists have experimentally examined the kinds of mistakes people make in deductive reasoning, the ways that people form and apply concepts, the speed of people thinking with mental images, and the performance of people solving problems using analogies. Our conclusions about how the mind works must be based on more than "common sense" and introspection, since these can give a misleading picture of mental operations, many of which are not consciously accessible. Psychological experiments that carefully approach mental operations from diverse directions are therefore crucial for cognitive science to be scientific.

Although theory without experiment is empty, experiment without theory is blind. To address the crucial questions about the nature of mind, the psychological experiments need to be interpretable within a theoretical framework that postulates mental representations and procedures. One of the best ways of developing theoretical frameworks is by forming and testing computational models intended to be analogous to mental operations. To complement psychological experiments on deductive reasoning, concept formation, mental imagery, and analogical problem solving, researchers have developed computational models that simulate aspects of human performance. Designing, building, and experimenting with computational models is the central method of artificial intelligence (AI), the branch of computer science concerned with intelligent systems. Ideally in cognitive science, computational models and psychological experimentation go hand in hand, but much important work in AI has examined the power of different approaches to knowledge representation in relative isolation from experimental psychology.

Although some linguists do psychological experiments or develop computational models, most currently use different methods. For linguists in the Chomskyan tradition, the main theoretical task is to identify grammatical principles that provide the basic structure of human languages. Identification takes place by noticing subtle differences between grammatical and ungrammatical utterances. In English, for example, the sentences "She hit the ball" and "What do you like?" are grammatical, but "She the hit ball" and "What does you like?" are not. A grammar of English will explain why the former are acceptable but not the latter. Later chapters give additional examples of the theoretical and empirical work performed by linguists in both the Chomskyan tradition and others.

Like cognitive psychologists, neuroscientists often perform controlled experiments, but their observations are very different, since neuroscientists are concerned directly with the nature of the brain. With nonhuman subjects, researchers can insert electrodes and record the firing of individual neurons. With humans for whom this technique would be too invasive, it has become possible in recent years to use magnetic and positronic scanning devices to observe what is happening in different parts of the brain while people are doing various mental tasks. For example, brain scans have identified the regions of the brain involved in mental imagery and word interpretation. Additional evidence about brain functioning is gathered by observing the performance of people whose brains have been damaged in identifiable ways. A stroke, for example, in a part of the brain dedicated to language can produce deficits such as the inability to utter sentences. Like cognitive psychology, neuroscience is often theoretical as well as experimental, and theory development is frequently aided by developing computational models of the behavior of sets of neurons.

Cognitive anthropology expands the examination of human thinking to consider how thought works in different cultural settings. The study of mind should obviously not be restricted to how English speakers think but should consider possible differences in modes of thinking across cultures. Chapters 12 and 13 describe how cognitive science is becoming increasingly aware of the need to view the operations of mind in particular physical and social environments. For cultural anthropologists, the main method is ethnography, which requires living and interacting with members of a culture to a sufficient extent that their social and cognitive systems become apparent. Cognitive anthropologists have investigated, for example, the similarities and differences across cultures in words for colors.

With a few exceptions, philosophers generally do not perform systematic empirical observations or construct computational models. But philosophy remains important to cognitive science because it deals with fundamental issues that underlie the experimental and computational approaches to mind. Abstract issues such as the nature of representation and computation need not be addressed in the everyday practice of psychology or artificial intelligence, but they inevitably arise when researchers think deeply about what they are doing. Philosophy also deals with general questions such as the relation of mind and body and with methodological questions such as the nature of explanations found in cognitive science.

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,

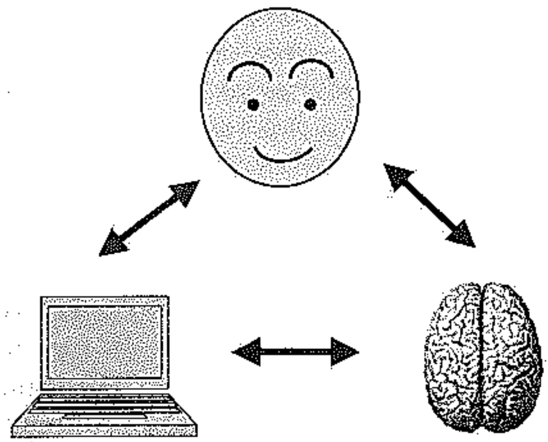

Figure 1.1
Three-way analogy between minds, computers, and brains.

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.

Box 1.1
Criteria for evaluating theories of mental representation.

(1) Representational power
(2) Computational power
  (a) Problem solving
    (i) Planning
    (ii) Decision
    (iii) Explanation
  (b) Learning
  (c) Language
(3) Psychological plausibility
(4) Neurological plausibility
(5) Practical applicability
  (a) Education
  (b) Design
  (c) Intelligent systems
  (d) Mental illness

Evaluating Approaches to Mental Representations

We can now be more specific about what to expect of a theory of mental representation. Box 1.1 lists five complex criteria for evaluating a particular account of the representations and computations that can be claimed to explain thought. Chapters 2–7 use these criteria to evaluate six different approaches to mental representation: logic, rules, concepts, images, cases, and connections (artificial neural networks).

Each of the approaches described in chapters 2–7 proposes a particular kind of representation and a corresponding set of computational procedures. The first criterion, representational power, concerns how much information a particular kind of representation can express. For example, a university calendar urges: "Once admitted to the University, students are advised to preregister for their courses well in advance of the beginning of lectures." Students who take such advice seriously will need to represent it internally in a form that leads to further inferences, such as the conclusion that they should get over to the registrar's office to sign up for next term's courses. We will see that different proposed kinds of mental representation vary greatly in representational power.

Mental representations are important not only for what they express, but especially for what you can do with them. We can evaluate the computational power of an approach to mental representation in terms of how it accounts for three important kinds of high-level thinking. The first is problem solving: a theory of mental representation should be able to explain how people can reason to accomplish their goals. There are at least three kinds of problem solving to be explained: planning, decision making, and explanation. Planning requires a reasoner to figure out how to get from an initial state to a goal state by traversing various intermediate states. Planning problems include mundane issues such as how to get to the airport before your flight leaves, to the sort of exercise students are commonly posed in their textbooks and their exams. In these questions, students are given some information and need to figure out how to calculate the answer. The starting state involves what the student knows and the information in the problem description, and the goal state includes having an answer. The student has to find a solution by constructing a successful sequence of calculations.

In decision making, people are faced with a number of different means for accomplishing their goals and need to select the best one. For example, a student about to graduate may need to choose among looking for a job, going to graduate school, or attending a professional school such as law or business. Such decisions are very difficult, since they require students to identify their goals and figure out which course of action will best accomplish those goals. In planning problems, the task is to find a successful sequence of actions, whereas in decision problems the task is to choose the best plan from among a number of possible actions.

Explanation problems are ones that require people to figure out *why* something happened. They range from mundane questions such as why a friend is late for dinner, to deep scientific questions such as why human language has evolved. Every minimally intelligent human being is capable of planning, decision making, and generating explanations. A cognitive theory must have sufficient computational power to offer possible explanations for how people solve these kinds of problems.

The computational power of a system of representations and procedures is not just a matter of how much the system can compute, but must also 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 sixty applications would take 2^{60} 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

Artificial intelligence on the Web: http://aima.cs.berkeley.edu/ai.html

Biographies of major contributors to cognitive science: http://mechanism.ucsd.edu/~bill/research/ANAUT.html

Cognitive Science dictionary, University of Alberta: http://web.psych.ualberta.ca/~mike/Pearl_Street/Dictionary/dictionary.html

Cognitive Science Society: http://www.cognitivesciencesociety.org/

Cogprints (archive of papers on cognitive science): http://cogprints.ecs.soton.ac.uk/

Dictionary of Philosophy of Mind: http://www.artsci.wustl.edu/~philos/MindDict/

Science Daily (mind and brain news): http://www.sciencedaily.com/news/mind_brain.htm

Yahoo! Cognitive Science page: http://dir.yahoo.com/Science/cognitive_science/

Notes

Discussions of thinking as computation often begin with an abstract model of computation such as the Turing machine, a simple device that consists of a tape and a mechanical head that can write symbols on spaces on the tape. Although it can be proven mathematically that such a machine can in principle do anything that any other computer can, the Turing machine is an excessively abstract analog of thinking, which is much better discussed in terms of higher-level computational ideas such as data structures and algorithms.

For more on explanation schemas and patterns, see Kitcher 1993, Leake 1992, Schank 1986, and Thagard 1999.

2 Logic

Although formal logic has not been the most influential psychological approach to mental representation, there are several reasons for beginning our survey with it. First, many basic ideas about representation and computation have grown out of the logical tradition. Second, many philosophers and artificial intelligence researchers today take logic as central to work on reasoning. Third, logic has substantial representational power that must be matched by other approaches to mental representation that may have more computational efficiency and psychological plausibility.

Formal logic began with the Greek philosopher Aristotle more than two thousand years ago. He systematically studied such inferences as

All students are overworked.
Mary is a student.
So, Mary is overworked.

Such patterns of inference, with two premises and a conclusion, are called *syllogisms*. In addition to cataloging many different kinds of syllogism, Aristotle showed how they can be analyzed purely in terms of their form. For the conclusion in the example to follow from the two premises, it does not matter that the syllogism is about overworked students. We can substitute "sausage" for "student," "orange" for "overworked," and "Marvin" for "Mary," and the conclusion that Marvin is orange follows from the revised premises even if it makes little sense. Aristotle initiated the use of symbols to show the form of the inference:

All *S* are *O*.
M is *S*.
So, *M* is *O*.