

Problem Solving and Reasoning

Kevin Dunbar & Jonathan Fugelsang: To appear in E.E. Smith & S. Kosslyn an introduction to cognitive psychology

Learning Objectives

1. The Nature of Problem Solving
 - 1.1. The Structure of a Problem
 - 1.2. Problem Space Theory
 - 1.3. Strategies and Heuristics
 - 1.4. The Role of Working Memory and Executive Processes
 - 1.5. How Experts Solve Problems
2. Analogical Reasoning
 - 2.1. Using Analogies
 - 2.2. Theories of Analogical Reasoning
 - 2.3. Beyond Working Memory
3. Inductive Reasoning
 - 3.1. General Inductions
 - 3.2. Specific Inductions
 - 3.3. Critical Brain Networks
4. Deductive Reasoning
 - 4.1. Categorical Syllogisms
 - 4.2. Conditional Syllogisms
 - 4.3. Errors in Deductive Thinking

DEBATE: Errors and Evolution

- 4.3.1. Form Errors
- 4.3.2. Content Errors

A CLOSER LOOK: Logic and Belief

- 4.4. Theories of Deductive Reasoning
- 4.5. Linguistic versus Spatial Basis

Revisit and Reflect

You're lucky—it's a quick, relaxing, and pleasant walk from your front door to the psychology building, where an exam awaits you this morning. Past the running track, across the small park, around the administration building, and you're there. You've crossed the park, but now, with the psychology building in sight, you also see barricades and police vehicles and officers. Your way is blocked: what's going on? Ah—you remember—the governor is coming today; protests must be expected. You'll have to find another way. You figure out that you can reach the psych building by first turning away from it and looping around on another path, avoiding the administration building entirely—and you won't lose much time. Success! Now for success on the exam. . . .

Human animal that you are, you have just engaged in problem solving and reasoning. "I think, therefore I am"—with this simple statement the French philosopher René Descartes (1596–1650) famously captured what many believe is the essence of what it is to be human. But what does it mean to think? Philosophers have grappled with this profound question for millennia; in recent decades it has come under the scrutiny of cognitive psychologists and

neuroscientists. **Thinking** is usually considered to be the process of mentally representing some aspects of the world (including ourselves) and transforming these representations so that new representations, useful to our goals, are generated. Thinking is often regarded as a conscious process, in which we are aware of the process of transforming mental representations and can reflect on thought itself. Problem solving and reasoning are two key types of thinking. **Problem solving** encompasses the set of cognitive procedures and thought processes that we apply to reach a goal when we must overcome obstacles to reach that goal. **Reasoning** encompasses the cognitive procedures we use to make inferences from knowledge and draw conclusions. (Reasoning can be part of problem solving.)

These are not isolated cognitive abilities, but build on and give input to each other and to other cognitive processes, including categorization (Chapter 4), imagery (Chapter 11), and decision making (Chapter 9). Furthermore, problem solving and reasoning depend on attention (Chapter 3), working memory (Chapter 6), executive processes (Chapter 7), and language (Chapter 12). Although problem-solving research and reasoning research are often carried out independently of each other, they are obviously related and, as such, we will cover them both in this chapter. This chapter sets out to describe how we come to make complex decisions about the world using a variety of problem-solving and reasoning tools. The successful use of these cognitive tools relies on many brain networks including the frontal and parietal cortices. We specifically address six questions:

1. What is the nature of problem solving?
2. How do we use heuristics or “mental shortcuts” to solve problems?
3. How do we use analogies to solve new problems?
4. What is the difference between induction and deduction?
5. How do our knowledge and beliefs influence “logical” reasoning?
6. How do our brains coordinate the vast amount of processing involved in problem solving and reasoning?

1. THE NATURE OF PROBLEM SOLVING

In the context of cognitive psychology, a *problem* is a situation in which there is no immediately apparent, standard, or routine way of reaching a goal. The determination of the goal and the degree of difficulty you face are both important: if you don’t care whether you get to the psychology building in time for the exam (or at all), or if a satisfactory detour is obvious, no problem faces you. Some problems, such as those that arise between parents and children as they try to get along with one another, may have emotional content; others, such as mathematical problems, are less emotional, but may involve emotions (e.g., anxiety) in certain circumstances, such as when math problems appear on an exam). Research on problem solving generally makes use of problems that are less emotional in nature, but it is thought that the types of strategies we use are similar for both emotional and nonemotional problems.

Problem solving, then, is a process of surmounting obstacles to achieve a goal. Knowing how to get the lights on in your apartment is not a problem when there is

power, but it is a problem when there is a power outage. So routine situations with routine answers are not regarded as problems. There must be novelty or nonstandard solutions that the problem solver must discover. Because problem solving is such an ubiquitous part of our lives, it has become an important area of research that is of both theoretical and practical importance.

The overarching goal of research on problem solving has been to identify the strategies we use when we are confronted by a novel situation and must decide on a course of action. The problem solver must identify the problem, find a way of representing it, and choose a course of action that will make it possible to achieve the goal. Because many different steps and different types of cognitive processes, including memory, attention, and perception, are involved. Many parts of the brain are involved in problem solving.

Research has been conducted on various aspects of problem solving and by various methods, both behavioral and brain based. Investigators have studied scientists such as Albert Einstein (Figure 10–1) as they solved problems (Wertheimer, 1945),



Figure 10–1 A famous problem solver

The Gestalt psychologist Max Wertheimer initiated one of the first investigations of problem solving by scientists in his landmark book, *Productive Thinking* (1945). Wertheimer spent a considerable amount of time corresponding with Albert Einstein, attempting to discover how Einstein generated the concept of relativity. This photograph shows Einstein around the time he developed relativity theory, about a century ago.

molecular biologists as they made discoveries (Dunbar, 2000), artists such as Picasso as they paint (Weisberg, 1993), and participants as they solved puzzles and problems in logic (Greeno, 1978; Klahr, 2000).

1.1. The Structure of a Problem

At its most basic level a problem can be thought of as having three parts. First is the **goal state**: this is where you want to be, at the solution to the problem. Second is the **initial state** (or **start state**): this is where you are now as you face the problem that needs to be solved. Third is the set of operations that you can apply—that is, the actions you can take—to get from the start state to the goal state. A simple example problem that has been frequently used in problem-solving research is the Tower of Hanoi task (Figure 10–2), also described in Chapter 7. In the initial state, all the disks are on peg 1, in increasing order of size from top to bottom. In the goal state, all three disks are on peg 3 in the same size order. The only operation permitted is the moving

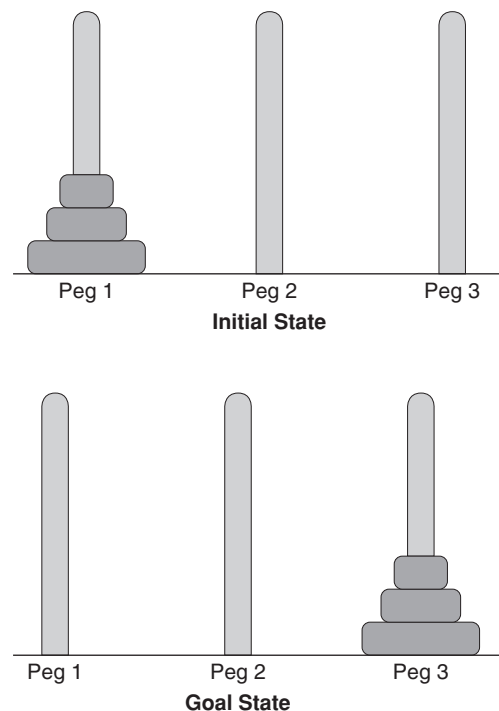


FIGURE 10–2 The Tower of Hanoi

This famous puzzle was invented by the French mathematician Edouard Lucas in 1883 and has been frequently used in problem-solving studies. In the three-disk version of the task, disks must be moved from peg 1 (the initial state) to peg 3 (the goal state) under specific constraints: only one disk may be moved at a time, and a larger disk may never be put on a smaller one.

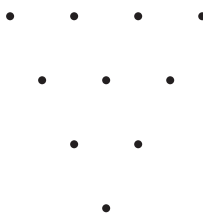
of a disk from one peg to another. The rules: you may move only one disk at a time, and you may not put a larger disk on a smaller one. To add to the challenge, try to solve the problem in your head, as in a typical experiment, with the minimum number of moves.

Try one of the playable electronic versions of the Tower of Hanoi on various Web sites that tally your moves, or you can use paper and pencil. How many moves did it take you to get to the goal state? Is there another set of moves that would have been successful? The Tower of Hanoi is thought to be representative of many types of problems because the solution is not obvious, many different strategies can be used to solve the problem, and many of the same problem-solving strategies are used in real-life problems (such as finding an alternative route to the psychology building).

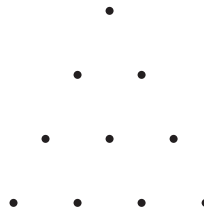
Problems in which, as in the Tower of Hanoi, the initial state and the goal state are clearly defined and the possible moves (and the constraining rules) are known are called **well-defined problems**. Many games, no matter how complicated the rules and how great the number of operations, are well-defined problems; chess is a good example.

But sometimes it is not possible to be sure about the rules, the initial state, the operations, or even the goal of a problem; such a problem is described as **ill defined**. A Hollywood producer looking to win next year's Oscar has an enormous ill-defined problem: Which of the thousands of scripts to choose? Go with a star or a brilliant unknown? What's the likely public reaction? Will production be too costly to do the story justice? Is the idea of the script exciting and original or off-the-wall and distasteful? Is it a tried-and-true formula, or hackneyed and overworked? Many real-world situations present ill-defined problems, in which there are no clearly defined initial or goal states, and the types of operations that are used to reach a goal are not highly constrained by rules. The solution of ill-defined problems presents an additional challenge to the problem solver: finding the *constraints* (i.e., the restrictions on the solution or means whereby it can be achieved) that apply to the particular situation.

A special case of an ill-defined problem is known as the **insight problem**, to which, despite all the unknowns, the answer seems to come all of a sudden in a flash of understanding. Many scientists and artists have reported working on a problem for months or years, making no progress; then, when they are relaxing or their attention is elsewhere, they suddenly arrive at the solution. Insight problems are often used in problem-solving research; anagrams and riddles are examples of insight problems. Try this one: by moving only three dots, invert the triangle



so it looks like this:



1.2. Problem Space Theory

Today, the main theory underlying research on problem solving is *problem space theory*, developed by Allen Newell and Herbert Simon and described in their book, *Human Problem Solving* (1972). Problem solving, in this view, is a search within a **problem space**, which is the set of states, or possible choices, that faces the problem solver at each step in moving from an initial state to a goal state. The problem solver moves through the space from state to state by various operations. Thus, the problem space includes the initial state, the goal state, and all possible intermediate states. The problem space for the three-disk Tower of Hanoi problem, shown in Figure 10–3,

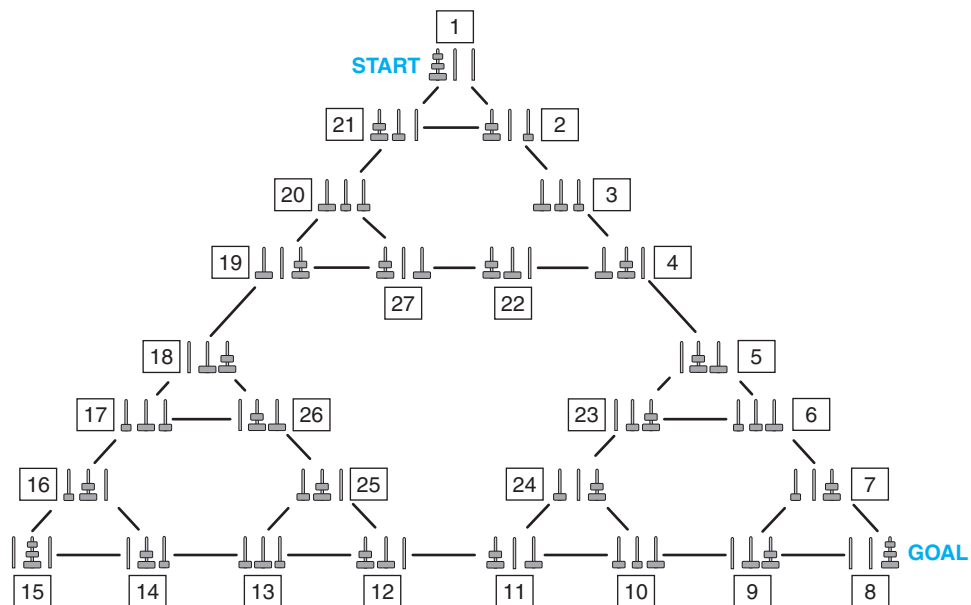


FIGURE 10–3 Problem space for the three-disk version of the Tower of Hanoi task

This figure depicts all possible states in the Tower of Hanoi task—but only seven moves are in fact necessary to get from the start state at 1 to the goal state at 8. The numbers are for identification of the various disk arrangements only; they do not indicate a sequence or an ideal spatial layout between steps. Rather, this figure shows how many operations are required between given states.

consists of 27 possible states. Two things are apparent from the diagram: there are a number of routes that can take you from the initial state to the goal state, and the shortest requires seven operations (this is the path along the right-hand side of the diagram—is it the one you took?). If two disks are added, so the tower is five disks high, the problem space includes 64 possible states.

In Figure 10–3, the states are numbered, to identify them and to provide a total count of states; they do not indicate a sequence. Nor is the spatial layout important. What is important is the distance between states—for instance, that state 5 is one operation away from state 6, 23, or 4. Newell and Simon's original work on problem solving consisted of writing computer programs that calculated the best moves to make on the basis of nonspatial criteria such as the similarity of the selected move to the goal state, and comparing the output of these programs to the moves made by human participants.

Problems such as the Tower of Hanoi and chess are highly constrained, with specific rules and clearly defined initial and goal states. Does problem space theory apply in less constrained domains—that is, can solving more complex problems and more ill-defined problems be characterized as searching a problem space? Much of the more recent research on problem solving has explored complex areas such as science, architecture, and writing (Dunbar, 2002; Klahr, 2000). The results indicate that in complex domains, such as molecular biology and cancer research, problem solvers search in more than one problem space. Thus, when working on scientific problems, for example, the problem space theory of problem solving has been expanded to include spaces such as a hypothesis space to formulate theories, an experiment space to design experiments, and a data space to interpret results (Klahr, 2000). Less constrained domains are of course more akin to situations in the real world than the Tower of Hanoi, and one of the goals of researchers who study problem solving is to understand the different types of problem spaces that we use in different tasks in real-life circumstances.

1.3. Strategies and Heuristics

A surefire way to solve a problem is to use an *algorithm* (discussed in Chapter 9), a set of procedures for solving a given type of problem that will always, sooner or later, produce the correct answer. The rules for finding a square root or performing long division are algorithms; so is a recipe for chocolate cake. But algorithms are often very time consuming, and they make great demands on both working memory and long-term memory.

Researchers using problems such as the Tower of Hanoi task in their studies have found that rather than using algorithms, participants often use specific strategies or heuristics when trying to solve problems. A **heuristic** is a rule of thumb that usually gives the correct answer—but not always. (Heuristics were also discussed in Chapter 9.) A commonly used heuristic in problem solving is “always move toward the goal.” This approach often provides the solution, but sometimes—such as when you're trying to get to the psychology building in the face of a temporary obstruction—it is necessary to move away from a goal and then move back toward

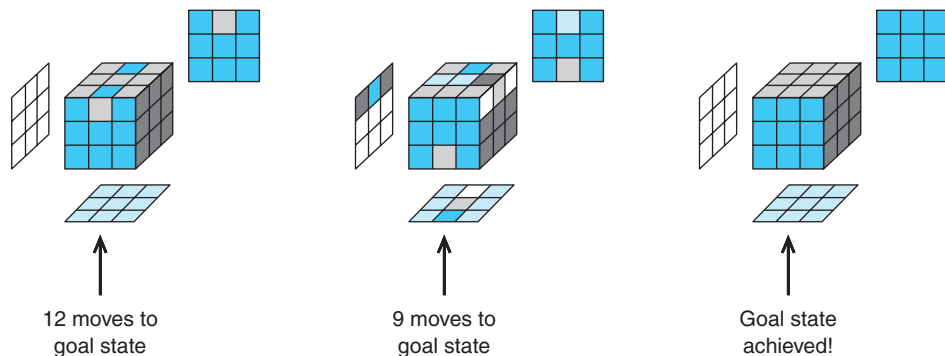


FIGURE 10-4 Solving Rubik's cube

Sometimes you have to move away from the goal in order to achieve it. The three same-color sides shown at the left will all have to be broken before all six sides of the cube can be made uniform.

it to arrive at the correct solution. (Folk wisdom recognizes the truth of this in the saying “The longest way round is the shortest way home.”) The twistable puzzle known as the Rubik's cube is a problem for which the heuristic of making moves that are similar to the goal state will not work. Each side of the cube is made up of nine colored squares (six colors are used). In the goal state, all the squares on a side are the same color. (There are playable electronic versions of Rubik's cube on the Web.) The cube is an interesting problem to examine because many steps are required to solve it, and because it is often necessary to move away from the goal state before finally achieving it. In the initial state of play shown in Figure 10-4, for example, when you are 12 moves from the goal, three of the six sides are of uniform color; but to arrive at the goal from this configuration, all six sides will have to be scrambled before you achieve success in 9 more moves.

Other commonly used problem-solving heuristics are *random search*, *hill climbing*, and *means-ends analysis*. The simplest and cognitively least demanding of these three is **random search** (also known as *generate and test*), which is essentially a process of trial and error: the problem solver randomly picks a move and tests to see whether the goal state is achieved. Researchers in the behaviorist tradition, among them E. L. Thorndike (1874–1949), proposed that animals and people solve problems by trial and error and randomly select moves until they achieve the goal state. Although random search appears to be a really inefficient and often unsuccessful strategy, we frequently resort to it when everything else has failed. It is hard to resist randomly pressing keys when your computer freezes and none of the known solutions work—and *sometimes* it does the trick. Random search is a fall-back heuristic that we use when other heuristics either do not work or are cognitively too demanding.

More often, we use other, more knowledge-dependent heuristics, such as **hill climbing**, in which the problem solver looks one move ahead and chooses the move that most closely resembles the goal state. As in climbing an actual hill, each step seems to move you closer to the goal; but if you are in thick fog (as you are

figuratively if you can see only one move ahead), you may end up at the top of a small rise and not the big hill that you are trying to climb. What happens if you apply hill climbing to the three-disk Tower of Hanoi task? Look at Figure 10–3 again and imagine yourself at state 5. Your only possible choices are state 23 and state 6. State 23 more closely resembles the goal state than does state 6—there are more disks on peg 3 in state 23 than in state 6—so if you’re using hill climbing you’d move to state 23. But state 23 is actually further away from the goal state than state 6, so hill climbing in this case is not an effective problem-solving strategy. Hill climbing is often a more reliable heuristic than random search, but it can lead the problem solver astray.

A classic problem that often leads participants to use the hill-climbing strategy is the water jug task (Atwood & Polson, 1976; Colvin et al., 2001). You have three water jugs of different capacity—8 ounces, 5 ounces, and 3 ounces. In the initial state the 8-ounce jug is full, the other two are empty. Your task is to transfer water from the large jug to the others so that you end up with 4 ounces of water in the large jug, 4 ounces of water in the medium-size jug, and the small jug is empty—this is the goal state. Moreover, whenever you pour water into another jug, you must fill it completely. The task and the problem space for possible moves are shown in Figure 10–5.

If you use the hill climbing strategy you might select a move that will put water in the large and medium jugs in amounts close to the goal state. So you pour water from the large jug into the medium one, and now you have 3 ounces of water in the large jug and 5 ounces of water in the medium jug. This is certainly closer to the goal than was the initial state, but now hill climbing fails you. Whatever move you take next will take you, as it were, *down* the hill, farther away from the goal state than your current state. Often at this point participants attempt to start the task again, this time moving farther away from the goal state in their next move by pouring water from the large jar into the small one. Thus, the hill climbing strategy, like the random search strategy, is an easy but often inefficient way of trying to solve a problem.

A more demanding, but more successful, strategy is **means–ends analysis**, in which the problem is broken into subproblems. If a subproblem at the first stage of analysis is not solvable, then the problem can be further broken into other subproblems until a soluble subproblem is found. Applying means–ends analysis to the three-disk Tower of Hanoi task, you would define your main goal as getting all three disks on peg 3 (look again at Figure 10–2). To accomplish this you must get the large disk on peg 3, but in the initial state you can’t move it: the medium disk is in the way. Try a further subgoal: get the medium disk out of the way; but again the move is not possible, this time because the small disk is in the way. The next subgoal? Move the small disk. This can be done: nothing blocks the small disk. But where should it be moved—to peg 2 or to peg 3? You must look ahead to the consequences of each of those moves. If you move the small disk to peg 3 (state 2 in Figure 10–3), then the medium disk can be moved to peg 2 (state 3), and so on. What you are doing in this process is setting up goals and subgoals until the entire problem is solved.

Until neuroimaging techniques became available in the late 1980s, three main methods were used to understand how we solve problems. First, and most obvious,

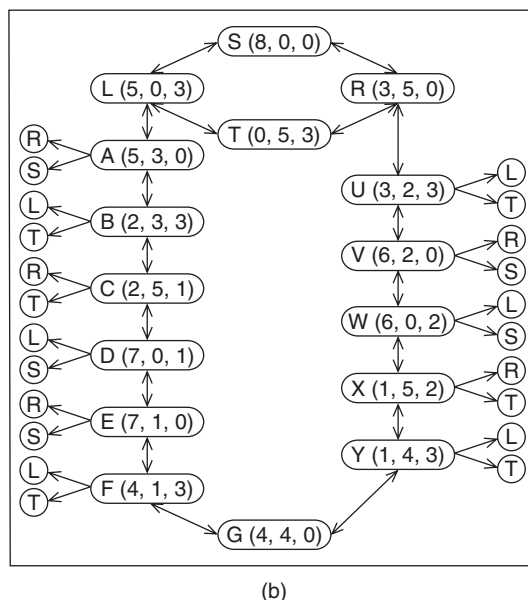
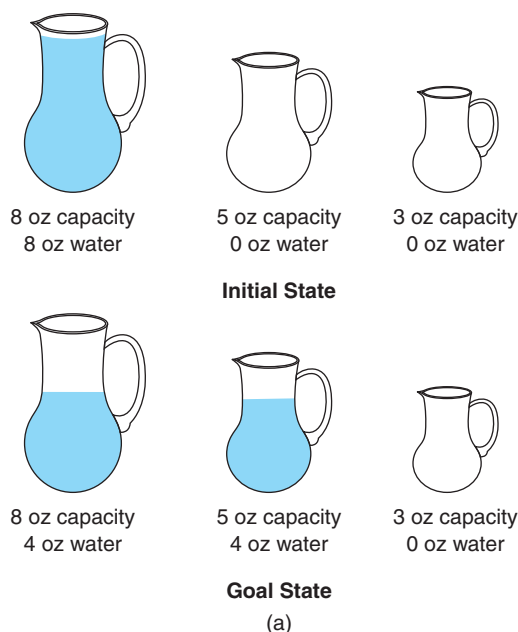


FIGURE 10-5 The water jug task

(a) Initial and goal states for the water jug task. (b) The problem space: the possible moves between start and goal. The notation gives the amount of water in each container as shown left to right: thus the start state is indicated as $S(8, 0, 0)$ and the goal as $G(4, 4, 0)$. The letters to the left of the parentheses identify the current state; the letters that follow the arrows from a current state name the possible following states, for example, from current state $B(2, 3, 3)$, possible moves are to $L(5, 0, 3)$ or to $T(0, 5, 3)$.

(Mary Kathryn Colvin, Kevin Dunbar & Jordan Grafman, The Effects of Frontal Lobe Lesions on Goal Achievement in the Water Jug Task, *Journal of Cognitive Neuroscience*, 13:8. (November 15, 2001), pp. 1129–1147. © 2001 by the Massachusetts Institute of Technology. Reprinted by permission.)

is to record problem-solving behavior. Researchers can record, in sequence, every move that a problem solver makes in the course of arriving at a solution. Using this method, in each case researchers can chart how long it takes to solve a given problem and the different types of moves that problem solvers take.

Another behavioral approach, developed in the 1970s, is **verbal protocol analysis**, which is the analysis of the thought process of the problem solver as described aloud by the solver in the course of working on the problem (Ericsson & Simon, 1984). The problem solver is recorded either in video or audio. Researchers then transcribe the protocol and analyze the transcript to determine the ways in which the problem solver represented the problem and the sequence of steps employed, to infer the participant's problem space.

A third approach looks to computers, building programs that embody the strategy that people presumably use to solve a problem, and then comparing the computer output to the moves made by a person. Computer models make it necessary to state explicitly every step in the problem-solving process, and have been used to simulate all three problem-solving heuristics: random search, hill climbing, and means-ends analysis. Many researchers, particularly the late Herbert Simon (himself a Nobel prize winner) and his colleagues at Carnegie Mellon University, have used all three methods of investigation to understand problem solving. The problems that they have explored range from the Tower of Hanoi to the discovery of the urea cycle (by an analysis of the laboratory notebooks of the discoverer, the Nobel prize-winning biochemist Hans Krebs) (Kulkarni & Simon, 1988; Newell & Simon, 1972).

1.4. The Role of Working Memory and Executive Processes

What regions of the brain might be involved in problem solving? Using PET, ERP, and fMRI techniques (see Chapter 1) with both brain-damaged and unimpaired participants, investigators have sought to discover how cognitive procedures used in problem solving are represented in the brain. Consider the different processes that are used in a task such as the Tower of Hanoi. The problem solver must determine what operations are required to reach the goal state. This requires keeping in mind the goals and subgoals. Such a task places considerable demands on working memory, and so we would expect significant activations in areas typically involved in working memory (for example, dorsolateral prefrontal cortex; see Chapter 6). This prediction is confirmed by neuroimaging research. Using a modified version of the Tower of Hanoi task with healthy participants, investigators found that brain activation in the right dorsolateral prefrontal cortex, bilateral parietal cortex, and bilateral premotor cortex increased as the task became more complex (Finchan et al., 2002) (Figure 10–6). These regions have been strongly implicated in working memory and executive processes (see Chapter 7), thus underscoring the strong relationship among problem solving, working memory, and executive processing.

Several studies have also examined problem solving in patients with localized brain damage. Patients with frontal-lobe lesions have great difficulty using means-ends analysis to solve the Tower of Hanoi problem (Goel & Grafman, 1995). Frontal patients also have difficulty in applying the hill climbing heuristic to the

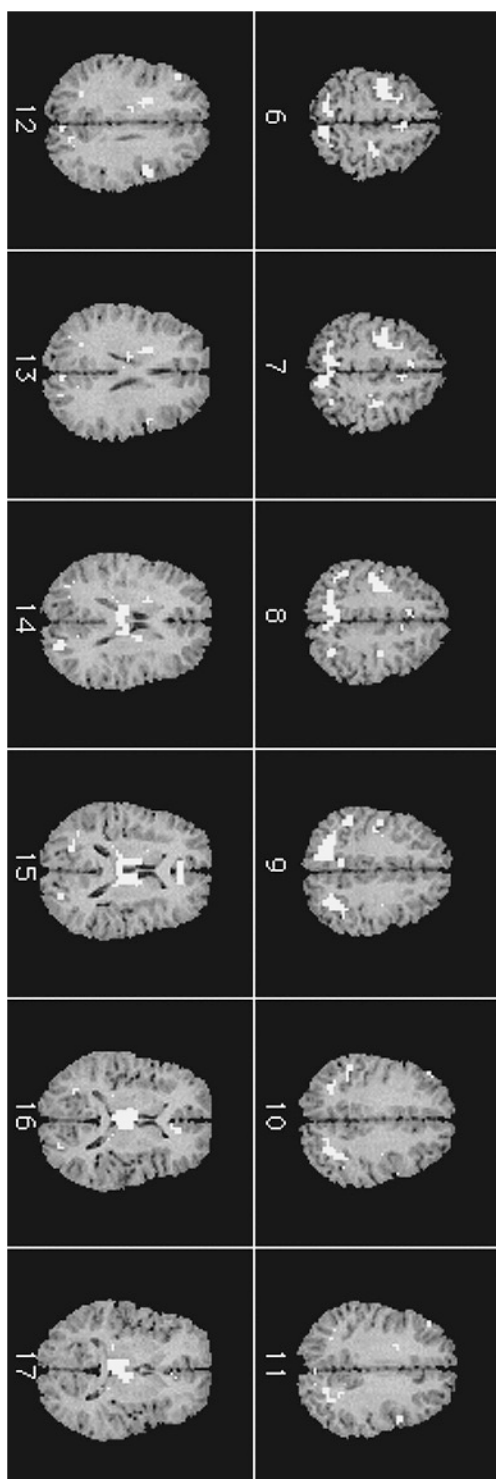


FIGURE 10-6 fMRI data on goal-directed processing

Goal-directed processing is distributed among prefrontal cortex, parietal cortex, cingulate gyrus, and subcortical structures (right caudate nucleus and thalamus). The slices of the brain shown are ordered from top (#6) to bottom (#17); The left side of each image is the right side of the brain and the right side of each image is the left side of the brain. Activation increased as the problems became more complex.

(Fincham, J. M., Carter, C. S., van Veen, V., Stenger, V. A. & Anderson, J. R. (2002). Neural mechanisms of planning: a computational analysis using event-related fMRI. *Proceedings of the National Academy of Sciences*, 99, 3346–3351. Reprinted by permission.)

water jug task: they find it hard to remember the moves that they have already made and cannot learn the moves that they must avoid (Colvin et al., 2001). Because these patients do not learn the moves to avoid, they cycle through the same moves again and again, never coming closer to a solution. From these studies, it is clear that the frontal lobes are involved in the long-term storage of information during problem solving, in working memory during problem solving, and in the execution of plans to solve a problem.

1.5. How Experts Solve Problems

A considerable amount has also been learned about problem solving by comparing experts and novices. Experts know more than do novices in their field of expertise, and presumably that's a big help in solving problems in their field. An interesting question is whether expertise provides more than simply additional information; that is, do experts have specialized problem-solving strategies that novices don't? The answer, apparently, is yes.

The first additional information is experts' organization of knowledge in their field, which is different from that of novices. Novices in a field often organize concepts in terms of surface similarity, whereas experts organize their knowledge in terms of deeper abstract principles. In a classic study of expertise (Chi et al., 1981), beginning and advanced physics students were presented with different types of physics problems and asked to sort the problems into similar categories (Figure 10–7). The novice physics students grouped problems that involved the same physical characteristics, putting tasks with blocks in one category and tasks involving springs in another. The graduate physics students performed the task very differently: they sorted the problems in terms of physical concepts, such as conservation of energy.

A second sort of additional information used by expert problem solvers is seen in encoding: experts and novices do not encode information in the same way. In an

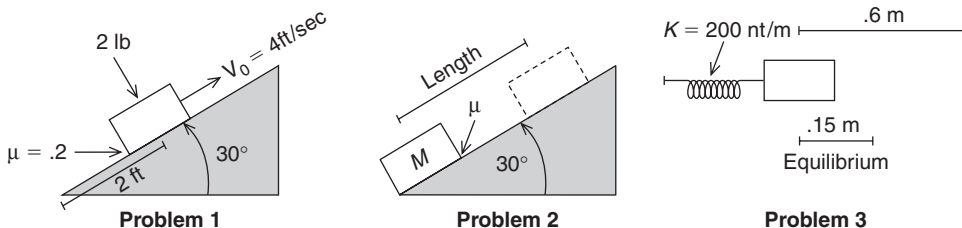


FIGURE 10–7 Which problem is the odd man out?

"Novices" (physics undergraduates) grouped problems 1 and 2, based on the superficial visual similarity to "blocks" on an "inclined plane" in both drawings. Graduate physics students, however, cut to the chase and recognized that problems 1 and 3 are the ones to be grouped: both deal with the conservation of energy.

(Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorization of physics problems by experts and novices. *Cognitive Science*, 5, 121–152. Reprinted by permission.)

important study of expertise, chess players of various proficiency were presented, for 5 seconds, with a chessboard with chess pieces in various positions; then the board was removed and the players were asked to reconstruct the arrangement of the pieces on a second board (Chase & Simon, 1973). The investigators found that experts were better than novices at reconstructing the arrangement only when the original configuration of pieces was taken from a real game; experts were no better than novices when the chess pieces were arranged randomly. Also, chess experts were able to recall positions from a real game much better than chess pieces that had been arranged randomly. Beginning chess players, however, did not perform any better on real positions than on a random layout. The investigators argued that chess experts performed better because they were able to encode the position of many pieces as a unit or “chunk”—and they were able to do this because arrangements from a real game “make sense” in terms of the rules of chess, whereas random layouts do not. This ability to chunk information is a hallmark of experts, from architects to zoologists. Experts chunk information and can access related chunks of knowledge, thus making their problem solving more efficient.

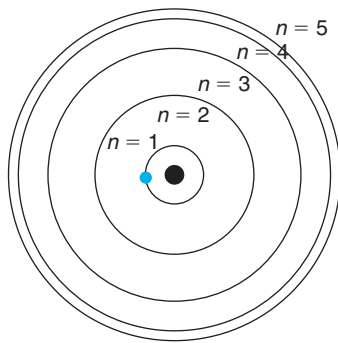
Yet another difference between expert and novice problem solvers involves the direction of the search and the problem space. Experts tend to employ a **forward search**, that is, they search from the initial state to the goal. An experienced physician, for example, works from the symptoms to diagnosis. A medical student generally uses a **backward search**, from the goal of diagnosis to the symptoms that constitute the initial state (Arocha & Patel, 1995). It is also possible for both people and computers to work simultaneously from both the goal state (backward search) and the initial state (forward search). Many computer programs such as the IBM chess program Deep Blue use a combination of search strategies to choose a move.

✓ Comprehension Check:

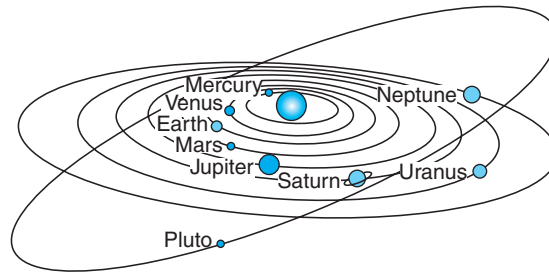
1. What is the difference between using a hill climbing versus a means–ends problem-solving strategy?
2. What is a problem space?

2. ANALOGICAL REASONING

People do not always try to solve a problem by using the detailed heuristics described in the preceding section. Instead people sometimes try to think of a solution to a similar problem. If you go to use your laptop, which has a physical lock, but you have lost your key, what do you do? Do you try a random search, or do you notice the similarity between the computer lock and a bicycle lock. Then you remember that bicycle thieves are using ballpoint pen tubes to open bicycle locks. Aha, you think, I’ll try a pen—and presto, you have your computer unlocked. This is *analogical reasoning*. In analogical reasoning, rather than beginning from scratch using heuristics such as means–ends analysis, you try to think of a problem with similar characteristics that



The Bohr Model of the Hydrogen Atom



The Solar System

FIGURE 10–8 The atom and the solar system

Perhaps one of the most famous scientific analogies is the comparison of the atom to the solar system, originally formulated by the Danish physicist Niels Bohr (1885–1962). He explained the hydrogen emission spectrum by picturing the hydrogen atom as a solar system with the nucleus—a sole proton—as the sun and its electron as a planet orbiting the nucleus at various fixed distances and energies; energy is emitted as the electron leaps from an outer to an inner orbit. (Modern quantum mechanics has replaced the orbits with energy levels, but the power of Bohr’s analogy transformed scientific thinking about the atom.)

has been solved before and use or adapt that solution in the present instance. Here the question is “Can a solution that worked for one problem map onto another?” Thus, **analogical reasoning** is often a process of comparison, using knowledge from one relatively known domain (“the source,” such as the bicycle thieves in the earlier situation) and applying it to another domain (“the target,” or your locked computer) (Clement & Gentner, 1991; Spellman & Holyoak, 1996). A famous analogy in science, though in some of its aspects now outdated, is shown in Figure 10–8.

2.1. Using Analogies

Let’s take another example from the world of computers that provides a good understanding of how analogical reasoning works. Consider the problem of developing a way to protect computers from viruses. Vaccines have been developed to protect humans from viral infections; might it be possible to develop a vaccine for computers? The answer is yes: reasoning by analogy with the action of biological viruses, computer scientists have done just that. There are of course differences between computer viruses and biological ones: a virus doesn’t give a computer a runny nose or a fever. But, in their underlying, or structural, characteristics, computer viruses and biological viruses have important elements in common: they are contagious or infectious, they can self-replicate by means of the host, and they can cause damage to the host. (Structural information typically specifies a relation between different entities, such as between a virus and its host.) Thus, analogical reasoning involves identifying and transferring structural information from a known system (in this

example, biological viruses) to a new system (in this example, computer viruses). Many researchers have argued that this structural similarity is one of the defining features of analogy (Gentner & Markman, 1997).

Analogical reasoning is generally thought to comprise five subprocesses:

1. *Retrieval*: Holding a target (such as the computer virus) in working memory while accessing a similar, more familiar example (such as a biological virus) from long-term memory.
2. *Mapping*: While holding both the source and the target in working memory, aligning the source and the target and mapping features of the source (such as “contagious,” “replicating,” and “harmful”) onto the target.
3. *Evaluation*: Deciding whether or not the analogy is likely to be useful.
4. *Abstraction*: Isolating the structure shared by the source and the target.
5. *Predictions*: Developing hypotheses about the behavior or characteristics of the target from what is known about the source (for example, predicting from the behavior of biological viruses that computer viruses can change their surface features to avoid detection).

These five components of analogical reasoning have been extensively investigated during the past 25 years and have led to many important experiments and computational models. One of the first cognitive investigations of analogical reasoning was that of Mary Gick and Keith Holyoak, carried out in 1980. These investigators presented participants with a problem, and its solution, in the form of a story. A few minutes later, after an intervening, irrelevant task, students were given a second story problem, but this time the story contained no solution. The first story was about a general who was planning to lead his army to attack a dictator who lived in a fortress. A number of roads converged on the fortress, but the dictator had mined all of them so that any large army traveling on a road would be blown up. The attacking general broke his army up into small groups that were not heavy enough to set off the mines and sent one of these small units along each of the roads. Arriving safely at the fortress, the soldiers regrouped to full force and captured the dictator.

The second problem story was about a patient with a stomach tumor. Doctors have a powerful laser beam that can burn the tumor out, but the beam is so strong that it will burn the healthy tissue as well. Participants were asked to suggest a solution for the doctors that would spare the healthy tissue but destroy the tumor. (What’s your suggestion? Of course, the fact that mention of this study appears in a discussion of analogical reasoning provides a considerable hint, one that was not available to the participants.) The solution that Gick and Holyoak (1980) were looking for was a convergence solution analogous to that in the first problem story: just as the general broke his army into smaller units so as to be able to travel swiftly and then regroup into a powerful force, the laser beam might first be broken up into a set of less powerful rays, all focused on the diseased spot, where they converge to burn out the tumor. Only 20 percent of the participants who read the army–dictator problem judged the radiation–tumor problem as similar and came up with the convergence solution. When the surface features of the source story were more similar

to those of the target problem, 90 percent of the participants came up with the convergence solution (Holyoak & Thagard, 1995). It seems that when a source problem shares only structural similarity with the target, far fewer participants recognize the analogy; when surface features also are similar, many more participants are able to retrieve a relevant analog.

Another approach to the study of analogy, used by Dedre Gentner and her colleagues, has been to investigate the factors that influence the retrieval of sources from memory. For example, participants are given a series of stories to read (Gentner et al., 1993). In one of them a hawk is attacked by a hunter but befriends the hunter by giving him some of her feathers. After this generous act, the hunter spares the hawk. A week later the participants read other stories. Some of these new stories had the same underlying structure, others shared only superficial features with the original story. One of the structurally similar stories was about a country, “Zerdia,” that is attacked by the country of “Gagrach;” “Zerdia” offers to share its computers with “Gagrach” and the two countries become allies. One of the stories that was superficially similar, but structurally dissimilar, to the hawk–hunter story was about a hunter who loved to fish and eat wild boar. When participants were asked which stories reminded them of the original one, most of them chose the stories sharing superficial features, not the ones that shared underlying sets of structural relations. But when the same participants were asked which stories were analogically similar, they named the stories with similar underlying structures.

Dunbar and Blanchette (2001) have found that scientists, politicians, and students can use both structural and superficial features of a problem, but are more likely to use structural features when they generate their own analogies, and superficial features when they receive ready-made analogies. Furthermore, Blanchette and Dunbar (2002) have found that students will make analogies automatically and without awareness.

2.2. Theories of Analogical Reasoning

A number of influential theories of analogical reasoning have been proposed, all of which can be expressed as computer models that make explicit the mechanisms thought to be involved. Two of the most important are the structure mapping theory (SMT) (Falkenhainer et al., 1989; Gentner, 1983) and the learning and inference with schemas and analogies (the LISA) model (Hummel & Holyoak, 1997, 2003). Both models treat analogical reasoning as the mapping of elements from a source to a target, and both propose a search of long-term memory for a source that has a similar underlying structure to that of the target.

The *SMT model* has two stages. In the first stage, long-term memory is searched for items that have a superficial feature that is contained in the target (for example, a computer virus). Thus, in the computer virus–biological virus analogy, memory might be searched for items such as *keyboard*, *mouse*, *not working*, *electrical*, and *infectious*. The second stage is evaluation: how good a match exists between what was retrieved in the first stage and the target. The SMT model as simulated by a computer behaves much like human participants: there will be many superficial matches, most of them having to do with functioning and nonfunctioning computers, and

possibly a useful one such as *infectious*. The main assumption of the SMT model is that although, structural similarity is the key component of analogical reasoning, the human cognitive system looks for superficial matches when searching memory, and we find it difficult to retrieve true relational analogs.

The *LISA model* accounts for the same type of data, but uses very different computational mechanisms that are like the neural networks discussed in earlier chapters, in that features of both source and target may be considered nodes in a network. Thus, the target is represented in terms of the activations of features of the source: *computer virus* would activate, for example, the features *malfunctioning*, *harmful*, and *self-replicating*. This simultaneous activation of a number of features in working memory results in the activation of similar constellations of features (rather than individual superficial features, as in SMT) in long-term memory, leading to the retrieval of a source analog such as *flu virus*.

There are a number of other models of analogical reasoning, each with slightly different assumptions and mechanisms. It is difficult to determine what types of models most accurately capture analogical reasoning, but brain-based research may provide the answer.

2.3. Beyond Working Memory

From behavioral analyses, we now know that analogical reasoning is highly demanding of attention and memory. First, we must attend to the appropriate superficial and structural similarities between source and target; then we must search long-term memory for the appropriate analog. Are brain regions that are involved in attention and searching memory—specifically, the prefrontal cortex—highly involved in analogical reasoning? A PET scanning study was designed to answer this question (Wharton et al., 2000).

On each trial, participants were presented with, first, a source picture, and then a target picture. There were two conditions, *analogy* and *literal comparison*. In the analogy condition, participants were to decide whether the target picture was an analog to the source picture. On each trial in which source and target were indeed analogous, the pictures contained different objects but shared the same system of relations. In the literal comparison condition, participants were simply asked to decide whether the source and the target pictures were identical. When they compared activations in the literal condition to those in the analogy condition, the investigators found significant activations in the middle frontal cortex and the inferior frontal gyrus (both of which are portions of prefrontal cortex), as well as the anterior insula and the inferior parietal cortex. The prefrontal cortex and parietal cortex are known to be highly involved in tasks that require attention and working memory.

But is analogical reasoning simply a product of attention and working memory and nothing more? How could we answer this question? One way would be to look for a specific neural correlate that dissociated the relational component of analogical reasoning from working memory. This approach was taken by investigators who conducted an fMRI study in which working-memory load and structural complexity in comparable tasks were varied independently (Kroger et al., 2002). As expected,

increasing the working-memory load resulted in increased activations in the dorso-lateral prefrontal and parietal cortices. Further, significant and unique activations in the left anterior prefrontal cortex were found when structural complexity was increased while the working-memory load was held constant. These data demonstrate that the relational component of analogical reasoning represents a cognitive capacity that recruits neural tissue over and above that of attention and working memory. Here is a nice example of how neuroimaging technology can provide us with new and informative data about cognition. By examining the neuroanatomical correlates of the subcomponent processes of analogical reasoning (that is, working memory and abstraction), we can begin to decompose how such reasoning is accomplished.

Comprehension Check:



1. What are the five subprocesses of analogical reasoning?
2. What is the role of working memory in analogical reasoning?

3. INDUCTIVE REASONING

Any thought process that uses our knowledge of specific known instances to draw an inference about unknown instances is a case of **inductive reasoning**. Common types of inductive reasoning often rely on **category-based induction**: either generalizing from known instances to *all* instances (which is a general induction), or generalizing from some members of a category known to have a given property to other instances of that category (which is a specific induction). If you have seen three violent football games and conclude that all football games are violent, you have made a *general induction*. If you see Alma Mater College play a violent game this weekend and therefore believe that College of the Gridiron, on the schedule for next Saturday, will also play violently, you have made a *specific induction*.

No inductive process can ever be certain: we cannot know all the instances that may exist, any of which may disprove the generalization. In both these types of induction, we are using our inference to add new knowledge, *which may be incorrect*, to our existing knowledge.

3.1. General Inductions

In the early 1980s, medical researchers were trying to identify the cause of a mysterious new disease soon called AIDS. It attacked a number of very different populations: young gay men, intravenous drug users, hemophiliacs, Haitians, infants, and recipients of blood transfusions. The only common factor among all these patients was a dramatic decrease in T lymphocytes, a type of white blood cell (Prusiner, 2002). From these instances, the prediction was made that all AIDS patients have decreased numbers of T cells, and it was proposed that the cause of the disease is an infectious agent that attacks these T lymphocytes. The researchers' approach to solving the AIDS problem employed a general induction from a number of instances.

This type of induction occurs in solving very different sorts of problems, from choosing a mate (of necessity—you can't possibly have knowledge of all possibilities!) to the origins of traffic jams to scientific discovery. Cognitive psychologists have investigated both the strategies we use to make such generalizations and the errors we may fall into when we do.

Research on generalization began in earnest during the 1950s. In an early study, investigators devised a task much like the game Mastermind: from feedback provided by the experimenter, participants had to discover rules for the game, thereby making generalizations (Bruner et al., 1956). (To get a feel for the task, you might want to try a few rounds of Mastermind; there are playable versions available on the Web.)

The task employed a deck of cards that varied along four dimensions, with three possibilities for each attribute: color (red, green, or black); number of items on a card (one, two or three); shape of item (circle, cross, or square); and number of borders (one, two, or three). Thus, there were $3 \times 3 \times 3 \times 3$ possible combinations of attributes, and so the deck consisted of 81 cards, or instances (Figure 10–9).

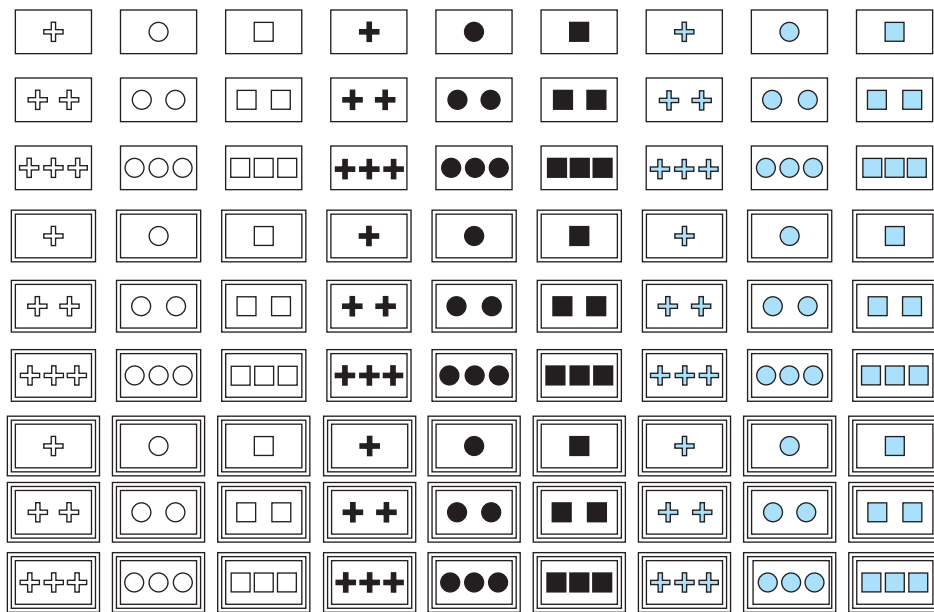


FIGURE 10–9 “Pick a card . . . discover a rule”

This is the deck of 81 cards, varying in the shape, color, and number of pictured objects, and in the number of borders, that was used in one of the first studies exploring how we generalize. Participants either selected, or were presented with, one card at a time. The rule for a positive or negative instance was known only to the experimenter. For each trial participants stated their view and were told only whether they were right or wrong—not what the rule was. They then moved on to the next trial. The goal: to discover the rule.

(Bruner, J. S., Goodnow, J. J. & Austin, G. A. (1956). *A study of thinking*. New York, NY: Science Editions. Reprinted by permission.)

In one version of the task, the cards were laid out face up and the experimenter arbitrarily determined a rule—for instance, “red and square”—but did not tell it to the participant. Instead, the experimenter pointed to a card that was red and square and told the participant that this card was an example of the rule. The participant then pointed to various cards in turn and the experimenter would state in each case whether or not that card was also an example of the rule. With each choice of card, the participant must offer a conjecture as to the general rule. The investigators varied the (unstated) rules, and found that simple rules (such as “red”) were very easy for participants to discover, conjunctive rules (such as “red and square”) were a little harder, and disjunctive rules (such as “red or square”) were difficult. Negative rules, such as “not red,” were very difficult to discover, and disjunctive negative rules, such as “not red or cross,” were the most difficult of all.

Why did participants find some general rules easier to apply inductive reasoning to than others? The contribution of Bruner and colleagues (1956) was to present an explanation in terms of the different reasoning strategies that participants used. In the *successive scanning* strategy, participants picked cards that differed by only one feature from the example card given by the experimenter (“red and square”). For example, the example card might contain three red crosses and one border; on the first trial the participant might pick a card with three red crosses and *two* borders. If that card fulfills the rule, the participant knows that the number of borders is not relevant and that attribute may be disregarded; if it doesn’t, the participant must continue to look at combinations of all three attributes when suggesting a card. Another approach, the *focus gambling* strategy, maintains one feature while changing all the others. A participant who knows from the experimenter that a card with three red crosses and one border fulfills the rule might next choose a card with two green triangles and one border. If the experimenter says “yes,” then the participant knows that rule is based on the number of borders; but if the experimenter says “no,” then the participant has not gained any new knowledge about the rule. Because both strategies involve first testing for single features, it is not surprising that participants found it easier to discover simple rules (such as “red”) than conjunctive rules (such as “red and square”).

The work of Bruner and colleagues (1956) led to two major developments in cognitive psychology. The first was concerned with the nature of categories. By focusing on rules rather than on arbitrarily named features that may not be recognized by all, Bruner’s work led to an examination of category formation that was relevant to categories that we use in real life (E. E. Smith & Medin, 1981). Bruner’s work also led to investigation of the way we test a hypothesis. A hypothesis is an idea or proposition that we can evaluate or test by gathering evidence to support or refute it. How does a participant who has inferred a rule through generalization discover whether or not that generalization is correct? What strategies are used to test a hypothesis? These questions became a central focus of research (Tweney et al., 1981).

A well-known task used to investigate the way we test hypotheses is the Wason 2–4–6 task, devised by the English psychologist Peter Wason (1924–2003). The task is structurally simple and easy to administer; try it on your friends. The experimenter states that the sequence 2–4–6 is a triad of numbers that is consistent with a rule.

The participant's goal is to discover the rule by generating hypotheses about it, as well as generating new triads of numbers to test the hypothesis. The experimenter tells the participant whether or not each proposed triad is consistent with the rule. The participant may "announce" an induced rule at any time, and is told by the experimenter whether the generalization is correct. If it is, the experiment is over. If it isn't, the participant is told to generate more triads of numbers. This procedure continues until the participant either announces the correct rule or gives up.

Participants typically begin by hypothesizing that the rule is "even numbers increasing by 2," and most test the hypothesis by generating triads of numbers that are consistent with it, suggesting, for example, the triad 8–10–12. Generally, participants propose three or four more such triads and are told that they are correct; they then announce, "even numbers increasing by 2" as the rule. But they then are told that this rule is not correct. Most participants then induce a more general hypothesis: "any set of numbers increasing by 2," suggesting triads like "1–3–5" or "7–8–10." But when they announce this new hypothesis they are told that this also is incorrect. At this point something interesting happens: most participants switch from trying to *confirm* their hypothesis to trying to *disconfirm* it (i.e., switch from generating triads consistent with their hypothesis to generating triads inconsistent with it) (Gorman et al., 1987). They might propose the triad 2–6–4; if so, they are told that this triad is inconsistent with the rule. Once participants have negative feedback to work with, they usually discover that the correct rule is simply "numbers of increasing magnitude."

Participants are sometimes resistant to taking into account information that is inconsistent with the rule that they have formed. For instance, if they are told that there is a probability of error in the feedback that they receive, then they attribute to error all feedback that is inconsistent with the rule they have induced (Gorman, 1989). Furthermore, even when participants are encouraged to look for disconfirming evidence when performing the Wason 2–4–6 task, their performance is not significantly improved (Tweney et al., 1981).

How representative is the 2–4–6 task of real-world situations in which we must generalize over a set of instances? Certainly not entirely. Participants show little **confirmation bias**, the predisposition to weight information in ways consistent with preexisting beliefs (Figure 10–10), when they are given a nonarbitrary concept to discover (Dunbar, 1993). Furthermore, in studies of scientists' reasoning, little evidence was found that these participants attempted to confirm their hypotheses by not considering other possibilities, or by ignoring inconsistent data (Dunbar, 1997, 1999).

3.2. Specific Inductions

To assume that if one member of a category has a particular feature another member will also have it is to make a specific, category-based induction. Of course, there's an obvious trap: the feature involved may not be common to all category members. Nonetheless, specific, category-based induction often allows us to make useful inferences about a new or unknown category member. In this way we can update our knowledge without having to find out, instance by instance, whether this particular information is true of all category members. If you hear that crows in the northeastern



FIGURE 10–10 Under a full moon . . .

Confirmation bias is an ubiquitous phenomenon: we tend, preferentially, to find evidence of what we already believe. Someone who believes that the crime rate increases when the moon is full will notice news reports of crimes committed then, but be less attentive to crime stories at other times. The result? “What did I tell you? Three burglaries this weekend—and I know why!”

United States have been dying of West Nile virus, you might induce that robins will die of West Nile virus; that would be making a specific, category-based induction. (The starting fact—in this case what you hear about crows—is often called the *premise*, analogous to the premise of an argument; the inference, which is about robins in this case, is the *conclusion*.) Would you also think that flamingos, pheasants, and ducks would die from the virus? Ornithologists did, and in August 1999, that is exactly what happened in the Bronx Zoo. The West Nile virus killed many species of birds.

Cognitive psychologists have been investigating specific, category-based induction since the mid-1970s (e.g., Rips, 1975). Such research has shown that we follow a number of heuristics in making category-based inductions. First, the more similar the premise instance is to the conclusion instance, the greater the likelihood that the feature mentioned in the premise will be attributed to the conclusion. Second, the more typical the premise instance is of its category, the more likely the conclusion instance will be judged to have the feature of interest. A third heuristic was identified by investigators who found that if the category involved is thought to be relatively homogeneous (for example, cats), we are willing to make stronger inferences by

projecting the feature (for example, tails) from one instance to other instances of the category (although in this case, we would be wrong—Manx cats have no tails). If, however, the category is thought to be more heterogeneous (for example, animals), then we are unwilling to make strong inferences to other instances of the category (Nisbett et al., 1983). Variability within the category containing the premise and conclusion instances can have a large effect on judgments (see also Heit, 2000).

A general theory of category-based induction, known as the *similarity-coverage model*, has been developed (Osherson et al., 1990). In this view, the similarity of members of categories is not sufficient to explain all phenomena observed in category-based induction. Rather, the model proposes that underlying the typicality effects observed in inductive reasoning—that the more typical the premise instance, the more readily its feature is mapped to the conclusion—is the notion of coverage. “Coverage” is defined as the *average* maximum similarity between the instances in the premise and each exemplar of that category in the conclusion. To illustrate, consider the following two cases:

<i>Premise:</i>	Dogs have a liver.	<i>Premise:</i>	Dogs have a liver.
<i>Premise:</i>	Cats have a liver.	<i>Premise:</i>	Whales have a liver.
<i>Conclusion:</i>	Mammals have a liver.	<i>Conclusion:</i>	Mammals have a liver.

Which argument do you think is stronger? If you are like most participants in the original experiments, you chose the argument on the right. The investigators explained the effect by pointing out that, although the argument on the left contains terms (“dog” and “cat”) that, to the non-zoologist, are more typical members of the category “mammal,” the argument on the right contains terms that, between them, have more coverage of the category—that is, at least one of the exemplars should be relatively similar to *any* other instance of that category (and it is the *maximum* of the two similarities that determines induction).

Clearly, working memory is involved in the inductions: we must hold in memory the information from which we generalize. Induction also involves the executive functions needed to propose the induced rule. It has been demonstrated that the frontal lobes play a dominant role in inductive inference.

3.3. Critical Brain Networks

Both studies of patients with various types of brain damage and neuroimaging studies of neurologically healthy participants point to the role of the frontal lobes in inductive reasoning. A standard test for frontal-lobe damage, discussed in Chapter 7, the Wisconsin Card Sort test (WCST) is a test of inductive reasoning in that the goal is to induce a rule. Participants are asked to match test cards to reference cards according to the color, shape, or number of stimuli on the cards. Feedback is provided after each trial, enabling the participant to learn (or induce) the correct rule for classifying the cards (e.g., sort the cards on the basis of color). After 10 or so correct trials, the rule is changed. Unimpaired, normal participants have little difficulty noting that the rule has changed. However, frontal-lobe patients, particularly those with

damage to the left dorsolateral prefrontal cortex, have great difficulty switching rules even when they have overwhelming evidence that the rule they are continuing to use is incorrect (Dunbar & Sussman, 1995).

These data have been corroborated by fMRI investigations with unimpaired, normal participants (see Monchi et al., 2001). In their study, participants responded to a computer-based WCST that was similar to the traditional version. They found significant activations in the mid-dorsolateral prefrontal cortex when participants received positive and negative feedback while performing the card sorting task. The researchers argued that these regions of the prefrontal cortex were activated because the participants had to compare the current feedback information to prior trials held in working memory. These data are consistent with the notion that generalization, at least with respect to that used in the Wisconsin Card Sort test, involves the active monitoring of events in working memory. In addition, the researchers found that a combined cortical/subcortical network of regions, including the ventrolateral prefrontal cortex, caudate, and thalamus, was activated when participants received only negative feedback. These regions have been shown to be involved in a number of tasks that require the updating and modification of behavior on the basis of negative feedback.

Several studies using PET and fMRI have examined the neural underpinnings of category-based induction. In one of these, participants were asked to judge the probability of a stated conclusion from a given set of premises, similar to other studies we have described (Parsons & Osherson, 2001). The investigators found activation in portions of the left hemisphere, including in the medial temporal and parahippocampal regions and in large sections of the frontal lobes. These data extend the observations in the patient studies, showing that the frontal lobes are part of a more distributed network of brain regions that together support inductive inference. As discussed in Chapter 5, it is widely agreed that the medial temporal lobes are involved in memory, including both storage and retrieval. Given this information, we can then envision how category-based induction demands the active retrieval of relevant information from long-term memory and holding of that information, and these processes demand resources supported by the frontal and temporal lobes.

A further question focuses on the influence of experience: a key characteristic of inductive inference is that the underlying cognitive processes can change with experience. In the 2–4–6 task, for example, participants began in ignorance of the rule. As they worked on the task, proposing triads of numbers and rules and receiving feedback, they began to develop specific hypotheses, and some participants learned the rule. How does the brain change during this type of learning?

To answer this question, investigators presented unimpaired, normal participants with a simple task: they were required to sort abstract drawings into two groups, according to their two unseen, but strongly related, prototypes (Seger et al., 2000). (As discussed in Chapter 4, the prototype is the “central” member of a category.) The investigators found that during early trials brain activations were limited to frontal and parietal regions in the right hemisphere. As learning progressed, activation in left-hemisphere regions began to be seen, specifically in the left parietal lobe and left

dorsolateral prefrontal cortex (Figure 10–11 on Color Insert A). What does this suggest? It appears that when the participants began to classify the drawings, they did so by processing the visual patterns of the stimuli. As learning progressed, however, they probably began to formulate an abstract rule. Reasoning from abstract rules is generally thought to be the realm of the left hemisphere. Like the neuroscience research discussed in the section on analogical reasoning, this is a nice example of how the use of neuro-imaging technologies can inform our understanding of higher level cognition.

Through the use of neuroimaging, researchers have recently been able to probe deeper into the underlying mechanisms involved in complex scientific hypothesis testing. For example, Fugelsang and Dunbar (2005) conducted an fMRI experiment examining the mechanisms by which we integrate data when testing specific hypotheses. Participants were asked to test specific hypothesis about the effect of various drugs designed to influence mood. The hypotheses could either be plausible or implausible. For example, the plausible hypotheses contained descriptions of drugs known to affect mood, whereas the implausible hypotheses contained descriptions of drugs known to have little to no effect on mood, for example, antibiotics. Data relevant to these hypotheses were then provided to participants in a trial-by-trial format where they viewed multiple trials of evidence for each type of drug. This evidence could be consistent or inconsistent with the hypothesis being tested. The researchers found that when participants were examining data relevant to a plausible hypothesis, regions in the caudate and parahippocampal gyrus were preferentially recruited. In contrast, when participants were examining data that were relevant to an implausible hypothesis, regions in the anterior cingulate cortex, precuneus, and left prefrontal cortex were selectively recruited.

What do these activations of different brain networks tell us about hypothesis testing? Let's consider first the caudate and parahippocampal gyrus activations, found with plausible hypotheses. These regions of the brain are typically thought to be involved in learning, memory, and the process of integrating information. Given that, these data suggest that we may be more inclined to learn and integrate new information if it is consistent with a plausible hypothesis. The anterior cingulate cortex, one of the regions activated when participants were examining data that were relevant to an implausible hypothesis, has been largely implicated in detecting errors and conflict (as discussed in Chapter 7). Do participants treat data relevant to implausible hypotheses as error? These data suggest this may be the case! Taken together, the findings of Fugelsang and Dunbar (2005) suggest that during inductive reasoning the human brain may be specifically tuned to recruit learning mechanisms when evaluating data that are consistent with preexisting hypotheses, and to recruit error detection mechanisms when evaluating data that are not consistent with hypotheses.

Again like the neuroscience work in analogical reasoning, this is a nice example of how the use of neuroimaging technologies can inform our understanding of cognition. By understanding the underlying brain networks that are involved in various complex tasks, we can begin to understand how the subcomponents of inductive reasoning (for example, attention, error processing, conflict monitoring, and working memory) interact.

Comprehension Check:

1. What is the difference between general and specific category-based induction?
2. What are the proposed roles for the frontal cortex and temporal lobes in category-based induction?

4. DEDUCTIVE REASONING

You're in the market for a new car. Lucky you, money is no object—but speed is. You go to the nearest Porsche dealership. You see that Porsche has developed a new model called the Boxster. From your knowledge of automobiles, you have come to the conclusion that all Porsches are reliable automobiles. Given that the Boxster is a Porsche, you expect the new Boxster to be reliable. (You have just made a valid logical deduction.) So you take the new Porsche Boxster out for a test drive and it breaks down after only 10 minutes on the road. The only logical conclusion you can make is that one of your premises must be false: either premise 1—"all Porsches are reliable"—is false (which is possibly the case), or premise 2—"the Boxster is a Porsche"—is false (which is highly unlikely). You have just done a fine piece of deductive reasoning.

Many theorists, from Aristotle on, have believed that deductive reasoning represents one of the highest achievements of rational thought. Deductive reasoning tasks are therefore one of the fundamental tools used by cognitive psychologists in the quest to understand some of the mysteries of human rationality.

One tool used to study deductive reasoning is the **sylllogism**, an argument that consists of two statements and a conclusion. The conclusion may be either true or false. A conclusion that follows from given premises by the laws of deductive logic is a *valid* conclusion. Your conclusion that the Boxster is a reliable vehicle was valid; nonetheless, it turned out not to be true, because one or the other of your premises was false. In studies of deductive reasoning, a participant is given two premises and a conclusion and then asked to state whether the conclusion necessarily follows—in other words, whether it is valid. The basic idea of deductive reasoning is that a valid conclusion follows from the premises as a matter of logical necessity (which is not the case in inductive reasoning, where the conclusion is not *necessarily* true).

4.1. Categorical Syllogisms

The relations between two categories of things can be described by a **categorical syllogism**. Stated formally, your reasoning at the Porsche dealership looked like this:

Premise 1: All Porsches are reliable.

Premise 2: The Boxster is a Porsche.

Conclusion: The Boxster is reliable.

In the language of logic, premise 1 is the major premise, premise 2 is the minor premise. The categorical syllogism can be generalized:

Premise 1: All A are B.

Premise 2: C is an A.

Conclusion: C is B.

The relationship between two terms in a categorical syllogism can be described by four types of statements:

Universal affirmative (UA): All A are B.

Universal negative (UN): No A are B.

Particular affirmative (PA): Some A are B.

Particular negative (PN): Some A are not B.

These relationships between two terms are often represented in Venn diagrams, named for the English mathematician and logician John Venn (1834–1923). The diagrams are graphical depictions, by means of overlapping circles, of the relationship between two or more terms. The terms are represented as circles, and the categorical relationship between them is denoted by the degree of overlap. Figure 10–12 shows the four types of categorical syllogisms as Venn diagrams.

Note that there are two possible ways to represent the universal affirmative assertion “all A are B.” A might represent a subset of B: the assertion that all jellybeans are red does not necessarily imply that there are no other things in the universe that are red. Alternatively, A and B might be equivalent, in which case everything that is red is a jellybean. This raises an important point to consider about judging the logical validity of categorical syllogisms; in order for a conclusion to be logically valid, it must be the only possible conclusion that follows from the premises. A total of 512 syllogisms can be constructed from all possible combinations of the quantifiers “all,” “some,” and “none,” two premises, and a conclusion. Of these 512 possible syllogisms, only 27 have been found to be valid (Johnson-Laird & Steedman, 1978).

4.2. Conditional Syllogisms

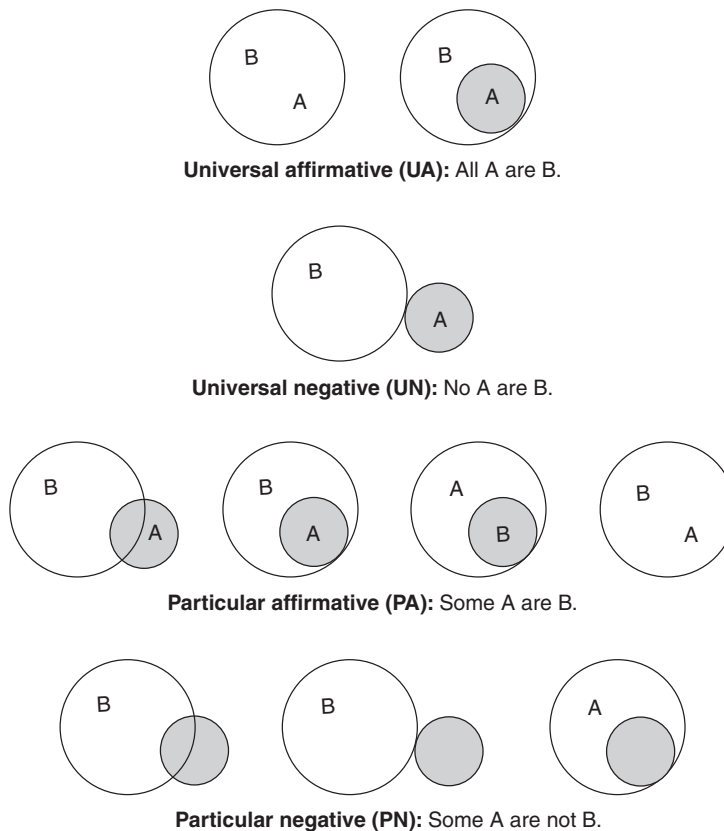
The occurrence of an event may be conditional on the occurrence of another: this relationship between events can be described in a **conditional syllogism**. Like categorical syllogisms, conditional syllogisms consist of two premises and a conclusion. The first premise of a conditional syllogism is a statement of the form “if p , then q ,” where p is some antecedent condition and q is some consequent condition. The second premise can take one of four forms:

Affirmation of the antecedent (AA): p is true.

Denial of the antecedent (DA): p is not true.

Affirmation of the consequent (AC): q is true.

Denial of the consequent (DC): q is not true.

**FIGURE 10–12** Venn diagrams

The possible categorical relationships between the variables A and B are shown here as Venn diagrams. The universal negative has only one representation, but note that the other assertions can be expressed in more than one way. Seeing the various possibilities makes it clear why premises containing particulars are much more difficult to reason with than premises containing universals.

Your car-buying reasoning can be cast as follows in the form of a conditional syllogism:

Premise 1: If the automobile is a Porsche, then it is reliable.

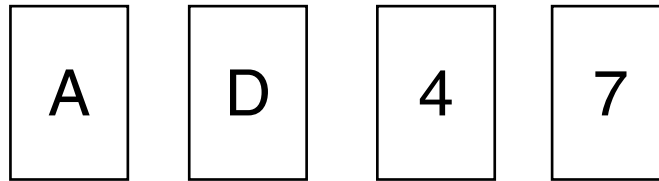
Premise 2: The Boxster is a Porsche.

Conclusion: The Boxster is reliable.

Premise 1 is of the form “if p then q ,” “Porsche” is the antecedent and “is reliable” is the consequent; premise 2, in this case, affirms the antecedent; the conclusion, “is reliable,” logically follows.

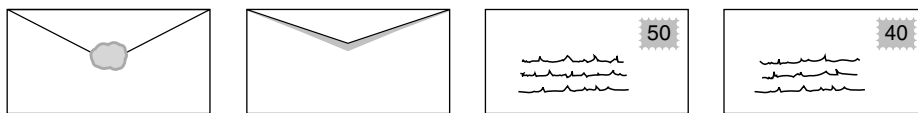
One of the most common tasks used to study conditional reasoning is the Wason selection task, a deceptively simple task in which typically fewer than 10 percent of participants make logically correct responses. An example problem in the task is

The rule: If a card has a vowel on one side, then it has an even number on the other side.



(a)

The rule: If a letter is sealed, then it has a 50-lira stamp on it.



(b)

FIGURE 10-13 The Wason selection task

(a) An abstract version and (b) one closer to life experience. The problem is the same in both: what is the least necessary number of cards (or envelopes) you have to turn over to establish whether or not the rule is true? Which ones would you turn over? Try both versions of the task; which one is easier?

shown in Figure 10-13. Four cards are laid out before the participant, bearing the letters A and D and the numbers 4 and 7. The participant is given this conditional rule: “If a card has a vowel on one side, then it has an even number on the other side.” The task: to determine whether the rule is true by turning over the least necessary number of cards. All right, try it: Which cards do you think you will need to turn over to decide whether the rule is true? Think about this. You could flip over card A to see whether there is an even number on the other side; if there is an odd number on the other side, then the rule has been falsified. But if you find an even number on the other side, the rule has been affirmed—so far.

Are we finished now? Well, you could turn over the 4 card to make sure there was a vowel on the other side—and if you chose this option, you responded like 46 percent of the participants in Wason’s original experiment, if that’s any comfort. Because where does this get you? The rule you were given made no reference to what you should expect on the other side of a card with an even number on it—it *doesn’t matter what’s on the reverse*, and you’ve wasted a move. Similarly, flipping over the D card will provide no useful information because the rule provided no information about what cards with consonants should have on the other side; so it doesn’t matter. The correct inference is to choose the A card and the 7 card. Why the 7 card? Because turning over the 7 card allows you to test the negative of the if-then statement that was offered as the rule: if you find a vowel on the other side of the 7 card, then, and only then, could you know whether the rule is true or false.

The finding that typically fewer than 10 percent of participants perform logically on the Wason selection task paints a relatively bleak picture of our ability to be logical. The version of the Wason task that is presented here, however, is quite abstract:

asking someone to make decisions about cards with even numbers and vowels does not draw on any relevant real-world knowledge. When versions of the task with “real-world” scenarios and combinations are presented (“If you borrow my car, then you have to fill up the tank with gas”), performance improves considerably.

4.3. Errors in Deductive Thinking

Reasoning deductively is not always a simple matter. In fact, most of us make erroneous judgments when reasoning both categorically and conditionally. The types of errors we make have provided a wealth of information to researchers interested in developing theories of deductive reasoning.

We make two main types of errors when reasoning deductively: *form errors* and *content errors*. **Form errors** result from errors in the structural form or format of the premise–conclusion relationship. **Content errors** result when the content of the syllogism is overly influential. The accompanying *Debate* box also discusses errors in deductive reasoning.

DEBATE

Errors and Evolution

Why do we make errors in deductive reasoning? Most theories of deductive reasoning are based on the assumption that errors in reasoning are due to limitations of the cognitive system, such as limited working memory capacity. Another theory, however, suggests that social and evolutionary factors are the cause of some deductive reasoning errors (Cosmides & Tooby, 1992). This view starts with the idea that humans are sensitive to rules for social reasoning—that is, the interpretation of social situations—because we have adapted, through evolution, to be sensitive to certain aspects of our social environment. In particular, this theory posits that humans possess a specialized brain “module” (i.e., self-contained system) for detecting those of their species who cheat in social exchanges (Stone et al., 2002).

This hypothesized evolutionary adaptation can account for performance on certain conditional deductive reasoning tasks. For example, researchers studied a patient, R.M., who had severe damage to the basal ganglia and the temporal pole, a brain structure that is thought to be a source of input into the amygdala, which is critical for processing emotional and social information. The lack of a functional temporal pole renders the amygdala largely disconnected and incapable of processing this information. R.M. was given different versions of the Wason selection task. He performed normally on tasks that required him to determine whether someone was breaking a precautionary rule (such as “if you engage in hazardous activity X, you must take precaution Y”), but performed poorly on logically identical tasks that required him to determine whether someone was cheating on a social contract (such as “if you receive benefit X, you must fulfill requirement Y”). The researchers argue that R.M.’s unique pattern of correct and erroneous reasoning could not occur if detecting cheaters depended solely on the application of general reasoning rules. Instead, his selective deficit suggested that detecting social cheaters requires specialized neural circuitry.

The evolutionary hypothesis stands in contrast to accounts that suggest that we solve different deductive reasoning problems (such as precautionary rules versus social contracts) by applying logical rules or by applying mental models. In all these cases, detecting cheaters would require no privileged circuitry above and beyond other reasoning domains. Again, more research is needed; the jury is still out on this issue.

4.3.1. Form Errors

A common form error in categorical reasoning is to accept a conclusion as valid if it contains the same quantifier—“some,” “all,” or “no”—as appears in the premises. This error is called the **atmosphere effect**: the use of these terms in the two premises conveys an overall mood, or atmosphere, that leads participants to accept a conclusion containing the same term (Woodworth & Sells, 1935). For example, it is easy to see that the conclusion “all As are Cs” necessarily follows from the two premises “all As are Bs” and “all Bs are Cs.” Consider now what happens when we replace the previous quantifier “all” with “no” or “some”:

Premise 1: No As are Bs.

Premise 2: No Bs are Cs.

Conclusion: No As are Cs.

It may not be intuitively obvious that this conclusion is invalid. Let’s replace the abstract A, B, and C with some concrete terms and see what the categorical syllogism looks like:

Premise 1: No humans are automobiles.

Premise 2: No automobiles are doctors.

Conclusion: No humans are doctors.

It is apparent now that the conclusion is invalid.

A related form error, in conditional reasoning this time, is known as **matching bias**, that is, accepting a conclusion as valid if it contains the syntactic structure of the premises or some of the terms of the premise. For example, in the Wason selection task (Figure 10–13), this error occurs when people erroneously turn over the 4 card because it is mentioned in the stated rule. Both the atmosphere effect with categorical syllogisms and the matching bias with conditional syllogisms point to the strong impact of syntactic structure. In both cases, we are strongly influenced by the quantifiers used in the premises. Why might this be the case?

One possibility is that certain objects in categorical and conditional statements—such as the formal quantifiers—draw our attention. It has been argued that we simply expect the information we receive to be relevant (Evans, 1989), and so we expect the quantifier to be critical. Thus, the bias to attend to the quantifier words in the premises and accept them in the conclusion arises because in fact most of the time the information we are given *is* relevant. Another reason we may have difficulty in reasoning with more complex categorical and conditional statements has to do with the troublesome nature of negative quantifiers. We do not always spontaneously convert negative statements (for example, “not an even number”) to positive statements (for example, “an odd number”). Finally, limitations on working memory could be at the root of many of the errors we make in deductive reasoning, and indeed, all contemporary theoretical accounts of deductive reasoning recognize the significant role that working memory plays in the processing that underlies such reasoning.

4.3.2. Content Errors

Logical deductions should be influenced only by the structure of the premises: the laws of logic are abstract and are independent of the content of the syllogism. But we human beings are embedded in a world where the content of something—what it is—is generally informative. A common content error is to focus on the truth or falsity of individual statements in the syllogism (while ignoring the logical connection between statements). This error was demonstrated in a study in which participants were presented with a number of invalid syllogisms whose conclusions sometimes contained true statements (Markovits & Nantel, 1989). Consider the following two examples:

Premise 1: All things that have a motor (A) need oil (B).

Premise 2: Automobiles (C) need oil (B).

Conclusion: Automobiles (C) have motors (A).

and

Premise 1: All things that have a motor (A) need oil (B).

Premise 2: Opprobines (C) need oil (B).

Conclusion: Opprobines (C) have motors (A).

Is either of these two conclusions valid? Which one? Most participants said that the first example was valid; in fact, they are both invalid. The first two premises say nothing about a relationship between C and A, which is what the conclusion is about. Nonetheless, participants accepted the first conclusion as valid more than twice as often as they did the second. We are apparently more likely to accept as logically valid an invalid conclusion if the premises and conclusion are true statements.

The **belief-bias effect**—the tendency to be more likely to accept a “believable” conclusion to a syllogism than an “unbelievable” one—is perhaps the most prevalent content effect studied in deductive reasoning (for a review, see Klauer et al., 2000). Consider the following.

Premise 1: No cigarettes (A) are inexpensive (B).

Premise 2: Some addictive things (C) are inexpensive (B).

Conclusion: Some addictive things (C) are not cigarettes (A).

About 90 percent of participants presented with this syllogism judged the conclusion to be valid. The conclusion is both logical (it necessarily follows from the premises) and believable (there are many addictive things that are not cigarettes). What happens when we rearrange the content of the syllogism?

Premise 1: No addictive things (A) are inexpensive (B).

Premise 2: Some cigarettes (C) are inexpensive (B).

Conclusion: Some cigarettes (C) are not addictive (A).

Only about 50 percent of participants recognize this conclusion as valid. But of course it is: the conclusion logically follows from the premises. The conclusion,

however, is no longer believable. The unbelievable content of the problem influences the ability of many participants to make a valid logical deduction.

Much research has found that both belief and logical validity influence our judgments of validity in an interactive fashion. By presenting participants with prose passages containing categorical syllogisms that varied in terms of validity and believability, Evans et al. (1983) found that the effects of logic were greater for unbelievable than for believable conclusions—that is, participants were more likely to ignore the logical structure of the syllogism if the conclusion was believable (see the accompanying *A Closer Look* box). This interaction between logical structure and content is one of the most tested phenomena in deductive reasoning, and contemporary theories of deductive reasoning generally take great care to address it.

4.4. Theories of Deductive Reasoning

There are several important theoretical accounts of deductive reasoning. One prominent class of theories of deductive reasoning proposes that deduction depends on formal rules of inference akin to those of a logical calculus (Braine & O'Brian, 1991; Rips, 1994). These theories propose that humans naturally possess a logical system that enables us to make deductions. In this view, we evaluate deductive syllogisms by constructing and verifying a “mental proof” in working memory. In other words, we attempt to solve deductive reasoning problems by generating sentences that link the premises to the conclusion and then determine whether the conclusion necessarily follows from the premises. It is thought that we assess the validity of the premise and conclusion by linking their representations in working memory with the logical rules people naturally possess. Rule-based approaches do very well at accounting for certain effects of logical form in reasoning. For example, the time it takes to solve conditional and categorical problems in deductive reasoning increases with the number of inferential steps needed as well as to the nature of the rules required to solve the problem.

Rule-based approaches also acknowledge content effects in deductive reasoning. How might knowledge or expectations influence the application of internalized logical rules? One possibility is that reasoning that ignores logical rules may occur because of limitations on working memory (Rips, 1994). As noted earlier, we commonly use heuristics to solve problems, and in deductive reasoning, for better or worse, we use many heuristics to aid in making logical inferences that put too much of a load on working memory. One such heuristic—developed because we have experienced valid believable examples in the past—may lead to the belief-bias effect: that believable conclusions are more likely to be valid than unbelievable ones (Rips, 1994).

Another view is the theory of *mental models* (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). Mental models are internal representations of real or imaginary situations that can be derived from something perceived in the environment or from information such as that in syllogisms. Deductive reasoning, in this view, occurs in three stages. First, a mental model is constructed that best represents the information in the premises. This requires comprehension of the terms in the

A CLOSER LOOK

Logic and Belief

An influential study by Jonathan Evans, J. L. Barston, and P. Pollard examined the relationship among logical processes, beliefs, and expectations; the results were reported in 1983 in a paper titled “On the Conflict between Logic and Belief in Syllogistic Reasoning,” published in *Memory and Cognition*, 11, 295–306.

Introduction

The investigators were interested in examining how one’s beliefs and expectations influence our adherence to the rules of logic. Do we reason “rationally,” ignoring the content of a given problem and focusing only on the logical structure of the arguments?

Method

In the experiment, 24 participants were presented with 80-word prose passages containing categorical syllogisms that were (1) logically valid and had a believable conclusion, (2) logically valid but had an unbelievable conclusion, (3) logically invalid but had a believable conclusion, and (4) logically invalid and had an unbelievable conclusion. The logical structure for the valid arguments had the following form:

Premise 1: No A are B.
 Premise 2: Some C are B.
 Conclusion: Some C are not A.

The invalid arguments had the form:

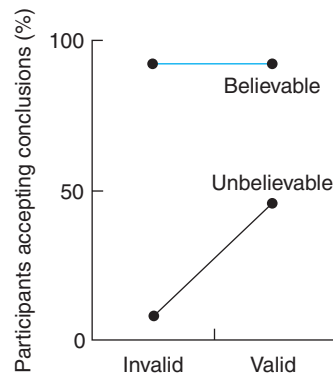
Premise 1: No A are B.
 Premise 2: Some C are B.
 Conclusion: Some A are not C.

The content of the arguments contained either believable conclusions (for example, “Some religious people are not priests”) or unbelievable conclusions (for example, “Some deep sea divers are not good swimmers”). Each participant evaluated four passages, one for each believability-by-validity condition.

Results

The data are plotted on the graph. First, as is evident, acceptance of a conclusion as valid was influenced by the logical validity of the categorical syllogism: when a conclusion was logically necessary, the proportion of participants who accepted it as valid increased. Second, acceptance of a conclusion as valid was also influenced by its believability: when a conclusion was believable, the proportion of participants who accepted it as valid increased. It is important to note, however, the interaction between logical validity and belief: the effects of logic were greater for unbelievable (46 percent versus 8 percent) conclusions than for believable conclusions (92 percent versus 92 percent); in fact, the participants in this experiment appeared to ignore completely the logical structure of the arguments when they considered the conclusion believable.

(continued)



Whether participants accepted a conclusion appeared to be unaffected by the validity of the conclusion if they regarded it as believable.

(Evans, J. St. B. T., Barston, J. L. & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11, 295–306.)

Discussion

The finding that beliefs strongly influence deductive reasoning challenges traditional views that argue that we humans reason on the basis of abstract, “content-free” rules of logic.

premises and of the relationship between them. For example, being told “All As are Bs” and “All Bs are Cs,” you might construct a model in which three mental objects are labeled “C,” two of them are also labeled “B,” and one of the latter is also labeled “A.” Second, a tentative conclusion is generated and evaluated so it can be determined whether it is consistent with the model derived in the first stage. In our example model, a tentative conclusion would be “All As are Cs.” Third, and this is the most controversial aspect of the theory, the conclusion must be validated. This involves the search for alternative models that are consistent with the premises but not with the conclusion. (In our example, any alternative model is consistent with the premises and conclusion.) If such an alternative model can be generated, then the conclusion is invalid and an alternative conclusion must be generated and evaluated, and so on. A conclusion is valid only if there are no alternative models available to falsify it.

The mental models theory provides a good account for both form and content errors in deduction. For example, the extent to which we successfully evaluate conditional and categorical syllogisms has been shown to be directly related to the number of models required—a matter of logical form (Johnson-Laird, 1983). The theory also provides an explanation of how knowledge or expectations influence reasoning: the believability of a conclusion, a product of knowledge and experience, may determine the extent to which alternative models are generated and verified. If the initial conclusion is believable, we may not search for alternative models and thus run the risk of accepting a believable but invalid conclusion.

4.5. Linguistic versus Spatial Basis

Work with patients who have brain damage, and neuroimaging of normal healthy brains has made it possible to study the neural underpinnings of deductive reasoning. This research has provided new insights into fundamental questions that have perplexed cognitive psychologists for decades. One such question that has received much attention is whether deductive and inductive reasoning are linguistically or spatially based. A linguistic model would propose that, because deductive reasoning involves linguistic properties of representations, we should see activation of the left-hemisphere language structures such as the frontal and inferior temporal regions (see Chapter 12). A spatial model of deductive reasoning, on the other hand, would suggest that in order to reason we create spatial representations (i.e., particular types of mental models) of linguistic information. In that case, we would expect to see activation of the visual-spatial-perception neural structures, such as those in the parietal and occipital lobes, particularly in the right hemisphere.

To date, the results of research have been quite mixed. On the one hand, researchers have provided data that support linguistic models. For example, patients with left-hemisphere brain damage were found to be severely impaired in a simple deductive reasoning task, whereas patients with damage in comparable regions in the right hemisphere were only minimally impaired relative to unimpaired, healthy controls (Read, 1981). Furthermore, right-hemisphere patients have been shown actually to perform better than left-hemisphere patients and controls (Golding, 1981). These patient studies provide some support for a linguistic model of deductive reasoning. In addition, a neuroimaging study with unimpaired, healthy participants found significant activations for deductive reasoning in the left inferior frontal gyrus (as well as the left superior occipital gyrus) (Goel et al., 1998). These findings also support the hypothesis that deductive reasoning is linguistically mediated. On the other hand, however, researchers found significant activations in the *right* middle temporal cortex and right inferior frontal gyrus for a similar task of deductive reasoning (Parsons & Osherson, 2001). These findings are more consistent with spatial models. Why the difference?

To begin at the beginning: these two neuroimaging studies differed in both the types of syllogisms used and the content of those syllogisms. In the Goel et al. (1998) study, which showed left-hemisphere activation and thus supported a linguistic model, categorical syllogisms were taken from a military context and employed terms not necessarily familiar to the participants (for example, *officers, generals, privates*). The Parsons and Osherson (2001) study, which showed right-hemisphere activation and thus supported a spatial model, presented participants with conditional arguments that contained more generally familiar material (for example, *doctors, firefighters, teachers*). Could these differences in materials result in different patterns of brain activation? The answer is yes: it has been suggested that deductive reasoning with highly familiar material recruits relatively more neural tissue from the right hemisphere, whereas content-free deductive reasoning recruits neural tissue predominantly in the left hemisphere (Wharton & Grafman, 1998). This fact alone can explain some of the

discrepancy. But neuroimaging work examining reasoning is still in its infancy, and much research remains to be done.



Comprehension Check:

1. What are the differences between form errors and content errors in deductive reasoning?
2. What are the similarities and differences between the rule-based and mental model theories of deductive reasoning?

Revisit and Reflect

1. What is the nature of problem solving?

Problem solving is a process of surmounting obstacles to achieve a particular goal. To do this, we must identify what the problem is and choose a course of action that will make it possible to achieve the goal. At its most basic level a problem can be thought of as having three parts. First is the goal state: this is where you want to be, at the solution to the problem. Second is the initial, or start, state: this is where you are now, facing the problem that needs to be solved. Third is the set of operations that you can apply—that is, the actions you can take—to get from the start state to the goal state. This all sounds relatively straightforward; however, some problems (known as ill-defined problems) are hard to define and represent because their rules and constraints are unclear. On the other hand, well-defined problems, which have rules that are clear no matter how complicated, are typically easy to define. The initial state, goal state, and the intermediate operations are all thought to occur within a defined problem space, which is the set of states, or possible choices, that face problem solvers at each step as they go from an initial state to a goal state.

Think Critically

- Are well-defined problems always easier to complete than ill-defined problems? Why or why not?
 - Can the solution of all problems be characterized in terms of search in a problem space? Are there key aspects of solving a problem that this approach leaves out? For example, are the start state and set of operations necessarily specified completely from the outset?
2. How do we use heuristics or “mental shortcuts” to solve problems?

A heuristic is a rule-of-thumb that may offer a shortcut to solving a problem. Problem solvers have a number of heuristics at their disposal to help them solve a problem. Typically, a heuristic can help the reasoner to achieve the goal state faster than an algorithm, which is a set of procedures for solving a given type

of problem that will always produce the correct answer (for example, the steps in taking a square root or performing long division). One heuristic is random search, a process of trial and error, such as randomly hitting keys on the keyboard when a computer freezes. A problem solver using the hill-climbing heuristic looks ahead one move and chooses the move that most closely resembles the goal state. In solving the Tower of Hanoi, the hill-climbing problem solver may try to select each move so that it most closely resembles the final state when the three disks are on the third pole. In the means-ends analysis heuristic, the problem solver breaks the problem into a series of subproblems, for example, completing one side of a Rubik's cube as the first stage in solving the puzzle.

Think Critically

- Can you think of situations in which certain heuristics might lead a reasoner astray?
 - Which heuristics might work better for solving a well-defined problem? An ill-defined problem?
3. How do we use analogies to solve new problems?

When solving a novel problem, we often try to think of a solution to a similar problem—that is, we reason by analogy. Specifically, analogical reasoning involves using knowledge from one relatively known domain (the source) and applying it to another, less familiar, domain (the target). Analogical reasoning is generally thought to comprise five subprocesses: (1) retrieval of relevant (source) information, (2) mapping features of the source onto the target, (3) evaluating whether or not the analogy is valid, (4) abstracting the relevant feature shared by the source and the target, and (5) predicting behavior or characteristics of the target from what is known about the source.

Think Critically

- Can you think of an example when you used an analogy to solve a novel problem?
 - Can analogies sometimes lead to faulty assumptions about the underlying nature of objects or events?
4. What is the difference between induction and deduction?

Reasoning can be loosely defined as the ability to draw conclusions from available information. The processes that we adhere to when reasoning can be subdivided into two main inferential processes, inductive reasoning and deductive reasoning. Inductive reasoning involves using known information to draw new conclusions that are likely to be true. Inductive reasoning frequently involves categories, generalizing from known instances to all instances, or from some instances to another instance. Deductive reasoning, on the other hand, involves

using known information to draw conclusions that must be true. Categorical reasoning (reasoning about the relations between two categories of things) and conditional reasoning (determining the degree to which the occurrence of an event may be conditional on the occurrence of another) are forms of deductive reasoning.

Think Critically

- Can you think of specific scenarios that would demand both inductive and deductive reasoning?
 - Can you imagine a situation where deductive logic leads you to a valid conclusion but your knowledge of the world tells you that this conclusion is not true? What is the reason for this discrepancy?
5. How do our knowledge and beliefs influence “logical” reasoning?

Evidence from Wason’s 2–4–6 inductive reasoning task has indicated that we typically show a confirmation bias when asked to discover a rule. Here, in a variety of tasks reasoners have been shown to spend the majority of their efforts trying to confirm a rule that they believe to be correct as opposed to trying to disconfirm it. Much research on deductive reasoning has shown that we often focus on the truth or falsity of individual statements in the syllogism while ignoring the logical connection between statements.

Think Critically

- Is the finding that our beliefs influence our logical reasoning process necessarily a discouraging result?
 - How might any of a host of cognitive processes—attention, executive processes, working memory—contribute to the interaction between beliefs and logical processing?
6. How do our brains coordinate the vast amount of processing involved in problem solving and reasoning?

Many of the brain areas linked to perception, attention, and memory are also highly involved in reasoning and problem solving. There is a very good reason for this—reasoning and problem solving are typically highly demanding of attention and memory. You must determine the goal of the current problem; you must perceive and extract the relevant properties of the current stimulus that will assist you to meet this goal; and, while keeping the current goal active in working memory, you must determine how the current features of the stimulus relate to the current goal and which operations to perform next. Depending on the outcome of this third step, you may need to modify your short-term goal in order to meet your desired end state. Such processing involves attention and working memory, thus invoking resources in the dorsolateral prefrontal cortex and anterior cingulate (among other areas). The requisite preliminary visual feature analyses, object identification, and object location analyses would utilize resources in the occipital, temporal, and parietal lobes, respectively. The interplay between the analyses

of the current features of the problem and the current goal state of the problem solver would invoke a feedback loop between the attention/working memory-related brain structures (in particular, prefrontal and anterior cingulate cortices) and the perceptual/object identification/location-related structures (in particular, occipital/temporal/parietal cortices).

Think Critically

- How does neuroimaging inform theories of problem solving and reasoning?
- One of the most interesting findings about the brain and problem solving is that many regions of the brain that are involved in attention and memory are also involved in thinking and reasoning. Why is this the case?