Chapter 8
Problem Solving

Keith J. Holyoak

The ability to solve problems is one of the most important manifestations of human thinking. The range of problems people encounter is enormous: planning a dinner party, tracking deer, diagnosing a disease, winning a game of chess, solving mathematical equations, managing a business. This radical diversity of problem domains contrasts with the relative specificity of many human cognitive activities, such as vision, language, basic motor skills, and memory activation, which have a relatively direct biological basis and which all normal individuals accomplish with substantially uniform proficiency. In the course of normal development we all learn, for example, to speak a native language, but without specialized experience we will never acquire competence in deer tracking or chess playing.

On the other hand, all normal people do acquire considerable competence in solving at least some of the particular types of problems they habitually encounter in everyday life. We might therefore suspect that problem solving depends on general cognitive abilities that can potentially be applied to an extremely broad range of domains. We will see, in fact, that such diverse cognitive abilities as perception, language, sequencing of actions, memory, categorization, judgment, and choice all play important roles in human problem solving.

The ability to solve problems is clearly a crucial component of intelligence. Furthermore, the phenomena of problem solving present many intriguing puzzles that must be accounted for by a successful theory of problem solving. For example, consider the differences between the best computer programs for playing chess and the performance of the very best human players. Before selecting its next move, a top-ranked chess-playing program is likely to assess billions of alternative possible continuations of the game. In contrast, the human grand master may consider a mere dozen alternatives—and then proceed to select a better move than the program did. What mechanisms allow the best move to so readily “come to mind”
for the grand master? And what kind of learning processes allow this sort of expertise to be acquired from problem-solving experience? These and other questions about human problem solving are the focus of this chapter.

In order to understand the nature of human problem solving, it is useful to first consider the nature of problems. We can often learn a great deal about how problems are solved by considering how they could be solved. That is, a task analysis of problems can provide information about constraints that the nature of the problem imposes on the nature of the problem solver. We will also see that task analysis suggests that problem solving is intimately connected with learning. An intelligent problem solver uses the results of solution attempts to acquire new knowledge that will help solve similar problems more readily in the future.

We begin by characterizing the nature of problem solving and the fundamental theoretical issues the topic raises for cognitive science. We introduce the topic using a problem that has been investigated by psychologists over the past several decades. In addition, we briefly consider the implications of neuropsychological evidence regarding the basic components of problem-solving skill. We then examine in more detail one of the major concerns of recent research, the acquisition of expertise in particular problem-solving domains, such as chess or physics. Finally, we examine aspects of problem solving that seem to involve parallel and unconscious information processing.

### 8.1 The Nature of Problem Solving

#### 8.1.1 Problem Solving as Search

The Gestalt psychologist Karl Duncker (1945) performed a series of experiments in which he recorded what university students said as they "thought aloud" while attempting to solve the "radiation problem":

Suppose you are a doctor faced with a patient who has a malignant tumor in his stomach. To operate on the patient is impossible, but unless the tumor is destroyed, the patient will die. A kind of ray, at a sufficiently high intensity, can destroy the tumor. Unfortunately, at this intensity the healthy tissue that the rays pass through on the way to the tumor will also be destroyed. At lower intensities the rays are harmless to healthy tissue but will not affect the tumor, either. How can the rays be used to destroy the tumor without injuring the healthy tissue?

As well as stating the problem verbally, Duncker showed his subjects the sketch in figure 8.1, which illustrates a ray passing through a cross-section of the body with the tumor in the middle. Obviously, this arrangement
will not do. You may want to pause here for a few moments to try to think of possible solutions to the radiation problem.

What makes the doctor's situation a "problem"? In general, a problem arises when we have a goal—a state of affairs that we want to achieve—and it is not immediately apparent how the goal can be attained. Thus the doctor has the goal of destroying the tumor with the rays, without damaging the surrounding healthy tissue. Some valuable clues to the nature of problem solving can be found in the everyday metaphors we use to talk about it (Lakoff and Turner 1989). It is conventional to think of abstract states such as goals as metaphorical spatial locations, and event sequences as metaphorical paths leading from one state to another. This spatial conception permeates descriptions of problem solving. We speak of "searching for a way to reach a goal," "getting around roadblocks" encountered along the way, finding a "shortcut" solution, "getting lost" in the middle of a solution, "hitting a dead end" and being forced to "backtrack," "approaching the problem from a different angle," and so on.

This conception of problem solving as search in a metaphorical space, which underlies our common-sense understanding, has been elaborated to provide a rigorous theoretical framework for the analysis of problem solving. Although some of the theoretical ideas can be traced back to Gestalt psychologists such as Duncker (1945), the modern formulation of a general theory of problem solving as search through a space is due to Newell and Simon (1972). In their problem-space formulation, the representation of a problem consists of four kinds of elements: a description of
the initial state at which problem solving begins; a description of the goal state to be reached; a set of operators, or actions that can be taken, which serve to alter the current state of the problem; and path constraints that impose additional conditions on a successful path to solution, beyond simply reaching the goal (for instance, the constraint of finding the solution using the fewest possible steps).

The problem space consists of the set of all states that can potentially be reached by applying the available operators. A solution is a sequence of operators that can transform the initial state into the goal state in accord with the path constraints. A problem-solving method is a procedure for finding a solution. Problem solving is thus viewed as search: methods are used to find a solution path among all the possible paths emanating from the initial state and goal state. Figure 8.2 provides a graphical illustration of a search space. Each circle represents a possible state of affairs, and the arrows represent possible transitions from one state to another that can be effected by applying operators. A sequence of arrows leading from the initial state to the goal state constitutes a solution path.

How might search proceed in attempting to solve the radiation problem? In his analyses of subjects' think-aloud protocols, Duncker identified
three general "paths" toward solutions. One approach is to alter the direction from which the rays are applied so as to avoid contact with the healthy tissue. For example, people often suggested sending the rays down the esophagus, taking advantage of an "open passage" through the healthy tissue. This solution is impracticable, since the esophagus is not straight, but nonetheless represents a serious attempt to achieve the doctor's basic goal. A second approach is to desensitize or immunize the healthy tissue so that it will not be harmed by the rays. A third approach is to reduce the intensity of the rays along their path to the tumor. Duncker observed that subjects often developed increasingly refined solutions that could be reached by following one or more of these basic search paths.

8.1.2 Heuristic Search

The problem-space analysis yields a mathematical result with brutal implications for the feasibility of solving many problems, such as the problem of winning a chess game. If at each step in the search any of $F$ operators might be applied, and a solution requires applying a sequence of $D$ steps (that is, $D$ is the "depth" of the search), then the number of alternative operator sequences is $F^D$. As $F$ and $D$ get even modestly large, $F^D$ becomes enormous. A typical game of chess, for example, might involve a total of 60 moves, with an average of 30 alternative legal moves available at each step along the way. The number of alternative paths would thus be $30^{60}$, a number so astronomical that not even the fastest computer could play chess by exploring every possible move sequence. The fact that the size of the search space increases exponentially with the depth of the search is termed combinatorial explosion, a property that makes many problems impossible to solve by exhaustive search of all possible paths.

Human beings, with their limited working memory, are actually far less capable of "brute-force" search than are computers. (See chapter 7.) For example, human chess players are unable to "look ahead" more than three or four moves. Yet a human grand master can play superlative chess, better than any computer program yet devised. How can this be? The answer is that human beings use problem-solving methods that perform heuristic search—rather than attempting the impossible task of examining all possible operator sequences, people consider only a small number of alternatives that seem most likely to yield a solution. Intelligent problem solving, in fact, consists largely of using methods for heuristic search. Some heuristic search methods are very general and can be applied to virtually any problem; others are much more specific and depend on detailed knowledge of a particular problem domain. As we will see, the development of expertise is largely the acquisition of knowledge that restricts the need for extensive search.
The efficacy of heuristic search depends in part on the nature of the problem to be solved. A major distinction is whether the best possible solution is required, or whether any reasonable solution that achieves the goal will suffice. Heuristic methods are seldom of much use in solving "best-solution" problems. An example is the notorious "traveling-salesman" problem. This problem involves taking the locations of a number of cities (say, ten) and trying to find the shortest possible route that passes through each of the cities exactly once. Due to combinatorial explosion, this problem has an enormous search space of possible routes once the number of cities grows at all large. No one has found a method other than "brute-force" search of all possible routes that guarantees finding the shortest route. However, if the goal is simply to find a route that is reasonably short by some criterion, heuristic search methods may be useful. Human problem solvers are particularly good at what Simon (1981) calls satisficing: finding reasonably good but not necessarily optimal solutions.

8.1.3 Means-Ends Analysis

Search for a problem solution can proceed in either of two directions: forward from the initial state to the goal state, or backward from the goal state to the initial state. Forward search involves applying operators to the current state to generate a new state; backward search involves finding operators that could produce the current state. Duncker observed both forward and backward search in his studies with the radiation problem. For example, a subject might first seek a free path to the stomach, and then realize that the esophagus could serve this function. This procedure would exemplify backward search from a goal to a way of achieving it. In contrast, other subjects seemed to make a relatively "planless" survey of the situation, thinking of various body parts associated with the stomach. In this process the solver might "stumble upon" the esophagus, and then search forward to find a possible way to put it to use in generating an attempted solution. In general, it is most efficient to search in whichever direction requires fewest choices at each decision point. For example, if there is only one way to reach the goal state, it may be easiest to work backward from the goal. (Rips, chapter 9, also distinguishes between forward and backward strategies.)

Newell and Simon (1972) suggested a small number of general heuristic search methods. One of the most important of these, means-ends analysis, involves a mixture of forward and backward search. The key idea underlying means-ends analysis is that search is guided by detection of differences between the current state and the goal state. Specifically, means-ends analysis involves these steps:
1. Compare the current state to the goal state and identify differences between the two. If there are none, the problem is solved; otherwise, proceed.
2. Select an operator that would reduce one of the differences.
3. If the operator can be applied, do so; if not, set a new subgoal of reaching a state at which the operator could be applied. Means-ends analysis is then applied to this new subgoal until the operator can be applied or the attempt to use it is abandoned.
4. Return to step 1.

Suppose, for example, that you have the goal of trying to paint your living room. The obvious difference between the current state and the goal state is that the room is unpainted. The operator "apply paint" could reduce this difference. However, to apply this operator you need to have paint and a brush. If these are lacking, you now set the subgoal of getting paint and brush. These could be found at a hardware store. Thus you set the subgoal of getting to a hardware store. And so on, until the conditions for applying the operator are met, and you can finally reduce the difference in the original problem.

Duncker observed the use of a form of means-ends analysis in his studies of the radiation problem. For example, a subject might first think of the possibility of desensitizing the healthy tissue. But how is this to be done? The person might then form the new subgoal of finding some chemical that could be used to alter the sensitivity of the tissue.

Means-ends analysis illustrates several important points about intelligent heuristic search. First, it is explicitly guided by knowledge of the goal. Second, an initial goal can lead to subsequent subgoals that effectively decompose the problem into smaller parts. Third, methods can be applied recursively; that is, in the course of applying a method to achieve a goal, the entire method may be applied to achieve a subgoal. Thus, in step 3 of means-ends analysis, the method may be reapplied to the subgoal of reaching a state in which a desirable operator is applicable.

8.1.4 Planning and Problem Decomposition

The idea that the process of problem solving is a kind of search suggests a separation between the initial planning of a solution and its actual execution. It is usually advantageous to perform at least a partial search by "looking ahead" for a solution before actually applying any operators. The obvious advantage of planning is that by anticipating the consequences of possible actions, one can avoid the unfortunate consequences of making overt errors. To the extent that errors are irreversible, or reversible only with difficulty, planning is especially important. A doctor would not attempt to treat a malignant tumor without careful planning. By imagining the
consequences of actions prior to an overt solution attempt, one can identify dead ends without actually executing actions. In addition, planning provides information that can be used to monitor and learn from an overt solution attempt. By explicitly anticipating the consequences of applying operators, the problem solver generates expectations that can be compared to what actually happens when the operators are applied. If the actual effects of operators differ from their expected effects, this may trigger revision of the plan as well as revision of beliefs about what will happen in similar future applications of the relevant operators. Problem solving thus provides valuable information that can guide learning (Holland et al. 1986).

Planning often is combined with a process of problem decomposition, in which an overall problem is broken into parts, such that each part can be achieved separately. Suppose, for example, that you need to select a slate of officers to run a club. Rather than try to select people to fill the entire slate at once, it makes more sense to decompose this goal into several subgoals: select a president, a treasurer, and so on. Each of these subgoals defines a problem that can be attacked independently. Finding a solution to each subgoal will require fewer steps than solving the overall compound goal. Because search increases exponentially with the number of steps, solving all the subgoals, each of which requires a relatively small number of steps, is likely to require far less total search than would have been needed to solve the entire problem at once. Thus the sum of the steps required to solve the subgoals of selecting a good person to fill each of the officer positions, considered separately, is likely to be less than the number of steps that would be required to make an equally good selection of the entire slate, treating the task as a single undecomposed problem.

Unfortunately, realistic problems are seldom perfectly decomposable into parts that can be solved completely independently. For example, choices of officers for the various positions in a club interact in various ways. The same person cannot be both president and treasurer, and the various officers need to get along with each other—Sally might make a fine president and Joe a good treasurer, but if they dislike each other they would make a poor combination. But despite this lack of complete independence, total search may be minimized by first making some tentative decisions about each subgoal and then later working on integrating the components into a workable overall plan. That is, the problem solver can take advantage of the fact that some problems are partially decomposable (Simon 1981). This can best be done if foresight is used to form a coherent overall plan before actually beginning an overt solution attempt. Thus, before actually proposing a slate of officers, we could check for compatibility of the tentative list of choices and make corrections where needed. The general strategy is to first try to solve each subgoal independently, but to
note constraints on how the individual decisions interact, and then to check that these constraints are satisfied before finalizing the overall solution attempt. Planning is thus particularly important in effectively reducing search for partially decomposable problems.

8.1.5 Production-System Models of Problem Solving

Newell and Simon's problem-space analysis is highly abstract and is potentially compatible with a variety of specific representations and algorithms. In practice, however, their approach has been closely tied to a particular type of formal model, the production system (Newell 1973, 1990). The central component of a production system is a set of production rules (also termed condition-action rules). A typical production rule might be

IF you have a paint roller  
and you have paint  
and you have a surface ready to paint  
and the surface is large  
and your goal is to paint the surface  
THEN roll the paint onto the surface  
and expect the surface to be painted.

This rule represents the knowledge required for appropriate application of a problem-solving operator. The "then" portion of the rule specifies the action to be taken and the expected state change it will bring about; the "if" portion consists of a set of clauses describing when the operator could and should be invoked. Notice that the clauses in the condition of this rule are of two types. The first four describe preconditions that must be met before the operator can be applied—you need a roller before you can roll. The fifth clause specifies a goal for which the operator is useful—if you want to paint, consider using a roller. The goal restriction helps to limit search, because it means this rule will be considered only when the relevant goal has arisen.

A typical production system operates by cycling through these steps:

1. The conditions of rules are matched against the currently active contents of working memory (for instance, the representation of the current problem state) to identify those rules with conditions that are fully satisfied.
2. If more than one rule is matched, procedures for conflict resolution select one of the matched rules.
3. The selected rule is fired; that is, its action is taken.
4. Return to step 1.
Production-system models of problem solving have been extremely influential in the development of modern cognitive science. For example, Rips in chapter 9 shows how a production-system model of deduction can be used to account for many findings regarding human deductive reasoning. Within artificial intelligence such models have been used to develop expert systems that help perform such tasks as medical diagnosis and mineral exploration. In cognitive psychology, Anderson's ACT* model (1983) is predicated on the claim that human cognition is fundamentally a production system. The knowledge in a production system is encoded in a highly modular fashion. It is relatively easy to add new rules to the system without unduly disrupting the operation of the older rules. Production systems are therefore capable of modeling learning from problem-solving experience (Anderson 1983; Rosenbloom and Newell 1986).

8.1.6 The Brain and Problem Solving

Our discussion so far has focused on computational analyses (for example, the size of search spaces) and types of representations and processes (for example, production systems), rather than on evidence regarding the way problem solving is actually performed by the brain. As is the case for other major cognitive activities, however, it is valuable to consider the implications of neuropsychological evidence in characterizing the basic components of human problem solving. (See chapter 7 on working memory and chapter 1 on categorization.) We can attempt to understand the functional decomposition of problem-solving skills in terms of the functions of relatively localized brain areas, as has been done for language, vision, memory, and categorization.

This is not an easy task. Noninvasive imaging techniques are only beginning to be applied to the study of problem solving. Localization of functions is more often inferred either from lesion experiments with animals, which obviously differ radically from humans in their cognitive abilities, or from clinical studies of brain-damaged individuals, who seldom have injuries confined to a single clear anatomical region. Given its integrative nature, problem-solving ability is likely to be impaired to some degree whenever any major cognitive function, such as working memory, is disturbed. (See chapter 7.) It is therefore especially difficult to identify brain areas that are selectively implicated in problem solving or planning per se.

Nonetheless, there is some evidence that implicates the frontal lobes of the cerebral cortex as an area of special importance in problem solving (Stuss and Benson 1986). This large area at the front of the cortex appears to play a role in a broad range of cognitive and emotional responses. However, careful clinical observations and a few experimental studies have
revealed some interesting selectivity in the deficits that result from damage to this area. Part of the selectivity concerns what is not seriously affected. People with frontal lesions typically are not intellectually impaired as measured by traditional IQ tests. In some cases they are able to function reasonably well in professions in which they were experienced prior to incurring the injury. Nonetheless, some major decrements in cognitive abilities can be identified. A major source of difficulty is novelty: frontal-lobe patients may be able to perform well-learned tasks using old information, yet have great difficulty solving new types of problems. As the great Russian psychologist A. R. Luria put it, "When intellectual operations demand the creation of a program of action and a choice between several equally probable alternatives, the intellectual activity of patients with a marked 'frontal syndrome' is profoundly disturbed" (1969, 749).

Based on an extensive review of the literature on the effects of frontal lesions, Stuss and Benson (1986, 222) suggested several classes of deficits. Frontal damage leads to deficits in the ordering or handling of sequential behaviors; impairment in establishing, maintaining, or changing a mental "set"; decreased ability to monitor personal behavior; dissociation of knowledge from the direction of action; and various changes in normal emotional and motivational responses. Each of these classes of deficits is linked to problem solving. The ability to plan and execute sequences of actions is, of course, essential. Establishing, changing, and maintaining a set requires the ability to selectively attend and respond to goal-relevant information. On a categorization task, frontal-lobe patients are likely to repeat errors despite corrective feedback and be unable to shift from one basis of classification to another. For example, a person who has learned to sort a set of objects by color will have great difficulty sorting by shape instead. (See Delis et al. 1992 for a careful analysis of the deficits observed in sorting performance by frontal patients.) Some of these impaired functions appear to involve working memory. (See chapter 7 for a discussion of the role of the frontal cortex in working memory.)

In addition, frontal patients have difficulty monitoring their own behavior. They may behave in socially unacceptable ways even though they appear to understand that their behavior is wrong. Similarly, they have trouble translating verbal instructions into appropriate actions. A patient may be told to return to work, may express a desire to return to work, and yet fail to do so. Finally, abnormal emotional responses and attitudes are revealed by failure to set goals or care about the future. The person appears to lack "drive" or "motivation."

Shallice (1982) conducted a study that specifically examined the manner in which frontal-lobe patients approach novel problems requiring planning and organized sequential action. He tested patients with various forms of brain damage, as well as control subjects, on their ability to solve various
versions of the “Tower of London” puzzle (see figure 8.3). This puzzle consists of three differently colored beads and three pegs of different lengths. The experimenter places the beads in a starting configuration, and the subject must then move them into a new configuration defined by the experimenter, in a minimum number of moves. (The related “Tower of Hanoi” puzzle is discussed in chapters 7 and 9.) The number of moves required to achieve the goal defines the level of difficulty. Although all the groups of brain-damaged subjects in Shallice’s study were impaired in their performance relative to the control subjects, those with damage to the left frontal lobe showed the greatest decrement, particularly for the more difficult versions of the puzzle. The nature of the errors made by the frontal-lobe subjects indicated that they had difficulty in planning; they could not establish an appropriate order of subgoals. The deficit was not due to a general limitation of short-term memory, for variations in the patients’ performance on a digit-span test could not account for the differences in problem-solving success.

More recent evidence suggests that we should be cautious in interpreting the findings above (see Shallice 1988). In particular, although the overall frontal deficit on the Tower of London task has been replicated (Owen et al. 1990), the selective effect of left-hemisphere damage has not been found in other experiments. However, the role of the frontal cortex in problem solving has been confirmed for another task—chess playing—using PET activation measures (Nichelli et al., 1994). When chess players are asked to decide whether it is possible to achieve checkmate in one move (a task requiring planning), brain activity was selectively increased in regions of both the left and the right frontal cortex. This same study found that several posterior regions of the brain, especially those associated with generation of visual images, also play significant roles in chess playing.

Stuss and Benson (1986) argue that the frontal lobes are crucial in executive control of cognition. The frontal-lobe syndrome in large part appears to involve a loss in ability to control cognitive processes: the ability to select and maintain goals, to plan sequential activities, to anticipate the
consequences of actions, to monitor the effects of actions, and to revise plans on the basis of feedback. (See chapter 7 for a related discussion of executive processes.) These neuropsychological observations have several implications for theories of problem solving and planning, most of which are consistent with other evidence obtained with normal subjects. The crucial importance of selective attention and of the ability to organize sequential action would be expected on the basis of task analysis. The fact that the deficits are primarily observed when the patient faces novel problems suggests that expertise leads to a reduction in the requirements for executive control. The gap between verbal knowledge and action is consistent with the claim that developing skill in solving new problems involves a process of proceduralization: translating verbal knowledge into procedures, perhaps encoded as production rules (Anderson 1983). It appears that proceduralization is impaired by frontal-lobe damage.

The motivational component of the syndrome emphasizes the significance of an aspect of problem solving that is often neglected in computational approaches to problem solving. Unless the organism cares about the future, there is no clear basis for establishing or maintaining goals; and without goals, problem solving simply disintegrates.

8.2 Development of Expertise

Our survey of the nature of problem solving has raised a number of issues that an adequate theory must explain: namely, how goals are formed; how heuristic methods develop; how problems can be decomposed; and how planning is conducted. In addition, a theory must explain how learning takes place during problem solving, and how knowledge acquired in one problem situation is transferred to another. Many of these issues are related to central questions that have been addressed by research on problem solving. How does a novice problem solver become an expert? What makes expert problem solvers better than novices? Clearly, experts in a domain have had more training and practice than have novices, but what exactly is it that experts have learned? Have they learned how to reason better in general, or perhaps to become more skilled in applying heuristic search methods, such as means-ends analysis? Let us look at two domains in which a considerable amount of research has examined differences in the problem-solving methods of experts and novices: playing chess and solving textbook physics problems.

8.2.1 Expertise in Chess

The pioneering work on expertise in chess playing was reported by De Groot (1965). In order to determine what makes a master chess player
better than a weaker player, De Groot had some of the best chess players in the world “think out loud” as they selected chess moves. By analyzing these problem-solving protocols—transcripts of what the players said as they reached a decision—De Groot was able to observe some distinct differences in the ways in which masters and novices selected moves.

His results did not support any of the obvious hypotheses about the masters having superior general reasoning ability or greater proficiency in means-ends analysis. Nor was it that the masters performed more extensive search through the vast space of alternative possible moves. In fact, if anything the masters considered fewer alternative moves than did the weaker players. However, the master players spent their time considering relatively good moves, whereas the weaker players spent more time exploring bad moves. It appeared that the masters were able to exploit knowledge that led them very quickly to consider the best moves possible, without extensive search.

The most striking difference between the two classes of players was observed in a test of their perceptual and memory abilities. Chase and Simon (1973) extended De Groot’s results. In the test the player saw a chess position, drawn from the middle portion of an actual chess game, which was presented for just 5 seconds. An example board position is depicted in figure 8.4a. After the board was removed, the player was asked to reconstruct it from memory. In Chase and Simon’s experiment the subject was either an expert master player (M), a very good class A player (A), or a beginning player (B). The number of pieces correctly recalled over seven trials by each player is depicted in figure 8.5a. The results showed that the greater the expertise of the player, the more accurately the board was recalled.

![Figure 8.4](image.png)

Examples of chess configurations: (a) real middle game; (b) random counterpart. From Chase and Simon 1973.
One might suppose that this result indicates that master chess players have particularly good memories. However, this is not generally the case. To assess this possibility, Chase and Simon also performed the test using random board positions such as the one illustrated in figure 8.4b. As the results shown in figure 8.5b indicate, the master player's advantage was entirely eliminated in this condition.

On the basis of these and other related findings, Chase and Simon argued that master players have learned to recognize large, meaningful perceptual units corresponding to board configurations that tend to recur in real chess games. These units are stored as unitary chunks in long-term memory. Such chunks can be used to encode quickly and accurately the configuration of a novel but realistic board. They are useless, however, in encoding random positions, in which meaningful patterns are unlikely to occur. Chunks also serve as the conditions of production rules that suggest good moves. These rules would have the form, "IF Pattern P is present on the board; THEN consider Move M." Such specific rules would direct the master player quite directly to the relatively small number of alternative moves that are serious candidates, without having to search through large numbers of highly implausible possibilities. It seems likely that the development of specialized rules that are cued by perceptual units contributes to the acquisition of expertise in many domains other than playing board games.

Figure 8.5
Number of pieces recalled correctly by master (M), class A player (A), and beginner (B) over trials: (a) for actual board positions; (b) for random board positions. From Chase and Simon 1973.
8.2.2 Expertise in Physics

The conclusions derived from studies of chess have been confirmed and extended by work on expert–novice differences in solving physics problems. A study by Chi, Feltovich, and Glaser (1981) provided especially interesting results. These investigators asked experts and novices to sort physics problems into clusters on the basis of similarity. Novices tended to
base their sortings on relatively superficial features of the problem statements. (See chapter 1 for a discussion of the role of perceptual similarity in categorization.) For example, figure 8.6 depicts the diagrams for two problems that were often grouped together by novices. A glance at the diagrams associated with each problem indicates that they look very similar; the novices explained that both are “inclined-planes” problems. In fact, although both of these problems involve inclined planes, very different procedures are required to solve them. By contrast, figure 8.7 shows two problem diagrams that experts classified as belonging together. These look very different; however, the experts explained that both problems can be solved by the law of “conservation of energy.”

In general, the work of Chi, Feltovich, and Glaser (1981) and others indicates that experts have learned schemas for identifying important categories of problems. Problem schemas represent complex categories that are defined in part by patterns of relations between problem elements, rather than by the specific elements (such as inclined planes) themselves. Problem schemas in physics are based on more abstract relations than those included in the perceptual chunks available to the chess master; but like perceptual chunks, abstract problem schemas function to vastly reduce the amount of search required to find appropriate solutions. In general, expertise in problem solving is in large part the result of the development of sophisticated mental representations for categorizing problems in the domain.

8.2.3 How Does Expertise Develop?

Research comparing the performance of expert and novice problem solvers tells us a great deal about how to characterize the differences in the knowledge used by people at different skill levels; however, it tells us less about how an initially unskilled problem solver can eventually become an expert. A number of theoretical efforts have, however, attempted to describe learning mechanisms that might allow some combination of direct problem-solving experience, instruction, and exposure to solved examples—the obvious types of environmental inputs available to the learner—to produce increased expertise. (See also the discussion of knowledge reorganization in chapter 4.)

Most models of learning have assumed a production-system representation for procedural knowledge; accordingly, learning is mainly treated as the acquisition of new production rules. The general idea is that by inspecting the results of a solution attempt, learning mechanisms can encode important regularities into new rules. For example, Larkin (1981) has developed a computer simulation that can learn to solve physics problems more efficiently. The program starts by using means-ends analysis to find
unknown quantities by using equations. For example, to find the value of acceleration, $a$, it might use the equation $V_f = V_i + at$ (final velocity equals initial velocity plus acceleration times time). The learning mechanism could then form a new production rule that checks to see whether $V_i$, $V_f$, and $t$ are known (the condition), and if so asserts that $a$ can be found (the action). This new rule will then eliminate the need to apply means-ends analysis to solve future problems with this form. The result of this learning mechanism is a shift from a novice strategy of working backward from the goal, using means-ends analysis and subgoaling, to a more expert strategy of working forward from the givens to the unknown goal quantity. Protocol studies with human experts and novices in physics have found evidence for such a shift from backward to forward search (Larkin et al. 1980).

This type of change in strategy may not always result from forming new rules based on solutions initially found by means-ends analysis. In fact, Sweller, Mawer, and Ward (1983) found that use of a means-ends strategy can actually impair acquisition of expertise in solving mathematics problems. They argue that means-ends analysis focuses attention on the specific features of the problem situation required to reach the stated goal, reducing the degree to which other important aspects of problem structure are learned. Sweller et al. found that a forward-search strategy developed more rapidly when learners were encouraged to explore the problem statements more broadly, simply calculating as many variables as possible. They suggested that less directed exploration of the problems facilitated acquisition of useful problem schemas.

In addition to acquiring new rules and schemas, expertise may be improved by combining old rules in more efficient ways. For example, if two rules apply one after another, it may be possible to construct a single rule that combines the effects of both. This process is termed composition (Anderson 1983; Lewis 1978). As mentioned earlier, Anderson (1983, 1987) also stresses the role of a process of proceduralization, which uses very general productions for following verbal instructions to construct more specific productions to execute a solution procedure.

Finally, learning mechanisms can also make use of solved examples that are provided to illustrate solution procedures. Learners can use examples when they are first presented to actively construct useful rules and schemas. Chi et al. (1989) have found that good and poor learners use solved examples of physics problems in radically different ways. Good learners generate inferences that serve to explain why the examples can be solved in the illustrated manner, whereas poor learners encode them in a much more passive manner (for instance, they fail to ask questions while studying the examples). The development of expertise clearly depends not only on the nature of environmental inputs provided to problem solvers but also on the learning skills they bring to the task.
8.3 Restructuring and Parallelism in Problem Solving

8.3.1 Ill-Defined Problems

The search metaphor for problem solving, as elaborated into formal models by Newell and Simon and others, has clearly been extremely useful in understanding human problem solving. However, neither the metaphor nor the models derived from it capture the full richness of the mental processes that underlie problem-solving skill. The search perspective seems most appropriate when the problem solver has a clear goal, understands the initial state and constraints, and knows exactly what operators might be useful. Given an appropriate method, finding a solution is then indeed a search through a well-defined space of possibilities; if a solution path exists, it will eventually be found by patiently “grinding it out.”

Many of the most difficult problems that beset us, however, have a very different quality. For example, if your goal is to find a career that will bring you happiness, it may be very difficult to specify the state from which you are starting, the operators that might be applicable, or even to recognize when your goal has been achieved. Reitman (1964) observed that many problems are ill defined in that the representations of one or more of the basic components—the goal, initial state, operators, and constraints—are seriously incomplete. Ill-defined problems are usually hard, and not simply because the search space is large. Indeed, many ill-defined problems seem difficult, not because we are swamped by the task of searching through an enormous number of alternative possibilities, but because we have trouble thinking of even one idea worth pursuing.

Duncker’s radiation problem is in some ways quite ill defined. Very few subjects suggest the idea of irradiating the tumor from multiple directions simultaneously with low-intensity rays, focusing the rays so that the tumor receives a higher dosage than does the surrounding healthy tissue. Yet most people will agree that this is a rather good solution to the problem, once it is pointed out to them. To actually discover the solution, however, usually requires more than dogged search, because the operator “create multiple ray sources” is unlikely to even be considered. In fact, Duncker found that the sketch he usually provided to his subjects (figure 8.1) actually impeded discovery of the idea of using multiple rays, for the diagram shows only a single ray source. In this case the solution can be achieved only by first inventing a new operator.

8.3.2 Restructuring, Insight, and Analogies

To understand how ill-defined problems can be solved, it is useful again to look at everyday language. We speak of “looking at the problem in a new light,” having the solution “just pop out,” and realizing the answer was “staring me in the face all along.” These metaphors suggest that a solution
may not always be reached by a gradual serial-search process; rather, it may be achieved suddenly as the result of “seeing” the problem differently. The notion that problem solving shares certain important properties with perception was a major theme of the Gestalt psychologists such as Duncker (1945) and Maier (1930), who proposed that solutions sometimes require insight based on a restructuring of the problem representation. People do not always simply establish a representation of a problem and then perform search; rather, they sometimes change their representations in major ways. Although Newell and Simon’s treatment of search and means-ends analysis was foreshadowed by Duncker’s ideas, the Gestalt emphasis on the importance of restructuring has been less prominent in their theory (see Newell 1985).

There has been very little firm experimental evidence to support the notion that some problems are solved by restructuring and sudden insight. However, work by Metcalfe (1986a, b; Metcalfe and Wiebe 1987) has established several criteria that distinguish the process of solving “insight” problems from the process of solving “routine” problems. Her experiments compared people’s performance in predicting their own ability to solve algebra problems (routine) versus a variety of insight problems, such as this:

A landscape gardener is given instructions to plant four special trees so that each one is exactly the same distance from each of the others. How could the trees be arranged? ¹

One major distinction involved subjects’ ability to predict whether they would eventually be able to solve the problem, for those problems they could not solve immediately. For algebra problems, subjects’ “feelings of knowing” accurately predicted their eventual success; that is, people were able to tell which algebra problems they would be able to solve if they tried, and which would prove intractable. In contrast, subjects were completely unable to predict which insight problems they would be able to solve (Metcalfe 1986b).

A second major distinction was apparent in a measure of subjects’ changes in expectations during problem solving. Metcalfe and Wiebe (1987) had subjects rate how “warm” they felt as they worked on each problem: that is, how close they believed they were to finding a solution. If people were using means-ends analysis, or any heuristic-search method involving a comparison of the current state and the goal state, they should

¹. The gardener could build a hill plant one tree at the top, and plant three others in an equilateral triangle around the base of the hill. Assuming the hill is built to the appropriate height, the four trees will form a tetrahedron in which each of the four corners is equidistant from every other corner.
be able to report getting “warmer” as they approached the goal (because they would know that the difference between the current state and the goal was being progressively reduced). On the other hand, if a solution was discovered on the basis of a rapid restructuring of the problem, the problem solver would not be able to report increased warmth prior to the insight.

Figure 8.8 depicts the striking difference that emerged between the patterns of warmth ratings obtained for insight and for algebra problems. Each of the histograms in the figure shows the distribution of subjects’ warmth ratings at a particular time prior to achieving a solution. A rating of 1 indicates least warmth (feeling nowhere close to a solution), whereas a rating of 7 indicates maximal warmth (the problem seems virtually solved). The histograms in each column are ordered in time from bottom to top, from 60 seconds prior to the solution to the time a solution was actually found. For the algebra problems (right-hand column), the ratings shift gradually to the right of the histogram as the time of solution approaches, indicating that subjects accurately reported getting warmer and warmer as they neared a solution. For the insight problems (left-hand column), the results are utterly different. The ratings do not shift as the time of solution approaches. Rather, most subjects rate themselves as “cold” until they suddenly solve the problem, at which time their rating jumps from maximally cold to maximally warm.

Metcalfe’s results thus provide empirical evidence that insight is a real psychological phenomenon. For a problem that requires insight, people are unable to assess how likely they are to solve it, either in advance of working on the problem or as they actually are working on it. If they eventually succeed in finding a solution, they are genuinely surprised.

Although restructuring and insight do play a role in problem solving, we are far from a full understanding of the mechanisms involved. In some cases restructuring may be triggered by finding an analogy between the target problem at hand and some other situation (the source analog) from a very different domain. For example, Gick and Holyoak (1980, 1983) performed a series of experiments in which subjects first read a story about a general who wished to conquer a fortress located in the middle of a country. Many roads radiated out from the fortress, but these were mined so that although small groups of men could pass over them safely, any large group would detonate the mines. Yet the general needed to get his entire army to the fortress to capture it. He accomplished this by dividing his men into small groups, dispatching each to the head of a different road, and having all the groups converge simultaneously on the fortress.

Does this story remind you of any of the problems we have discussed? It is, of course, an analog to Duncker’s radiation problem. When college students were asked to use the fortress problem to help solve the radiation
Figure 8.8
Frequency histograms of warmth ratings for correctly solved insight and algebra problems. The panels, from bottom to top, give the ratings 60, 45, 30, and 15 seconds before solution. As shown in the top panel, a 7 rating was always given at the time of solution. From Metcalfe and Wiebe 1987.
problem, most of them came up with the idea of using converging low-intensity rays. In the absence of the source analog, very few subjects tested by either Duncker or by Gick and Holyoak proposed this variation of the “reduced-intensity” type of solution. Providing the analogy allowed subjects to restructure their representation of the target problem so that the operator of creating multiple ray sources was constructed and used. (Of course, analogies may also be used to help solve problems that do not require restructuring.)

How can a useful analogy be found? It is often difficult. Gick and Holyoak found that many subjects would fail to notice on their own that the fortress story was relevant to solving the radiation problem even when the analogs were presented in immediate succession. Accessing a source analog from a different domain than the target is yet more difficult when the source has been encoded into memory in a different context (Spencer and Weisberg 1986). In the absence of guidance from a teacher, analogical access requires that elements of the target problem must serve as retrieval cues, which will activate other related situations as the result of activation spreading through semantic memory along the paths that link similar concepts. The greater the similarity of the elements of the two analogs, the more likely it is that the source will be retrieved (Holyoak and Koh 1987).

Of course, there is no guarantee that a source analog will help rather than hinder solving the target problem. For example, Duncker suggested that people may sometimes be misled by a false analogy to the radiation problem: seeing the rays as analogous to a syringe that produces an injection only after the needle is inserted. This source analog might suggest that the rays could be turned on at full strength only after they had reached the tumor. But of course, the intensity of rays cannot be altered once they have been emitted by the ray source. Analogy can provide ideas about how to solve a novel problem, but these ideas are only plausible conjectures, not firm deductions.

8.3.3 Parallel Constraint Satisfaction

One of the hallmarks of Newell and Simon’s approach to problem solving is an emphasis on the serial nature of the solution process. A problem is typically decomposed into subproblems; then each subgoal is solved, one by one. For any particular subgoal, alternative operators are tried sequentially in the search for a solution path. Parallel processing is certainly not entirely excluded; in particular, the process of matching the current problem state against the conditions of production rules is typically assumed to be performed in parallel. Nonetheless, the serial aspects of the solution process are theoretically most central.

The notion of restructuring, in contrast, suggests that parallel (and largely unconscious) information processing may have a major impact on
problem solving. The role of spreading activation in the retrieval of potential source analogs is one example of how parallel access to information stored in long-term memory may redirect conscious problem solving. It is also possible that the way in which active information is used to construct a solution may sometimes involve parallel integration of knowledge rather than strictly sequential processing. Indeed, this is the intuition that appears to have led the Gestalt psychologists to claim that problem solving was similar to perception. The following quotation from Maier (1930, 116) illustrates this connection (as well as the notorious vagueness that left Gestalt theories of problem solving in ill repute):

First one has one or no gestalt, then suddenly a new or different gestalt is formed out of the old elements. The sudden appearance of the new gestalt, that is, the solution, is the process of reasoning. How and why it comes is not explained. It is like perception: certain elements which one minute are one unity suddenly become an altogether different unity.

One of the major advances of modern cognitive science has been to build much more explicit models of how parallel processes are used in perception; consequently, we can now begin to delve more deeply into what it might mean for problem solving to have the perception-like quality that a unified interpretation of an input typically emerges from parallel integration of information at an unconscious level. A key idea is the concept of parallel constraint satisfaction, which is illustrated in the work of Marr and Poggio (1976) on vision, and described in more general terms by Rumelhart et al. (1986).

The idea of finding a solution that satisfies the constraints of the problem is, of course, familiar by now. For example, in section 8.1.4 we looked at the problem of selecting a slate of officers, in which it is necessary to consider interactions between the decisions about each position (for instance, the person selected as president must get along with the treasurer). We saw that the overall problem of choosing a slate can be decomposed into the subproblems of filling each position. These subproblems can be solved separately, with a subsequent check to make sure no interactive constraints are violated. This form of constraint satisfaction is not inherently parallel.

Sometimes, however, the interactive constraints are so pervasive that it is not feasible to solve each subgoal separately and only then check that all constraints are satisfied. In addition, satisfying a constraint is not always an all-or-nothing matter. For example, a possible president-treasurer pair may be compatible to some degree; if each person is individually an excellent choice for the position, the pair may be satisfactory despite some interpersonal tension. When the solution of each subgoal depends in a major way
Figure 8.9 depicts a striking perceptual example of when parallel constraint satisfaction is important. If our problem is to recognize a word, search can be sharply reduced if the problem is decomposed into the subproblems of recognizing each constituent letter. But as figure 8.9 illustrates, the subgoals of identifying the individual letters cannot always be solved independently. You probably recognize the top word in the figure as RED, even though each of the letters is partially obscured. In fact, as the other three words in the figure show, each letter in RED is ambiguous: the R could be a P, the E could be an F, and the D could be a B. Clearly, our interpretation of each individual letter is affected by the interpretations we give the others.

At first glance, the recognition of RED in figure 8.9 seems to create a paradoxical “chicken-and-egg” problem: you need to identify the letters to recognize the word, but you need to identify the word to recognize the letters. This recognition problem can, however, be solved by parallel constraint satisfaction (McClelland et al. 1986). Each subgoal of identifying a letter is influenced not only by the constraints provided by the visual input at that position but also by the constraints imposed by the surrounding letters. The solution process is highly interactive, with information about
possible letters at each position being integrated to form the optimal "gestalt."

Although it is clear that perception involves parallel constraint satisfaction, we need to consider whether similar processes might be involved in higher-level problem solving. Is there any reason, beyond Gestalt intuitions, to suppose that parallel constraint satisfaction also plays a role in the kinds of restructuring we discussed earlier? In fact, there is. One clear possibility arises in the process of solving a target problem by analogy, as discussed earlier. For example, how could a person make use of the fortress problem to help solve the radiation problem? Clearly, part of the person's task will be to find the best mapping, or set of correspondences, between the elements of the source and target. That is, the person must realize that the general's goal of capturing the fortress corresponds to the doctor's goal of destroying the tumor and that the army's ability to do the capturing is like the rays' capacity to destroy. How can this mapping be established? The problem of finding the best overall mapping can be decomposed into the subproblems of finding the best mapping for each of the constituent elements (just as the problem of word recognition can be decomposed into the subproblems of identifying the constituent letters). But as in the perceptual example in figure 8.9, the subgoals of mapping elements cannot be accomplished in isolation. Why, for example, should we map the fortress onto the tumor, rather than, say, onto the rays? After all, neither pair is highly similar in meaning.

Holyoak and Thagard (1989) have proposed a model of how analogical mappings can be constructed by parallel constraint satisfaction. The basic idea is simple, as illustrated in figure 8.10. The nodes in this figure represent some possible mappings between elements of the radiation and fortress problems, and the arrows represent positive and negative relations between possible decisions. One major constraint on analogical mappings is that each pair of mapped elements in the source and target should play consistent roles. This constraint is termed structural consistency (Falkenhainer, Forbus, and Gentner 1989). Suppose, for example, that "capturing" maps onto "destroying." If the analogy is structurally consistent, then the capturer in the fortress problem would have to map onto the destroyer in the radiation problem ("army = rays"), and the object of capturing would map onto the object of destruction ("fortress = tumor"). In fact, as illustrated in figure 8.10, the intuitively correct mappings between elements in the two problems form a mutually consistent set and hence support each other.

A closely related constraint on mapping is that the mapping should be one to one: if an element of one analog maps onto a particular element in the other analog, it probably doesn't also map onto a different element. Thus, the mapping "fortress = tumor" contradicts "fortress = rays" (as
Figure 8.10
A simplified constraint-satisfaction network for finding an analogical mapping between elements of the "fortress" and "radiation" problems.

does "army = rays"). The structurally consistent mappings thus not only support each other but tend to discredit alternative mappings as well. Holyoak and Thagard found that a constraint-satisfaction model of analogical mapping provided a good account of a wide range of data regarding when people find the mapping process easy or difficult. Although many hurdles remain, there is reason to hope that work in cognitive science is beginning to establish a firmer basis for the Gestalt intuition that human perception and thinking have a fundamental unity.

Suggestions for Further Reading
A number of books can be explored for more detailed discussions of aspects of problem solving. Newell and Simon (1972) is a classic, but difficult. Ginsberg (1993) provides a good introduction to artificial intelligence that includes a discussion of problem solving and search. The most highly developed production-system models of cognition are ACT* (Anderson 1983) and SOAR (Newell 1990); both of these books include chapters on learning in the context of problem solving. Klahr, Langley, and Neches (1987) is a collection of papers on learning within production systems, and Michalski, Carbonell, and Mitchell (1983, 1986) are two volumes of papers on machine learning, several of which involve problem solving. A thorough treatment of the analysis of verbal protocols as a method of
studying problem solving is provided in Ericsson and Simon (1984). Ericsson and Smith (1991) have edited a book with chapters on expertise in many domains, including chess and physics.

A detailed survey of frontal-lobe functions is provided in Stuss and Benson (1986). Polya (1957) and Wickelgren (1974) discuss useful problem-solving heuristics in an informal manner. A discussion of relations among learning, categorization, analogy, and problem solving is contained in Holland et al. (1986). The Gestalt notion of restructuring is articulated in Duncker (1945). Holyoak and Thagard (1995) provide a general discussion of the role of analogy in thinking. There are many books on creative thinking; among the best is that of Boden (1992). The papers cited in section 8.3.3 can be consulted for a general introduction to parallel constraint satisfaction.

Problems

8.1 Can you think of any way in which means-ends analysis might lead a problem solver away from the goal in some situations?
8.2 How does problem decomposition reduce the size of the search space?
8.3 What qualities make a problem suitable for solution by parallel constraint satisfaction?
8.4 A robot in an office can perform a small number of actions: it can PUSH an object, PICK-UP an object, CARRY an object, PUT-DOWN an object, or WALK by itself. It can PICK-UP an object only if that object has nothing else on it and if its own arm is empty. It can PUSH an object even if that object has something else on it.
   a. Write a production rule appropriate for using the operator PICK-UP. Include relevant preconditions and a goal.
   b. Suppose the robot is in room B and a table with a typewriter on it is in room A. The robot is instructed to move the table into room B. List the subgoals that would be established by means-ends analysis in the course of solving this problem, using the fewest possible operators.
8.5 A problem can be solved in five steps. At each step any one of ten operators can be applied. How many possible paths are there in the search space?

Questions for Further Thought

8.1 Does research on expertise provide any useful suggestions about how best to teach novices?
8.2 What defines an "insight" problem?

References


