C-I-P-1-6-5

HERBERT A. SIMON
Carnegie Institute of Technology

Scientific Discovery and the Psychology of Problem Solving

> The very fact that the totality of our sense experiences is such that by means of thinking (operations with concepts, and the creation and use of definite functional relations between them, and the coordination of sense experiences to these concepts) it can be put in order, this fact is one which leaves us in awe, but which we shall never understand. One may say "the eternal mystery of the world is its comprehensibility." It is one of the great realizations of Immanuel Kant that the setting up of a real external world would be senseless without this comprehensibility.
>
> —Albert Einstein
> *Out of My Later Years*

In the previous chapter a theory of human problem solving was put forward with references to some of the evidence for its validity. The theory has been formalized and tested by incorporating it in programs for digital computers and studying the behavior of these programs when they are confronted with problem-solving tasks.

The thesis of the present chapter is that scientific discovery is a form of problem solving, and that the processes whereby science is carried on can be explained in the terms that have been used to explain the processes of problem solving. In particular, I shall undertake to show how the theory of problem solving described in the previous chapter can account for some of the principal reported phenomena of scientific discovery.

For a description of these phenomena, the analysis will draw heavily upon previous published accounts. Discussions of scientific discovery have always been highly anecdotal, most of our specific information on the subject deriving from reports of specific examples, recorded in some instances by historians and philosophers of science, in some instances by psychologists, but often by the discoverers themselves. The classics in the latter category are Henri Poincaré's celebrated lecture, translated as "Mathematical Creation" (New York: The Science Press, 1913), and the delightful essay by Jacques Hadamard, *The Psychology of Invention in the Mathematical Field* (Princeton: Princeton U. Press, 1945). Chapter 10 of Max Wertheimer's *Productive Thinking* (New York: Harper & Row, enlarged ed., 1959) reports a series of interviews with Albert Einstein on the course of events that led to the invention of the theory of special relativity.

The literature on the topic produced by philosophers of science is substantial, but has been for purposes of this analysis, on the whole, less useful. (I will mention two important exceptions in a moment.) The reason is that philosophers of science tend to address themselves to the normative more than to the descriptive aspects of scientific methodology. They are more concerned with how scientists *ought to* proceed, in order to conform with certain conceptions of logic, than with how they *do* proceed. Notions of how they ought to proceed focus primarily on the problem of induction: on how generalizations might validly arise from data on particulars and on the degree to which a corpus of data logically confirms a generalization. These are interesting questions of philosophy, but they turn out to have relatively little relation to the actual behavior of scientists—and perhaps less normative value than has been supposed.

In the past few years, two philosopher-historians of science, both originally trained in physics, have made particularly significant contributions to the psychology and sociology of scientific discovery. Both have been quite explicit in distinguishing the processes of discovery from the traditional canons of "sound" scientific method. I shall make considerable use of their work and ideas. One of these men, Norwood Russell Hanson, has set forth his views most extensively in *Patterns of Discovery* (Cambridge: Cambridge University Press, 1958). The other, Thomas S. Kuhn, has produced an original and stimulating account of *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

To explain scientific discovery is to describe a set of processes that is sufficient—and *just* sufficient—to account for the amounts and directions of scientific progress that have actually occurred. For a variety of reasons, perhaps best understood by psychoanalysis, when we talk or write about scientific discovery, we tend to dwell lovingly on the great names and the great events—Galileo and uniform acceleration, Newton and universal gravitation, Einstein and relativity, and so on.[1] We insist that a theory of discovery postulate processes sufficiently powerful to produce these events. It is right to so insist, but we must not forget how rare such events are, and we must not postulate processes so powerful that they predict a discovery of first magnitude as a daily matter.

On the contrary, for each such event there is an investment of thousands of man-years of investigation by hundreds of talented and hard-working scientists. This particular slot machine produces many stiff arms for every jackpot. At the same time that we explain how Schrödinger and Heisenberg, in 1926, came to quantum mechanics, we must explain why Planck, Bohr, Einstein, de Broglie, and other men of comparable ability struggled for the preceding twenty years *without* completing this discovery. Scientific discovery is a rare event; a theory to explain it must predict innumerable failures for every success.

The great events do not, of course, represent sudden leaps forward, unrelated to previous exploration. While modern quantum mechanics clearly did not exist in 1924, and clearly did in 1926, the approach to it was gradual and steady, involving all the illustrious scientists mentioned in the previous paragraph and many hundreds more. And the particular advance that we identify as "the discovery" was followed by many man-years of exploitation and consolidation, just as it was preceded by man-years of exploration and anticipation. The central point remains: scientific discovery, when viewed in detail, is an excruciatingly slow and painful process.

Related to the rarity of great discoveries—and relevant to our understanding of the process—is the rarity of great discoverers. If there are only a few great discoveries, and if a great discoverer is someone who makes a great discovery, then such persons must be rare by definition. But there is a substantive question too. Does science depend, for its major progress, upon heroes who have faculties not possessed by journeymen scientists? Or are the men whose names we associate with the great discoveries just the lucky ones—those who had their hands on the lever at the precise moment when the jackpot showered its rewards.

A case could be made for either view, and my own hunch is that the truth lies somewhere between. If it is luck, a few men in each generation appear more skillful in wooing the goddess than are their fellows. On the other hand, I have encountered no evidence that there exist significant differences between the processes that great scientists use in achieving their discoveries and the processes used by those men we regard merely as "good" scientists.

The theory of scientific discovery I propose to set forth rests on the hypothesis that there are no qualitative differences between the *processes* of revolutionary science and of normal science, between work of high creativity and journeyman work. I shall not claim that the case can be proven conclusively. My main evidence will be data indicating that the processes that show up in relatively simple and humdrum forms of human problem solving are also the ones that show up when great scientists try to describe how they do their work. How convincing the evidence is can better be judged at the end of the chapter.

Let us return, then, to the problem-solving theory proposed in the last chapter and confront that theory with the recorded phenomena of scientific discovery.

The problem-solving theory asserted that thinking is an organization of elementary information processes, organized hierarchically and executed serially. In overall organization, the processes exhibit large amounts of highly selective trial-and-error search using rules of thumb, or heuristics, as bases for their selectivity. Among the prominent heuristic schemes are means-end analysis, planning and abstraction, factorization, and satisficing. Our task is to show how a system with these characteristics can behave like a scientist.

Selective Trial-and-Error Search

The prominence of selective trial-and-error processes in accounts of scientific discovery makes an extended discussion of this phenomenon unnecessary.[2] Examples of such accounts that come immediately to mind, out of a multitude that could be cited, are Hanson's analysis of the development of Kepler's theories (*Patterns of Discovery*, pp. 73–84), and Wertheimer's report of his conversations with Einstein on the theory of special relativity (*Productive Thinking*, Chapter 10).

Wertheimer's book is particularly interesting in this connection, because he can be regarded as a hostile witness. As a Gestaltist he maintains the greatest skepticism about the processes, like trial-and-error, postulated by associationists to account for problem solving. In fact, he almost never uses the phrase "trial and error" without prefixing the adjective "blind." His chapter certainly provides no evidence that Einstein engaged in "random" search. It does provide ample evidence that processes of which the thinker is aware. It assumes, further, that the organization of the totality of processes, conscious and unconscious, is fundamentally serial rather than parallel in time.

Our examination of the phenomena of incubation and illumination and their explanation will proceed in several stages. First, I shall describe briefly the phenomena themselves. Second, I shall consider the question of why the phenomena should be regarded as surprising and in what sense they require special explanation. Finally, the information-processing theory of problem solving will be applied to provide an explanation of the main features of incubation and illumination.

The phenomena themselves are relatively simple, and their occurrence is well documented. In the case of many important scientific discoveries (we do not know in what proportion of all cases), the discoverer reports three main stages in the progress of his inquiry. The first stage, which Hadamard calls "preparation," involves conscious, prolonged investigation that is more or less unsuccessful in solving, or sometimes even satisfactorily framing, the problem. Ultimately, frustration becomes intense, and the problem is dropped from conscious attention. Some time later, often suddenly and with little or no warning (as in the instance reported by Poincaré), or immediately upon awakening from sleep, the central idea for the solution presents itself to the conscious mind, only the details remaining to be worked out. The period between this illumination and the preceding preparation is the incubation period.

While there is little question about the phenomena, they provide no clues as to what goes on during incubation. In the absence of a full-fledged theory of problem solving, one can fill that period with almost any imaginable activity. Illumination is a vivid experience for the person who experiences it, because he is given no hint as to what occasioned the problem solution. Worse, since the incubation processes apparently go on independently of his conscious efforts to solve the problem (and best after these efforts have ceased), the experience gives him few cues as to what he should do when he next encounters a difficult problem—other than to "sleep on it." He must wait until the god decides to seize him.

We can see readily why the phenomenon should be puzzling and surprising to the illuminatee. The solution to a problem that has resisted his hardest efforts suddenly, and without further work, reveals itself to his conscious mind. The notions of continuity in space and time are intrinsic to most of our ideas of causation, and illumination appears to violate this continuity. One must say "appears" because, of course, the laws are only violated in the way they are violated when a magician produces a ible. If the third condition, superior heuristics, is chiefly responsible for the discovery, no more trial-and-error search will be present than would appear normal in cases of less creative activity.

The evidences of a high degree of persistence in pursuing fundamental problems are numerous in the biographies of creative scientists. Persistence does not always mean continual conscious preoccupation with the problem, or orderly, organized pursuit, but concern with the problem over a considerable period of years, indicated by recurrent attention to it. One could conjecture that while the biographies of "journeyman" scientists might reveal persistent attention to a problem *area* over comparable periods of time, the activity would more likely than in the case of highly creative scientists represent attacks upon, and solutions of, a whole series of relatively well-structured problems within the general area (e.g., determinations of structures of a number of molecules, or of the parameters of a system under a range of experimental conditions). However, the data on this point remain to be gathered.

A good deal less conjectural is the hypothesis that superior problem solvers in a particular area have more powerful heuristics and that they will produce adequate solutions with less search, or better solutions with equivalent search as compared with less competent persons. A. de Groot, for example, compared the searches of grandmasters and ordinary chess players for a good move in a middle-game position. Both classes of players searched for about the same length of time (which was partly an artifact of the laboratory situation), and examined approximately the same number of branches of the game tree. In fact, it was impossible to distinguish, from the statistics of the search, between the grandmasters and the ordinary players. They were easily distinguished by one datum, however: In the particular position examined, all five grandmasters attained better solutions to the problem (chose moves that could be shown to be objectively better) than any of the solutions attained by the ordinary players. While the grandmasters did not engage in more search than the others, their superior selective heuristics allowed them to search more significant and relevant parts of the game tree.[4]

Whence do the superior heuristics, the secret weapons, of the creative scientist come? Frequently, they derive from his possession of a superior technique of observation or of representation. Examples of the former are commonplace: Leeuwenhoek and his microscope, Galileo and his telescope, Lawrence and his cyclotron, and so on. God is on the side of the highest resolutions. The classic example of the interaction between apparatus for symbolizing or representation and scientific discovery is the relation of the calculus to the birth and growth of Newtonian mechanics. One might ask how the creative scientist comes to possess superior techniques. The answer would again be in terms of luck, persistence, and superior heuristics. The answer is not really circular, for it is quite legitimate, in dynamic systems, to explain chickens by the hatching of eggs, and eggs by the laying processes of chickens.

The theory of problem solving set forth in these two chapters itself provides an example of apparatus and representation as sources of heuristic. The idea that problem solving is a process of selective trial and error is an old one. The idea remained vague and largely untested until a formalism became available (list-processing language for computers) that was powerful enough to state the theory formally and precisely and until an instrument became available (the digital computer) that was powerful enough to draw out the implications and predictions of the theory for human problem-solving behavior. The scientists who have been active in developing and testing this theory were all in one way or another —sometimes in very "accidental" ways—thrown into contact with computers soon after these instruments were invented.

Incubation and Unconscious Processes in Discovery

The phenomena of incubation and sudden illumination have held immense fascination for those who have written on scientific discovery. Poincaré's experience on boarding the bus at Coutances takes its place in the annals of illumination along with Proust's madeleine dipped in tea:

> Just at this time I left Caen, where I was then living, to go on a geological excursion under the auspices of the school of mines. The changes of travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry.[5]

Hadamard places particular emphasis on the role of the unconscious in mathematical invention. While he proposes no specific theory of the processes that go on during incubation, he argues strongly that these are active processes and not merely a forgetting of material generated during conscious work that is inhibiting the problem solution.

The theory of problem solving proposed in the last chapter does not assign any special role to the unconscious—or, for that matter, to the conscious. It assumes, implicitly, that the information processes that occur without consciousness of them are of the same kinds as the processes of which the thinker is aware. It assumes, further, that the organization of the totality of processes, conscious and unconscious, is fundamentally serial rather than parallel in time.

Our examination of the phenomena of incubation and illumination and their explanation will proceed in several stages. First, I shall describe briefly the phenomena themselves. Second, I shall consider the question of why the phenomena should be regarded as surprising and in what sense they require special explanation. Finally, the information-processing theory of problem solving will be applied to provide an explanation of the main features of incubation and illumination.

The phenomena themselves are relatively simple, and their occurrence is well documented. In the case of many important scientific discoveries (we do not know in what proportion of all cases), the discoverer reports three main stages in the progress of his inquiry. The first stage, which Hadamard calls "preparation," involves conscious, prolonged investigation that is more or less unsuccessful in solving, or sometimes even satisfactorily framing, the problem. Ultimately, frustration becomes intense, and the problem is dropped from conscious attention. Some time later, often suddenly and with little or no warning (as in the instance reported by Poincaré), or immediately upon awakening from sleep, the central idea for the solution presents itself to the conscious mind, only the details remaining to be worked out. The period between this illumination and the preceding preparation is the incubation period.

While there is little question about the phenomena, they provide no clues as to what goes on during incubation. In the absence of a full-fledged theory of problem solving, one can fill that period with almost any imaginable activity. Illumination is a vivid experience for the person who experiences it, because he is given no hint as to what occasioned the problem solution. Worse, since the incubation processes apparently go on independently of his conscious efforts to solve the problem (and best after these efforts have ceased), the experience gives him few cues as to what he should do when he next encounters a difficult problem—other than to "sleep on it." He must wait until the god decides to seize him.

We can see readily why the phenomenon should be puzzling and surprising to the illuminatee. The solution to a problem that has resisted his hardest efforts suddenly, and without further work, reveals itself to his conscious mind. The notions of continuity in space and time are intrinsic to most of our ideas of causation, and illumination appears to violate this continuity. One must say "appears" because, of course, the laws are only violated in the way they are violated when a magician produces a rabbit from a hat. When we watch the magician, we do not cease to believe in the spatial and temporal continuity of causation, but only in our ability to observe the connections. The same distinction applies to illumination.

If illumination is surprising to a scientist who experiences it, it is less easy to see why it should surprise a psychologist.[6] It is commonplace that many, if not most, of the processes of the central nervous system are inaccessible to consciousness. The subconscious plays a major role in modern theories of motivation, emotion, and psychopathology. There is no a priori reason, then, to assign the problem-solving processes to the conscious rather than the unconscious. From the phenomenal evidence, they in fact belong to both.

I have been using the terms "conscious" and "unconscious" (or "subconscious"—for present purposes, no distinction is made between unconscious and subconscious) to distinguish between what a person is aware of and can report, and what he is not aware of and cannot report. The reports of illumination contain numerous instances that occurred immediately on awakening, but also numerous others that occurred when the discoverer had been awake for some time. Hence, "unconscious" is a more comprehensive term than "asleep." For the sake of parsimony, we shall assume that unconscious processes of the same kinds can occur both in the sleeping and waking states.

It has sometimes been argued that the evidence for unconscious processes is evidence that the information processing in the brain is parallel rather than serial. This argument only has force, of course, for unconscious processes that occur in the waking state when, presumably, they are operating in parallel with the conscious processes and are capable (viz., the Poincaré episode) of interrupting the latter. One can show, however, that a serial system is capable (through a "time-sharing" organization of its processing) of behaving in the observed manner, and the explanation I shall propose for illumination is compatible with either a serial or a parallel organization of cognitive processing.

With these preliminaries out of the way, let us return to incubation and illumination. I should like to describe two mechanisms currently employed in the information-processing theories that appear to go a long way toward accounting for these phenomena. The first of these mechanisms is called *familiarization,* the second is called *selective forgetting.* The familiarization mechanism emerged in the course of constructing a theory of human rote memory, the forgetting mechanism in the course of trying to discover why the organization of the first theorem-proving program, the Logic Theorist, was more effective in solving problems than the organization of early versions of the General Problem Solver. Neither mechanism was devised, then, with incubation and illumination in mind; they were introduced into the theory to meet other requirements imposed by the data on problem solving.

1. Familiarization. Thinking processes make use of certain means in the central nervous system for holding symbols in short-term or "immediate" memory. Little is known of the neurophysiological substrate of immediate memory, but a good deal is known about its phenomenal characteristics. Most important, the number of symbols that can be stored in immediate memory is severely limited—in George Miller's words, "seven, plus or minus two." But a "symbol" can serve as the name for anything that can be recognized as familiar and that has information associated with it in permanent memory. Thus "*a*" is a symbol; so is "Lincoln's Gettysburg Address." For most native speakers of English "criminal lawyer" is a symbol, but for a person just learning the language, the phrase may constitute a pair of symbols denoting a lawyer with certain antisocial tendencies.

The important facts are (1) that only about seven symbols can be held and manipulated in immediate memory at one time and (2) that anything can become a symbol through repeated exposure to it, or familiarization. Familiarization involves storing in *permanent* memory information that allows the symbol to be recognized and a single symbol or "name" to be substituted for it.

Since immediate memory can only hold a few symbols at a time, complex structures can only be acquired by gradually building them up from substructures which are formed, in turn, from still smaller substructures. As each substructure is learned and stored in permanent memory, the symbol that serves as its "name" internally can be used in immediate memory as a single chunk when combining it with other substructures. Thus, a total structure of unlimited size can be assembled without the need for holding more than a few symbols in immediate memory at any given moment. Lincoln's Gettysburg Address is memorized by assembling phrases out of words (which are already familiar units), sentences out of phrases, paragraphs out of sentences, and so on.

Familiarization processes, for reconciling the limits of immediate memory with the needs for storing information structures of unlimited size and complexity in permanent memory, are incorporated in the information-processing theory of memorization called EPAM (Elementary Perceiver and Memorizer), a program that has successfully accounted for a wide range of laboratory data on human memorizing.[7] We will assume here that these same processes go on during complex problem solving, so that in later stages of problem solving complex units are available that existed only as disconnected particulars at an earlier stage.

In proving mathematical theorems it is common first to introduce and prove some subsidiary theorems, or lemmas, which then enter as premises in the proof of the final theorem. The lemma serves to sum up a whole segment of the proof so that the name of the lemma can be used as premise in place of that segment. It should not be assumed that all or most familiarization is as deliberate or conscious as this use of lemmas by mathematicians, but the processes are analogical and perform the same function.

2. *Selective Forgetting.* A second mechanism to be found in information-processing theories of problem solving that is essential to our proposed explanation of incubation and illumination involves more rapid forgetting of some memory contents than of others. The selective forgetting rests, in turn, on the distinction between forms of short-term and long-term memory.

In the typical organization of a problem-solving program, the solution efforts are guided and controlled by a hierarchy or "tree" of goals and subgoals. Thus, the subject starts out with the goal of solving the original problem. In trying to reach this goal, he generates a subgoal that will take him part of the way (if it is achieved) and addresses himself to that subgoal. If the subgoal is achieved, he may then return to the now-modified original goal. If difficulties arise in achieving the subgoal, sub-subgoals may be erected to deal with them.

The operation of such a process requires the goal hierarchy to be held in memory. If a subgoal is achieved, it can be forgotten, but the tree of unattained goals must be retained. In human problem solvers this retention is not always perfect, of course. When part of the structure is lost, the subject says, "Where am I?" or "Now why was I trying to get that result?" and may have to go over some of the same ground to get back into context—i.e., to locate himself in that part of the tree that has been retained in memory. If we were designing such a system, instead of probing the one that human beings possess, we would specify that the goal tree be held in some kind of temporary memory, since it is a dynamic structure, whose function is to guide search, and it is not needed (or certainly not all of it) when the problem solution has been found. Our hypothesis is that human beings are also constructed in this way—that the goal tree is held in a relatively short-term memory.

During the course of problem solving, a second memory structure is being built up. First of all, new complexes are being familiarized, so that they can be handled by the processing system as units. In addition, the problem solver is noticing various features of the problem environment and is storing some of these in memory. If he is studying a chess position, for example, in the course of his explorations he may notice that a particular piece is undefended or that another piece is pinned against the queen.

This kind of information is perceived while the problem solver is addressing himself to particular subgoals. What use is made of it at the time it is noted depends on what subgoal is directing attention at that moment. But some of this information is also transferred to more permanent forms of memory and is associated with the problem environment—in this example, with the chess position. This information about the environment is used, in turn, in the processes that erect new subgoals and that work toward subgoal achievement. Hence, over the longer run, this information influences the growth of the subgoal tree. To have a short name for it (since it is now a familiar unit for us!), I will call the information about the task environment that is noticed in the course of problem solution and fixated in permanent (or relatively long-term) memory the "blackboard."

The course of problem solving, then, involves continuous inter-action between goal tree and blackboard.[8] In the course of pursuing goals, information is added to the blackboard. This information, in turn, helps to determine what new goals and subgoals will be set up. During periods of persistent activity, the problem solver will always be working in local goal contexts, and information added to the blackboard will be used, in the short run, only if it is relevant in those contexts.

What happens, now, if the problem solver removes himself from the task for a time? Information he has been holding in relatively short-term memory will begin to disappear, and to disappear more rapidly than information in long-term memory. But we have hypothesized that the goal tree is held in short-term memory, the blackboard in long-term memory. Hence, when the problem solver next takes up the task, many or most of the finer twigs and branches of the goal tree will have disappeared. He will begin again, with one of the higher level goals, to reconstruct that tree—but now with the help of a very different set of information, on the blackboard, than he had the first time he went down the tree.

In general, we would expect the problem solver, in his renewed examination of the problem, to follow a quite different path than he did originally. Since his blackboard now has better information about the problem environment than it did the first time, he has better cues to find the correct path. Under these circumstances (and remembering the tremendous differences a few hints can produce in problem solution), solutions may appear quickly that had previously eluded him in protracted search.

There is almost no direct evidence at the present time for the validity of this explanation of incubation and illumination. (I have been able, introspectively, to account for my most recent illumination experience quite simply in these terms, but perhaps my introspections are compromised as witnesses.) It invokes, however, only mechanisms that have already been incorporated in problem-solving theories. It does leave one aspect of the phenomena unaccounted for—it does not explain how the problem that the problem solver has temporarily (consciously) abandoned is put back on the agenda by unconscious processes. It does, however, account for the suddenness of solution without calling on the subconscious to perform elaborate processes, or processes different from those it and the conscious perform in the normal course of problem-solving activity. Nor does it postulate that the unconscious is capable of random searches through immense problem spaces for the solution.

It is difficult, in brief compass, to give an actual example of the tree-blackboard scheme in operation, but a schematized hypothetical example will show in general how the mechanism operates. Suppose that we assign "values" to nodes on the goal tree, the values representing estimates of the reward that could be achieved by searching further from the corresponding nodes. The purpose of the search is to find a node with a value of at least 20—such a node represents a solution of the problem (Figure 1).

A reasonable search rule, starting from any given node, would be to search next from the subbranch with the highest value. Thus, if the problem solver were at node *G* he would pick up branch *J*, with value 12, next, then the subbranch *P* (value 15) of that branch, the sub-subbranch *Q* (value 8), and so on.

Suppose that, in addition, each time a new node was generated, its name and value were added to a list on a blackboard, and that as soon as the subnodes of that node had been generated, the name and value of the node was erased. The blackboard would then contain, at any moment, the names and values of all nodes that had been generated but had not yet been explored. A possible search rule, different from the one previously mentioned, would be always to pick for next exploration the node on the blackboard with the highest value.

Using the first search rule, the search of this particular hypothetical tree would proceed: *A-B-E-G-J-P-Q-* . . . Using the second search rule, the search of the tree would proceed: *A-B-E-C-F-I-M*, reaching the solution. For, the branch *C* with value 11, generated at the same time as *B*, but not immediately investigated, would be selected from the blackboard in preference to the subgoal *G*, with value only 9, of goal *E*.

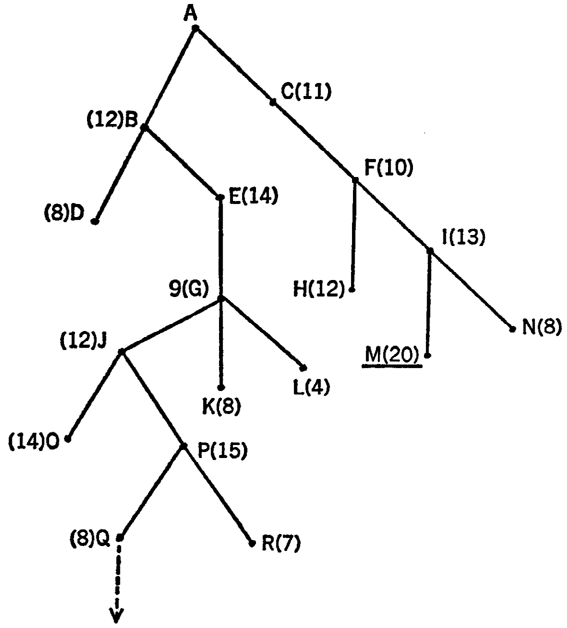

Figure 1

Now our theory of incubation and illumination derives from the hypothesis that during continued attention to a problem, search tends to be context-determined and to follow something like the first rule. During incubation, the tree disappears, leaving the blackboard, and when search resumes, it begins on the basis of the second rule.

Experiments with programs for discovering mating combinations in chess have shown that very different exploration trees are generated in game situations by the two rules, and that the second rule usually finds the mating combinations with far less search than the first. It would be easy, then, to reproduce incubation and illumination phenomena with these programs—by starting a search with a program using the first rule, but maintaining a blackboard, then at some point switching for a short time to the second rule.

As was mentioned earlier, the same point is demonstrated by comparison of the problem-solving power of the Logic Theorist with the power of early versions of the General Problem Solver. Without going into detail, it can simply be stated that the Logic Theorist used a form of the tree-and-blackboard scheme, while search in the General Problem Solver was always determined in the local context of current goals.[9]

Problem Formulation in Scientific Discovery

The theories described in the previous section postulate organized systems of processes which, when a problem of an appropriate kind is posed, will go to work on that problem and attempt to solve it. Scientific development involves not only solving problems, but posing them as well. In some discussions of creativity, asking the right questions is regarded as the crucial creative act; answering questions, in this view, is a relatively routine activity once the questions have been properly posed.

The view that question asking rather than question answering is the critical part of the creative process would be hard to defend in its extreme form. Perhaps it even illustrates itself, for in implying a sharp boundary between question asking and question answering, it may be posing the wrong question. If the issue were properly stated, we would see, first, that reformulation of questions—more generally, modification of representations—is one of the problem-solving processes; second, that the task of formulating a problem can itself be posed as a problem to a problem-solving system.

In exploring the relation of question asking to question answering, Thomas Kuhn's distinction between normal and revolutionary science becomes relevant. Normal science, he argues, does not have to pose its own questions. These questions have already been formulated for it by previous scientific revolutions. The textbooks and classics of science, incorporating the revolution, served "for a time implicitly to define the legitimate problems and methods of a research field for succeeding generations of practitioners." They can do this for two reasons: "Their achievement [is] sufficiently unprecedented to attract an enduring group of adherents away from competing modes of scientific activity. Simultaneously, it [is] sufficiently open-ended to leave all sorts of problems for the redefined group of practitioners to resolve."

Kuhn refers to achievements that share these two characteristics as "paradigms," and he defines normal science as scientific activity within the framework of received paradigms, revolutionary science as scientific activity that establishes new paradigms.[10] Within Kuhn's theory, it is easy to state who poses the problems for investigators engaged in normal science: Their problems come from the paradigms themselves. We must either define "creativity" so that it does not imply question asking as well as question answering, or we must conclude that creativity is not involved in normal science. The choice is one of definition.[11]

Is it necessary to adduce entirely new mechanisms to account for problem formulation in revolutionary science? Kuhn argues that it is not, for the paradigms of any given revolution arise out of the normal science of the previous period. Normal science, in Kuhn's account, leads to the discovery of anomalies, of facts that are difficult or impossible to reconcile with the accepted paradigms. The new problem then—the problem to which the prospective revolutionists address themselves—is to modify the paradigm, or replace it with another that is consistent with the facts, including the new anomalous ones.

In sum, we do not need a separate theory of problem formulation. A problem-solving system of the kind we have been considering—capable of generating subproblems from an initial problem, and capable of testing the adequacy of its solutions by generating new data about the environment—such a system will continue indefinitely to create new problems for itself. Problem formulation in science is to be understood by looking at the continuity of the whole stream of scientific endeavor.

A theory of scientific discovery adequate to explain revolutionary as well as normal science must account not only for the origins of problems, but for the origins of representations, of paradigms, as well. I do not underestimate the importance of this topic, but I shall not undertake to deal with it at any length here. In a previous paper, my colleagues A. Newell and J. C. Shaw, and I have made some general observations about it to which I refer the reader.[12] I shall add just a few more comments.

New representations, like new problems, do not spring from the brow of Zeus, but emerge by gradual—and very slow—stages. The caution stated in the opening pages of this chapter may be recalled: We must not overestimate the capacity of the human mind to invent new representations. The number of such inventions in human history has been very small.

Problem solvers use representations of the spatial relations of objects (engineering drawings are a relatively sophisticated and formalized example). They use representations of the abstract relations of objects (as, for example, in flow charts, genealogical charts, and chemical formulae). They use representations of programs (for example, systems of differential equations, computer programs). One can list a small number of other basic forms of representation and a somewhat larger number of specialized formats within each of these basic forms. The list is not long, and it is hard to find items for it whose history does not go back beyond the Christian era. (The *program* is probably the most recently developed broad form of representation, but it must not be forgotten that a recipe is a program, as is an algorithm like Eratosthenes' sieve. The differential equation represents a highly important subclass within this broad class of representations.)

Thus, our general answer to the question, "Where do representations come from?" is the same as our answer to the question, "Where do problems come from?" Representations arise by modification and development of previous representations as problems arise by modification and development of previous problems. A system that is to explain human problem solving and scientific discovery does not need to incorporate a highly powerful mechanism for inventing completely novel representations. If it did contain such a meachanism, it would be a poor theory, for it would predict far more novelty than occurs.

Conclusion

Theories are now available that incorporate mechanisms sufficient to account for some of the principal phenomena of problem solving in at least certain relatively well-structured situations. The aim of this chapter has been to ask how much these theories need to be modified or extended in order to account for problem solving in science. The general tenor of the argument has been that problem solving in science, like problem solving in the psychological laboratory, is a tedious, painstaking process of selective trial and error. Our knowledge of it does not suggest the presence of completely unknown processes far more powerful than those that have been observed in the laboratory.

Several kinds of objections can be raised, and have been, against this "minimalist" theory. One objection is that it does not account for striking phenomena like incubation and illumination. To meet this objection, a mechanism has been proposed that is believed sufficient to produce exactly these kinds of phenomena.

Another objection is that the theory only explains how problems are solved that have already been stated and for which there exist well-defined representations. This objection has not been answered in detail, but an answer has been sketched in terms of the broader social environment within which scientific work takes place. Most scientific activity goes on within the framework of established paradigms. Even in revolutionary science, which creates those paradigms, the problems and representations are rooted in the past; they are not created out of whole cloth.

We are still very far from a complete understanding of the whole structure of the psychological processes involved in making scientific discoveries. But perhaps our analysis makes somewhat more plausible the hypothesis that at the core of this structure is the same kind of selective trial-and-error search that has already been shown to constitute the basis for human problem-solving activity in the psychological laboratory.

NOTES

The preparation of this chapter has been aided by research grants from The Carnegie Corporation and the National Institutes of Health (MH-07722-01). Most of the ideas in it have been developed in collaboration with my colleague Allen Newell. See his "Some Problems of Basic Organization in Problem-Solving Programs," in *Self-Organizing Systems,* eds. Yovits, Jacobi, and Goldstein (New York: Spartan Press, 1962).

1. Obviously, I am not immune to this tendency.
2. For further discussion, see the previous chapter and A. Newell, J. C. Shaw, and H. A. Simon, "The Processes of Creative Thinking" in Gruber *Contemporary Approaches to Creative Thinking,* eds. Gruber, Terrell, and Wertheimer (New York: Atherton Press, 1962).
3. There are numerous anecdotes, some true, some fictitious, about the role of luck in invention. It is clear, however, that chance events played a role in: discovering vulcanization of rubber, the sweetening power of saccharine, developers in photography, and many other discoveries. See Joseph Rossman, *The Psychology of the Inventor* (Washington: The Inventors Publishing Co., 1931), Chap. 7.
4. A. de Groot, *Thought and Choice in Chess* (Amsterdam: Mouton, 1965).
5. Henri Poincaré, *Mathematical Creation,* reprinted in *The World of Mathematics,* ed. James R. Newman, IV, 2041–50.
6. Mary Henle begins her essay on "The Birth and Death of Ideas" with the sentence, "Perhaps the most astonishing thing about creative thinking is that creative thinkers can tell us so little about it" (in *Contemporary Approaches to Creative Thinking,* Chap. 1). Why astonishing? Would we say: "Perhaps the most astonishing thing about neurotic behavior is that persons suffering from neuroses can tell us so little about it?" Why would we expect, a priori, self-consciousness to be more characteristic of the one than of the other?
7. For an introduction to EPAM see E. A. Feigenbaum, "The Simulation of Verbal Learning Behavior," pp. 297–309 in *Computers and Thought,* eds. Feigenbaum and Feldman (New York: McGraw-Hill, 1964).
8. The role of goal tree and blackboard in the organization of problem solving have been discussed by Allen Newell, in "Some Problems of Basic Organization."

9. See A. Newell, J. C. Shaw, and H. A. Simon, "Empirical Explorations of the Logic Theory Machine," and A. Newell and H. A. Simon, "GPS, A Program that Simulates Human Thought," in Feigenbaum and Feldman, eds., pp. 109–33, 279–93.
10. Kuhn, pp. 10–12.
11. This account elides some important details. Generating subgoals from a more general goal is a form of question asking also, which is a part both of normal science and of our problem-solving theories. Since this process has already been considered, our only present concern is with problems whose generation cannot be explained in this way.
12. "The Processes of Creative Thinking," pp. 98–104.