Title
A computational model of scientific insight

Authors
Langley, Pat
Jones, Randolph

Publication Date
1986-12-31

Peer reviewed

A Computational Model of Scientific Insight

Pat Langley
Randolph Jones
Irvine Computational Intelligence Project
Department of Information and Computer Science
University of California, Irvine, CA 92717

Technical Report 87–01

December 31, 1986

This work was supported by Contract N00014–84–K–0345 from the Information Sciences Division, Office of Naval Research. We would like to thank Bernd Nordhausen, Donald Rose, Rogers Hall, and Jaime Carbonell for discussions that led to the ideas in this paper.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 9	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* A Computational Model of Scientific Insight		5. TYPE OF REPORT & PERIOD COVERED Interim Report 7/86–12/86
		6. PERFORMING ORG. REPORT NUMBER UCI–ICS Technical Report 87–01
7. AUTHOR*(s)* Pat Langley Randolph Jones		8. CONTRACT OR GRANT NUMBER*(s)* N00014–84–K–0345
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Information & Computer Science University of California, Irvine, CA 92717		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Information Sciences Division Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE December 31, 1986
		13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

To appear in R. Sternberg (Ed.), *The Nature of Creativity*, Cambridge University Press.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

cognitive modeling
scientific discovery
analogy
spreading activation
machine learning
scientific insight
retrieval

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

OVER

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. ABSTRACT

Scientific discoveries often take the form of *insight*, in which previously unseen and unexpected connections suddenly reveal themselves to the mind. In this paper, we present a computational theory of this phenomenon. We recount a number of well-known examples of the process, along with some early and recent theories that attempt to explain the phenomenon. However, our reservations about these theories have led us to develop an alternative model. We explain insight as the sudden retrieval of an analogy from long-term memory. To model this retrieval, we use a form of *spreading activation.* In constructing our framework, we have built on research in two separate lines of research in cognitive science — on reasoning by analogy and qualitative models. Thus, we also review some work in these areas.

Introduction

Creativity lies at the heart of the scientific process. Although much of science involves the dreary application of well-worn methods, true progress requires an act of discovery. In some cases, these discoveries take the form of *insight*, in which previously unseen and unexpected connections suddenly reveal themselves to the mind.

Introspectively, the moment of insight often contains a 'mystical' quality, and this has led many to assume the process lies outside the realm of human understanding. Early theories of scientific insight shared in this feeling, relying heavily on notions of unconscious (and thus noninspectable) processing. But in the past few decades, cognitive psychology and artificial intelligence have made significant strides in understanding the nature of human cognition. It seems only natural to apply their methods to develop a process explanation of this intriguing phenomenon.

In this paper, we present just such a computational theory of scientific insight. We begin by recounting some well-known examples of the process, along with some early theories that attempted to account for the phenomenon. We also review some more recent attempts to explain insight in process terms, but our reservations about these models have led us to develop an alternative theory. Our framework builds on two separate lines of research in cognitive science – on reasoning by analogy and on qualitative mental models. Thus, we also review some work in these areas before moving on to the details of our model.

The Advantages of Cognitive Simulation

Before addressing the substantive issues, we should briefly consider our methodological assumptions. One of our basic tenets is that the construction of *cognitive simulations* can improve our understanding of human behavior. A cognitive simulation is simply a computer program that is intended to model human cognitive processes in some area. This approach has proved successful in a wide variety of domains, including problem solving (e.g., Newell & Simon, 1972), vision (e.g., Marr, 1982), natural language (e.g., Schank & Abelson, 1977), and memory (e.g., Anderson & Bower, 1973).

The cognitive simulation approach has a number of advantages over more traditional psychological methods. First, the act of constructing a running computer program ensures that one's theory is internally consistent. Second, one can determine the consequences of changing a theory by adjusting the computational model and observing the new behavior. Most important, it forces one to think in terms of specific representations of knowledge and to explicitly specify processes for manipulating those representations. This leads to more specific – and thus more testable – models of cognitive behavior. We refer the reader to Newell and Simon (1972) and Anderson (1976) for additional discussion of this methodology.

The goal of our research is to construct a running cognitive simulation of scientific insight. Although we have not yet achieved that goal, we believe the very act of thinking in process terms has revealed aspects of insight that we would otherwise have missed.

The Problem Space Hypothesis

Much of the research within the cognitive simulation approach relies on what Newell (1980) has called the *problem space hypothesis.* This states that all cognitive behavior involves search through some problem space. A problem state is composed of a set of *problem states*, including the initial state from which search begins. New states are generated by applying *operators* to existing states, letting one systematically explore the space until the goal state has been reached.

As an example, suppose we wanted to solve some problem in linear algebra. The initial state might be a set of n equations in n unknowns, such as

$$2x + 3y = 8$$
$$3x - 6y = -9$$

In this case, our goal would be to find some value for each unknown. There are two operators for generating new states – adding two equations together and multiplying an equation by a constant. An intermediate state for the above problem might include the equations

$$4x + 6y = 16$$
$$3x - 6y = -9$$

By applying the right operators in the right order, we would eventually reach the goal state, which would tell us that $x = 1$ and $y = 2$.

Unfortunately, the problem space for most interesting tasks are combinatorial in nature, so that many alternative paths present themselves. One response is to carry out an exhaustive search of the problem space, but this rapidly becomes unmanageable for even simple domains. A more reasonable approach is to carry out a *heuristic* search of the problem space, using rules of the thumb to suggest likely states to expand and likely operators to select. This approach is not guaranteed to find an optimal solution, but it is likely to produce an acceptable solution in reasonable time. Humans problem solvers appear to rely heavily on heuristic search methods.

The problem space hypothesis has been quite successful within artificial intelligence and cognitive science, and we will see later that most explanations of insight have been formulated within this framework. In fact, the problem space approach has become so popular in some circles that many view it as 'truth' rather than as an hypothesis. Nevertheless, one can imagine competing frameworks for describing cognition, and as we will see, our theory of scientific insight incorporates such an alternative approach, based on the joint notions of mental models and reasoning by analogy.

The Phenomenon of Scientific Insight

The popular view of science assumes that progress occurs through the methodical collection of data and careful inferences from those observations. Although certain scientific work occurs in this mode, real progress often seems to require a 'leap of intuition' or a 'flash of insight', in which an old problem is suddenly seen in a different light. Let us consider some examples of this phenomenon.

Probably the most famous instance of scientific insight is Archimedes' discovery of the principle of displacement (Dreistadt, 1968). The Greek scholar had been given the problem of determining whether the king's crown was pure gold, or whether the gold was mixed with silver. Knowing the density of gold and the weight of the crown, he needed only to find its volume in order to check for purity. But the crown's shape was irregular, and he could not measure its volume without melting it down again. Archimedes worked on the problem for some time without finding a solution. Then, as he lowered himself into a bath, he noticed that the water level rose simultaneously. With this came the realization that any object displaces its own volume when submerged in a liquid, and that this provided the means for measuring irregular volumes.*

Another well-known example of scientific insight is Louis Kekulé's discovery of the ring structure of the benzene molecule. The scientist tried for some time to identify a structural model that would account for benzene's chemical makeup. Finally, he sat down by the fire and began to doze (Dreistadt, 1968; Farber, 1966). In his sleepy state he watched the smoke rising from the fire, 'twisting in a snakelike motion'. At this point, one of the snakes took its own tail in its mouth, creating a ring. In a sudden flash, Kekulé realized the molecule must be structured as a ring.

Insights seem to be fairly common in mathematics, and the eminent French mathematician Henri Poincaré (1952) reported a number of his own insights in a lecture at the Société de Psychologie in Paris. In one particularly striking example, he detailed his discovery of an expression for Fuchsian functions:

> At this moment I left Caen, where I was then living, to take part in a geological conference arranged by the School of Mines. The incidents of the journey made me forget my mathematical work. When we arrived at Coutances, we got into a [bus] to go for a drive, and, just as I put my foot on the step, the idea came to me, though nothing in my former thoughts seemed to have prepared me for it, that the transformations I had used to define Fuchsian functions were identical with those of non-Euclidean geometry. (p. 53)

This case differs from our earlier examples in the lack of any obvious external stimulus that is closely related to the insight. We will return to this issue later, since it bears on our theory.

Hadamard's Theory of Scientific Insight

Hadamard (1949) gives us a splendid discussion of the phenomenon of insight. In addition to reviewing numerous instances from the history of science, he identifies four distinct stages that seem to occur in every documented case of scientific insight – *preparation*, *incubation*, *illumination*, and *verification*. These stages and their characteristics constitute a set of

* It is said that Archimedes' joy at this insight was so great that he leaped from his bath and ran naked through the streets of Syracuse, exclaiming 'Eureka!', or 'I have found it!'.

empirical generalizations relating to insight, and any successful theory must account for their existence.

The preparation phase involves intense effort in attempting to solve some problem. In some cases this attempt leads directly to a solution, but for especially difficult problems one eventually 'gives up'. This abandonment constitutes the entry into the incubation stage, during which the problem solver devotes his conscious processing to other issues. Depending on the situation, incubation can last anywhere from seconds to years, but eventually the solution 'proposes itself' during the illumination stage, which occurs both unexpectedly and very rapidly. This is the 'aha' experience that produces exclamations like the 'Eureka' of Archimedes. However, these 'leaps of intuition' are not always valid, and sometimes lead to 'false insights'. Thus, one must still check the details during the final verification phase.

In addition to describing these four stages and their relation to one another, Hadamard also proposes a theory of insight which gives a major role to unconscious reasoning. His explanation assumes three levels of the mind which work together during the process of discovery – the fully conscious, the fringe conscious, and the unconscious. The first refers to our everyday mode of thought, in which we are aware of the mental steps we traverse. The unconscious refers to thought processes that are not available to introspection, of which we are not even aware. The fringe conscious occupies the gray area between these two extremes, in which we are aware of ideas but not focusing on them. One can view this as the 'peripheral vision' area of the mind.

Hadamard's theory states that the preparation stage involves only conscious thought. However, the mental activity during preparation serves to 'stir up' ideas relevant to the problem at hand. During the incubation phase, the unconscious mode takes charge and considers alternative solutions that incorporate the ideas produced during the earlier preparation. When the unconscious encounters an especially promising combination, it deposits the result into the fringe conscious. The mind seizes upon this new idea and experiences the flash of insight as it enters full consciousness. Finally, one continues in the conscious mode while the result is checked.

Clearly, most of the action in this theory is occurring at the unconscious level, and it is natural to ask how this mechanism manages to sift through so many ideas and distinguish the profitable ones from others. Hadamard argues that the unconscious is able to generate combinations of ideas that are specific enough to be fruitful and yet general enough not to miss the solution completely. This process is likened to the scattering of a hunting cartridge. The pellets are spread enough so that one does not miss the target, yet not so much that it is useless to aim. Hadamard concludes that great mathematicians differ from ordinary people in the selective ability of their unconscious, which lets them generate ideas that are aesthetically pleasing or interesting.

Ohlsson's Restructuring Theory

Ohlsson (1984a, 1984b) has proposed a computational model of insight by attempting to integrate ideas from Gestalt psychology with the problem space framework. In the Gestalt paradigm, every situation was characterized by some *structure* in the mind. These structures

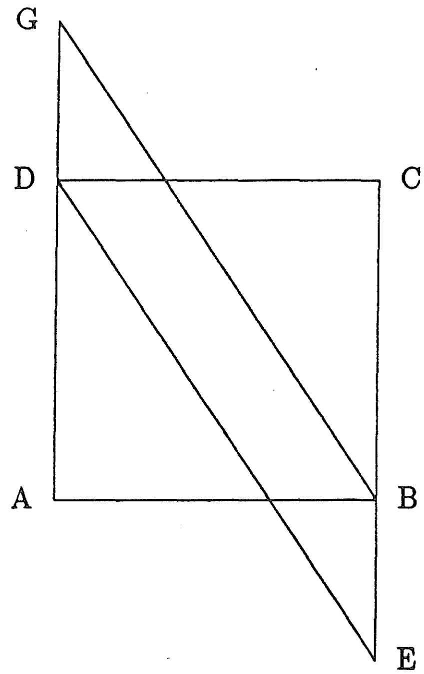

Figure 1. Find the sum of the areas of square $ABCD$ and parallelogram $EBGD$, given $\overline{AB} = a$ and $\overline{AG} = b$.

were influenced by *forces* which could become unbalanced and introduce gaps. An unsolved problem was viewed as some situation in which gaps existed between one's current state and the goal state. When forces became unbalanced enough, *restructuring* occurred and some new configuration was produced. Gestalt psychologists claimed that these restructuring events are more likely to occur when the problem solver has carefully analyzed the problem, carefully analyzed the goal, and made a series of unsuccessful attempts at solving the problem.

According to the problem space hypothesis, normal problem solving involves a search through a problem space. Ohlsson claims that restructuring requires search through the *description space* for a problem. That is, restructuring involves finding a different way to *look* at the problem, rather than trying to solve the problem in a straightforward manner. He further assumes that humans are able to 'look ahead' a few steps, and that this lets them know when they are near their goal. When a problem solver encounters an impasse, he attempts to view the problem in a different light. This can lead to a new representation for the problem, which constitutes restructuring. In some cases, this representational shift leads to a state that is only a few steps from the goal; the shift combines with the look ahead ability to produce a flash of insight.

As an example, Ohlsson presents the problem shown in Figure 1, in which one must compute the sum of the areas of a square and an overlapping parallelogram. The straightforward solution is to calculate the area of the square and the area of the parallelogram (which requires calculating the base of the parallelogram) and adding them together. Most people do not know the formula for the area of a parallelogram, and so cannot solve the problem with the information provided. This causes an impasse and this in turn leads to restructuring.

One such alteration lets the problem solver view the picture instead as two overlapping triangles (DCE and GAB). Given this representation, one can calculate the areas of the two triangles and add the results. These operations are simple, since the base and height of the triangles are given in the problem statement. The feeling of insight might or might not occur in this case, depending on whether the problem solver can look ahead the required three steps.

Another restructering simplifies the problem even further. If one notices that the triangles can be 'slipped apart' to form a rectangle, then one need only calculate the area of that rectangle, using the base and height already given. In this case, the feeling of insight is almost certain to occur, since the goal state is only two steps away from the initial state in this new space.

Simon's Theory of Familiarization and Selective Forgetting

Simon (1977) has also proposed a computational explanation of Hadamard's four stages of insight. The theory combines models of human memory with information processing models of problem solving. Research on human short-term memory has shown that its capacity is severely limited, but it has also shown that this limitation can be offset by experience. When asked to remember 'artificial' input such as digits or letters, people can hold only about seven symbols in memory at one time. However, given sufficient experience in a domain, humans form *chunks* to describe regularities in that domain. Simon calls this the process of *familiarization.* Memory experiments show that subjects can hold approximately the same number of chunks in short-term memory, regardless of the complexity of the chunks. Thus, one can remember around seven letters, seven familiar words, or even seven familiar sentences. An extreme example is the Gettysburg Address, which is composed of sentences, which are in turn composed of phrases, which are themselves composed of certain words, and so forth.

Simon proposes that familiarization occurs during conscious problem solving of the type that characterizes Hadamard's preparation stage. As the problem solver carries out a heuristic search through the problem space, he also builds up higher level structures that describe regularities in that space. The goals and states that are generated during search are held in short-term memory, while these chunks are stored in long-term memory. On difficult problems, the problem solver can easily get lost and be forced to start over from the beginning. Meanwhile, he is becoming more familiar with the structure of the space and its components. Eventually, he may decide the problem is too difficult and abandon his efforts. At this point, the structures in short-term memory will fade rapidly, but the chunks that have been stored in long-term memory will remain. Simon terms this process *selective forgetting.*

Later the problem solver may reexamine the troublesome problem. However, this time he has a powerful repertoire of chunks available in long-term memory, and these lead the search down a quite different path than on earlier occasions. These chunks may allow the person to move directly to the goal, and in some cases this may occur so rapidly as to produce the experience of illumination. The process of familiarization, combined with the mechanism of selective forgetting, gives the problem solver a new approach, thus transforming a difficult problem into a straightforward one.

Commentary

Let us consider the similarities and differences between Hadamard's, Ohlsson's, and Simon's theories of the insight process. All assume that Hadamard's four-stage model provides a reasonable description of the phenomena, and concentrate on explaining the processes that underly the different stages. Furthermore, all agree that the preparation and verification stages involve conscious problem solving, though Ohlsson and Simon give more detail, since they can build on the results of modern cognitive psychology.

The theories differ in their treatment of the incubation and illumination stages. Although he does not cast it in quite these terms, Hadamard argues that incubation involves a search through the space of idea combinations. This search is carried out by unconscious mechanisms, which employ measures of interestingness or elegance, both to select promising candidates and to decide when a likely solution has been found. Illumination is secondary in this framework, serving only to notify the conscious mind of the solution. Most of the interesting action occurs in the unconscious during incubation, though the preparation stage also serves to 'stir up' the ideas that are used by the unconscious.

However, developments in cognitive psychology strongly suggest that search of this kind requires conscious attention. Thus, Simon rejects the notion of an unconscious that can selectively search large problem spaces of the sort required for many scientific discoveries. He replaces Hadamard's unconscious search scheme with two much simpler unconscious processes – familiarization and selective forgetting. The first of these occurs during the preparation stage, while the second occurs during incubation. Together, they clear the way for conscious problem solving mechanisms to find a solution during illumination. This stage occurs so quickly because the chunks acquired in the preparation phase make the search process trivial. This explanation is much more distributed than Hadamard's, assigning significant roles to each stage.

In Ohlsson's theory the major action occurs during illumination, when the problem solver restructures the problem description so that its solution becomes obvious. This explanation does not attempt to account for the role of incubation, and in fact this stage is not even mentioned in the theory. Presumably, Ohlsson would argue that in some cases restructuring does not occur until some time after an impasse is reached, but this does not explain what causes restructuring when it does occur.

Although each of these theories of insight have their attractions, we are not satisfied that any provides an adequate explanation of the phenomena. Hadamard attributes powerful search capabilities to the unconscious that contradict the findings of cognitive psychology. Simon invokes the more plausible mechanisms of familiarization and selective forgetting, but he does not explain why the problem solver returns to the problem when he does. Finally, Ohlsson posits a restructuring process that generates a new, simpler problem space. Like Simon's theory, this framework is consistent with our knowledge of the human information processing system, but it does not explain the incubation stage.

In the following pages, we propose an alternative theory of scientific insight that diverges from the existing theories along a number of dimensions. One difference is that it does not rely on the problem space hypothesis, as do the approaches of Simon, Ohlsson, and even Hadamard. Rather, we assume that insight is a memory-related phenomenon that centers on mechanisms of indexing and retrieval. Another distinguishing feature of our framework is the central role played by analogy. Given the importance of this mechanism to our work, we will diverge slightly to review some earlier work on the topic.

Research on Reasoning by Analogy

Looking back on our examples of scientific insight, it becomes apparent that all involved some form of *analogy*. Archimedes formed an analogy between his body submerged in the bath and the king's crown submerged in a container of known volume. Kekulé formed a mapping between a snake biting its own tail and the benzene ring. Finally, Poincaré's insight was based on an analogy between the Fuchsian transformations and non-Euclidean transformations.

Polya (1945), Sternberg (1977), and others have argued for the importance of analogy in human cognition. Thus, it would not be surprising to find analogy occurring in scientific discovery. We will argue that this mechanism plays an important role in many (though not necessarily all) cases of insight, and analogy occupies a central position in our theory of that phenomenon. But before describing this theory, let us first review some previous work on analogy itself.

Dreistadt's Analogy-Based Theory of Insight

Dreistadt (1968) has also noted the role analogy in historical examples of insight. He summarizes his own theory of insight in the following words:

> This writer explains insight as occurring when one finds a stimulus pattern (the analogy) in which parts of the form or structure are like the structure of the problem-situation and the rest of the structure of this stimulus pattern (the analogy) indicates how to organize the unintegrated materials of the problem or how to reorganize the problem by putting the parts that are out of place into their correct place, or both, thereby completing the whole which is then the solution of the problem. (p. 111)

In other words, an insight occurs when the problem solver finds some analogy that is similar to the current problem, and that suggests a different view on the problem that makes its solution clear. In this framework, most of the action occurs during the illumination stage, which involves the discovery of a suitable analogy.

To test this hypothesis, Dreistadt (1969) performed a number of experiments to determine the influence of analogies and incubation periods on subjects' ability to solve problems. One group of subjects was given twenty minutes in which to solve a set of tricky problems. A second group was given the same amount of time to solve the same problems, but was also presented with pictures which contained analogical 'hints' to help find the solution. However, this group was not told the purpose of the pictures. A third group was allowed five minutes to concentrate on the problem, then was given an eight minute 'incubation' period (involving a distracting activity), and finally was given seven more minutes to solve the problem. A final group was presented with the pictorial analogies *and* given an incubation period.

Dreistadt measured both the number of correct solutions in each group and the closeness of their incorrect answers. He also interviewed subjects about their impressions of the problem solving task. He found that pictorial analogies significantly aided the solution process, even though subjects were not always aware they had been given a hint. Incubation alone did not seem to help in problem solving, but there was some evidence that incubation enhanced the effect of the pictorial analogies. These results lend credibility to the belief that analogies are important in scientific insight, and we will return to this view later in the paper.

Hall's Framework for Analogy

Although Dreistadt presented evidence for the role of analogy during insight, he did not suggest details for this process. However, a number of researchers within AI and cognitive science have described computational models of analogy. Hall (1986) provides an excellent review of these alternative approaches and suggests an organizing framework for research on analogy. This framework includes four components – recognition, elaboration, evaluation, and consolidation.

Reasoning by analogy involves mapping from some existing structure, the *source*, onto some new structure, the *target*. One typically begins with an incomplete description of the target. The first step involves retrieving a plausible source from long-term memory; this is the *recognition* process. Once a likely source has been identified, one must *evaluate* the analogy to ensure that it is reasonable. Assuming the mapping is acceptable, one then carries over relevant aspects of the source to fill out the target description; this is the *elaboration* stage. Finally, for successful analogies one may want to store an abstract description in memory to simplify retrieval in future situations; this is the *consolidation* process.

We will use this framework in our discussion of the three particular computational models of analogy that we consider below. We should note that much of the work on analogy focuses on learning tasks, and the consolidation stage plays an important role in this context. However, our focus is on scientific discovery and insight, and consolidation seems less relevant for this domain. Also, Hall's framework downplays the need to store and index experiences in long-term memory before recognition/retrieval can occur. We will include this earlier step in our treatment of analogy.

Gentner's Structure Mapping Theory

Gentner (1983) has put forth a *structure mapping* theory that attempts to distinguish useful analogies from poor ones. This framework assumes that memory contains representations of objects linked together by predicates. Some predicates accept only one argument, while others relate two or more arguments. The attribute *red* is an example of the former; the relation *larger* is an example of the latter. Gentner makes a further distinction between first-order predicates, which relate objects, and second-order predicates, which relate other predicates. The relation *larger* is an example of the first, while *cause* is an example of the second.

The structure mapping theory claims that single-argument attributes are useful when noting *similarities* between two situations, but relations are more important for drawing *analogies.* For example, the statement 'The X12 star system in the Andromeda galaxy is like our solar system' involves a similarity, implying that the X12 star is yellow, hot, about the same size as Sol, and so forth. In contrast, the statement 'The hydrogen atom is like our solar system' involves an analogy. In this case, we certainly do not mean that the hydrogen atom is hot and yellow, but we do mean that certain objects (electrons) revolve around the atom, more or less as planets revolve around the sun.

Gentner's theory does not address the issue of retrieving or recognizing analogies, but it does provide criteria for evaluating their quality and it does suggest principles for carrying out elaboration. The theory can be summarized by three mapping rules:

1. Disregard attributes of objects, such as size or color;
2. Try to preserve relations between objects;
3. In deciding which relations to preserve, select those which retain consistency among higher-order relations.

Gentner refers to the third rule as the *systematicity principle.* The reasoning behind this principle is that the best analogies retain the highest-order relations.

As an example, consider the partial representation of the solar system shown in Figure 2 (a). If we state that 'the atom is like the solar system', the structure mapping theory predicts that only those relations presented in Figure 2 (b) would be carried over. In this case, the sun corresponds to the nucleus of the atom and the planet maps into the electron. Notice that none of the sun's attributes are carried along, nor is the fact that the sun is hotter than its planet, since this relation is not involved in the higher-order *cause* relation.

Winston's Theory of Analogy

Winston (1980) has proposed an alternative theory that focuses on different aspects of the analogical reasoning process. As in Gentner's framework, memory consists of objects linked together by relations, and together these form schemas describing some connected set of events. But Winston provides much more than this; his theory also addresses the issues of indexing and retrieval.

When a new event or description is stored in memory, it is indexed by the type of object it contains. For example, if one reads a story about a wicked stepmother and a beautiful girl, the schema summarizing the story would be indexed through those concepts. Later, when one reads another story involving a wicked stepmother or a beautiful girl, one would be reminded of the earlier schema. The actual process is both more complex and more general than this account suggests. Winston organizes concepts in is-a hierarchies, so that 'wicked stepmother' would be stored as a subtype of 'stepmother', this would be stored as a subtype of 'parent', and so forth.

Thus, a story involving any type of parent might remind the reader of the original schema,

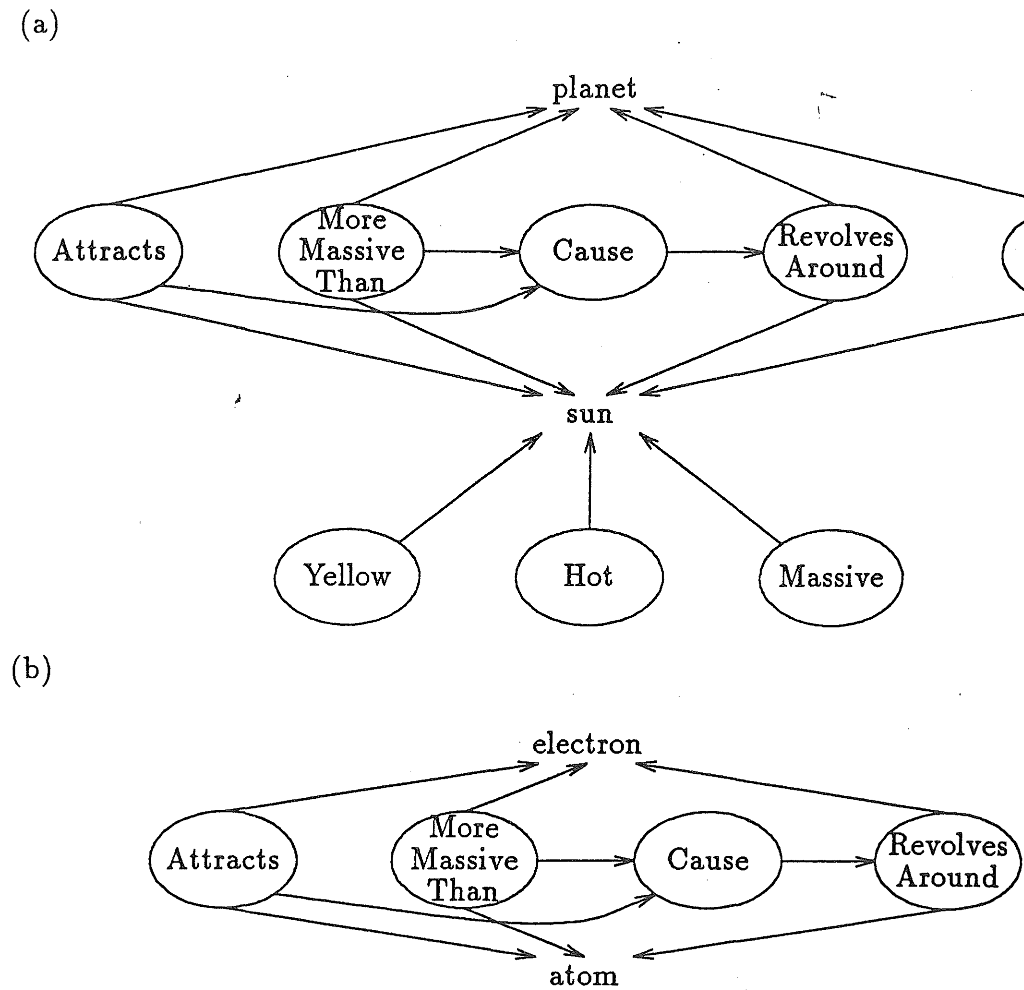

Figure 2. Creating a representation for the atom from the statement, "The atom is like the solar system." Higher order relations are carried over and simple attributes are ignored.

though to a lesser extent. But this extension means that, potentially, any new experience could remind one of any earlier experience. Winston responds to this issue by preferring connections that are more discriminating during the retrieval process. For instance, fewer stories would be indexed by 'wicked stepmother' than by the more general 'parent' concept. As a result, a new story containing a wicked stepmother would be more likely to remind one of an earlier story with a wicked stepmother than stories containing other kinds of parents. This approach has some similarities to 'spreading activation' models of retrieval.

Once a plausible source for the analogy has been established in this manner, Winston's model compares all possible mappings between the source and the target, and then evaluates them according to their degree of match. This evaluation process gives preference to higher order relations, but it also takes objects into account. The approach also differs from Gentner's in that the elaboration process carries over both relations *and* attributes. Finally, the method consolidates its analogically-based findings by transforming them into production rules, but this process need not concern us here. Winston has tested his mechanism in a number of domains, including story understanding and electric circuits.

Carbonell's Theory of Derivational Analogy

Carbonell (1986) has explored the use of analogy in problem solving. When one encounters a completely novel problem, the only choice is to employ weak problem solving methods such as heuristic search or means-ends analysis. However, when the current problem shares features with an earlier problem for which the solution is already known, one can use this knowledge to direct search on the current task.

But Carbonell argues that superficial similarities in the problem structure and solution path are less important than the *reasons* why particular steps were taken. He suggests that weak problem solving methods lay down a derivational trace in memory. This trace contains not only the final solution path, but includes failed paths and the reasons why these paths did not lead to a solution. It also includes subgoals and the reason for their creation.

When a problem solver encounters a new problem, he begins applying weak methods to search for a solution. If the initial reasoning steps are similar to those used in another problem, then he retrieves the earlier problem and its derivational trace. Once a plausible analogy has been identified, the mapping is elaborated by 'replaying' the derivations from the earlier problem and checking to see that similar derivations carry over to the new problem. In cases where the same reasoning holds for both problems, the analogous path is followed. In other cases the analogous derivation does not hold, and some different justification must be found for problem solving to proceed. As in Gentner's theory, the emphasis here is on evaluation and elaboration, though Carbonell also addresses the problems of indexing, retrieval, and consolidation.

Commentary

Before moving on, we should attempt to extract some useful lessons from the earlier work on analogy. One immediate observation is that these theories are weak on the side of indexing and retrieval. Gentner does not even address this issue, and Carbonell must carry out some search before being reminded of an earlier problem. Winston provides the most coherent model of retrieval, employing a mechanism similar to Anderson's (1983) spreading activation theory. This simple mechanism has two interesting properties. First, it seems quite plausible that activation could spread in parallel, and that this process could occur at unconscious levels. Second, the mechanism may sometimes suggest very poor analogies, since it focuses on the types of objects shared by two situations and not on their relations. Later, we will see that these features of spreading activation play an important role in our theory of insight.

All three theories share a concern with the elaboration process, in which one constructs a detailed mapping between source and target and carries over relevant structures. However, they disagree significantly on methods for determining relevance. Winston uses causal relations to select an optimal mapping, but once this has been established he carries along all consistent structures. In contrast, Gentner explicitly abandons single-argument predicates and retains only those relations that occur as arguments of higher-level relations. This naturally leads to a more abstract description of the target than does Winston's approach. However, Gentner's emphasis on the number and nature of arguments seems somewhat ad hoc. Carbonell's claim is more elegant – that one carries over only those structures that were used in *deriving* the source structures. We will reinvoke this idea later as well.

An Alternative Theory of Insight

We are now ready to present an alternative theory of scientific insight. In many ways, our framework is an extension of Dreistadt's analogy-based scheme, with computational ideas borrowed from Carbonell, though Gentner and Winston have also influenced our thinking. Below we present an overview of the theory and its differences from earlier frameworks. After this, we consider the details of the theory and the phenomena it covers.

An Overview of the Theory

The basic tenet of our theory is that insight does not result from search through a problem space, but rather is a *memory-related* phenomenon. Moreover, the process of insight often involves the recognition, evaluation, and elaboration of analogies. We will not claim that all insights can be explained in this manner, but we believe that many examples from the history of science have this quality.

However, one can instantiate this general view in different ways. One argument, following Hadamard's line of reasoning, is that unconscious processes lead to the recognition of analogies during incubation. These mechanisms would search the space of possible mappings and, when a suitable analogy had been found, would deposit the result in the fringe consciousness. Presumably this search would be directed by heuristics of elegance (or even by Gentner's notion of systematicity). This scheme is identical to Hadamard's, except that his 'combination of ideas' has been replaced by the recognition of useful analogies.

We will argue instead that the process of analogical retrieval occurs entirely during the illumination stage, and that this retrieval is usually cued by some external event. This explanation requires a very rapid mechanism, presumably one that occurs in parallel. A spreading activation process (like that used by Winston) has precisely these characteristics, and this will form another cornerstone of our theory of insight.

But if analogies are formulated during illumination, what purpose is served by the incubation stage? Our theory provides a simple answer to this question: there is *nothing* occurring during incubation. There are no unconscious processes selectively searching a problem space, as Hadamard suggests, nor is there even significant forgetting, as Simon proposes. Instead, the memory system is simply biding its time, waiting for some cue to initiate retrieval of a promising analogy.* When this occurs, it takes place very rapidly, giving the flash of insight that so many scientists have experienced.

Note that we are not rejecting the notion of unconscious processes. Both indexing and

* This suggests the need for a fifth group in Dreistadt's experiment, which would be supplied with pictorial analogies only *after* the incubation period. Our theory predicts that this group's problem solving ability would be comparable with the group that had analogies for the entire time.

Table 1. A comparison of four theories of insight.

	Hadamard	Ohlsson	Simon	Current
Preparation	'stirring up' of ideas	generation of impasse	familiar-ization	indexing of useful structures
Incubation	unconscious search for ideas	nothing	selective forgetting	nothing
Illumination	noticing a solution	restructuring, look ahead	informed problem solving	retrieval of analogy
Verification	checking solution	checking solution	checking solution	checking solution

retrieval are unconscious in that the problem solver has no conscious control over them. We have only rejected (along with Simon) the concept of unconscious *reasoning*. Such a process is inconsistent with currently accepted theories in cognitive psychology, which assume that problem solving requires conscious attention. We have replaced it with the much weaker mechanism of unconscious spreading activation. This component is completely consistent with recent work, and some widely recognized theories (e.g., Anderson, 1983) rely heavily on such a mechanism.

Our theory makes less controversial claims about the preparation stage. There is little doubt that conscious problem solving occurs during this period, and that this lays the foundation for the later illumination. Like Simon, we believe that preparation serves to index useful structures in memory, and differ from his view only in the purpose to which these structures are later put. Our emphasis on analogy makes the particular form of indexing very important, and we will have more to say about this below.

The verification stage also clearly involves conscious checking of the insight's validity, but our model of this stage also differs from earlier theories. Recall that Hall's framework distinguishes between evaluation and elaboration. The first of these components is closest in spirit to traditional notions of verification, since it involves checks on the quality of the proposed analogy. But the process does not stop there. One must also decide which aspects of the source one should carry over to the target situation; this is the elaboration process. As we will see below, our model of elaboration incorporates Carbonell's notion of derivations.

Table 1 presents the major differences of our theory from earlier models. To summarize, we claim that insight is a memory-based phenomenon, in contrast to the search-based theories of Hadamard, Ohlsson, and Simon. The theory further states that insight is a form of reasoning by analogy, requiring indexing during the preparation stage, retrieval during the illumination stage, and elaboration during the verification stage. Finally, the theory states that no significant processes occur during the incubation period.

Process Models and Behavioral Descriptions

We have seen that structures are indexed in memory during the preparation stage, and that this lays the foundation for retrieval during illumination. However, we have not yet stated what type of structures are stored, nor what indexing scheme is used. There are many possible responses to these questions, but the one we have taken builds upon Forbus' (1984, 1986) *qualitative process* theory. Thus, we should review this framework briefly before moving on.

Like other researchers* in the area of qualitative physics, Forbus has noted that people often reason about physical processes in a qualitative manner. For example, if asked to describe the process by which water boils, one might say, 'If you heat water, its temperature will increase until it reaches the boiling point; after this the water turns into steam.' Note that this statement does not mention specific quantities of heat, temperature, or rates of change. Instead, it focuses on the *qualitative* changes.

Forbus' framework centers on the notion of *processes* that produce changes over time. Heating, boiling, evaporation, and fluid flow can all be described as such qualitative processes. Each process can be described in terms of the objects involved, the conditions under which it occurs, and its influences or effects. When a set of objects meet the conditions on a process, that process becomes 'active' and leads to changes in those objects. Consider the process of boiling water as an example. In this case, the objects consist of a heat source, a container, and some water in the container. The condition on this process states that the water's temperature must be greater than the boiling point of water. When this occurs, two changes result: the amount of water decreases and the amount of steam increases.

Given a set of processes and an initial state, Forbus shows how one can generate an *envisionment* for that physical system. An envisionment specifies all possible qualitative states that the system can enter, along with the order relations between those states. Each state contains only qualitative information, such as whether a quantity is increasing, decreasing, or remaining constant. In some situations an envisionment will contain nondeterministic branches, and in these cases one cannot predict which behavior will actually occur.

For instance, the envisionment for heating water in a closed container would include three possibilities. First, if the temperature of the heat source is less than the boiling point, then the water will get hotter until it reaches this temperature and then remain constant. Second, if this temperature is greater than the boiling point, then at some point the water will begin turning into steam, and this will continue until all the water is gone. In the third case, the pressure within the container builds up sufficiently to cause an explosion. Each

---

* We refer interested readers to de Kleer and Brown (1983), Kuipers (1984), and Iwasaki and Simon(1986) for alternative approaches to qualitative physics. We find Forbus' framework the most consistent with our goals, but we do not have the space to discuss the reasons here.

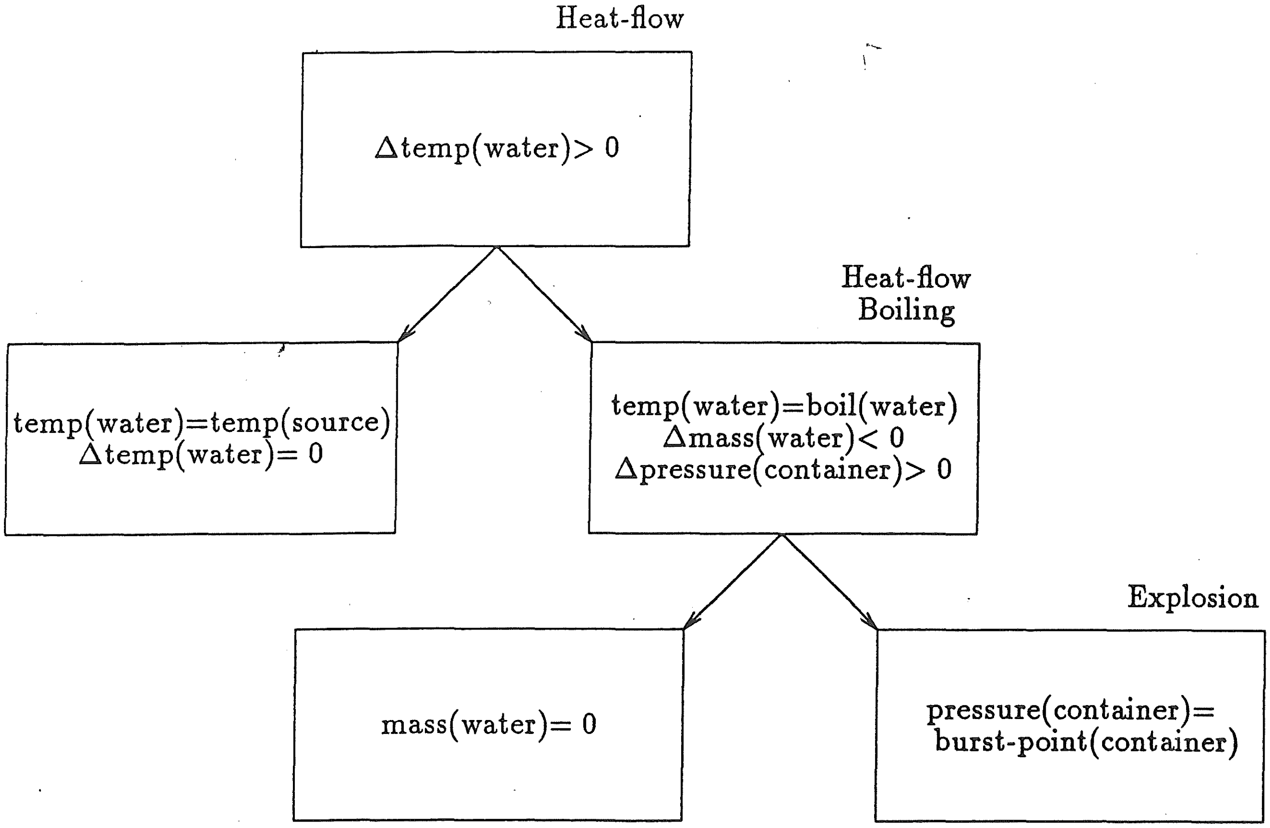

Figure 3. An envisionment for boiling water.

of these situations corresponds to a different path in the envisionment, which we present graphically in Figure 3.

In summary, Forbus' qualitative process theory describes physical systems at two levels – a *theoretical* level in terms of processes and structures and a *behavioral* level in terms of envisionments. We will see that this distinction has important implications for our theory of insight.

The Task of Theory Formation

Now that we have explained the distinction between process-structural models and envisionments, we can clearly describe a task that commonly confronts the scientist. Since one can use a process-structural model to *derive* an envisionment, we can view the former as an *explanation* of the latter. But in many cases, the scientist can induce a behavioral description for a physical system by observing its behavior, and must then infer some process model that accounts for that envisionment. We will call this the task of *qualitative theory formation.**

* Of course, there are many other facets to scientific discovery, but we do not have the room to consider them here. We refer interested readers to Lenat (1983) and Langley, Simon, Bradshaw, and Zytkow (1986) for computational studies of some other aspects of discovery.

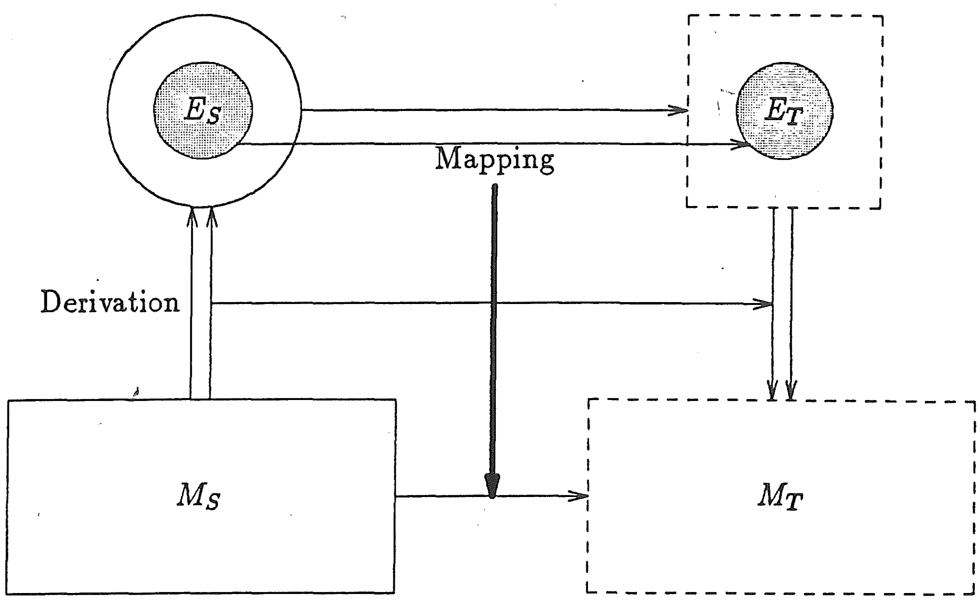

Figure 4. The mappings found from E_S to E_T are applied to M_E, thus inferring M_T. New objects and relations are inferred in E_T and M_T based on the derivation of E_S from M_S.

This problem lends itself readily to analogy-based solutions, as can be seen from Figure 4. Let M_S represent the process-structural model for some known phenomenon, such as the flow of fluid through a pipe, and let E_S stand for the envisionment derived from that model. Given some new behavioral description E_T that has been induced from observations, such as the behavior of an electric circuit, one can form an analogy between E_S and E_T. Once this mapping has been established, we can use M_S to infer an analogous process-structural model M_T. If we have been careful in the mapping process, then M_T will constitute an explanation of E_T in the same sense that M_S explains the behavioral description E_S.

We will limit our model of scientific insight to the task of qualitative theory formation. This means that it will not account for Poincaré's insight about the Fuchsian transformations, since this did not involve the construction of a qualitative process model to explain observed behavior. However, many examples of scientific insight (including Archimedes' and Kekulé's discoveries) do take on this form, and we will focus on these in the remainder of the paper. We believe our theory can be extended to cover other aspects of insight, but we have no firm evidence to present in defense of that claim. Now that we have defined the class of phenomena that we hope to model, let us turn to the details of our theory.

Indexing, Spreading Activation, and Retrieval

According to Hall's framework, the first step in reasoning by analogy involves retrieving a candidate structure. However, before this can happen such structures must have been stored in long-term memory, and they must be indexed in ways that allow their retrieval. Scientists undoubtedly index their domain knowledge in many ways, but our theory states that the most important indexing scheme for analogical retrieval centers on behavioral descriptions or envisionments.

For example, consider a situation in which two containers have different levels of liquid. If we connect these containers with a pipe, fluid will flow from the one with a higher level of liquid to the one with a lower level. This process will continue until the system has reached equilibrium. One very important aspect of this system is that it began with two unequal quantities of the same type, and these quantities changed over time until they became equal. Thus, one could index this envisionment through the concept of *equilibrium*, though certainly other features would also come into play.

One promising mechanism for explaining retrieval centers on the notion of *spreading activation.* In this framework, memory is viewed as a large semantic network (Quillian, 1968) consisting of nodes connected by labeled links. Some nodes correspond to general concepts such as *water* and *height.* These may be *activated* by interaction with the environment, and when this occurs, activation 'spreads out' from the source nodes in concentric rings.* Theoretically, this process occurred unconsciously and in parallel.

Anderson (1976, 1983) incorporated the notion of spreading activation in his ACT theory to model human fact retrieval. In this framework, the semantic network played the role of long-term memory. As new symbols entered short-term memory, activation spread out from these symbols through the semantic network, causing portions to enter short-term memory. In order to explain various memory phenomena, Anderson hypothesized that, as one spread out from a given source node, the initial activation was split between the links emanating from that node. Thus, nodes with a large 'fan' would have less impact on retrieval than nodes with a smaller fan. In addition, activation was divided proportionally according to the *trace strength* of the links involved. This trace strength increased the more often a link was stored or accessed.

Now let us see how the spreading activation approach can be used to model the illumination process. As before, assume that the scientist already has stored knowledge of many physical situations as schemas in long-term memory, and that he has indexed these situations through features of their envisionments. Now the scientist encounters a new situation and constructs an envisionment from his observations. Activation will spread out from this description, possibly pushing the activation of an existing schema above threshold and depositing the structure into short-term memory. Presumably, human memory contains thousands of such schemas, many having features in common with the new situation. But if we assume that activation is divided proportionally according to trace strength, then well-stored schemas will be greatly preferred. And schemas that have been given significant attention in the recent past – during the preparation stage – will have very high trace strengths indeed.

---

* Quillian focused on finding intersecting paths between source nodes. Recently Charniak (1986), Granger, Eiselt, and Holbrook (1986), and Norvig (1985) have reinvoked this mechanism to model natural language understanding and inference. We have not used this particular approach in modeling insight, though it may provide a plausible alternative. The approach we have chosen is more similar to that used by Holland, Holyoak, Nisbett, and Thagard (1986) in their work on analogy.

This explanation accounts for insights like that of Archimedes, in which some new experience itself becomes the source of the analogy. But it does not explain another form of insight, in which the analogy maps between two structures already in long-term memory. In such cases, the scientist already has the necessary schema E_S in memory during the preparation stage, but he does not retrieve it until later, after some appropriate cue occurs. At first glance this seems odd, but the explanation becomes apparent on closer inspection. During the preparation stage, the scientist's attention is focused on the current schema E_T, which is gradually being laid down in memory with ever higher trace strengths. Activation spreads out from this structure, but due to the high fan and low trace strengths of analogous schemas, none rise above threshold. Finally the scientist gives up on the problem.

Later, some cue enters memory from the environment that reminds the problem solver of the older, better-understood schema E_S. If the cue itself has little fan, this schema will be retrieved even with its low trace strength. Once E_S enters short-term memory, activation spreads from it to associated schemas. This time the fan is large, but one schema among the many competitors has a very high trace strength – E_T. Illumination occurs as this schema enters short-term memory and the mapping between E_S and E_T becomes apparent. Insight occurs at this moment, rather than during preparation, because activation does not spread away from schemas with high trace strengths; rather, it spreads *towards* them.

Consider the situation in which two objects with different temperatures are brought into contact. Over time, the temperature of one object will increase and the other decrease, until the temperatures stabilize when equilibrium is reached. Suppose we spread activation out from the resulting envisionment through indices such as having two objects in *contact* and two quantities reaching *equilibrium.* One promising analogy for explaining the behavior of the temperatures is the envisionment for fluid flow that we described earlier. This mapping would not suggest itself during preparation because there are many schemas that share some features with the temperature situation. But later, some environmental cue such as a waterfall might cause retrieval of the fluid flow schema, and this in turn would retrieve the temperature schema.

This explains the nature of the retrieval process and its sudden character, but presumably such retrievals occur in everyday life without the *Eureka* experience. Yet the rarity of such events follows naturally from the notion that activation occurs at different levels. In normal situations, we retrieve relevant schemas and deposit them in short-term memory, but at a relatively low level of activation. In true cases of insight, the retrieved schema has been stored so strongly that, when finally retrieved, it receives a major influx of activation. If we assume limited amounts of such activation, then the retrieved schema effectively becomes the center of attention, flushing all other structures almost instantaneously. This rapid reorganization of the contents of short-term memory gives us the 'aha' feeling we associate with true illumination.

Elaboration through Derivation

In Hall's framework, once a potential analogy has been recognized and evaluated, the mapping must still be elaborated. But how does one decide which aspects of the source to carry over to the target? Examples from the history of science provide some constraints on this process. When Dalton proposed the atomic theory of matter, it was presumably based on an analogy with macroscopic objects that could be decomposed into parts. Yet notice that, though macroscopic objects have features like color and smell, Dalton did not endow atoms with these attributes. Similarly, the caloric theory of heat was based on a fluid analogy, but its authors did not carry along aspects of fluids such as taste or viscosity.

The point here is that, at least in scientific analogies, the elaboration process is quite selective. Only certain characteristics of the source situation are carried over into the target description. This is one of Gentner's main arguments for her structure mapping theory; in her framework, higher level predicates determine which structures will be elaborated and which will be ignored. Although we could simply have borrowed this solution from Gentner, the presence of Forbus-style qualitative processes in our theory suggests a more elegant way to achieve the same effect.

Recall that Carbonell's theory of analogy relied heavily on the notion of *derivations.* A structure from the source (e.g., portions of a search tree) were carried over to the target only if an analogous derivation held for that structure in the target. What is interesting about Forbus' qualitative process theory is that it provides mechanisms for *deriving* behavioral descriptions (envisionments) from process and structural descriptions. We can then use these derivations to decide which process components and structural aspects of the source situation should be elaborated in the target situation.

For instance, suppose we have a qualitative process model of fluid flow toward equilibrium, as described earlier. The structural description of this situation may contain many features, such as the color of the liquids, their taste, their relative heights, and so forth. However, only some of these attributes are used to derive the envisionment that predicts the fluids will move toward equilibrium. Now suppose we form an analogy between this envisionment and the observation that adjacent objects with different temperatures also move toward equilibrium. We would like to infer some structural description of the new situation, but which attributes and objects should we carry over?

Our theory states that one should carry over only those aspects that were actually used in the original derivation. In this case, the heights of the liquids were used, so their analog (temperature) would be included in the new description. Features like taste and color were not used in the derivation, so these would be omitted from the new model. Note that new objects as well as attributes may be inferred. One cannot have the height of a fluid without first having the fluid, so one cannot have temperature without having some analogous fluid. Early heat theorists called this substance *caloric.*

Returning to Figure 4, we can see the relation between derivational elaboration and the other mechanisms in the theory. A Forbus-like derivation process is used to explain the source envisionment (E_S) given the source model (M_S). The entire structure is indexed by characteristics of this envisionment, and these links are used by the spreading activation process during retrieval. Given an envisionment (E_T) for the target (presumably inferred by observation), this retrieval mechanism suggests the source envisionment as a likely analogy. If this mapping is consistent, then the elaboration process plays the source derivation 'in reverse' to determine the relevant objects and attributes of the target model (M_T).* When this stage is complete, the scientist has generated an abstract process-structural model that explains the observed behavior by analogy with another situation.

Predictions of the Theory

Let us examine how our theory accounts for some insight-related phenomena. First, recall that the duration of the incubation stage can vary widely. According to the theory, this occurs because the problem solver must wait until an appropriate cue appears. This might be readily available (as in Dreistadt's experiments) or it take weeks or months to present itself. In the former case, illumination would occur almost immediately; in the latter, illumination would be delayed until the cue arises. In some cases, no useful cue ever appears, and the problem remains unsolved. Presumably this is a common occurrence, but we seldom hear about failed insights because they are not very newsworthy.

A second phenomenon is that different people have different levels of ability in problem solving and discovery. At first glance our theory cannot explain such individual differences. If cues appear randomly, then anyone should be able to make great discoveries by simply being in the right place at the right time. However, this analysis ignores the importance of the indexing that occurs during the preparation stage. Appropriate indexing is all important to the process of analogical retrieval, and different levels of domain knowledge can lead to quite different indexing schemes. Thus, experts in a given domain are more likely to store problems in ways that will let them be easily retrieved later. And the more attention given to a problem during the preparation stage, the more ways in which it will be indexed and the more firmly will its links be established. This explanation contrasts sharply with that given by Hadamard, which attributes differences in creativity to differences in unconscious reasoning processes.

We have seen that dreams play an important role in some insights, and this also seems to cause difficulty for the theory. If insights rely on external cues to initiate analogical retrieval, then they should never occur during dream states. However, this objection also disappears on closer inspection. There is no inherent reason why the retrieval cues must be *external*; they might also be internally generated during periods of free association, and this is exactly what dreams provide. But since the chains occurring in dreams are semi-random, they provide little more direction than chance external cues. Thus, dream-based illuminations may be delayed as long as those based on interactions with the environment. Kekulé did not dream of a snake because his unconscious mind was gnawing at the benzene problem. Rather, he dreamed of a snake by chance and this cued a useful mapping to the problem.

The theory does not provide a satisfactory account of the Poincaré episode. In this case, illumination came to the mathematician as he was stepping onto a bus, and he does not report any external cue that seems related to the problem. However, this does not mean that such a cue was not present. Recall that the retrieval process itself is unconscious, so it seems quite plausible that some cue was available and reminded Poincaré of the non-Euclidean

---

* Note that one can also infer missing components of the envisionment. However, this process is purely structural and does not rely on the notion of derivations.

transformations, even though he was not aware of it. Dreistadt's (1969) experiment suggests that this situation is relatively common. Many of his subjects did not think they had been aided by the analogies, and those who did were not clear as to how the pictures had helped.

The final phenomenon predicted by the theory is the occurrence of false insights. These tend to be overlooked in the discovery literature, since they are usually rejected soon after generation. But most scientists will admit that some of their most promising insights have failed to stand up under scrutiny. Recall that the spreading activation process responsible for retrieval in our theory is not very selective. In many cases it will propose analogies that will not carry through when examined more closely. Nor would we expect more than this from such a rapid, unconscious recognition process. Of course, the ratio of false insights to useful ones will depend on the indexing scheme and the particular connections formed, and this will depend on the problem solver's level of expertise and the effort spent during preparation. In general, we would expect expertise and hard work to increase the proportion of true insights.

Conclusions

In the previous pages, we reviewed some examples of scientific insight and Hadamard's four-stage description of this process. We examined some earlier theories that have attempted to account for the roles played by Hadamard's stages – preparation, incubation, illumination, and verification. However, we found these search-based explanations lacking and proposed an alternative theory that viewed insight as a memory-based phenomenon. The new theory is based on the dual notions of analogy and qualitative mental models; it draws heavily on concepts developed by earlier researchers in these areas, but links them together in a new organization. Finally, we saw that this theory accounts for many of the features normally associated with insight.

Although our focus has been scientific insight, we should briefly consider our framework's implications for other forms of creativity. We will not argue that analogy is the only path to creative thought, but we find it plausible that this mechanism does underlie many forms of original thinking. To the extent that this holds, many of the mechanisms contained in our theory of scientific insight will carry over into these other areas. These include our two main biases – that insight is explained better in terms of memory processes than search methods, and that retrieval mechanisms such as spreading activation account for illumination effects without the need for unconscious reasoning processes. Although we place significant limits on our theory as a whole, we believe that these two claims have great generality and could be profitably applied to explain other forms of creative behavior.

Before closing, we should address two traditional concerns of the literature on creativity. The first involves methods for measuring creative ability. On this count, our theory takes an extreme stance. Humans possess no general creativity factor, so no such component exists to be measured. Instead, humans possess a wealth of knowledge structures indexed by concepts that person judges important. The level of creativity that a person exhibits will depend on his knowledge, his indexing scheme, and the particular situation in which he finds himself.

But this claim suggests a useful response to the second traditional concern of creativity work – methods for improving creative ability. We have seen the important role that preparation plays in scientific insight, and presumably any creative act must have substantial knowledge structures on which to build. One cannot expect to be creative in any domain until one has achieved knowledge of that domain. Knowledge of an area may also improve retrieval abilities by leading one to index structures through concepts useful to that domain. Creativity is more than simple retrieval, involving the application of old ideas to new situations, but retrieval plays an essential part in the creative experience.

At this point, we have presented only the vague outline of a theory. The next step is to specify enough details to let us implement the framework as a running cognitive simulation. Whether our theory will hold together against this harsh test, only time and experience will tell. Most likely, the process of constructing this model will reveal many problems and inconsistencies, but these will lead us to refine and improve the model, until we achieve a real understanding of the nature of scientific insight.

References

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Washington, DC: V. H. Winston.

Carbonell, J. G. (1986). Derivational analogy: A theory of reconstructive problem solving and expertise acquisition. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach, volume 2* (pp. 371–392). Los Altos, CA: Morgan Kaufmann.

Charniak, E. (1986). A neat theory of marker passing. *Proceedings of the fifth national conference on artificial intelligence* (pp. 584–588). Philadelphia, PA: Morgan Kaufmann.

de Kleer, J., & Brown, J. S. (1983). Assumptions and ambiguities in mechanistic mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 155–190). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dreistadt, R. (1968). An analysis of the use of analogies and metaphors in science. *The Journal of Psychology, 68*, 97–116.

Dreistadt, R. (1969). The use of analogies and incubation in obtaining insights in creative problem solving. *The Journal of Psychology, 71*, 159–175.

Farber, E. (1966). Dreams and visions in a century of chemistry. In R. F. Gould (Ed.), *Kekulé centennial* (pp. 129–139). Washington, DC: American Chemical Society.

Forbus, K. D. (1984). *Qualitative process theory* (Technical Report Number 789). Cambridge, MA: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Forbus, K. D. (1986). *Interpreting observations of physical systems.* (Technical Report Number UIUCDCS–R–86–1248). University of Illinois at Urbana–Champaign, Department of Computer Science.

Gentner, D. (1983). A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Granger, R. H., Eiselt, K. P., & Holbrook, J. K. (1986). Parsing with parallelism: A spreading-activation model of inference processing during text understanding. In J. L. Kolodner & C. K. Riesbeck (Eds.), *Experience, memory, and reasoning* (pp. 227–246). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hadamard, J. (1949). *The psychology of invention in the mathematical field.* Princeton, NJ: Princeton University Press.

Hall, R. P. (1986). *Understanding analogical reasoning: Computational approaches.* (Technical Report Number 86–11). University of California, Irvine, Department of Information and Computer Science.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.

Iwasaki, Y., & Simon, H. A. (1986). *Theories of causal ordering: Reply to de Kleer and Brown.* (Technical Report Number CMU–CS–86–104). Pittsburgh, PA: Carnegie–Mellon University, Department of Computer Science.

Kuipers, B. (1984). Commonsense reasoning about causality: Deriving behavior from structure. In D. G. Bobrow (Ed.), *Qualitative reasoning about physical systems* (pp. 169–203). Amsterdam, The Netherlands: Elsevier Science.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. (1986). *Scientific discovery: Computational explorations of the creative processes.* Cambridge, MA: MIT Press.

Lenat, D. B. (1983). The role of heuristics in learning by discovery: Three case studies. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 243–306). Los Altos, CA: Morgan Kaufmann.

Marr, D. (1982). *Vision.* San Francisco, CA: W. H. Freeman.

Newell, A. (1980). Reasoning, problem solving, and decision processes: The problem space hypothesis. In R. Nickerson (Ed.), *Attention and performance VIII.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A., & Simon H. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Norvig, P. (1985). Frame activated inferences in a story understanding program. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence* (pp. 624–626). Los Angeles, CA: Morgan Kaufmann.

Ohlsson, S. (1984a). Restructuring revisited: Summary and critique of the gestalt theory of problem solving. *Scandinavian Journal of Psychology, 25*, 65–78.

Ohlsson, S. (1984b). Restructuring revisited: An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology, 25*, 117–129.

Poincaré, H. (1952). *Science and method* (F. Maitland, Trans.). New York, NY: Dover.

Polya, G. (1945). *How to solve it.* Princeton, NJ: Princeton University Press.

Quillian, M. R. (1968). Semantic memory. In M. L. Minsky (Ed.), *Semantic information processing.* Cambridge, MA: MIT Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, H. A. (1977). *Boston studies in the philosophy of science: Volume 54. Models of discovery.* Boston, MA: Reidel.

Sternberg, R. J. (1977). *Intelligence, information processing and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Winston, P. H. (1980). Learning and reasoning by analogy. *Communications of the ACM, 23*, 689–703.