Psychological Bulletin
1999, Vol 125. No 5. 524-543

0033-2909/99/$3.00

Studies of Scientific Discovery: Complementary Approaches and Convergent Findings

David Klahr and Herbert A. Simon
Carnegie Mellon University

This review integrates 4 major approaches to the study of science—historical accounts of scientific discoveries, psychological experiments with nonscientists working on tasks related to scientific discoveries, direct observation of ongoing scientific laboratories, and computational modeling of scientific discovery processes—by viewing them through the lens of the theory of human problem solving. The authors provide a brief justification for the study of scientific discovery, a summary of the major approaches, and criteria for comparing and contrasting them. Then, they apply these criteria to the different approaches and indicate their complementarities. Finally, they provide several examples of convergent principles of the process of scientific discovery.

The central thesis of this article is that although research on scientific discovery has taken many different paths, these paths show remarkable convergence on key aspects of the discovery processes, allowing one to aspire to a general theory of scientific discovery. This convergence is often obscured by the disparate cultures, research methodologies, and theoretical foundations of the various disciplines that study scientific discovery, including history and sociology as well as those within the cognitive sciences (e.g., psychology, philosophy, and artificial intelligence).

Despite these disciplinary differences, common concepts and terminology can express the central ideas and findings about scientific discovery from the various disciplines, treating discovery as a particular species of human problem solving. Moreover, we may be able to use these concepts and this vocabulary over an even broader domain to converge toward a common account of discovery in many areas of human endeavor: practical, scientific, and artistic, occurring both in everyday life and in specialized technical and professional domains.

The doing of science has long attracted the attention of philosophers, historians, anthropologists, and sociologists. More recently, psychologists also have begun to turn their attention to the phenomena of scientific thinking, and there is now a large and rapidly growing literature on the psychology of science. (A good description of the field in its infancy can be found in Tweney, Doherty, & Mynatt, 1981, and a recent summary of topics and findings from investigations of the developmental, personality, cognitive, and social psychology of science can be found in Feist & Gorman, 1998).

Our review links four major approaches to the study of science—historical accounts of scientific discoveries, laboratory experiments with nonscientists working on tasks related to scientific discoveries, direct observation of ongoing scientific laboratories, and computational modeling of scientific discovery processes—by viewing them through the lens of the theory of human problem solving. First, we provide a brief justification for the study of scientific discovery. Then, we summarize the major approaches and provide criteria for comparing and contrasting them. Next, we apply these criteria to the different approaches and indicate their complementarities. Finally, we provide several examples of convergent principles of the process of scientific discovery.

Why Study Scientific Discovery?

What accounts for the appeal of science as an object of study? The answer varies somewhat according to discipline. From our perspective as cognitive scientists, we see five reasons for studying science: for its human and humane value, to understand its mythology, to study the processes of human thinking in some of its most creative and complex forms, to gain insight into the developmental course of scientific thinking, and to design artifacts—computer programs and associated instrumentation—that can carry out some of the discovery processes of science and aid human scientists in carrying out others.

Value

The nature of human thinking is one of the "big questions"—along with the nature of matter, the origin of the Universe, and the nature of life. The kind of thinking we call scientific is of special interest, both for its apparent complexity and for its products.

Scientific thinking has enhanced our ability to understand, predict, and control the natural forces that shape our world. As the myths of Prometheus and Pandora forewarned, scientific discoveries have also provided foundations for a technical civilization fraught with opportunities, problems, and perils; and we call on science increasingly to help solve some of the very problems it has inadvertently created. The processes that produced these outcomes are irresistible objects of study. Indeed, the same forces that motivate physicists, chemists, mathematicians, and biologists to understand the important phenomena in their domains drive historians, philosophers, sociologists, and psychologists to investigate science itself.

Correspondence concerning this article should be addressed to David Klahr or Herbert A. Simon, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent to klahr@cmu.edu

Mythology

A rich mythology about the ineffability of the scientific discovery process, producing a paradoxical view of science as magic, has captured the romantic imagination. As Boden (1990) puts it in her book on the psychology of creativity,

> The matters of the mind have been insidiously downgraded in scientific circles for several centuries. It is hardly surprising, then, if the myths sung by inspirationalists and romantics have been music to our ears. While science kept silent about imagination, antiscientific songs naturally held the stage. (p. 288)

At times, the mythology of the inspirationalists and romantics has been promulgated by eminent scientists. For example, Einstein, in one of his discussions with Wertheimer (1945) reveals his uncertainty about "whether there can be a way of really understanding the miracle of thinking" (p. 227). Nevertheless, Einstein does allow that the same processes that support everyday thought also support scientific thought:

> The scientific way of forming concepts differs from that which we use in our daily life, not basically, but merely in the more precise definition of concepts and conclusions; more painstaking and systematic choice of experimental material, and greater logical economy. (Einstein, 1936/1950, p. 98)

We believe that Einstein was wrong about the first claim (that thinking is miraculous) and correct about the second (that scientific concept formation is not qualitatively different from the everyday variety). The study of scientific discovery aims at specifying how normal cognitive processes enable humans to generate the precise definitions, systematic choice of experimental material, and logical economy that Einstein identifies as the hallmarks of scientific thought.

Pressing the Limits

A common scientific strategy for understanding a complex system is to explore its behavior at the boundaries. Pushing the envelope allows researchers to test whether the same mechanisms that account for normal performance can account for extreme performance. For example, the same forces that account for lift in an airplane's airfoil also account for stalls, but the mechanisms of subsonic flight cannot fully account for the dynamics of supersonic flight.

In human cognition, the products of scientific thinking lie at the boundaries of thought capabilities. They epitomize the systematic and cumulative construction of the view of the world around us (and in us). Scientific knowledge represents, as Perkins (1981) puts it, "the mind's best work." There are other manifestations of cognitive excellence, but science has an internal criterion of progress that sets scientific discovery somewhat apart from other complex human thought, such as the creation of new art or political institutions.

Although the products of science reach one of the limits of human thought, it remains an open question whether or not the processes that support creative scientific discovery are widely different from those found in more commonplace thinking. We hypothesize that they are not. Sir Francis Crick's (1988) reflections on the processes leading to discovery of the structure of DNA concur with this view:

> I think what needs to be emphasized about the discovery of the double helix is that the path to it was, scientifically speaking, fairly commonplace. What was important was *not the way it was discovered*, but the object discovered—the structure of DNA itself. (p. 67; emphasis added)

Although Crick views the thinking that led him and Watson to make this remarkable discovery as fairly commonplace scientific thinking, he stops short of the additional claim—made by Einstein—that commonplace scientific thinking is not basically different from everyday thinking. Combined, the two claims support the view that one can study scientific thinking by, for example, examining the thought processes of participants in psychology experiments. Even if the discoveries that participants make in such experiments—the products of their inquiries—are of no scientific significance, the explication of the processes used to make those discoveries can add to the understanding of real-world scientific discovery.

The Paradox of Children's Thinking and Its Development

The similarities between children's thinking and scientific thinking have an inherent allure and an internal contradiction. The allure resides in the inescapable wonder and openness with which both children and scientists approach the world around them.

> Children are born scientists. From the first ball they send flying to the ant they watch carry a crumb, children use science's tools—enthusiasm, hypothesis, tests, conclusions—to uncover the world's mysteries. But somehow students seem to lose what once came naturally. (Parvanno, 1990, as quoted in Elder, 1990, p. 20)

The paradox comes from the fact that investigations of children's thinking have produced evidence that partially supports diametrically opposing views of their scientific reasoning skills. In support of the child-as-a-scientist position, the developmental literature is replete with reports showing that very young children can formulate theories (Brewer & Samarapungavan, 1991; Karmiloff-Smith, 1988; Wellman & Gelman, 1992), reason about critical experiments (Samarapungavan, 1992; Sodian, Zaitchik, & Carey, 1991), and evaluate evidence (Fay & Klahr, 1996). Many psychological studies also show that adults often exhibit systematic and serious flaws in their reasoning (Kuhn, Amsel, & O'Loughlin, 1988; Schauble & Glaser, 1990) even after years of formal scientific training (Mitroff, 1974). In contrast, many investigators (e.g., Kern, Mirels, & Hinshaw, 1983; Kuhn et al., 1988; Kuhn, Garcia-Mila, Zohar & Andersen, 1995; Siegler & Liebert, 1975) have demonstrated that trained scientists, and even untrained lay adults, commonly outperform children on a variety of scientific reasoning tasks.

The conflicting theoretical claims and empirical results emerging from the child-as-scientist debate have important implications for science education. Moreover, the debate raises some deep questions about how both scientists and children really think—questions that can only be approached through more precise formulation and study of the empirical and the theoretical aspects of scientific thinking and discovery.

Machines in the Scientific Process

The final reason for studying science is that such study may lead to better science. What researchers learn about the science of science leads into a kind of engineering of science in which—as in other areas—knowledge of a natural process can be used to create an artifact that accomplishes the same ends by improved means.

This transition from scientific knowledge to engineered artifact has already happened for scientific discovery, as computational models used as theories of discovery in specific domains have been transformed into computer programs that actually do some of the discovery in these domains. An early example of this transition from psychological model to expert system was the DENDRAL program that, taking mass spectrogram data as input, identified the molecules that had produced the spectra Among the descendants of DENDRAL are programs that carry out automatically much of the analysis for genome sequencing in biology and programs that, independently or in association with human scientists, discover plausible reaction paths for important chemical reactions.

Recent examples include Valdés-Perez's (1994a, 1994b, 1994c) systems for discoveries in chemistry and physics, Fajtlowicz's in mathematics (Erdos, Fajtlowicz, & Staton, 1991), Hendrickson's program for the synthesis of organic compounds (Hendrickson & Sander, 1995), and Callahan and Sorensen's (1992) systems for making discoveries in the social sciences (see Valdés-Perez, 1995; & Darden, 1997, for brief reviews of recent work in the field).

Today, the conception and design of expert systems that can collaborate with human scientists in making discoveries is a vigorous and growing field of artificial intelligence research, enlisting the efforts of both computer scientists and natural scientists in the disciplines where the applications are being made. We do not cover this activity in this article but call attention to it because ideas from the expert-systems research are relevant for the theory of human scientific thinking; and vice versa, components of human discovery processes may be embedded in practical expert systems.

Conclusion: Why Study Scientific Discovery?

Scientific discovery is a highly attractive area for research because it possesses several consequential features: relevance to one of the great scientific questions, a mythology, a domain for testing theory at the limits, a developmental paradox, and direct applicability to expert system design. How does one go about doing such research?

Approaches to the Study of Science

Empirical investigations of science fall into five overlapping categories: (a) historical accounts, (b) laboratory studies, (c) observations of ongoing discovery, (d) computational or simulation models, and (e) sociological studies. In this section, we describe each approach briefly, and in the following section, we discuss their complementarity and convergence.

Historical Accounts

Historical accounts of scientific discovery usually aim at describing the cognitive and motivational processes of persons who have made major contributions to scientific knowledge. To varying degrees, they examine both the processes germane to the scientific problems themselves (internalist accounts) and the interaction of these processes with the broader social environment (externalist accounts). These accounts—which are based on analyses derived from diaries, scientific publications, autobiographies, lab notebooks, correspondence, interviews, grant proposals, and memos—have been provided not only by historians (e.g., Galison, 1987; Holmes, 1985) but also by philosophers (e.g., Gooding, 1990; Nersessian, 1992; Thagard, 1992) and psychologists (e.g., Feist, 1991, 1994; Gruber, 1974; Terman, 1954). The analyses are sometimes further enriched by retrospective interviews with the scientists—for example, Wertheimer's (1945) classic analysis of Einstein's development of special relativity or Thagard's (1998) analysis of the recent discovery of the bacterial origin of stomach ulcers.

Laboratory Studies

Another way to study science is to observe people's problem-solving processes in situations crafted to isolate one or more essential aspects of real-world science. These studies are typically carried out in the psychology laboratory under the standard rubrics of experimental design, with experimental and control conditions, and the use of statistical significance tests. Participants in such studies have included young children (Schauble, 1990; Siegler & Liebert, 1975; Sodian et al., 1991), college sophomores (Mynatt, Doherty, & Tweney, 1977, 1978; Qin & Simon, 1990; Schunn, 1995), lay persons (Klahr, Fay, & Dunbar, 1993; Kuhn, 1989), and practicing scientists (Schunn & Anderson, in press). The tasks have ranged from some that bear only an abstract relation to real scientific tasks to others in which a real scientific discovery was stripped to its essential features so that it could be replicated in the psychology laboratory.

Examples of abstract tasks include the discovery of the physics of an artificial universe (Mynatt et al., 1977, 1978), the discovery of an unknown function on a programmable device (Klahr & Dunbar, 1988; Schunn, 1995), and the discovery of arbitrary rules and concepts (Bruner, Goodnow, & Austin, 1956; Wason, 1960) Examples of laboratory simplification of real scientific discoveries include Dunbar's (1993) simulated molecular genetics laboratory, in which participants were challenged to replicate Jacob and Monod's (1961) discovery of genetic control, and Schunn and Anderson's (1999) comparison of experts' and novices' ability to design and interpret memory experiments.

However, laboratory experimentation plays a much broader role in science than simply as a tool for testing hypotheses that someone has proposed, and experimentation on the discovery process itself can and should have similar breadth. In science there is an important, and extremely common, form of experiment, at times referred to somewhat dismissively as exploratory, that is guided by no specific hypothesis to be tested, and no clear control condition, but only a vague and general direction of inquiry. The goal of exploratory experiments is to permit phenomena to appear that will invite exploration or suggest whole new forms of representation or generate new hypotheses (Simon & Kotovsky, 1963). Exploratory experiments are of great importance in science, even though they are often treated as second-class citizens in textbooks on research methodology.

Michael Faraday's discovery of electromagnetic induction in 1831 came out of just such experiments, which were guided by no hypothesis more specific than "if, as Oersted showed, electric currents can generate magnetism, then there should be circumstances under which magnetism will generate electric currents."[1] There was no control condition, but just a great deal of skillful manipulation of apparatus, each new experiment being suggested by the outcomes of the previous ones, to find an arrangement that would produce the hoped-for phenomenon and contribute to understanding the conditions under which it would appear.

Similar comments can be made about Krebs' experiments that led to the discovery of the ornithine cycle for the in vivo synthesis of urea. The experiments were largely driven by the broad idea that amino acids and ammonia were likely sources of the nitrogen in urea, which provided a reason for experiments with various amino acids. However, it was a lucky accident that the key catalyst for the reaction was an amino acid, ornithine, which, not being a source of the nitrogen at all, was tested for the wrong reason. The experiments do not fit the textbook paradigm of control and hypothesis testing, but they were clearly exploratory in nature, and Krebs himself saw them in that way.

In the realm of the psychology of discovery, examples of such exploratory laboratory studies include Qin and Simon's (1990) experiment in which college sophomores were presented with data on planetary distances and periods to see if they could discover Kepler's third law and Klahr and Dunbar's (1988) initial study with the BigTrak. (In fact, in an interesting recursive twist, one of the most important results produced by Klahr & Dunbar's exploratory study was that their participants frequently did experiments in the absence of any hypothesis but with the goal of generating some interesting behavior of the device they were exploring.) We describe a few other such studies in later sections of this article.

In studying scientific discovery, laboratory experiments need not be limited to replicating the processes scientists have used to discover hypotheses to fit data, or to test hypotheses. Exploratory experiments can also be used to help interpret other forms of evidence about discovery, and we give examples, especially from the work of Gooding (1990), of original and insightful efforts in this direction. For instance, reconstructing the instruments and methods used in historical discoveries permits the student of discovery to re-experience the processes of generating and interpreting the original data in its physical context, thereby casting light on the difficulties faced by the original investigators in arriving at the meanings of what they saw.

Observation of Ongoing Discovery

The most direct way to study science is to study scientists as they ply their trade. The observer records the important activities of day-to-day lab meetings, presentations, pre- and postmeeting interviews, lab notes, and paper drafts. The raw data are then coded and interpreted within the framework of psychological constructs.

Both pragmatic and substantive factors make direct observation extraordinarily difficult, and therefore the least common approach to studying science. It requires the trust and permission of the scientists to allow an observer in their midst. The observer must be sufficiently well versed in the domain of investigation to understand deeply what is happening and what the fundamental issues, problems, and solutions are.[2] Moreover, it is extremely time consuming. Finally, such investigations require a bit of luck, because the outcome of such a study is of much more interest if a major discovery is made during the period of observation than if nothing of any great importance happens. One recent exemplary case of this approach can be found in Dunbar's studies of four different world-class labs for research in molecular genetics (Dunbar, 1994, 1997; Dunbar & Baker, 1994). Giere's (1988, chapter 5) account of how a high energy physics lab is organized provides a somewhat similar example of the in vivo approach, in contrast to the in vitro approach of laboratory experiments.

Computational Models of Discovery: Artifact and Explanation

A theory of scientific discovery processes can sometimes be cast in the precise terms of a computational model that simulates these processes and re-enacts discoveries. From a mathematical standpoint, such a model is a set of difference equations that describes and predicts the dynamic path the discovery system will follow from the time it takes up a problem until it solves or abandons it. Thus, it is highly similar to the systems of differential equations used to model theories in the natural sciences.

The goal of such a model is to replicate the key steps in the cognitive processes of scientists as they made important discoveries (see Shrager & Langley, 1990, for an introduction to this literature). Modeling draws on the same kinds of information as do the historical accounts. However, it goes beyond the historical record to hypothesize cognitive mechanisms that are sufficiently specific to make the same discoveries the human scientist made, following the same path.

Computational modeling of concept formation and rule induction—activities relevant to scientific discovery—has a long history (Hovland & Hunt, 1960; Simon & Kotovsky, 1963). More recent (and much more complex) examples of the approach include computational models of the cognitive processes used by Kepler, Glauber, Dalton, Krebs, and others in making historically important scientific discoveries (Cheng & Simon, 1992; Gooding, 1990; Grasshoff & May, 1995; Kulkarni & Simon, 1990; Langley, Simon, Bradshaw, & Zytkow, 1987; Simon, Langley, & Bradshaw, 1981).

Sociological Approaches

In recent years, the sociology of science has directed most of its attention to externalist accounts of discoveries, which seek to explain discovery as a product of political, anthropological, or social forces (Bloor, 1981; Latour & Woolgar, 1986; Pickering, 1992; Shadish & Fuller, 1994). In these approaches, the mechanisms linking such forces to actual scientific practice are usually motivational, social–psychological, or psychodynamic, rather than

[1] For a contrary view, see Williams (1965), but Gooding (1990) develops in detail essentially the position taken here.

[2] However, those interested in sociological factors sometimes prefer instead to be ignorant of the substantive knowledge and scientific conventions of the laboratory under investigation. "Latour's knowledge of science was non-existent; his mastery of English was very poor; and he was completely unaware of the existence of the social studies of science. Apart from (or perhaps even because of) this last feature, he was thus in a classic position of the ethnographer sent to a completely foreign environment" (Latour & Woolgar, 1986, p. 273).

cognitive (Bijker, Hughes, & Pinch, 1987). An interdisciplinary amalgam of such studies has developed under the rubric of social studies of science (e.g., see Laudan, 1977; Mahoney, 1979). This orientation to the study of science provides some important insights on how social and professional constraints influence scientific practices, although these accounts tend to treat cognitive processes at a large grain size relative to the questions addressed in this article (Latour & Woolgar, 1986)

Some of this work has taken an extreme deconstructionist turn that has been rejected (Gross & Levitt, 1994) and parodied (Sokal, 1996). Unfortunately, such extremes have led many researchers in the physical and biological sciences to mistakenly conclude that all social science approaches—including cognitive psychology—have little to contribute to a better understanding of science. However, the work surveyed here is not subject to this criticism, for it conforms to the same canons of science as the research it undertakes to describe and explain. In any event, because of our internalist emphasis, we do not have much to say, in this article, about the sociology of science, either in its defensible or indefensible forms.

Assessing the Approaches

Each of the four approaches that are used in research on scientific discovery has its particular strengths and weaknesses. In this section, we summarize the criteria generally used for evaluating research methods in general and psychological research methods in particular, and then we assess, in general terms, the extent to which the different approaches satisfy these criteria.

Criteria for Evaluating Research Methods

The effectiveness of a research strategy for addressing a particular problem of scientific discovery can be assessed in terms of eight criteria: (a) face validity, (b) construct validity, (c) temporal span and resolution of data, (d) fruitfulness for discovering new phenomena, (e) rigor and precision, (f) control and factorability of variables, (g) external validity, and (h) social and motivational context.

Face validity. A study has face validity if it measures what it is supposed to measure. Research on scientific discovery has face validity to the extent that the phenomenon being investigated is clearly an instance of something being discovered by a scientist. The farther the study is from science or discovery, the lower the face validity. For example, research on a discovery of historical importance—such as Faraday's discovery of the magnetic induction of electricity (Duncan & Tweney, 1997) or his work on acoustics (Ippolito & Tweney, 1995)—has very high face validity because there is no question that the behavior being studied really did lead to a discovery. On the other hand, research that asks college students to list as many uses of a brick as they can think of (Finke, Ward, & Smith, 1992, pp. 183–184) or to discover a rule about number triples of which 2-4-6 is an example (Wason, 1960) has lower face validity because the extent to which the discovery has anything in common with genuine scientific discoveries is open to question. In an appropriate research design, the way in which participants approach these kinds of laboratory tasks may, in fact, reveal something about their scientific skills, but this is not evident unless researchers have independent evidence for similarity between thinking in the psychologist's lab and thinking about solving a real scientific problem

Construct validity. As phenomena in a domain begin to acquire a theory, theoretical terms are generally introduced, referring to entities that are not directly observable. The ability to evaluate theoretical terms and test theories containing them depends on the operations and instruments used to measure presence and magnitude of such terms indirectly and convergently (Simon, 1974, 1983, 1985). Thus, construct validity evaluates how well the measures being used are good operationalizations of the underlying theoretical constructs about scientific discovery.

Temporal span and resolution of data. Two important and related criteria for assessing how well a research method captures important processes are the temporal span of the discovery episode that is studied and the temporal resolution of the data that describe behavior during that span. Typically, longer spans of episodes produce data of lower resolution. Thus, a particular problem may occupy a scientist for an hour, a day, a week, a year, or decades, whereas at the other end of the scale, insights and acts of recognition might require only fractions of a second of cognitive processing.

For example, while working primarily on electrochemical problems, Faraday brooded intermittently about electromagnetism for an entire decade—from 1821, when he first learned of Oersted's induction of magnetism by an electric current, to 1831, when he made his initial key discovery of induction of currents by magnets. Throughout this period, he kept a meticulous diary that described each experiment and its outcome. Thus, the grain size during periods when he was doing electromagnetic experiments was on the order of a few hours, but there were often gaps of months, or even years, between such experiments.

More than 2 decades passed between the time when Kepler published (Kepler, 1596/1937b) an early (and erroneous) version of his Third Law and the time when he returned to the problem and got the right answer after a few weeks' further work (Kepler, 1619/1937a). However, here we have a much rougher grain size because our only record of his work on the problem are (a) a few published paragraphs when he announced the erroneous law and (b) the pages in the volume in which he published his successful second set of calculations, with a few comments on how the work had been stretched out over several months because of the computational errors he made.

Of course, not all discoveries extend over such long periods of time: Planck achieved his revision of Wien's law of blackbody radiation in a single evening in 1900 (Langley et al., 1987); and when we move from real science to laboratory studies of scientific thinking, we enter the realm of tasks that typically require from tens to hundreds of minutes. For example, participants in Wason's rule-discovery tasks usually take about 20 to 30 min to discover the rule; Klahr and Dunbar's (1988) participants spent about 30 min to make their discoveries; Qin and Simon's (1990) participants took about an hour, on average, to discover Kepler's Third Law; participants in Schunn's (1995) milk-truck microworld took up to 90 min before they discovered one of its complex rules; and some participants in Mynatt et al.'s (1978) artificial-universe task worked at it for up to 10 hr. A graduate student was given the same data that Balmer used to discover in 1885, after several months of search, his formula for the hydrogen spectrum. In about 6 weeks of half-time work, the student rediscovered the formula (Simon, personal communication).

Although these discoveries occur over relatively brief spans, the data resolution is correspondingly finer grained so that participants' hypotheses, representations, insights, and impasses can be recorded over extremely short durations, sometimes as little as a fraction of a second. Observation of such situations enables the researcher to track the initial interpretation of problem instructions, the creation of initial representations, impasses, and revised representations. Data collected from such investigations often include extensive verbal and behavioral protocols that can be analyzed at varying levels of aggregation.

Fruitfulness for discovering new phenomena. The contemporary literature on research methodology is dominated by the notion, promulgated by Popper (1959) among others, that the purpose of observation in general, and experiment in particular, is to test hypotheses in order either to falsify or validate them. In contrast to this position, we have argued that much of the important empirical work in science is undertaken—to use Reichenbach's phrase—in the context of discovery rather than the context of verification (see Simon, 1973). That is, a major goal of empirical work in science is to discover new phenomena and generate hypotheses for describing and explaining them and not simply to test hypotheses that have already been generated. Indeed, theories cannot be tested until they have been created, and creation takes place in the context of discovery, not verification. In his *Patterns of Discovery* (1958), Hanson took a pioneering step toward giving discovery equal time with verification in the study of science.

The distinction between finding new phenomena (e.g., Oersted's unexpected discovery that an electric current created a magnetic field orthogonal to it) and testing a theory that explains it (e.g., Michelson and Morley's experiment showing that the velocity of light was independent of its motion through an aether) is closely related to the distinction in computational models of discovery between data-driven and theory-driven systems (Karp, 1990; Langley et al., 1987) and to the distinction, in psychological studies of discovery, between experimenters and theorists (Klahr & Dunbar, 1988).

Hence, among our criteria for evaluating methods for research on scientific discovery, we include their fruitfulness for discovery. We use this criterion to evaluate the extent to which a particular approach to research on scientific discovery is likely to uncover new phenomena about the discovery process. This criterion is related to construct validity because how well theoretical constructs are operationalized depends, in part, on how such terms are generated in the first place (Langley et al., 1987).

Rigor and precision. At one end of the scale, we have data (usually numerical) that can be reported with precision (e.g., reaction times, proportion of correct answers, changes in hypotheses, number of trials to solution) or events recorded in a reproducible coding scheme (e.g., rigorous coding of verbal protocols); at the other end, we have descriptions (usually verbal and informal) of complex events, involving interpretation and summarization of the source data beyond the resources of a precise coding scheme.

Typical examples of the former include the vast literature on discovery-related cognitive processes: concept-learning studies, investigations of complex problem solving, and the simulated science investigations cited earlier, where speed of solution, errors along the way, and other statistics are measured. Typical examples of the latter include the analyses of Faraday's diaries (Duncan & Tweney, 1997) or Darwin's notebooks containing successive versions of what became his theory of evolution by natural selection (Gruber, 1974). Unless qualitative data are coded and analyzed according to unambiguous and objective criteria, it is difficult to make precise predictions and hence to test theories rigorously against the data.

Control and factorability of variables. Experiments, at least those described in textbooks on research methods, aim to separate the effects of specific variables on the phenomena of interest, to control for the effects of other variables, and to minimize the systematic effects of uncontrolled sources of error. The so-called experimental method, adorned with all of these attributes, is often taken as the quintessence of science.

However, we have already noted that much science is, and must be, observational rather than experimental in this special sense—that is, aimed at discovering new and interesting phenomena and using them to stimulate the generation of hypotheses, rather than at testing hypotheses formulated prior to observation. There is a literature on observational (Webb, Campbell, Schwartz, & Sechrest, 1966), quasi-experimental (Cook & Campbell, 1979), case studies (Barlow & Hersen, 1984; Kratochwill, 1978), and naturalistic methodologies. There is less than full consensus about the role of these methodologies in the processes of science. An important goal in research on scientific discovery is to assess and evaluate the respective roles of orthodox experiments and other kinds of observation in the processes of discovery and verification.

A great deal can be learned about the discovery process by examining failures as well as successes. Although the historical record of scientific discovery focuses more on the latter than on the former, there is a literature that attempts explain why, in situations where a race for some discovery occurred, one lab succeeded while others failed (Crick, 1988; Judson, 1979). Moreover, in every success story, a period, sometimes very long, of failure to reach the goal precedes the final success. Although a great deal of information is therefore available on the conditions that have to be satisfied to turn failure into success, the myriad failures of scientific discovery remain underreported and underexamined. Laboratory studies can be of special value in this connection, for they allow investigators to control precisely the variables that are hypothesized to affect success and failure (e.g., Dunbar, 1993; Gorman, 1992; Penner & Klahr, 1996) or to examine some of the detailed differences in processes used by successful and unsuccessful participants.

External validity. To what extent the results of a study can be broadly generalized is the question of external validity. In any particular instance of empirical research on discovery, only a limited task being addressed by a specific population (or person, in the case of historical studies) is investigated. This is true not only for laboratory studies but also for historical studies and in vivo investigations. As the totality of domains in which discoveries may be made is boundless, each piece of research may be regarded as a case study. The challenge is to bring such studies together so that a theory of discovery can emerge from them.

External validity is related to the child-as-scientist debate mentioned earlier. Although there is no question about children's inadequacy relative to adults when it comes to skills that are domain specific, laboratory study or natural observation may reveal, in specific contexts, that children have surprising competence in some aspects of scientific reasoning. Nevertheless, until flexible and adaptive deployment of such competencies is demonstrated, the external validity of such findings remains in question

Social and motivational context. In this article, we are concerned primarily with internalist accounts of discovery, but even if external factors are not the focus of study, they may still provide alternative explanations for phenomena that have been attributed to internal factors. At a minimum, no study of discovery can disregard the central fact that individual scientists or groups of scientists are always part of a wider social environment, inside and outside science, with which they are in constant communication and which has strongly shaped their knowledge, skills, resources, motives, and attitudes.

The interaction between social and cognitive factors is beautifully illustrated in Thagard's account of the discovery of the bacterial origins of stomach ulcers (Thagard, 1998). This account demonstrates how the weight of empirical evidence can overwhelm even the most entrenched and socially accepted scientific beliefs. The initial view that ulcers were caused by bacteria was viewed as preposterous when first proposed in 1983 but has, by now, achieved nearly universal acceptance. The reasons for both the initial and final positions clearly involve important social mechanisms Nevertheless, as Thagard cautions,

> It is important not to succumb to the slogan that science is a social construction. Proponents of that slogan tend to ignore both the psychological processes of theory construction and acceptance . . and the physical processes of interaction with the world via instruments and experiments Undoubtedly interests and social networks abound in the ulcers case as in other episodes in the history of science But explaining scientific change solely on the basis of social factors is as patently inadequate as purely logical and psychological explanations. (p. 134)

Thus, acknowledgment of the role of noncognitive factors in the process of scientific discovery does not plunge such research into the social constructivist pit We fully concur with Giere's (1988) wry critique of extreme versions of the doctrine of cultural relativism sometimes embraced by constructivists.

> Constructivists have no qualms about assuming the reality of *other people*. They are perfectly willing to explain Jones' actions by reference to a conversation Jones had with Smith Is that not assuming too much? Should we not rather say that Jones believed he had a conversation with Smith? Surely that would be silly Restricting explanations of physicists' activities to invoking only their beliefs about protons, rather than protons themselves, is just as silly. (p 127)

Strengths of the Several Methodologies

How do the different approaches to studying discovery fare on the evaluative criteria described above? In Table 1, we provide a succinct depiction of their relative merits. However, the table requires some interpretation, for matters are not as simple as it makes them appear In order to give a coherent appraisal of each approach, our interpretation proceeds column-wise (by approach), keeping in mind the fact that the row-wise (by criterion) comparisons are particularly informative (e.g., the relative face validity of historical study vs. laboratory experiment). In the following discussion, we limit the explanation of our evaluation mainly to cells in Table 1 with entries that are high (***), low (*), or empty. At this level of analysis, any more precise comparisons would be difficult to justify.

Historical studies. We give historical studies high marks for face validity because, by definition, they investigate the very phenomenon that they seek to explain. Topic and scope are chosen after the fact, so there is no doubt that a real discovery by a real scientist is the focus of investigation. The temporal resolution of such investigations depends on the sources of data available. Resolution is quite coarse when the primary sources are publications, but it can become much finer—on a scale of days, say—to the extent that laboratory notebooks and correspondence are available. Historical studies seldom permit the study of anything approaching minute-by-minute or even hour-by-hour sequences of the scientists' thoughts.

Given the unique and often idiosyncratic aspects of the interaction between a particular scientist, a particular state of scientific knowledge, and a particular discovery, historical studies are highly likely to generate new phenomena. The down side of this potential

Table 1
Dimensions of Strength of Four Approaches to Research on Scientific Discovery

	Type of approach				
		Laboratory studies			
Evaluative criteria	Historical studies	Exploratory	Controlled	Direct observation	Computational modeling
Face validity	***	*		***	*
Construct validity	*	*	**	**	**
Temporal span & resolution					
Short & fine-grained	*	***	***	**	***
Long & coarse-grained	***			*	**
New phenomena	***	***	*	***	**
Rigor & precision	*	**	***	**	***
Control & factorability		*	***	*	***
External validity	*	*	**	*	**
Social & motivational factors	***			***	*

Note Each approach is evaluated on the criteria as either very high (***), high (**), modest (*), or poor (no entry) for novelty is low external validity, with respect to generalizability to all science and all scientists Such careful investigations as Gooding's analyses of Faraday's extensive and meticulous notebooks (Gooding, 1990) are of necessity limited to a single scientist However, as such studies cumulate, they can be treated collectively as a sample of events that can be compared and contrasted to find the underlying general laws

We give this approach low marks on rigor and precision because of the subjective and unverifiable nature of much of the raw data. Even when based on daily lab notebooks, the data are subject to all of the self-reporting biases that make retrospective verbal reports less rigorous than concurrent verbal protocols (Ericsson & Simon, 1993). And when the accounts are based on the recollections and introspections provided in autobiographies of great scientists (e.g., Hadamard, 1945; Poincaré, 1929), or by systematic interviews with scientists about their work (e.g., Rowe, 1953), reliability is always in doubt: "But did they *really* think this way?" asks Nersessian (1992) of her own analysis of Maxwell's discovery of electromagnetic field theory. "In the end we all face the recorded data and know that every piece is in itself a reconstruction by its author" (p. 36).

Finally, historical studies rank high on their tendency to address both social and motivational factors surrounding the discovery process This is true, in part, because historical studies predate the emergence of the cognitive sciences, and to the extent that they treated psychological variables at all, they tended to concentrate on noncognitive types of psychological variables

Laboratory studies Enough has already been said about laboratory studies to indicate their strengths. Their chief limitations are in face validity—it is seldom possible to study in the laboratory discoveries like those that fill the histories of science, although we discuss cases where this has been done. Controlled laboratory studies are especially adapted to the needs of verification in a theory-driven paradigm. As usually designed, with concern for clear tests of well-defined hypotheses using experimental controls, the controlled laboratory study reduces, although it does not exclude, the likelihood of wholly unanticipated outcomes that involve new variables or new phenomena that were not considered in the experimental design. Exploratory laboratory studies sacrifice the rigor and precision of controlled experiments in favor of their potential for producing interesting new phenomena. Both types of laboratory studies tend to generate fine-grained data over relatively brief periods, and they typically ignore or attempt to minimize the effects of social and motivational factors on the discovery process.

Direct observation Direct observations of ongoing science have many of the characteristics of historical data—in particular, high face validity and potential for detecting new phenomena. However there are two important differences between historical and direct approaches. First, the observations may achieve much finer-grained temporal resolution of ongoing research processes than historical research Second, direct observation provides a level of rigor, precision, and objectivity that is lacking in retrospective accounts by scientists of their discoveries.

Computational modeling Our evaluation of computational modeling derives from our view of it not as a method for gathering data but as a medium for generating theories and for representing and testing them against data that have been obtained by the other methods. Hence, it is clearly not a substitute for the others but complementary to them One of its important applications is to provide tests of the *sufficiency* of the mechanisms postulated in a theory of discovery to actually produce the discovery. The model will be unable to achieve the discovery unless it does possess a sufficient set of mechanisms, appropriately organized.

The modeling approach can achieve high external validity by using a single model to simulate behavior in a whole range of discovery tasks (Kulkarni & Simon, 1990). To this end, a model of discovery, like any theory, must be factored into two components: (a) its basic mechanisms, retained without alteration from one application to another, and (b) specific knowledge of the content and research methods of each task domain to which it is applied. The first component reveals the extent to which general methods can account for discoveries over a range of domains, and the second component indicates the extent to which a discovery relies on domain-specific knowledge and methods A general theory of discovery can emerge from the components of such models that are common to many tasks.

Of course, this separation of the general from the specific is not limited to formal simulation models but extends to general theories of discovery, however expressed. What is special about simulation models is the rigor with which they can be stated and the clarity with which general and task-dependent elements in the system can be distinguished. Simulations provide us with powerful methods for comparing the theoretical implications of the data from particular case studies, interpreting the data in a common formal language that can reveal both the identity or similarity of processes that were involved in each case and the differences among them.

The construct validity of modeling depends on the construct validity of the tasks that are modeled. Modeling enables us to express a theory rigorously, and generally, to simulate phenomena at whatever temporal resolution and for whatever durations are relevant. The effects of changes in particular variables can be studied, and other variables can be held constant. Theoretical variables are clearly specified so that construct validity is high.

Social variables can, in principle, be incorporated in models, although, with the important exception of including the specialized and socially acquired knowledge of the domain expert, this has not usually been done in modeling discovery. On a large scale, it would be possible to model a scientific community rather than a single scientist, the members of the community being linked, for example, by the blackboard of publication. Limits on speed of computation may require a trade-off between fine temporal resolution and long durations.

Assessing the Approaches: Summary

It should be clear from this brief exercise in comparative assessment that there is no single best way to study the discovery process. Research on the science of discovery is subject to the same inevitable trade-offs that characterize research methodologies and paradigms in all scientific disciplines. But these trade-offs do not imply that the different approaches are incompatible. To the contrary, the fundamental thesis in this article is that the findings from these diverse approaches, when considered in combination, can advance our understanding of the discovery process more than any single approach. In the final two sections of this article, we attempt to show how the methods complement one another and how their complementarities are beginning to produce consistent convergent evidence about the process of scientific discovery. But first, in order to provide a common language for describing these complementarities and convergences, we introduce, in the following section, a set of concepts and terms that have been used to characterize human problem solving.

Scientific Discovery as Problem Solving

We argued earlier, in citing Francis Crick's account of the discovery of DNA, that major scientific discoveries are so labeled because the knowledge that they produce is important and not because they derive from any unusual thought processes. Psychologists have been making the case for the nothing-special view of scientific thinking for many years (e.g., Simon, 1966), pointing out that information-processing theories of human problem solving could account for many of the unique characteristics of scientific discovery. This view has been elaborated more recently—particularly with respect to the issue of creativity—by several others, including Boden (1990), Perkins (1981), Simon et al. (1981), and Weisberg (1993). The success of computational models like BACON and KEKADA—based as they are on a small set of relatively straightforward heuristics for finding regularity in existing data sets—provides further support for this position.

This view does not imply that the average person could walk into a scientist's lab and proceed to make discoveries. Practitioners of a scientific discipline must acquire an extensive portfolio of relatively particular methods and techniques during their long professional training and must apply their skills in the context of an immense, cumulative base of shared knowledge about the discipline's phenomena, theories, procedures, instrumentation, experimental paradigms, and data-analytic methods, not to mention its history, funding procedures, social and political implications, institutional structure, and even its publication practices (see Bazerman, 1988).

These components of expertise constitute the strong or domain-specific methods. The processes we are focusing on are the weak methods: domain-general, universal, problem-solving processes. Although the strong methods used in scientific problem solving distinguish the content of scientific thinking from everyday thought, we claim that the weak methods invoked by scientists as they ply their trade are the same ones that underlie all human cognition. In this section, we sketch very briefly some of the basic components of the general theory of problem solving to provide a common language for discussing the convergence of different approaches to the study of discovery.

Problem Solving, Search, and Weak Methods

A problem consists of an initial state, a goal state, and a set of operators for transforming the initial state into the goal state by a series of intermediate steps. Operators have constraints that must be satisfied before they can be applied. The set of states, operators, goals, and constraints is called a problem space, and the problem-solving process can be characterized as a search for a path that links initial state to goal state (Newell & Simon, 1972).

Initial state, goal state, operators, and constraints can each be more or less well defined. For example, one could have a well-defined initial state and an ill-defined goal state and set of operators (e.g., make something pretty with these materials and tools), or an ill-defined initial state and a well-defined final state (e.g., prove a particular mathematical conjecture). But well-definedness depends on the familiarity of the problem space elements, and this, in turn, depends on an interaction between the problem and the problem solver. More specifically, it rests on the process of recognition. Before any search process can be applied, its relevance must be recognized by the detection of appropriate patterns in the situation. Observation of such patterns evokes information about the situation that can help guide the search. As this information usually is domain specific, the recognition mechanism tends to make it available just where it is potentially relevant (i.e., instances of positive transfer). Such recognition is not always productive, as witness cases of negative transfer, functional fixedness (Dunker, 1945), and Einstellung (Luchins, 1942).

Although scientific problems are much less well defined than the puzzles commonly studied in the psychology laboratory, they can be characterized in these terms. In both cases, well-definedness and recognition depend not only on the problem but also on the knowledge that is available to the problem solver. For that reason, much of the training of scientists is aimed at increasing the degree of well-definedness of problems in their domain.

In all but the most trivial problems, the search process can be quite demanding. If we represent the problem space as a branching tree of m moves with b branches at each move, then there are b^m moves in the full problem space. As soon as m and b exceed small values, exhaustive search of the space is beyond human capacity. Thus, effective problem solving depends in large part on processes that constrain search judiciously to the exploration of a few branches.

Search constraint processes may include weak methods and strong methods. Weak methods, although requiring little knowledge of the problem structure, are correspondingly unselective in searching the problem space. Strong methods may find solutions with little or no search. For example, someone who knows the calculus and is seeking the maximum of a function applies a known algorithm (taking the derivative and setting it equal to zero), finding the answer without search. But it is up to the recognition process to detect the fit between a given problem and the maximization of a continuous function. We describe five major weak methods: generate and test, hill climbing, means–ends analysis, planning, and analogy.

Generate and test. This method, often called trial and error, consists simply of applying some operator to the current state and then testing to determine if the goal state has been reached and the problem solved. If it hasn't, then some other operator is applied. An example of a dumb generating process would be searching in a box of keys for a key to fit a lock and tossing failed keys back into the box without noting anything about the degree of fit. A slightly smarter generator would, at the least, try each key only once.

Hill climbing. In hill climbing, one makes a tentative step in each of several directions and then heads off in the direction that has the steepest gradient. More generally, the method computes progress in the direction of the goal. The move that shows most progress is chosen, and then the process iterates from the new state. Hill climbing uses more information than generate and test about the direction and distance of the goal and uses that information to constrain search in the problem space.

Means–ends analysis. Means–ends analysis compares the current state and the goal state and describes the differences between them. Then it searches for an operator that is designed to reduce the most important differences (Dunker, 1945; Newell & Simon, 1972). If the conditions for the applicability of the operator are not met, a subgoal is formulated to reduce the difference between the current state and a state in which the desired operator can be applied. Thus, the method attempts to solve the subproblem recursively.

Planning. Planning involves (a) forming an abstract version of the problem space by omitting certain details of the original set of states and operators, (b) forming the corresponding problem in the abstract problem space, (c) solving the abstracted problem by applying any of the methods listed here (including planning), (d) using the solution of the abstract problem to provide a plan for solving the original problem, and (e) translating the plan back into the original problem space and executing it (Newell & Simon, 1972).

Preparing a meal is a complex problem-solving task. A plan might be the following: go into the kitchen, select a menu, prepare each dish, set the table, serve the meal. Because planning suppresses much of the detail in the original problem space, it is not always possible to implement the plan, for some of the steps in planned solution paths may not be achievable. For example, an essential ingredient for one of the items on the planned menu may not be on the shelves.

Analogy. Analogy involves mapping a new target domain onto a previously encountered base domain (Vosniadou & Ortony, 1989). Mappings vary widely in complexity. In their simplest manifestation, they involve simply the recognition that the current problem can be solved by a known procedure. At the other extreme, the mapping process may be quite elaborate (Gentner & Jeziorski, 1989; Halford, 1992) and, like the other weak methods, not guaranteed to produce a solution. Analogy can be viewed as one method for changing the given problem space to another that is more effective.

Analogical mappings thus provide the principal bridge between weak and strong methods when the source of the analogy is a well-defined procedure. Used in conjunction with domain-specific knowledge, analogy may enable the search process to be greatly abridged when patterns are noticed in the current problem state. Prestored knowledge can be evoked and used to plan the next steps toward solution of the problem, provide macros to replace whole segments of step-by-step search, or even suggest an immediate problem solution. The recognition mechanism (with its associated store of knowledge) is a key weapon in the arsenal of experts and a principal factor in distinguishing their performance in the domain of expertise from that of novices.

Problem Solving: Summary

Scientific practice applies a plethora of strong methods, such as standard experimental paradigms, established theories and known parameter values, specialized instrumentation, and even highly constrained publication formats. Strong methods can admit, as in the maximization example above, the direct application of a method with little or no search. Weak methods, though less effective when strong methods are available, are of special interest for a theory of scientific discovery because they are applicable in a wide variety of contexts and because fewer and fewer strong methods remain available as the scientist approaches the boundaries of knowledge. Moreover, analogy, though a weak (i.e., very general) method, draws on all the domain-specific knowledge and skill stored in memory.

Especially central to the weak methods are the processes of selective (heuristic) search and the habit of storing in long-term memory large bodies of domain-specific information, indexed by recognizable patterns, so that its relevance will be evoked in particular situations and the information will become accessible. To study scientific discovery, we have to find out how to observe these and the other weak methods at work on scientific problems or to evoke them experimentally in contexts that mimic some of the richness of actual research contexts, while at the same time maintaining the objectivity that supports sound inference.

Complementarity of Approaches

Our theoretical framework views scientific discovery as a type of complex problem solving. This framework provides a common language that can be used to describe both complementarity and convergence in the various approaches to the study of scientific discovery. A powerful way to exploit complementarity is to study the same scientific discovery using more than one approach. In this section, we give several examples of how the strengths of different approaches can be complemented and their weaknesses attenuated or eliminated by using them together. We begin with historical and laboratory studies, whose strong complementaries are revealed by Table 1.

Combining Historical With Laboratory Studies

The discovery of genetic control. In the late 1950s, Jacques Monod and François Jacob discovered the mechanisms by which the synthesis of lactose is controlled in bacteria by control genes (Jacob & Monod, 1961). For this discovery, Monod, Jacob, and their mentor André Lwoff were awarded the Nobel prize in 1965. A substantial historical literature examines the discovery (e.g., Judson, 1979, 1996), including an autobiographical account (Jacob, 1988). As is sometimes the case in historical analyses of scientific discovery, cognitive processes are given an important role in Judson's account of Jacob and Monod's work, although the historian necessarily treats these processes at a very general—almost metaphorical—level.

> Scientists reach an extreme, sustained identification of the patterns of their thought with the patterns that they perceive in, and project into, the phenomena they are trying to elucidate. . . . In each of the discoveries he (Monod) made in the ensuing ten years there was a moment of total absorption as he resolved the bacterial cell like a partly cut diamond slowly in the light, then a gleam of perception so quick—and so quickly resolved into its place in the sequential pattern—that to Monod himself, on his testimony at least, there had been nothing much to call an intuitive leap, merely an extension, subject to test, of the inevitable logic of the system itself. (Judson, 1996, p. 387)

But for a cognitive psychologist, to characterize Monod's discoveries in terms of a gleam of perception is to not describe them at all. Instead, the goal is to identify specific and well-understood cognitive processes and then to determine their role in the discovery. In the case of the discovery of the mechanism of genetic control, perhaps the most important cognitive processes involved were the representational changes that enabled Jacob and Monod to replace their entrenched idea that genetic control must be some kind of activation mechanism with the discovery that it was, instead, an inhibition mechanism. As Dunbar (1993) put it,

> They made the novel and unexpected discovery that groups of genes control other genes and keep many genes inhibited until particular enzymes are needed. This was the novel concept of an "operon" that forced a "radical restructuring" of the concept of how genes work. (p. 398)

To better understand the cognitive processes involved in this important discovery, Dunbar (1993) created a laboratory task that he used to study the behavior of college students faced with a problem that captured some of the essential elements of the discovery problem faced by Monod and Jacob, while eliminating many others.[3] Dunbar's simplifications were primarily aimed at constraining the search space and at keeping the depth of search within reasonable limits. His three principle simplifications were (a) to have participants discover how genetic regulation worked (whereas Monod and Jacob first had to discover that there was such a thing as genetic regulation of some genes by other genes), (b) to provide participants with a highly constrained experiment space (whereas Monod, Jacob, and their colleagues had to invent many new procedures), and (c) to limit the necessary discovery to the particular instance at hand, rather than to a broad-based concept of genetic control.

In summary, Dunbar placed his participants in an experimental context that simulated Monod and Jacob's problem at the point where the idea of control genes had occurred to them and they had developed a basic experimental procedure for testing alternative control-gene hypotheses to explain the lactose phenomena. The students were asked to design and run (simulated) experiments to discover the lactose control mechanism. Using a real scientific task increased the typically low face validity of a controlled laboratory study. Although the actual task was simplified for purposes of the experiment, some of the basic components—the problem, the givens, the research methods permitted by known kinds of experiments, the structure of the solution—were all preserved. With good control of the variables, the laboratory data could cast light on the size and structure of the problem spaces that Monod and Jacob searched and on some of the conditions of search that were necessary or sufficient for success.

> Despite the differences between the original discovery of Monod and Jacob and that observed in the studies reported here, clear similarities exist between the conceptual processes employed by the subjects and those employed by Jacob and Monod. . . . The problem for the subjects was to conceive of a new mechanism that could be applied to genetic regulation. That is, the subjects generated a new concept of mutually interacting genes that regulate enzyme production by inhibition. These subjects behaved just like Monod and Jacob. Furthermore, just like the subjects in the experiments reported in this article, Monod and Jacob had difficulty in formulating the concept of inhibitory control due to their belief in activation: In the 1940's, Monod began investigating the conditions under which *E. coli* could be *induced* to produce certain enzymes. Monod hypothesized that this induction of enzymes was an activation mechanism, or a positive process. (Dunbar, 1993, p. 431)

Planck's Law. In at least three other cases, historically important discoveries that have been studied extensively by historical methods have been the subject of complementary experiments in the psychology laboratory. In 1900, Max Planck, having published a theory of blackbody radiation that accounted for the fit of the observed data with Wien's Law (an exponential function), learned that new data in the infrared range showed a large departure from Wien's Law. The data still looked exponential in the higher frequencies but passed nearly linearly through the origin. On the evening of the same day on which he learned about the new infrared data, Planck revised Wien's Law into what we now call Planck's Law (and, in the process of providing an explanation for the new law during the next several months, more or less accidentally discovered the quantum).

To learn more about how the first step was accomplished, some mathematicians and physicists of National Academy stature were approached with the following question: "I have some very smooth and noise-free data relating two variables, x and y. For large values of x, y appears to be an exponential function of x; but for small values of x, the function passes through the origin and is nearly linear. Can you suggest what function might fit these data?" (Langley et al., 1987, pp. 47–53).

Five out of the eight scientists who were asked the question answered it, and in well under 3 min. The answer was always essentially Planck's Law. When asked how they reached their answer, the respondents were able to report either (a) that they expanded the exponential into a Taylor's Series and noted that if they subtracted unity from it, it would have the desired properties; or (b) that they visualized the graph of the function and saw it cutting the y-axis at $y = 1$, then subtracted unity from it. In no case did they notice the relation of their problem and solution to Planck's Law of blackbody radiation, although all of them were thoroughly familiar with that law.

Planck published an account of how he himself found the function that fit the new data. His path is slightly different from those given above but is equally simple. Looking at his previous theory, he quickly found that he could get the desired result by adding a quadratic term to a particular linear equation in the derivation. He then spent 2 or 3 months trying to rationalize the change in terms of the physics of the situation. (His numerical law of blackbody radiation is still accepted, but his physical explanation bears little relation to today's quantum mechanics.) This complementarity of history with experiment greatly reduces the mystery of how Planck's initial step—the change in the key function—could be achieved by Planck in a few hours. Neither the respondents in the experiment nor Planck made use of the physics of the situation in taking that step; it was pure numerology.

Balmer's Law. A second example is provided by Balmer's Law, a simple algebraic formula for the frequencies of successive lines in the spectrum of hydrogen: $x = kn^2/(n^2 - 4)$, where x is a frequency, k is a constant, and n is the sequence number of the line with that frequency ($n \geq 3$). The law was discovered by Balmer, a geometry teacher who was thoroughly innocent of the physics of the problem, after seeing the values of the first four spectral lines (Banet, 1966).

[3] Clearly, no laboratory simulation could present the full challenge presented by scientific problems. Monod and Jacob worked together for several years before they made their discovery, and most college students would prefer to spend less time than that in the psychologists' laboratory.

To learn more about the problem space in which this solution was found, a graduate student in engineering was hired for a summer, given the spectral data (simply as a sequence of four values of x and n), and asked to find a pattern that could be extrapolated to higher values of n. He found the equivalent of Balmer's law after about 6 weeks, working perhaps 10 to 20 hr per week on the problem. Here, as with Planck's Law, an important physical law was discovered by a pure exercise in pattern finding, without physical motivation or theory. (Thirty years later, Balmer's Law served as the principal evidential base for Rutherford's quantum theory of the hydrogen atom.) In both cases, the evidence from a laboratory study complements the less detailed historical evidence in revealing the nature of the discovery path.

Faraday's experiments. Laboratory experiments of quite a different kind are replications by historians of science (not subjects) of the historical experiments that are being studied. Gooding (1990) has emphasized that experiments don't interpret themselves. Even the process of describing laboratory experiments and representing the instruments and the observations in language that permits other scientists to understand and replicate them is a problem, often difficult, that has to be solved by the investigator.

Between late August and early December 1831, Faraday carried out the key experiments that demonstrated the phenomena of induction of currents by magnets. As he began to draft his first paper on this work, he discovered that he had difficulty in interpreting his own experiments, as recorded in his lab diary, and especially the directions of the magnetic forces and currents. He took about a week to replicate some of his experiments and to develop a consistent way of representing and describing his manipulations and findings so that he could communicate them clearly to his fellow scientists.

Gooding (1990) has given us a careful and insightful analysis of the same process when Faraday, in 1821–1822, published a survey article on the status of electromagnetic research shortly after Oersted had made his surprising discovery of the generation of a magnetic field by an electric current. Gooding points out that Faraday repeated most of the experiments he was reviewing to understand, represent, and describe the manipulations and findings unambiguously and that he sometimes supplemented his written communications to his colleagues with small examples of the experimental apparatus that enabled them to repeat the experiments themselves. And most instructive for our present discussion, Gooding himself found it highly informative to repeat again many of the experiments to understand the problems of experimentation and communication faced by Faraday and his contemporaries.

In all these cases of complementarity between historical and laboratory approaches, we see that the use of a historically important scientific discovery as the substance of the task solves the problem of face validity for the laboratory data and that the laboratory provides valuable new information about the processes of discovery at a high level of temporal resolution—on the order of minutes and seconds. In combination, these approaches cut the Gordian Knot of face validity and external validity.

Combining History With Modeling

Our next example illustrates the complementarity of historical and model-building approaches. The path that the biochemist, Hans Krebs, followed in his discovery of the reaction path for the in vivo synthesis of urea has been the subject of a very careful and thorough historical study by Holmes (1991), who used not only the published papers but also the lab notebooks of Krebs and his assistant Henseleit, and conducted interviews with Krebs (some 40 years after the discovery was made).

Recently, two computer models of scientific discovery (KEKADA, by Kulkarni & Simon, 1988; CDP, by Grasshoff & May, 1995) have been applied to modeling the urea synthesis case. After the models proposed an experiment and were given its outcome, they then proposed another experiment, using the knowledge of previous outcomes to select the search path. Both programs, using no more knowledge of biochemistry than Krebs possessed at the outset of his work, succeeded in discovering the reaction path discovered by Krebs, following fairly closely the same lines of experimentation.

The simulations showed that the experimentation could be steered by very general hypotheses (e.g., hypotheses, already widely accepted, that ammonia and amino acids were likely sources for the nitrogen in urea) so that experimental outcomes generally guided theorizing, rather than theory guiding experimental design. The simulations also showed that surprise at unexpected experimental outcomes could provide powerful heuristics for choosing the next steps in search. The models sharpened up considerably the ambiguities in choice of strategy that Holmes had detected at various points in Krebs' search. The KEKADA program has also simulated some aspects of Faraday's 1831 discovery of induction of electricity by magnetism, including the effects of the surprise experienced at the outcome of his first experiment.

There has been extensive computer modeling of other important historical discoveries, but with less detailed comparison than in the examples just cited between the paths of discovery inferred from the historical materials and those followed by the simulation programs. In this research, the emphasis has been on finding general mechanisms of discovery that are effective over a wide range of tasks, in the sense of being able to reproduce the product, if not always the (largely unknown) details of the process, of the corresponding historical discoveries.

BACON, already mentioned, is a widely known model of this kind which has been used to simulate such discoveries as Boyle's law, Kepler's Third Law, Ohm's law, Coulomb's law, Archimedes' law, Black's laws of heat, various laws of chemical combination, the law of gravitation, conservation of momentum, and Snell's law of refraction, as well as to perform concept attainment and series extrapolation tasks (Langley et al., 1987, Part II). Although BACON exists in a half dozen variant forms, all of them depend on a small set of heuristics that guide the generation of mathematical functions to fit the given data to a good approximation. Hence BACON is a heuristic search system that uses a generate-and-test subsystem to generate hypotheses for comparison with data, the hypothesis generator being sensitive to feedback of the results of trying to fit the previous hypotheses.

BACON's heuristics, which are strictly weak methods, contain no information about the meaning of the data, so that the program's discoveries are describable as data driven rather than theory driven. Because BACON carefully designs each new function it generates with the aid of information about previous fits or misfits, it typically generates only a few functions before finding one that fits the data. Thus, BACON throws light on those discoveries of science, especially frequent in the early years of any new field, where little or no theory is initially available to guide experiment.

BACON is just one of a family of models of various kinds of data-driven discovery. Others include STAHL, GLAUBER, and DALTON (Langley et al., 1987, Part III). A characteristic of this line of research is that it begins to fill in the large lacuna in the literature of scientific methodology, which has tended to neglect data-driven discovery, observation, and exploratory experiments, and has paid attention almost exclusively to controlled experiments as a means for testing theories that have already been fully formulated. An important product of the modeling has been to complement historical studies where the science is in large measure data driven but where historical data are not sufficiently fine-grained to show how observations of phenomena can guide experimentation (e.g., the studies of the work of Krebs and Faraday).

History, Lab, and Model

One example can be cited where historical data, a model, and a laboratory experiment have all been used to provide complementary analyses of the same historically important discovery: Kepler's discovery of his Third Law of Planetary Motion, which states that the periods of revolution of the planets about the Sun vary with the 3/2 power of their mean distances from the Sun. The historical record of this discovery is very sparse (Gingerich, 1975), consisting largely of Kepler's own published accounts, first of his discovery of an erroneous law (that the periods vary with the squares of the distances), then of his discovery of the correct law 2 decades later.

From the history, we know some of the difficulties he encountered (especially errors in arithmetic), and the distractions of his life when the problem lay fallow. We know that he did not have logarithms available at the time he found the law. We are almost wholly lacking fine-grained temporal data on the stages of discovery. There was little physical motivation for the law, although Kepler held some notions about the sun as the source of force that produced the revolutions, and these notions may have pointed him to the square law—certainly not to the $D^{3/2}$ power law. But the central question is what led him, with only a few weeks of active search, to this particular mathematical function—$P = D^{3/2}$—out of all the functions he might have generated? His problem space of possible functions was enormous, and that he found this particular one relatively quickly calls for explanation.

When BACON is given the same data that Kepler had, and nothing more, it finds the correct law as the third or fourth function that it generates, the exact order depending on the precise heuristics it uses. In either case, BACON's heuristics guide it almost directly to the answer after trying no more than one or two inadequate alternatives. What is perhaps more remarkable is that the second function it tries is the quadratic—the false answer that Kepler proposed in his initial publication on the law. Given the sketchy historical evidence, BACON does a remarkable job of tracing the original discovery path. Whether it followed that path for the same reasons that Kepler did cannot be answered with the data that are available.

To complement further the data available from history and from the BACON simulation, Qin and Simon (1990) conducted tests with laboratory participants, giving them Kepler's data (the variables identified only as x and y) and asking them to find a function that fit the data. Of 14 college students, 4 found Kepler's Third Law in an hour or less; the other 10 failed. Of those who failed, 4 had weak mathematical backgrounds and generated scarcely any functions but straight lines. The remaining 6 generated a variety of functions (not including the right one) but showed no evidence that the results from an unsuccessful attempt to fit a function had any influence on what function they chose to test next. On the other hand, the four participants who found the law all used selective heuristics to choose the next function for testing according to the nature of the misfits of those previously tried. Their heuristics had a close resemblance to those of BACON (with which they were not familiar).

The values of experimentation and modeling in providing evidence about feasible and likely search paths in the absence of much historical data to identify the path actually followed to make the original discovery are well illustrated by this example.

Laboratory Studies: Exploratory and Controlled

In some cases, the two types of laboratory studies—exploratory and controlled—can be focused on a common problem. In Klahr and Dunbar's (1988) original study of BigTrak, there were no control conditions. Instead, the project was conceived as an exploration of what would happen when participants were presented with a relatively complex (for a lab study) discovery task. The results of this study, as noted elsewhere, led to the discovery of two distinct strategies for approaching the discovery problem (i.e., the theorist–experimenter distinction). This exploratory study was then followed up with very carefully designed factorial experiments that controlled for the age and scientific background of participants and introduced different levels of plausibility of the suggested hypothesis that participants were asked to explore (Klahr et al., 1993). This was followed by a series of investigations (reported in Klahr, 1999) that alternated between exploratory and controlled laboratory experiments, yielding a number of interesting findings on the development of scientific reasoning processes.

Convergent Evidence of Principles of Discovery

In the previous section, we focused on the complementarities of the different approaches. In this section, we give several examples of the kind of convergent evidence obtained by using two or more approaches to study the same scientific discovery. We provide a few examples of how basic processes of discovery revealed in one situation can be tested and generalized to other situations. Such comparison takes us from the limitations of individual case studies to the construction and testing of general theories.

Surprise

In this century, the reigning theories of philosophy of science have generally taken hypotheses as first (or at least unexplained) causes that lead to experiments designed to test them. In this view, the hypotheses themselves derive from scientists' intuitions and are beyond scientific explanation (Popper, 1959). The history of science has taken a much less rigid position with respect to hypotheses and has included the question of their origins within the scope of its interests and methods.

For example, historical accounts of the discovery of radium by the Curies usually start with their project to obtain pure radioactive uranium from pitchblende to study the behavior and properties of uranium They were familiar with the level of radioactivity of uranium, and as they proceeded to process the pitchblende, they were surprised to find that the level of radioactivity began to exceed that of pure uranium. A surprise calls for an explanation, and the explanation that occurred to them was that the pitchblende contained a second substance (which they named radium) that was more radioactive than uranium. The test of this hypothesis consisted in extracting this substance, separating it from both the pitchblende and the uranium, and determining some of its key properties.

In this case, a phenomenon led to a hypothesis, rather than a hypothesis to experimental phenomena. This is not a singular case in scientific history but a frequent occurrence. Often, as was true in this instance, it is accompanied by surprise; that is, the observed phenomena was unexpected and unpredictable from the knowledge the scientists already held

A surprise can only occur when expectations that have been formed are violated. In observational studies and exploratory experimentation, phenomena are, from time to time, recognized as conflicting with previously stored knowledge and expectations about the problem domain. In the face of surprise, scientists frequently divert the path of exploration to ascertain the scope and import of the surprising phenomenon and to determine its mechanism (see Darden, 1992, and Darden & Cook, 1995, for an analysis of responses to anomalies based on the historical record, and Chinn & Brewer, 1998, for a laboratory investigation of how people respond to anomalous data).

We have seen that the KEKADA discovery model, already discussed in connection with modeling the research of Krebs and Faraday, addresses the surprise issue directly. When performing an experiment, the scientist simulated by KEKADA forms expectations, which are based on previous experience, about outcomes When the actual outcomes violate the expectations, the scientist is surprised and, in the KEKADA theory, takes steps to explain the surprising phenomenon These steps may include steps to discover the scope and generality of the phenomenon and then steps to discover its mechanism.

The KEKADA model permits an examination of surprise as it arises in different experimental environments, allowing a rigorous statement of a general theory of the role of surprise in discovery and of its mechanisms. In the case of Krebs, it shows how an unexpected, large yield of urea in the presence of a particular amino acid, ornithine, led Krebs to discover the catalytic role of ornithine in the production of urea from ammonia. Similarly, surprise at obtaining a transient flux of electricity in a circuit when a nearby magnet was activated led Faraday, through a long series of experiments aimed at understanding the mechanism of the surprising phenomenon, to discover the means of using magnets to produce continuous electric currents.

In a laboratory study in which participants had to discover the function of an unknown key on a simulated rocket ship, Klahr et al. (1993) investigated the effects of surprise by providing participants with suggested hypotheses about how the key worked. These hypotheses were designed to be either highly plausible or highly implausible. Adults and children had quite different reactions to implausible hypotheses. Adults usually proposed a competing hypothesis and then generated experiments that could distinguish between them. On the other hand, young children (third graders) tended to dismiss an implausible hypothesis and ignore evidence that it might be correct. Instead, they adopted a kind of engineering mode in which they attempted to demonstrate that their favored hypothesis was correct by showing that (under some circumstances) they could control the behavior of the device. These results imply that an important aspect of the development of scientific thinking is coming to accept, rather than deny, surprising results and to explore further the phenomenon that gave rise to them.

Here, we see historical studies, simulation models, and laboratory experiments all converging on a particular set of phenomena—in this case, reaction to surprise—thereby generating and testing new theories to describe and explain them. Observational studies can also contribute to this convergence if good fortune leads to the observation of an "aha" event or a surprise.

The BACON model, which we have mentioned several times, was built to test the veridicality and usefulness of the hypothesis that, especially in new fields of science, where theory is poorly developed or absent, observation typically precedes hypothesis construction: that the first step in progress is to find patterns (laws) in data. We have already described how BACON finds Kepler's Third Law on its third or fourth try and performs with similar efficiency in a large number of other tasks. Here, we see the emergence of a general theory of data-driven science based on heuristic search by a hypothesis generator of the space of possible problem solutions guided by feedback (knowledge of results) to the generator of successive hypotheses. The hypotheses are a combined product of the internal structure of the generator and of the observed data

To test the range of its applicability as a theory of discover, BACON, like KEKADA, can be applied to the findings of historical, laboratory, or observational studies of discoveries, and the acuteness of the test of its veridicality is only limited by the completeness and temporal resolution of the data that are available.

The Role of Analogy and Recognition

Although philosophers have long been interested in the role of analogy in science (Duhem, 1914/1954; Hesse, 1966), it is only in the past 25 years that analogy has assumed prominence in theories of problem solving and scientific discovery and that its underlying cognitive mechanisms have been studied in detail (Darden, 1980; Gentner, 1982). Holyoak and Thagard (1995) provide several examples of analogical problem solving in major scientific discoveries, ranging from a first-century analogy between sound and water waves to Turing's mind/computer analogy. As we argued earlier, analogy can be viewed as a complex form of recognition. Holyoak and Thagard also emphasize the role of analogical thinking in cognitive development (see Goswami, 1996, for an extensive review), and Nersessian (1984) documents its role in several of the major scientific discoveries of the 19th century. Analogical reasoning also plays a central role in recent analysis of the thinking processes of contemporary scientists working in their labs (Dunbar, 1994; Thagard, 1997; Ueda, 1997), and it is often the method of choice for formulating initial hypotheses and experiments in a variety of discovery contexts

Multiple Search Spaces

The reciprocal relation between hypotheses and phenomena that we have just observed has been noticed in a number of different approaches to scientific discovery, including laboratory studies, historical studies, and computational models of discovery. A problem-solving orientation enables us to use a common language in describing all of these in terms of search in multiple spaces. We begin this discussion with a characterization that includes only two distinct spaces, and then we expand the number of spaces as required.

The two-space view was first proposed by Klahr and Dunbar (1988) to account for the results of their laboratory study in which participants had to discover the functionality of a particular control button on a programmable toy vehicle. Klahr and Dunbar found that participants sometimes searched for experimental manipulations that would provide new information about the button's functions and sometimes searched for rules that explained the device's behavior in response to the manipulations. Noting that Simon and Lea (1974) had proposed that human concept formation uses searches in separate instance and hypothesis spaces, Klahr and Dunbar extended the dual-search notion to the domain of scientific discovery, where one has to coordinate search in two spaces: a space of experiments and a space of hypotheses.

Search in the hypothesis space. Generating new hypotheses is a type of problem solving in which the initial state consists of some knowledge about a domain, and the goal state is a hypothesis that can account for some or all of that knowledge in a more concise or universal form. Once generated, hypotheses are evaluated for their initial plausibility. Expertise plays a role here, as participants' familiarity with a domain tends to give them strong biases about the plausibility of hypotheses. Plausibility, in turn, affects the order in which hypotheses are evaluated: highly likely hypotheses tend to be tested before unlikely hypotheses (Klayman & Ha, 1987; Wason, 1968). Furthermore, participants may adopt different experimental strategies for evaluating plausible and implausible hypotheses.

Search in the experiment space. Hypotheses are both generated from and evaluated through experimentation. But it is not immediately obvious what constitutes a good or informative experiment. In constructing experiments, scientists are faced with a problem-solving task paralleling their search for hypotheses. However, in this case search is in a space of experiments rather than a space of hypotheses. If experiments are used to generate new information, then they should be designed to maximize the likelihood that they will reveal something of interest. If they are being used to test hypotheses, they should discriminate among rival hypotheses. Both uses of experimental outcomes involve search in a space of experiments that is only partially defined at the outset. Constraints on the search must be added during the problem-solving process.

The dual-search notion can be used to illustrate the convergence of several types of investigations of scientific discovery. In their laboratory studies, Klahr and Dunbar found that some participants (experimenters) focused on searching the space of possible manipulations, whereas other participants (theorists) focused on the space of possible explanations of the responses. Similar differences in preference between experiment-driven and theory-driven strategies have been noticed in other laboratory studies (Okada, 1994; Okada & Simon, 1997). Studies based on historical approaches can be interpreted in terms of the balance between hypothesis-space search and experiment-space search. For example, in most histories of Faraday's discovery of induction of electricity by magnets, much emphasis has been placed on the influence of Ampère's theory of magnetism on Faraday's thought, but a strong case can be made (Gooding, 1990) that Faraday's primary search strategy was to focus on experiment space search, yielding a discovery path that was driven largely by phenomena rather than theory.

The dual-space characterization reveals another type of convergence by allowing us to categorize computational models of discovery according to which space they emphasize. Some focus mainly on search in the hypothesis space: for example, the BACON models and variants (Langley et al., 1987), IDS (Nordhausen & Langley, 1993), PHINEAS (Falkenhainer, 1990), COPER (Kokar, 1986), MECHEM (Valdés-Perez, 1994b), HYPGENE (Karp, 1990), AbE (O'Rorke, Morris, & Schulenburg, 1990), OCCAM (Pizzani, 1990), and ECHO (Thagard, 1988). Other computational models focus mainly on the process of experiment generation and evaluation, for example, DEED (Rajamoney, 1993)and DIDO (Scott & Markovitch, 1993). A few deal with both processes, for example, KEKADA (Kulkarni & Simon, 1988), STERN (Cheng, 1990), HDD (Reimann, 1990), IE (Shrager, 1985), and LIVE (Shen, 1993). This categorization might provide a starting point for the integration of these different models into a very rich computational implementation of the dual-search framework.

Beyond two spaces. Although the two-space model was adequate to capture most of the behavior of participants in the Klahr and Dunbar work, analysis of participants' behavior in a more complex microworld laboratory (Schunn, 1995) necessitated the expansion from a two-space to a four-space model, depicted in Figure 1. In this model, the hypothesis space has been expanded to included both a data representation space and a hypothesis space. In the data representation space, representations or abstractions of the data are chosen from the set of possible features. What people search for in this space is an effective and informative way to represent the phenomena they are observing. Additional support

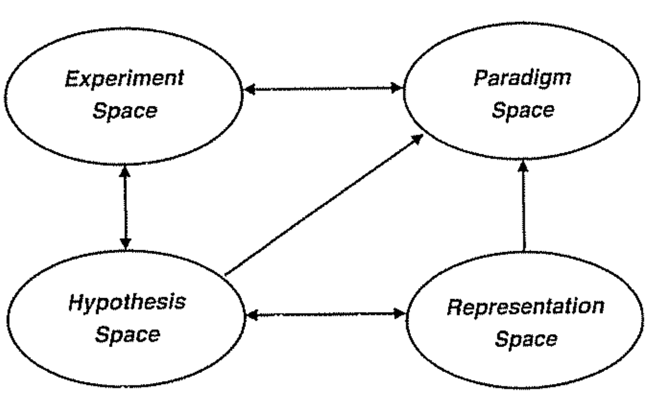

Figure 1. A four-space model of scientific discovery. The arrows indicate the direction of information flow among the four spaces. From "A 4-Space Model of Scientific Discovery" (p. 106), by C. D. Schunn and D. Klahr, 1995, in J. D. Moore and J. F. Lehman (Eds.). *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum. Copyright 1995 by C. D. Schunn. Reprinted with permission.

for the importance of a representation space comes from Cheng and Simon's (1992) analysis of Gallileo's research in which they compare the relative difficulty of mathematical and diagrammatic representations. Here, as in many areas of science, finding the right representation is crucial, and it requires heuristic search, with all of its associated weak methods, in a large space of possibilities.

Figure 1 also shows how the experiment space is now divided into an experimental paradigm space and an experiment space. In the experimental paradigm space, a class of experiments (i.e., a paradigm) is chosen that identifies the factors to vary and the components that are held constant. In the experiment space, the parameters settings within the selected paradigm are chosen.

It should be clear that there is no right number of spaces because that is entirely dependent on the nature of the discovery context (Langley et al., 1987).[4] In his analysis of the discovery of the bacterial origins of stomach ulcers, Thagard (1998) demonstrates the importance of search in at least three major spaces: hypothesis space, experiment space, and a space of instrumentation.

Search in the space of representations. One method for searching the representation space that has attracted considerable attention is analogy. A prominent example is Bohr's use of the solar system analogy in arriving at his quantum model of the hydrogen atom. He viewed the planetary electrons as orbiting the nucleus, ignoring the fact that, according to classical physics, the charged electrons would produce a magnetic field, thereby dissipating energy until they fell into the nucleus. Instead of abandoning the analogy, he borrowed Planck's quantum, which allowed energy to be dissipated only in leaps of quantum size, then showed that these leaps would produce a light spectrum corresponding exactly to the Balmer series of the hydrogen spectrum, introduced 30 years previously. Not an analogy so much as a thoroughly mixed metaphor, one might say, but a successful one, though it required much tinkering and radical additional representation changes (ultimately, Schrödinger's wave equations and Heisenberg's matrix mechanics) before it could be extended systematically beyond hydrogen and ionized helium to the other elements.

In the autumn and early winter of 1831–1832, when he was making his fundamental discoveries of induction of electric currents by magnetism, Faraday went through a whole series of representations of the phenomena he was seeing, which are revealed by his diary. Initially, he visualized the sudden energizing of a magnet as creating a current in any nearby closed metallic circuit, but this state was immediately terminated by the creation of an electrotonic state in the circuit that opposed the flow of current. He tried hard, but unsuccessfully, to find independent experimental evidence for the electrotonic state. Then he noticed that if a magnet moved continually in the neighborhood of the circuit, a continuous current would be produced—the phenomenon could be represented in terms of relative movement. Next, he visualized the lines of magnetic force (which he had long been familiar with as revealed by the arrangement of iron filings around a magnet) and theorized that the current was induced when the relative movement of circuit and magnet caused the former to cut the lines of magnet force around the latter. In Faraday's case, it was not an analogy that led to this final representation but a series of phenomena observed in the course of a long series of experiments (most of them not predicted before the experiments were run) combined with the experience of actually seeing the lines of force.

In chemistry at the end of the 18th century, a major event was the rather rapid shift from the phlogiston theory of combustion to the oxygen theory, in which a change in representation of the standard experiments played a major role. In earlier experiments on combustion, the main phenomena observed were (a) the solid materials of combustion and their residues and (b) the heat, flame, and smoke produced during the combustion process. The latter provided the basis for hypothesizing a phlogiston, which was supposed to be driven out of materials in the course of combustion. New methods of observing and measuring the volumes and pressures of gases showed that large quantities of gases were absorbed, and other gases produced, during combustion in air. For example, oxygen was often absorbed and carbon dioxide given off. A focus of attention on the gases instead of the heat and flame changed the representation of the combustion process and produced the new oxygen theory.

From these examples, we see that representations can derive from many sources: analogies, phenomena produced by, but not predicted by, experiments, and even new ways of seeing experiments triggered by new instruments of observation.

Search in the strategy space. Finally, changes in strategy, even while a fixed problem representation is maintained, may play an important role in discovery. Often the change in strategy results from, or leads to, the invention of new scientific instruments or procedures. Breeding experiments are a tool of genetics research that goes back to Gregor Mendel (and for the applied genetics of agriculture, many centuries further back than that). The productivity of such experiments depended on the rates at which mutations occurred. Müller, with the simple idea that x-rays could induce higher rates of mutation, substantially improved that productivity.

A number of issues regarding the relative efficacies of different strategies for research on scientific discovery have been discussed in this article: strategies of using observational studies and exploratory experiments versus controlled experiments, or choices among historical studies, laboratory experiments, and observations. These same kinds of choices must be made in the other domains of scientific research.

Relation Between Science Studies and More General Studies of Creativity and Problem Solving

We come now to our final generalization: the hypothesis that the theory of scientific discovery is a special case of the general theory of problem solving, the special features being supplied by the strong methods of each discipline and the knowledge and procedures that support them, while the ubiquitous weak methods supply the commonalities. In our exploration of scientific discovery, we have seen that (a) it is based on heuristic search in a set of problem spaces: spaces of instances, of hypotheses, of representations, of strategies, of instruments, and perhaps others; (b) the control structures for search are such general mechanisms as trial

[4] Even though some scientists, the first author included, have been drawn into debates about whether the magic number is two (Klahr & Dunbar, 1988; van Joolingen & de Jong, 1997), three (Baker & Dunbar, 1996; Burns & Vollmeyer, 1997), four (Schunn & Klahr, 1995, 1996), or *N* (Wolf & Beskin, 1996).

and error, hill climbing, means–ends analysis, and response to surprise; and (c) recognition processes, evoked by familiar patterns recognized in phenomena, evoke knowledge and strong methods from memory, thereby linking the weak methods to the mechanisms that are domain specific.

All of the constructs and processes mentioned above are also the constructs and processes that are encountered in problem solving in all the domains in which it has been studied. A painter is not a scientist, nor is a scientist a lawyer, a businessman, a machinist, or a cook. However, they share the same general approach to solving their respective problems, and they use the same weak methods. When their problem-solving activity is described at the level we have been using in this article, each can understand the rationale of the expert's activity, however abstruse and arcane the content of their special expertise may appear.

If we press to the boundaries of creativity, the main difference we see from more mundane examples of problem solving is that the problems become less well structured, recognition becomes less powerful in evoking prelearned solutions or powerful domain-specific search heuristics, and more, not less, reliance has to be placed on weak methods. The more creative the problem solving, the more primitive the tools. Perhaps this is why childlike characteristics, such as the propensity to wonder, are so often attributed to creative scientists and artists.

References

Baker, L. M., & Dunbar, K. (1996). Problem spaces in real-world science: What are they and how do scientists search them? In G. W. Cottrell (Ed.), *Proceedings of the eighteenth annual conference of the Cognitive Science Society* (pp. 21–22). Mahwah, NJ: Erlbaum.

Banet, L. (1966). Evolution of the Balmer series. *American Journal of Physics, 34,* 496–503.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change (2nd ed.).* New York: Pergamon.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science.* Madison, WI: University of Wisconsin Press.

Bijker, W. E., Hughes, T. P., & Pinch, T. (1987). *The social construction of technological systems: New directions in the sociology and history of technology.* Cambridge, MA: MIT Press.

Bloor, D. (1981). *Knowledge and social imagery.* London: Routledge and Kegan Paul.

Boden, M. A. (1990). *The creative mind: Myths and mechanisms.* London: Basic Books.

Brewer, W. F., & Samarapungavan, A. (1991). Child theories versus scientific theories: Differences in reasoning or differences in knowledge? In R. R. Horfman & D. S. Palermo (Eds.), *Cognition and the symbolic processes: Applied and ecological perspectives* (pp. 209–232). Hillsdale, NJ: Erlbaum.

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

Burns, B. D., & Vollmeyer, R. (1997). A three-space theory of problem solving. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual meeting of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Callahan, J. D., & Sorensen, S. W. (1992). Using TETRAD II as an automated exploratory tool. *Social Science Computer Review, 10,* 329–336.

Cheng, P. C.-H. (1990). *Modeling scientific discovery.* Unpublished doctoral dissertation, The Open University, Milton Keynes, UK.

Cheng, P. C.-H., & Simon, H. A. (1992). The right representation for discovery: Finding the conservation of momentum. In D. Sleeman & P. Edwards (Eds.), *Machine learning: Proceedings of the ninth international workshop* (vol. ML92, pp. 62–77). San Mateo, CA: Morgan Kaufman.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching, 35,* 623–654.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field studies.* Chicago: Rand McNally.

Crick, F. (1988). *What mad pursuit: A personal view of scientific discovery.* New York: Basic Books.

Darden, L. (1980). Theory construction in genetics. In T. Nickles (Ed.), *Scientific discovery: Case studies* (pp. 151–170). Boston, MA: D. Reidel Publishing Company.

Darden, L. (1992). Strategies for anomaly resolution. In R. N. Giere (Ed.), *Cognitive models of science. Minnesota studies in the philosophy of science* (Vol. 15, pp. 251–271). Minneapolis: University of Minnesota Press.

Darden, L. (1997). Recent work in computational scientific discovery. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual meeting of the Cognitive Science Society* (pp. 161–166). Mahwah, NJ: Erlbaum.

Darden, L., & Cook, M. (1995). Reasoning strategies in molecular biology: Abstractions, scans and anomalies. *Philosophy of Science, 2,* 179–191.

Duhem, P. (1954). *The aim and structure of physical theory* (part I, chapter IV). Princeton, NJ: Princeton University Press. (Original work published as *La theorie physique.* Paris, 1914)

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17,* 397–434.

Dunbar, K. (1994). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *The nature of insight* (pp. 365–395). Cambridge, MA: MIT Press.

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. Ward, S. Smith, & S. Vaid (Eds.), *Conceptual structures and processes: Emergence, discovery and change* (pp. 461–492). Washington, DC: APA Press.

Dunbar, K., & Baker, L. A. (1994). Goals, analogy, and the social constraints of scientific discovery. *Behavioral and Brain Sciences, 17,* 538–539.

Duncan, S. C., & Tweney, R. D. (1997). The problem-behavior map as cognitive-historical analysis: The example of Michael Faraday. *Proceedings of the nineteenth annual conference of the Cognitive Science Society* (p. 901). Hillsdale, NJ: Erlbaum.

Dunker, K. (1945). On problem solving. *Psychological Monographs, 58,* p. 270.

Einstein, A. (1950). Physics & reality. In A. Einstein, *Out of my later years* (pp. 59–111). New York: Philosophical Library. (Original work published 1936)

Elder, J. (1990, January 7). Hands on. *The New York Times* (Education Life suppl.), p. 20.

Erdos, P., Fajtlowicz, S., & Staton, W. (1991). Degree sequences in the triangle-free graphs. *Discrete Mathematics, 92,* 85–88.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Falkenhainer, B. (1990). A unified approach to explanation and theory formation. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation* (pp. 157–196). San Mateo, CA: Morgan Kaufman.

Fay, A., & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development, 67,* 689–716.

Feist, G. J. (1991). *The psychology of science: Personality, cognitive.*

motivational and working styles of eminent and less eminent scientists. Unpublished dissertation, University of California. Berkeley.

Feist, G. J. (1994). Personality and working style predictors of integrative complexity: A study of scientists' thinking about research and teaching. *Journal of Personality and Social Psychology, 67*. 474–484.

Feist, G. J., & Gorman, M. E. (1998). The psychology of science: Review and integration of a nascent discipline. *Review of General Psychology, 2*, 3–47.

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition*. Cambridge, MA: MIT Press.

Galison, P. (1987). *How experiments end*. Chicago: University of Chicago Press.

Gentner, D. (1982). Are scientific analogies metaphors? In D. S. Miall (Ed.), *Metaphor: Problems and perspectives* (pp. 106–132). Brighton, UK: Harvester.

Gentner, D., & Jeziorski, M. (1989). Historical shifts in the use of analogy in science. In B. Gholson, W. Shadish, R. Beimeyer, & A. Houts (Eds.), *The psychology of science: Contributions to metascience* (pp. 296–325). Cambridge, UK: Cambridge University Press.

Giere, R. N. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Gingerich, O. (1975). The origins of Kepler's Third Law. In A. Beer & P. Beer (Eds.), *Vistas in astronomy* (pp. 595–601). Oxford, UK: Pergamon.

Gooding, D. (1990). *Experiment and the making of meaning*. Dordrecht, the Netherlands: Nijhoff/Kluwer.

Gorman, M. E. (1992). *Simulating science: Heuristics, mental models, and technoscientific thinking*. Bloomington, IN: Indiana University Press.

Goswami, U. (1996). Analogical reasoning and cognitive development. *Advances in Child Development and Behavior. 26*, 91–139.

Grasshoff, G., & May, M. (1995). From historical case studies to systematic methods of discovery. *Working notes: AAAI spring symposium on systematic methods of scientific discovery* (pp. 46–57). Stanford, CA: AAAI.

Gross, P. R., & Levitt, N. (1994). *Higher superstition: The academic left and its quarrels with science*. Baltimore, MD: Johns Hopkins University Press.

Gruber, H. E. (1974). *Darwin on man: A psychological study of scientific creativity*. New York: E. P. Dutton.

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. New York: Dover.

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development, 35*. 193–217.

Hanson, N. R. (1958). *Patterns of discovery*. Cambridge, UK: Cambridge University Press.

Hendrickson, J. B., & Sander, T. (1995). COGNOS: A Beilstein-type system for organizing organic reactions. *Journal of Chemical Information and Computer Sciences, 35*, p. 251.

Hesse, M. B. (1966). *Models and analogies in science*. Notre Dame, IN: University of Notre Dame Press.

Holmes, F. L. (1985). *Lavoisier and the chemistry of life: An exploration of scientific creativity*. Madison: University of Wisconsin Press.

Holmes, F. L. (1991). *Hans Krebs: The formation of a scientific life, 1900–1933* (Vol. 1). New York: Oxford University Press.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Hovland, C., & Hunt, E. (1960). The computer simulation of concept attainment. *Behavioral Science, 5*, 265–267.

Ippolito, M. F., & Tweney, R. D. (1995). The inception of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 433–462). Cambridge, MA: MIT Press.

Jacob, F. (1988). *The statue within*. New York: Basic Books.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology, 3*, 318–356.

Judson, H. F. (1979). *The eighth day of creation*. New York: Simon & Shuster.

Judson, H. F. (1996). *The eighth day of creation: Makers of the revolution in biology* (Expanded ed.). Plainview, NY: Cold Spring Harbour Laboratory Press.

Karmiloff-Smith, A. (1988). A child is a theoretician, not an inductivist. *Mind and Language, 3*, 183–195.

Karp, P. (1990). Hypothesis formation as design. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation* (pp. 275–317). San Mateo, CA: Morgan Kaufmann.

Kepler, J. (1937a). Harmonice mundi [The harmony of the world]. In W. von Dyck & M. Caspar (Eds.), *Gesammelte werke [Complete works]* (Vol. 1). Munich, Germany: Beck. (Original work published 1619)

Kepler, J. (1937b). Mysterium cosmographicum [The mystery of the universe]. In W. von Dyck & M. Caspar (Eds.), *Gesammelte werke [Complete works]* (Vol. 6). Munich, Germany: Beck. (Original work published 1596)

Kern, L. H., Mirels, H. L., & Hinshaw, V. G. (1983). Scientists' understanding of propositional logic: An experimental investigation. *Social Studies of Science, 13*, 131–146.

Klahr, D. (1999). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*. 1–55.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 24*. 111–146.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Kokar, M. M. (1986). Determining arguments of invariant functional descriptions. *Machine Learning, 1*, 403–422.

Kratochwill, T. R. (Ed.) (1978). *Single-subject research. Strategies for evaluating change*. New York: Academic Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*. 674–689.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific reasoning skills*. Orlando, FL: Academic Press.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60* (3, Serial No. 245).

Kulkarni, D., & Simon, H. A. (1988). The process of scientific discovery: The strategy of experimentation. *Cognitive Science, 12*. 139–176.

Kulkarni, D., & Simon, H. A. (1990). Experimentation in machine discovery. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery. Computational explorations of the creative processes*. Cambridge, MA: MIT Press.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.

Laudan, L. (1977). *Progress and its problems*. Berkeley: University of California Press.

Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs, 54*, (6, Whole No. 248).

Mahoney, M. J. (1979). Psychology of the scientist: An evaluative review. *Social Studies of Science, 9*, 349–375.

Mitroff, I. I. (1974). *The subjective side of science*. New York: Elsevier.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology, 29*, 85–95.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology, 30*. 395–406.

Nersessian, N. J. (1984) *Faraday to Einstein. Constructing meaning in scientific theories*. Dordrecht, the Netherlands: Martinus Nijhoff.

Nersessian, N. J. (1992) How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Cognitive models of science. Minnesota studies in the philosophy of science* (Vol. 15, pp. 3–44). Minneapolis: University of Minnesota Press.

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

Nordhausen, B., & Langley, P. (1993). An integrated framework for empirical discovery. *Machine Learning, 12,* 17–47.

Okada, T. (1994). *Collaborative scientific discovery processes*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Okada, T., & Simon, H. A. (1997) Collaborative discovery in a scientific domain. *Cognitive Science, 21,* 109–146.

O'Rorke, P., Morris, S., & Schulenburg, D. (1990). Theory formation by abduction: A case study based on the chemical revolution. In J. Shrager & P. Langley (Eds.). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.

Penner, D. E., & Klahr, D. (1996) When to trust the data: Further investigations of system error in a scientific reasoning task. *Memory & Cognition, 24,* 655–668.

Perkins, D. N. (1981) *The mind's best work*. Cambridge, MA: Harvard University Press.

Pickering, A. (Ed.). (1992). *Science as practice and culture*. Chicago: University of Chicago Press.

Pizzani, M. J. (1990). *Creating a memory of causal relationships*. Hillsdale, NJ: Erlbaum.

Poincaré, H. (1929) *The foundation of science*. New York: The Science Press.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Qin, Y., & Simon, H. A. (1990). Laboratory replication of scientific discovery processes. *Cognitive Science, 14,* 281–312.

Rajamoney, S. A. (1993) The design of discrimination experiments. *Machine Learning, 12,* 185–203.

Reimann, P. (1990) *Problem solving models of scientific discovery learning processes*. Frankfurt am Main, Germany: Peter Lang.

Rowe, A. (1953). *The making of a scientist*. New York: Dodd, Mead.

Samarapungavan, A. (1992) Children's judgments in theory-choice tasks: Scientific rationality in childhood. *Cognition, 45,* 1–32.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31–57.

Schauble, L., & Glaser, R. (1990) Scientific thinking in children and adults. In D. Kuhn (Ed.). *Contributions to human development: Vol. 21. Developmental perspectives on teaching and learning thinking skills* (pp. 9–26). Basel, NY: Karger.

Schunn, C. D. (1995). *A goals/effort tradeoff theory of experiment space search*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Schunn, C. D., & Anderson, J. R. (in press) The generality/specificity of expertise in scientific reasoning. *Cognitive Science*.

Schunn, C. D., & Klahr, D. (1995) A 4-space model of scientific discovery. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the seventeenth annual conference of the Cognitive Science Society* (pp. 106–111). Mahwah, NJ: Erlbaum.

Schunn, C. D., & Klahr, D. (1996) Integrated yet different: Logical, empirical, and implementational arguments for a 4-space model of inductive problem solving. In G. Cottrell (Ed.), *Proceedings of the eighteenth annual meeting of the Cognitive Science Society* (pp. 25–26). Mahwah, NJ: Erlbaum.

Scott, P. D., & Markovitch, S. (1993). Experience selection and problem choice in an exploratory learning system. *Machine Learning, 12,* 49–67.

Shadish, W. R., & Fuller, S. (Eds.). (1994). *Social psychology of science*. New York: Guilford Press.

Shen, W-M. (1993). Discovery as autonomous learning from the environment. *Machine Learning, 12,* 143–165.

Shrager, J. (1985). *Instructionless learning: Discovery of the mental model of a complex device*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Shrager, J., & Langley, P. (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufman.

Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology, 10,* 401–402.

Simon, H. A. (1966). Scientific discovery and the psychology of problem solving. In R. Colodny (Ed.), *Mind and cosmos* (pp. 22–40). Pittsburgh, PA: University of Pittsburgh Press.

Simon, H. A. (1973). Does scientific discovery have a logic? *Philosophy of Science, 40,* 471–480.

Simon, H. A. (1974). How big is a chunk? *Science, 183,* 482–488.

Simon, H. A. (1983). Fitness requirements for scientific theories. *British Journal for the Philosophy of Science, 34,* 355–365.

Simon, H. A. (1985). Quantification of theoretical terms and the falsifiability of theories. *British Journal for the Philosophy of Science, 36,* 291–298.

Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review, 70,* 534–546.

Simon, H. A., Langley, P., & Bradshaw, G. L. (1981) Scientific discovery as problem solving. *Synthese, 47,* 1–27.

Simon, H. A., & Lea, G. (1974) Problem solving and rule induction: A unified view. In L. Gregg (Ed.). *Knowledge and cognition* (pp. 105–128). Hillsdale, NJ: Erlbaum.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62,* 753–766.

Sokal, A. (1996) Transgressing the boundaries: Towards a transformative hermeneutics of quantum gravity. *Social Text, 46/47,* 217–252.

Terman, L. M. (1954) Scientists and nonscientists in a group of 800 men. *Psychological Monographs, 68,* (Whole No. 378).

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

Thagard, P. (1997). Medical analogies: Why and how. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual meeting of the Cognitive Science Society* (pp. 739–763). Mahwah, NJ: Erlbaum.

Thagard, P. (1998). Ulcers and bacteria I: Discovery and acceptance. *Studies in History and Philosophy of Biology and Biomedical Science, 29,* 107–136.

Tweney, R. D., Doherty, M. E., & Mynatt, C. R. (Eds.) (1981). *On scientific thinking*. New York: Columbia University Press.

Ueda, K. (1997). Actual use of analogies in remarkable scientific discoveries. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual meeting of the Cognitive Science Society* (pp. 763–768). Mahwah, NJ: Erlbaum.

Valdés-Perez, R. E. (1994a). Algebraic reasoning about reactions: Discovery of conserved properties in particle physics. *Machine Learning, 17,* 47–68.

Valdés-Perez, R. E. (1994b) Conjecturing hidden entities via simplicity and conservation laws: Machine discovery in chemistry. *Artificial Intelligence, 65,* 247–280.

Valdés-Perez, R. E. (1994c). Human/computer interactive elucidation of reaction mechanisms: Application to catalyzed hydrogenolysis of ethane. *Catalysis Letters, 28,* 79–87.

Valdés-Perez, R. E. (1995). Some recent human/computer discoveries in science and what accounts for them. *AI Magazine, 16,* 37–44.

van Joolingen, W. R., & de Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science. 25,* 307–346.

Vosniadou, S., & Ortony, A. (1989). *Similarity and analogical reasoning.* Cambridge, UK: Cambridge University Press

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12.* 129–140.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology. 20,* 273–281.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive reserach in the social sciences.* Chicago: Rand McNally.

Weisberg, R. W. (1993). *Creativity: Beyond the myth of genius.* New York: W. H. Freeman

Wellman, H. M., & Gelman, S. (1992). Cognitive development: Foundational theories in core domains. *Annual Review of Psychology, 43.* 337–375.

Wertheimer, M. (1945). *Productive thinking.* New York: Harper & Row.

Williams, L. P. (1965) *Michael Faraday.* New York: Basic Books.

Wolf, D. F. II, & Beskin, J. R. (1996). Task domains in N-space models: Giving explanation its due. In G. Cottrell (Ed.), *Proceedings of the eighteenth annual meeting of the Cognitive Science Society* (pp. 27–28) Mahwah, NJ: Erlbaum.

Received August 4, 1998
Revision received January 15, 1999
Accepted January 26, 1999 ■