Minnesota Studies in the
PHILOSOPHY OF SCIENCE

RONALD N. GIERE, GENERAL EDITOR
HERBERT FEIGL, FOUNDING EDITOR

VOLUME XV
Cognitive Models of Science

EDITED BY
RONALD N. GIERE

UNIVERSITY OF MINNESOTA PRESS
MINNEAPOLIS

Minnesota Studies in the
PHILOSOPHY OF SCIENCE

RONALD N. GIERE, GENERAL EDITOR
HERBERT FEIGL, FOUNDING EDITOR

VOLUME XV
Cognitive Models of Science

EDITED BY
RONALD N. GIERE

UNIVERSITY OF MINNESOTA PRESS
MINNEAPOLIS

Published by the University of Minnesota Press
2037 University Avenue Southeast, Minneapolis, MN 55414
Printed in the United States of America on acid-free paper

Library of Congress Cataloging-in-Publication Data

Cognitive models of science / edited by Ronald N. Giere.
p. cm. — (Minnesota studies in the philosophy of science ;
v. 15)
"Volume grew out of a workshop ... held in October 1989 under the
sponsorship of the Minnesota Center for Philosophy of Science"—P.
Includes bibliographical references and index.
ISBN 0-8166-1979-4 (hc)
1. Science—Philosophy. 2. Cognition. I. Giere, Ronald N.
II. Minnesota Center for Philosophy of Science. III. Series.
Q175.M64 vol. 15
[Q175.3]
501 s—dc20
[501] 91-34528
CIP

Contents

PART I
MODELS FROM COGNITIVE PSYCHOLOGY

PART II
MODELS FROM ARTIFICIAL INTELLIGENCE

PART III
MODELS FROM NEUROSCIENCE

PART IV
BETWEEN LOGIC AND SOCIOLOGY

PART V
CRITIQUE AND REPLIES

CRITIQUE

REPLIES TO GLYMOUR

Acknowledgments

Major support for the Center Workshop on Implications of the Cognitive Sciences for the Philosophy of Science was provided by the National Science Foundation through its program in History and Philosophy of Science. Additional funding was provided by the Scholarly Conference Fund of the College of Liberal Arts and by the Graduate School of the University of Minnesota. The editor owes particular thanks to the secretary of the Center for Philosophy of Science, Steve Lelchuk, who oversaw the whole process from sending out initial invitations to contributors, through local arrangements for the workshop, to compiling the final manuscript for the University of Minnesota Press. He also prepared the index.

Ronald N. Giere

Introduction: Cognitive Models of Science

This volume grew out of a workshop on implications of the cognitive sciences for the philosophy of science held in October 1989 under the sponsorship of the Minnesota Center for Philosophy of Science. The idea behind the workshop was that the cognitive sciences have reached a sufficient state of maturity that they can now provide a valuable resource for philosophers of science who are developing general theories of science as a human activity. The hope is that the cognitive sciences might come to play the sort of role that formal logic played for logical empiricism or that history of science played for the historical school within the philosophy of science. This development might permit the philosophy of science as a whole finally to move beyond the division between "logical" and "historical" approaches that has characterized the field since the 1960s.

There are, of course, philosophers of science for whom the very idea of a "cognitive approach" to the philosophy of science represents a regression to ways of thinking that were supposed to have been decisively rejected by the early decades of the twentieth century. From the time of the classical Greek philosophers through the nineteenth century, logic and psychology were closely related subjects. Nineteenth-century writers spoke easily of the principles of logic as "the laws of thought." Under the influence of Frege and Russell, that all changed. Logic became an autonomous, normative discipline; psychology an empirical science. The idea that how people *actually* think might have any relevance to the question of how they *should* think was labeled "psychologism," and catalogued as an official "fallacy." This point of view was incorporated into logical empiricism, the dominant Anglo-American philosophy of science until the 1960s. But it has persisted in various forms even among the strongest critics of logical empiricism, Imre Lakatos (1971) being a primary example.

Part of Kuhn's (1962) contribution to the philosophy of science was to challenge the separation of psychology from the philosophy of science.

His account of science invoked notions from gestalt psychology and the early "new look" psychologists associated with Jerome Bruner (Bruner, Goodnow, and Austin 1956). N. R. Hanson (1958), inspired mainly by Wittgenstein, had reached similar conclusions somewhat earlier. By the end of the 1960s, Quine (1969) had made psychology the basis for a "naturalized epistemology." Similar ideas were championed in related fields by such pioneers as Donald Campbell (1959) in social psychology and Herbert Simon (1977) in economics, psychology, and computer science. Although very influential, these works did not quite succeed in making psychology a fundamental resource for the philosophy of science.

One of the main reasons the constructive psychologism of Kuhn and Quine did not have more impact was simply that neither of the psychological theories to which they appealed — gestalt psychology and behaviorism, respectively — was adequate to the task. That has changed. Since the 1960s the cognitive sciences have emerged as an identifiable cluster of disciplines. The sources of this development are complex (Gardner 1985) but include the development of computers and transformational approaches in linguistics.

The emergence of cognitive science has by no means escaped the notice of philosophers of science. Within the philosophy of science one can detect an emerging specialty, the philosophy of cognitive science, which would be parallel to such specialties as the philosophy of physics or the philosophy of biology. But the reverse is also happening. That is, the cognitive sciences are beginning to have a considerable impact on the content and methods of philosophy, particularly the philosophy of language and the philosophy of mind (Dennett 1983; Fodor 1987; P. M. Churchland 1989; P. S. Churchland 1986), but also on epistemology (Goldman 1986). The cognitive sciences are also now beginning to have an impact on the philosophy of science. Inspired by work in the cognitive sciences, and sometimes in collaboration with cognitive scientists, a number of philosophers of science have begun to use the cognitive sciences as a *resource* for the philosophical study of science as a cognitive activity.

The unifying label "cognitive science" in fact covers a diversity of disciplines and activities. For the purposes of this volume, I distinguish three disciplinary clusters: (1) artificial intelligence (itself a branch of computer science), (2) cognitive psychology, and (3) cognitive neuroscience. These clusters tend to be thought of as providing three different levels of analysis, with the functional units becoming more abstract as one moves "up" from neuroscience to artificial intelligence. Each of these disciplinary clusters provides a group of models that might be used in approaching problems that are central to the philosophy of science. I begin with cognitive psychology because it seems to me that the models being developed within cognitive psychology are, at least for the moment, the most useful for a cognitive approach to the philosophy of science.

1. Models from Cognitive Psychology

Nancy Nersessian provides a prototype of someone drawing on research in the cognitive sciences to solve problems in the philosophy of science. The focus of her research is a problem originating in the historical critique of logical empiricism. Logical empiricism made science cumulative at the observational level while allowing the possibility of change at the theoretical level. But any noncumulative changes at the theoretical level could only be discontinuous. The historical critics argued that science has not been cumulative even at the empirical level. But some of these critics, such as Kuhn and Feyerabend, also ended up with a view of theoretical change as being discontinuous, though for different reasons. Thus was born the problem of "incommensurability." Nersessian's project is to dissolve the problem of incommensurability by showing how the theoretical development of science can be continuous without science as a whole being cumulative.

The historical focus of her study is the development of electrodynamics from Faraday to Einstein. She argues that neither philosophers nor historians have yet done justice to this development. This includes those like Hanson and Kuhn who explicitly appealed to gestalt psychology. Nersessian argues that more recent work in cognitive psychology provides tools that are more adequate to the task, particularly for understanding the use of analogy, mechanical models, thought experiments, and limiting case analysis. Here I will note only two features of her study that have general implications for a cognitive approach to the philosophy of science.

Most historically minded critics of logical empiricism took over the assumption that scientific theories are primarily *linguistic* entities. The main exception is Kuhn, who gave priority to concrete exemplars over linguistically formulated generalizations. Nersessian adopts a theory of "mental models" as elaborated, for example, by Johnson-Laird (1983). On this approach, language, in the form of propositions, may be used not to describe the world directly but to construct a "mental model," which is a "structural analog" of a real-world or imagined situation. Once constructed, the mental model may yield "images," which are mental models viewed from a particular perspective. This interplay of propositions, models, and images provides a richer account of the representational resources of scientists than that employed by either logical empiricists or most of their critics. It may be thought of as an extension of the model-theoretic approach to the nature of scientific theories as elaborated, for example, by Suppe (1989), van Fraassen (1980, 1989), and myself (Giere 1988). In any case, the cognitive theory of mental models provides the main resource for Nersessian's account of the dynamics of conceptual change in science. Some such account of representation seems sure to become standard within a cognitive approach to the philosophy of science.

Another assumption shared by logical empiricists and most of their historically based critics is that the basic entities in an account of science are abstractions like "theories," "methods," or "research traditions" (which for both Lakatos [1970] and Laudan [1977] are explicitly characterized in terms of laws, theories, and methodological rules). Nersessian, by contrast, insists on including the *individual scientist* as an essential part of her account. Her question is not simply how the theory of electrodynamics developed from the time of Faraday to that of Einstein, but how Faraday, Maxwell, and Einstein, as individual scientists, developed electrodynamics. Theories do not simply develop; they are *developed* through the cognitive activities of particular scientists. It is the focus on scientists, as real people, that makes possible the application of notions from cognitive psychology to questions in the philosophy of science.

Nersessian's insistence on the role of human agency in science is strongly reinforced by David Gooding's analysis of the path from actual experimentation, to the creation of demonstration experiments, to the development of theory. Insisting that all accounts of scientific activity, even those recorded in laboratory notebooks, involve reconstruction, Gooding distinguishes six types, or levels, of reconstruction. Standard philosophical reconstructions, which Gooding labels "normative," are last in the sequence. The first are "cognitive" reconstructions, with "rhetorical" and "didactic" reconstructions being among the intermediate types. Gooding is particularly insistent on the importance of "procedural knowledge," such as laboratory skills, in the cognitive development of science.

Gooding focuses on the sequence of experiments that led to Faraday's invention and development of the world's first electric motor. He develops a notation for representing the combined development of experimental procedures and theory that led to Faraday's discovery. This notation is an elaboration of one developed earlier by Ryan Tweney. Utilizing this notation, Gooding exhibits a multiplicity of possible experimental pathways leading to the electric motor, noting that Faraday himself presented different reconstructions of the pathway in different contexts.

Gooding concludes his essay by arguing that the power of thought experiments derives in part from the fact that they embody tacit knowledge of experimental procedures. This argument complements Nersessian's analysis of how an "experiment" carried out in thought can have such an apparently powerful empirical force. She argues that conducting a thought experiment is to be understood as using a mental model of the experimental situation to run a simulation of a real experiment. The empirical content is built into the mental model, which includes procedural knowledge.

Ryan Tweney was among the first of recent theorists to advocate a cognitive approach to the study of science. And he has pursued this approach in both experimental and historical contexts. Here he explores some implications of the recent vogue for parallel processing for the study of science as a cognitive process. Tweney acknowledges the importance of having models that could plausibly be implemented in a human brain, but is less impressed by neuroscientific plausibility than by the promise of realistic *psychological* models of perception, imagery, and memory — all of which he regards as central to the process of science.

Rather than joining the debate between advocates of serial models and of parallel models, Tweney takes a third route that focuses attention on cognitive activities in natural contexts — leaving the question of which sort of model best fits such contexts to be decided empirically on a case-by-case basis. But it is clear that Tweney is impressed with the promise of parallel models, even though, as he points out, they have yet to be applied successfully to higher-level cognitive processes. Here he considers two applications: (1) an account of the memory aids used by Michael Faraday to index his notebooks, and (2) Paul Thagard's analysis of scientific revolutions using a parallel network implementation (ECHO) of a theory of explanatory coherence. He finds the concept of parallel processing useful in the first case but superfluous in the second.

For nearly two decades, sociologists of science have been gathering under a banner labeled "The Social Construction of Scientific Knowledge." The aforementioned essays suggest that we can equally well speak of "The Cognitive Construction of Scientific Knowledge." There are, however, two important differences in the ways these programs are conceived. First, unlike social constructionists, cognitive constructionists make no claims of exclusivity. We do not insist that cognitive construction is all there is. Second, social constructionists typically deny, or claim indifference to, any genuine representational connection between the claims of scientists and an independently existing world. By contrast, connections with the world are built into the cognitive construction of scientific knowledge. This is particularly clear in Gooding's paper, which emphasizes the role of procedural knowledge in science.

The historical movement in the philosophy of science made conceptual change a focus of research in the history and philosophy of science. It has subsequently become a research area within cognitive psychology as well, although Piaget had already made it a focus of psychological research in Europe several decades earlier. Indeed, one major strand of current research may be seen as an extension of Piaget's program, which used conceptual development in children as a model for conceptual development in science (Gruber and Vonèche 1977). This line of research is represented here by Susan Carey.

Carey works within the "nativist" tradition, which holds that at least some concepts are innate, presumably hard-wired as the result of our evolutionary heritage. The question is what happens to the conceptual structure possessed by a normal human in the natural course of maturation, apart from explicit schooling. An extreme view is that conceptual development consists only of "enrichment," that is, coming to believe new propositions expressed solely in terms of the original set of innate concepts. Another possible view is that humans also form new concepts by differentiation and combination. Objects become differentiated to include animals, then dogs. Colors become differentiated into red, green, blue, and so forth. Combination produces the concept of a red dog (an Irish setter). Carey argues that normal development also produces conceptual systems that are, in Kuhn's (1983) terms, "locally incommensurable" with earlier systems.

As an example of local incommensurability, Carey cites the differentiation of the concepts weight and density within a system originally possessing only a single undifferentiated concept. The earlier (child's) conceptual system and the later (adult) system are incommensurable because there remains no counterpart of the undifferentiated weight/density concept in the adult conceptual system. Carey presents evidence showing that children do indeed begin with an undifferentiated weight/density concept and gradually develop a system with the differentiated concept. She also describes earlier historical research (Wiser and Carey 1983) showing that seventeenth-century scientists at the Academy of Florence possessed a similarly undifferentiated concept of temperature/heat. She thus links research in the history of science with research in cognitive development.

Carey takes pains to argue that local incommensurability between children's and adults' concepts does not mean that adults and children cannot understand one another, that children do not learn language by interacting with adults, or that psychologists cannot explain the child's conceptual system to others. So the concept of incommensurability employed here has none of the disastrous implications often associated with philosophical uses of this notion. It seems, therefore, that philosophers and psychologists may at last have succeeded in taming the concept of incommensurability, turning it into something that can do useful work.

The shift from novice to expert provides another model recently exploited by cognitive psychologists to study conceptual change in science. Michelene Chi has been a leader in this research. Here, however, she treats conceptual change in more general terms. She argues that even Carey's notion of change between incommensurable conceptual systems is not strong enough to capture the radical nature of the seventeenth-century revolution in physics. That revolution, she argues, involved a more radical conceptual shift because there was a shift in *ontological* categories. In particular, the conceptual system prior to the scientific revolution mainly employed concepts within the ontological category of "material substance" whereas the new physical concepts were mainly relational, covering what she calls "constraint-based events." According to Chi's analysis, therefore, the difficulty people have moving beyond an undifferentiated weight/density concept is due to difficulty in conceiving of weight as relational rather than substantial. Density, being an intrinsic property of objects (mass per unit volume), is developmentally the more primitive concept.

Chi criticizes other studies of conceptual change, such as Carey's, for giving only a kinematics of change and not providing any account of the dynamical mechanisms of change. She locates the engine of change at the early stages of a scientific revolution when anomalies first begin to be recognized as such. This leads not so much to a change in concepts as to the construction of a rival conceptual system that for a while coexists with the original system. Completing a scientific revolution, or a process of conceptual change in an individual, is a matter of increasing reliance on the new system and disuse of the old.

The final two essays in Part I employ a cognitive approach to problems that were prominent among logical empiricists. Questions about the nature of observation and, more technically, measurement were high on the agenda of logical empiricism. That was in large part because of the foundational role of observation in empiricist epistemology. But even if one abandons foundationalist epistemology, there are still interesting questions to be asked about observation and measurement. Richard Grandy explores several such issues from the general perspective of cognitive agents as information processors.

One question is whether there is any place for the notion of an "observation sentence" in a cognitive philosophy of science. Developing suggestions first explored by Quine, Grandy argues that there is, although whether a sentence can serve as an observation sentence cannot be solely a matter of either syntax or semantics. It depends, among other things, on the typical reactions of a relevant community in a variety of contexts. That is something that can itself only be determined empirically.

Another topic Grandy explores is the relative information provided by the use of various types of measurement scales. Grandy demonstrates that the potential information carried by a measurement typically increases as one moves from nominal, to ordinal, to ratio scales. More surprising, he is able to show that what he would regard as observation sentences typically convey more information than ordinal-scale measurements, though not as much as ratio-scale measurements. This is but one step in a projected general program to analyze the contributions of new theories, instruments, and methods of data analysis in terms of their efficiency as information generators or processors. Such an analysis would provide a "cognitive" measure of scientific progress.

In the final essay of Part I, Wade Savage explores the possibility of using recent cognitive theories of perception to develop a naturalized foundationalist empiricism. He begins by distinguishing strong from weak foundationalism. Strong foundationalism is the view that some data provided by sensation or perception are both independent (not based on further data) and infallible (incapable of error). Weak foundationalism holds that only some data of sensation or perception are more independent and more reliable than other data. Savage's view is that weak foundationalism provides a framework for a naturalistic theory of *conscious* human knowledge and strong foundationalism provides a framework for a naturalistic theory of *unconscious* human knowledge. The mistake of the classical foundationalists, he claims, is to have assumed that strong foundationalism could be a theory of conscious knowledge.

Recent work in cognitive science, Savage claims, is particularly relevant to the theory of unconscious human knowledge. Drawing inspiration from works by people such as Marr (1982) and MacArthur (1982), Savage argues for a *presensational foundationalism* in which the basic data are unconscious "ur-sensations." His main arguments are directed toward showing that ur-sensations are both independent and infallible in an appropriate sense. Savage concludes by sketching a theory reminiscent of Minsky (1987) that explains the relationship between unconscious ur-sensations and the consciousness of higher cognitive agents.

2. Models from Artificial Intelligence

Among the many crosscurrents within the fields of computer science and artificial intelligence (AI) is a tension between those who wish to use the computer as a means to study the functioning of *human* intelligence and those who see the computer primarily as a tool for performing a variety of tasks quite apart from how humans might in fact perform those same tasks. This tension is evident in the original work on "discovery programs" inspired by Herbert Simon (1977) and implemented by Pat Langley, Simon, and others (Langley et al. 1987). This work has demonstrated the possibility of developing programs that can uncover significant regularities in various types of data using quite general heuristics. Among the prototypes of such programs are BACON, GLAUBER, and KEKADA (Kulkarni and Simon 1988). BACON, for example, easily generates Kepler's laws beginning only with simple data on planetary orbits.

One way of viewing such programs is as providing "normative models" in the straightforwardly instrumental sense that these models provide good means for accomplishing well-defined goals. This use of AI is exhibited in this volume by Gary Bradshaw and Lindley Darden. For a variety of other examples, see Shrager and Langley (1990).

Bradshaw, who began his career working with Simon and Langley, applies Simon's general approach to problem solving to invention in technology. He focuses on the much-discussed historical question of why the Wright brothers were more successful at solving the problem of manned flight than their many competitors. Dismissing a variety of previous historical explanations, Bradshaw locates the crucial difference in the differing heuristics of the Wright brothers and their competitors. The Wright brothers, he argues, isolated a small number of functional problems that they proceeded to solve one at a time. They were thus exploring a relatively small "function space" while their competitors were exploring a much larger "design space."

Darden proposes applying AI techniques developed originally for diagnosing breakdowns in technological systems to the problem of "localizing" and "fixing" mistaken assumptions in a theory that is faced with contrary data. Here she outlines the program and sketches an application to the resolution of an empirical anomaly in the history of Mendelian genetics. Darden is quite clear on the goal of her work: "The goal is not the simulation of human scientists, but the making of discoveries about the natural world, using methods that extend human cognitive capacities."

Programs like those of Darden and others are potentially of great scientific utility. That potential is already clear enough to inspire many people to develop them further. How useful such programs will actually prove to be is not something that can be decided a priori. We will have to wait and see. The implications of these sorts of programs for a cognitive philosophy of science are mainly *indirect*. The fact that they perform as well as they do can tell us something about the structure of the domains in which they are applied and about possible strategies for theorizing in those domains.

Others see AI as providing a basis for much more far-reaching philosophical conclusions. The essay by Greg Nowak and Paul Thagard and that by Eric Freedman apply Thagard's (1989) theory of explanatory coherence (TEC) to the Copernican revolution and to a controversy in psychology, respectively. Nowak and Thagard hold both that the objective superiority of the Copernican theory over the Ptolemaic theory is shown by its greater overall explanatory coherence, and that the triumph of the Copernican theory was due, at least in part, to the intuitive perception of its greater explanatory coherence by participants at the time.

Thagard, who advocates a "computational philosophy of science" (Thagard 1988), implements his theory of explanatory coherence in a connectionist program, ECHO, which utilizes localized representations of propositions. It has been questioned (for example, by Tweney and Glymour in this volume) whether ECHO is doing anything more than functioning as a fancy calculator, with all the real work being done by TEC. If so, it is a very fancy calculator, performing nonlinear optimization with several constraints, and containing various adjustable parameters that can materially affect the outcome of the calculation.

Freedman's study provides a further illustration of the operations of TEC and ECHO. He analyzes the famous controversy between Tolman and Hull over the significance of Tolman's latent-learning experiments. Applying TEC and ECHO, Freedman finds that Tolman's cognitive theory is favored over Hull's behaviorist theory. Yet Hull's approach prevailed for many years. By varying available parameters in ECHO, Freedman shows several ways in which ECHO can be made to deliver a verdict in favor of Hull. For example, significantly decreasing the importance of the latent-learning data can tip the balance in favor of Hull's theory. To Freedman, this provides at least a suggestion for how the actual historical situation might be explained. So ECHO does some work. But this study also makes it obvious that to decide among the possibilities suggested by varying different parameters in ECHO, one would have to do traditional historical research. ECHO cannot decide the issue.

Thagard's work shows that a deep division between cognitive psychologists and AI researchers carries over into the ranks of those advocating a cognitive approach to the philosophy of science. Most cognitive psychologists would insist that cognitive psychology is fundamentally the study of the cognitive capacities of human agents. For a cognitive philosophy of science this means studying how scientists actually represent the world and judge which representations are best. In Thagard's computational philosophy of science, all representations are conceptual and propositional structures, and judgments of which representations are best are reduced to computations based primarily on relationships among propositions. There is as yet little evidence that the required propositional structures and computations are psychologically real.

3. Models from Neuroscience

In Part III the relevance of models from the neurosciences to the philosophy of science is argued by the primary advocate of the philosophical relevance of such models, Paul Churchland. It is Churchland's (1989) contention that we already know enough about the gross functioning of the brain to make significant claims about the nature of scientific knowledge and scientific reasoning. In his essay he argues that a "neurocomputational" perspective vindicates (more precisely, "reduces") a number of claims long advocated by Paul Feyerabend. For example: "Competing theories can be, and occasionally are, *incommensurable*," and "the long-term best interests of intellectual progress require that we proliferate not only theories, but research *methodologies* as well."

It is interesting to note that Churchland's analysis supports Chi's conclusion that the dynamical process in theory change is not so much transformation as it is the construction of a new representation that then replaces the old representation. That, Churchland claims, is just how neural networks adapt to new situations.

Whatever one's opinion of Churchland's particular claims, I think we must all agree that the neurosciences provide a powerful and indisputable constraint on any cognitive philosophy of science. Whatever cognitive model of scientific theorizing and reasoning one proposes, it must be a model that can be implemented by humans using human brains.

4. Between Logic and Sociology

Except during momentary lapses of enthusiasm, no one thinks that a cognitive theory of science could be a *complete* theory of science. The cognitive activities of scientists are embedded in a social fabric whose contribution to the course of scientific development may be as great as that of the cognitive interactions between scientists and the natural world. Thus cognitive models of science need to be supplemented with social models. The only requirement is that the two families of models fit together in a coherent fashion.

There are those among contemporary sociologists of science who are not so accommodating. Latour and Woolgar, for example, are now famous for suggesting a ten-year moratorium on cognitive studies of science, by which time they expect to have constructed a complete theory of science that requires no appeal to cognitive categories. Such voices are not directly represented in this volume, but they do have supporters nonetheless.

Arthur C. Houts and C. Keith Haddock agree with the sociological critics of cognitivism in rejecting the use of cognitive categories like representation or judgment in a theory of science. But they insist there is need for a genuine psychology of science. From a cognitivist point of view, these are incompatible positions. For Houts and Haddock these positions are not incompatible because their psychology of science is based on the behaviorist principles of B. F. Skinner. In Skinnerian theory, the determinants of behavior are to be found in the environment, both natural and social, which provides the contingencies of reinforcement. There is no need for any appeal to "mental" categories such as representation or judgment. Several commentators, for example, Slezak (1989) and myself (Giere 1988), have criticized behaviorist tendencies in the writings of sociologists of science. For Houts and Haddock, these tendencies are not a basis for criticism but a positive virtue. They make possible a unified approach to both the psychology and the sociology of science.

Within cognitive psychology there is a tradition, already several decades old, in which scientific reasoning tasks are simulated in a laboratory setting. Michael E. Gorman reviews this tradition and compares it with the more recent tradition of computational simulation pioneered by Simon and represented in this volume by Thagard. He relies heavily on the distinction between externally valid and ecologically valid claims. A claim is *externally valid* if it generalizes well to other well-controlled, idealized conditions. A claim is *ecologically valid* if it generalizes well to natural settings, for example, to the reasoning of scientists in their laboratories. Gorman argues that while both laboratory and computer simulations may be externally valid, laboratory studies are more ecologically valid. Granting this conclusion, however, does little to remove doubts about the ecological validity of laboratory studies themselves.

Gorman proposes bridging the gap between cognitive and social studies of science by designing experimental simulations that include social interactions among the participants. Here experimental paradigms from social psychology are merged with those that have been used in the experimental study of scientific reasoning. Gorman's hope is that one might eventually develop experimental tests of claims made by sociologists as well as by more theoretical "social epistemologists" such as Steve Fuller.

Fuller himself questions a central presupposition of most cognitive approaches to the philosophy of science, namely, that the individual scientist is the right unit of analysis for any theory of science. Not that he advocates returning to abstract entities like theories. Rather he thinks that the appropriate unit will turn out to be something more like a biological species than an individual scientist. Bruno Latour's (1987) "actor network" may be a good example of the kind of thing Fuller expects might emerge as the proper unit of study. Fuller's argument is both historical and critical. He sketches an account of how the individual scientist came to be regarded as the basic entity for epistemology generally, and why this assumption has led to difficulties in several areas, particularly in analytic epistemology, but also in Churchland's neurocomputational approach.

5. Critique and Replies

Clark Glymour was among the first philosophers of science to grasp the possibility of deploying methods and results from the cognitive sciences, particularly artificial intelligence, to the philosophy of science itself. (Herbert Simon, who I definitely would wish to claim as a philosopher of science, must surely have been the first.) But as his contribution to this volume makes crystal clear, Glymour is quite disappointed with what some other philosophers of science have been doing with this strategy. In his essay in Part V he expresses his disappointment with work by three of the participants in the Minnesota workshop, Churchland, Thagard, and myself. By mutual agreement, Glymour's comments appear as he wrote them. They are followed by replies from each of the three named subjects of his remarks. Since my own reply is included, I will say no more here in my role as editor.

References

Bruner, J. S., J. Goodnow, and G. Austin. 1956. *A Study of Thinking*. New York: Wiley.

Campbell, D. T. 1959. Methodological Suggestions from a Comparative Psychology of Knowledge Processes. *Inquiry* 2:152–82.

Churchland, P. M. 1989. *A Neurocomputational Perspective*. Cambridge, Mass.: MIT Press.

Churchland, P. S. 1986. *Neurophilosophy*. Cambridge, Mass.: MIT Press.

Dennett, D. 1983. *Brainstorms*. Cambridge, Mass.: MIT Press.

Fodor, J. 1987. *Psychosemantics.*Cambridge, Mass.: MIT Press.

Gardner, H. 1985. *The Mind's New Science.* New York: Basic Books.

Giere, R.N. 1988. *Explaining Science: A Cognitive Approach.* Chicago: University of Chicago Press.

Goldman, A. 1986. *Epistemology and Cognition.* Cambridge, Mass.: Harvard University Press.

Gruber, H., and J.J. Vonèche, eds. 1977. *The Essential Piaget.* New York: Basic Books.

Hanson, N.R. 1958. *Patterns of Discovery.* Cambridge: Cambridge University Press.

Johnson-Laird, P.N. 1983. *Mental Models.* Cambridge, Mass.: Harvard University Press.

Kuhn, T.S. 1962. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

———. 1983. Commensurability, Comparability, Communicability. In *PSA 1982*, ed. P. Asquith and T. Nickles, 669–87. East Lansing, Mich.: Philosophy of Science Association.

Kulkarni, D., and H. Simon. 1988. The Processes of Scientific Discovery: The Strategy of Experimentation. *Cognitive Science* 12:139–75.

Lakatos, I. 1970. Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge*, ed. I. Lakatos and A. Musgrave, 91–195. Cambridge: Cambridge University Press.

———. 1971. *History of Science and Its Rational Reconstructions.* Boston Studies in the Philosophy of Science, ed. R.C. Buck and R.S. Cohen, vol. 8. Dordrecht: Reidel.

Langley, P., H.A. Simon, G.L. Bradshaw, and J.M. Zytkow. 1987. *Scientific Discovery.* Cambridge, Mass.: MIT Press.

Latour, B. 1987. *Science in Action.* Cambridge, Mass.: Harvard University Press.

Laudan, L. 1977. *Progress and Its Problems.* Berkeley: University of California Press.

MacArthur, D.J. 1982. Computer Vision and Perceptual Psychology. *Psychological Bulletin* 92:283–309.

Marr, D. 1982. *Vision.* San Francisco: Freeman.

Minsky, M. 1987. *The Society of Mind.* New York: Simon & Schuster.

Quine, W.V.O. 1969. Epistemology Naturalized. In *Ontological Relativity and Other Essays*, 69–90. New York: Columbia University Press.

Shrager, J., and P. Langley, eds. 1990. *Computational Models of Scientific Discovery and Theory Formation.* Hillsdale, N.J.: Erlbaum.

Simon, H.A. 1977. *Models of Discovery.* Dordrecht: Reidel.

Slezak, P. 1989. Scientific Discovery by Computer as Empirical Refutation of the Strong Programme. *Social Studies of Science* 19:563–600.

Suppe, F. 1989. *The Semantic Conception of Theories and Scientific Realism.* Urbana: University of Illinois Press.

Thagard, P. 1988. *Computational Philosophy of Science.* Cambridge, Mass.: MIT Press.

———. 1989. Explanatory Coherence. *Behavioral and Brain Sciences* 12:435–67.

van Fraassen, B.C. 1980. *The Scientific Image.* Oxford: Oxford University Press.

———. 1989. *Laws and Symmetry.* Oxford: Oxford University Press.

Wiser, M., and S. Carey. 1983. When Heat and Temperature Were One. In *Mental Models*, ed. D. Gentner and A. Stevens. Hillsdale, N.J.: Erlbaum.

PART I

MODELS FROM COGNITIVE PSYCHOLOGY

Nancy J. Nersessian

How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science

The Scene

August 19, 1861, a cottage in Galloway, Scotland.

The young Clerk Maxwell is sitting in a garden deep in thought. On the table before him there is a sheet of paper on which he sketches various pictures of lines and circles and writes equations.

The Question

What is he thinking? Is he trying to cook up a model to go with the equations he has derived already by induction from the experimental data and electrical considerations alone? Is he concerned that his mathematical results are not quite right and so is thinking how to fudge his analysis to make it look right in terms of the model? Is he searching for a way to make the notion of continuous transmission of actions in an electromagnetic "field" meaningful? And if so, what resources is he drawing upon? What is he doing?

The Problem

Do we have the means to understand what Maxwell is doing? What scientists like him are doing when they are creating new conceptions? Based on the record they leave behind, can we hope to fathom the creative processes through which scientists articulate something quite new? Or are these processes so mysterious that we are wasting our time by trying to understand them? And if we could, what possible profit could such understanding yield for the philosopher of science? the historian of science? others?

The Path to Solution

I hope to persuade the reader that we can formulate a more rigorous analysis of the creative processes of scientific discovery and give more satisfactory answers to long-standing, unresolved puzzles about the nature of conceptual change in science than we have now by combining two things that are usually kept apart. One is fine-structure examinations of the theoretical and experimental practices of scientists who have created major changes in scientific theory. The other is what we have been learning about the cognitive abilities and limitations of human beings generally. Creative processes are extended and dynamical, and as such we can never hope to capture them fully. But by expanding the scope of the data and techniques allowed into the analysis we can understand more than traditional approaches have permitted so far.

Recent developments in psychology have opened the possibility of understanding what philosophers and historians have been calling "conceptual change" in a different and deeper way. Through a combination of new experimental techniques and computer modeling, new theories about human cognitive functioning have emerged in the areas of representation, problem solving, and judgment. An interdisciplinary field of cognitive science has recently formed — a loose confederation of cognitive psychology, artificial intelligence, cognitive neurology, linguistics, and philosophy. It offers analyses and techniques that, if used with proper respect for their scope and their limitations, can help us develop and test models of how conceptual change takes place in science.

In this essay I set myself the following aims: (1) to propose a fresh method of analysis; (2) to recast the requirements of a theory of conceptual change in science; (3) to draw on new material from a heuristically fertile case study of major conceptual change in science to analyze some processes of conceptual change — analogical and imagistic reasoning and thought experiments and limiting case analyses; (4) to examine these in light of some work in cognitive science; and (5) to argue, more generally, for what philosophers and historians of science and cognitive scientists might gain from further application of the proposed method of analysis.

1. What Is "Cognitive-Historical" Analysis?

"Cognitive-historical" analysis in the sense employed here is not quite the same as what historians of science do in their fine-structure historical examinations of the representational and problem-solving practices scientists have employed to create new scientific representations of phenomena. Rather, it attempts to enrich these further by means of investigations of ordinary human representational and problem-solving practices carried out by the sciences of cognition. *The underlying presupposition is that the problem-solving strategies scientists have invented and the representational practices they have developed over the course of the history of science are very sophisticated and refined outgrowths of ordinary reasoning and representational processes.* Thus, the method combines case studies of actual scientific practices with the analytical tools and theories of the cognitive sciences to create a new, comprehensive theory of how conceptual structures are constructed and changed in science. The historical dimension of the method has its origins in the belief that to understand scientific change the philosophy of science must come to grips with the historical processes of knowledge development and change. This is the main lesson we should have learned from the "historicist" critics of positivism. Equally as important as problems concerning the rationality of acceptance — which occupy most philosophers concerned with scientific change — are problems about the construction and the communication of new representational structures. The challenging methodological problem is to find a way to use the history of scientific knowledge practices as the *basis* from which to develop a theory of scientific change.

The cognitive dimension of the method reflects the view that our understanding of scientific knowledge practices needs to be psychologically realistic. Putting it baldly, creative scientists are not only exceptionally gifted human beings — they are also human beings with a biological and social makeup like all of us. In a fundamental sense, science is one product of the interaction of the human mind with the world and with other humans. We need to find out how human cognitive abilities and limitations constrain scientific theorizing *and this cannot be determined a priori.* This point is not completely foreign to philosophers. It fits into a tradition of psychological epistemology beginning with Locke and Hume and making its most recent appearance with the call of Quine for a "naturalized epistemology." Why did earlier "psychologizing" endeavors fail? The main reason was their reliance on inadequate empiricist/behaviorist psychological theories. The development of cognitive psychology has paved the way for a much more fruitful synthesis of psychology and epistemology. Suggestions for how to frame such a synthesis are to be found, for example, in the work of Alvin Goldman (1986). Insights from cognitive psychology are beginning to make their way into investigations of scientific reasoning (see, e.g., Giere 1988; Gooding 1990; Gorman and Carlson 1989; Langley et al. 1987; Thagard 1988; and Tweney 1985). What is needed now is