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 to integrate these with the historical findings about the representational and problem-solving practices that have actually brought about major scientific changes.[1]

Philosophers this century have mostly been working under the assumption that analysis of science takes place within two contexts: justification and discovery. The former is traditionally within the province of philosophers; the latter, of historians and psychologists. Cognitive-historical analysis takes place in a new context — that of development — which is the province of all three. The context of development[2] is the domain for inquiry into the processes through which a vague speculation gets *articulated* into a new scientific theory, gets *communicated* to other scientists, and comes to *replace* existing representations of a domain. These processes take place over long periods of time, are dynamic in nature, and are embedded in social contexts.

This new context of development, in actuality, was opened up by the work of Hanson, Kuhn, and Feyerabend nearly thirty years ago, but they lacked the analytical tools to pursue it in depth. True, they attempted to integrate insights from psychology into their analyses. However, cognitive psychology, in the form of the "new look" psychology of Bruner and others, was in its infancy, whereas Gestalt psychology offered no understanding of the processes underlying the "gestalt switch." Since their vision predated the kind of psychological theory that would have helped them better express it, it would be unfair to fault them for not completing what they had begun. It is important, however, to see the continuing repercussions of the inadequate insights they did use.

First, drawing from these psychological theories led to an unfortunate identification of knowledge change with hypothesized aspects of visual perception. Second, and more important, the metaphor of the "gestalt switch" led them astray in a way that has had deeper and more lasting consequences. The metaphor does not support the extended nature of the conceptual changes that have actually taken place in science.[3] Thus, while calling for a historicized epistemology, Kuhn's and Feyerabend's own historical analyses offered in support of the "incommensurability" hypothesis were decidedly unhistorical in the following sense. By emphasizing the endpoints of a conceptual change (e.g., Newtonian mechanics and relativistic mechanics), the change of gestalt was made to appear artificially abrupt and discontinuous. Historically, however, we did not get to relativity without at least passing through electromagnetism and the theory of electrons; and this developmental process is central to understanding such questions as the nature of the relationship — or reason for lack of relationship — between, e.g., the different concepts called by the name "mass" in each theory. My earlier study of the construction of the concept of electromagnetic field, *Faraday to Einstein: Constructing Meaning in Scientific Theories* (1984), was an attempt to show how incorporating the dimension of development into the analysis gives a quite different picture of the nature of meaning change in science.

Significantly, although Kuhn does talk about discovery as an "extended process" (Kuhn 1965, pp. 45ff.) and, in his role as historian of science, has provided detailed examinations of such processes, in his role as philosopher of science he identifies conceptual change with "the last act," when "the pieces fall together" (Kuhn 1987).[4] Thus portrayed, conceptual change appears to be something that happens to scientists rather than the outcome of an extended period of construction by scientists. A "change of gestalt" may be an apt way of characterizing this last point in the process, but focusing exclusively on that point has — contrary to Kuhn's aim — provided a misleading portrayal of conceptual change; has reinforced the widespread view that the processes of change are mysterious and unanalyzable; and has blocked the very possibility of investigating how precisely the new gestalt is related to its predecessors. *In short, the metaphor has blocked development of the historicized epistemology being advocated.*

The ultimate goal of the cognitive-historical method is to be able to reconstruct scientific thinking by means of cognitive theories. When, and if, we reach that point, we may decide to call the method "cognitive analysis of science." At present, however, cognitive theories are largely uninformed by scientific representational and problem-solving practices, making the fit between cognitive theories and scientific practices something that still needs to be determined.[5] Cognitive-historical analysis is reflexive. It uses cognitive theories to the extent that they help interpret the historical cases — at the same time it tests to what extent current theories of cognitive processes can be applied to scientific thinking and indicates along what lines these theories need extension, refinement, and revision. In other words, the method is a type of bootstrapping procedure commonly used in science.

2. What Would a Cognitive Theory of Conceptual Change in Science Look Like?

2.1. Background

Much philosophical energy this century has been spent on the problem of conceptual change in science. The major changes in physical theory early in the century thrust the problem of how to understand the seemingly radical reconceptualizations they offered into the spotlight for scientists, historians, and philosophers alike. As we know, the comforting picture of conceptual change as continuous and cumulative offered by logical positivism itself suffered a revolutionary upheaval in the mid-1960s. The critics of positivism argued that major changes in science are best characterized as "revolutions": they involve overthrow and replacement of the reigning conceptual system with one that is "incommensurable" with it. The infamous "problem of incommensurability of meaning" dominated the literature for over a decade. Philosophers have by and large abandoned this topic. Those who work on scientific change tend now to focus on the problem of rational choice between competing theories. This shift in focus did not, however, come from a sense of having a satisfactory solution to the infamous problem, but more from a sense of frustration that the discussion and arguments had become increasingly sterile.

The crux of the original problem, however, is still with us. How, if in any manner at all, are successive scientific conceptualizations of a domain related to one another? The instinctive response of critics of incommensurability has always been that even though they are not simply extensions, the new conceptual structures must somehow grow out of the old. The view of knowledge change as a series of unconnected gestalt switches has a high intuitive implausibility. *In recasting the problem of conceptual change, the cognitive-historical method furnishes the means through which to turn these intuitions into solid analyses.*

In cognitive-historical analysis the problem of conceptual change appears as follows. It is the problem of understanding how scientists combine their human cognitive abilities with the conceptual resources available to them as members of scientific communities and wider social contexts to create and communicate new scientific representations of a domain. For example, the problem posed in the opening scene becomes that of understanding how Maxwell joins his human cognitive endowment with the conceptual resources of a Cambridge mathematical physicist living in Victorian England to construct and communicate a field representation of electromagnetic forces. Admittedly, this is a quite complex problem and we are only beginning to have the means to attack it. Nevertheless, I shall show in some detail that the cognitive-historical approach offers more possibility of achieving a solution than any we have yet considered.

Where the traditional philosophical approach views conceptual change as static and ahistorical, cognitive-historical analysis is able to handle the dynamic and historical process that it is. Customarily, conceptual change is taken to consist of the replacement of one linguistic system by another, and understanding conceptual change requires analyzing the logical relationships between propositions in the two systems.

In a cognitive theory conceptual change is to be understood in terms of the people who create and change their representations of nature and the practices they employ to do so. This opens the possibility of understanding *how it is that scientists build on existing structures while creating genuine novelty*. That is to say, a route is opened toward explaining the continuous and noncumulative character of conceptual change that is amply supported by results of individual studies of scientific creativity undertaken by historians of science.

2.2. Outline of a Cognitive Theory of Conceptual Change

Further on in this essay I will turn to an examination of some of the processes of conceptual change. We need first, however, to have some sense of how a full theory would look.

A scientific theory is a kind of representational system. Several forms of representation have been proposed by cognitive scientists. Although it is a point of some controversy as to whether there is any form of representation other than strings of symbols, I will be following an authoritative account by Johnson-Laird (1983) in assuming the existence of at least three: (1) "propositional" representation (strings of symbols such as "the cat is on the mat"), (2) "mental models" (structural analogs of real-world or imagined situations, such as a cat being on a mat or a unicorn being in the forest), and (3) "images" (a mental model from a specific perspective, such as looking down on the cat on the mat from above). I will also be assuming with him that even if at the level of encoding all representations are propositional, in reasoning and understanding people construct mental models of real and imaginary phenomena, events, situations, processes, etc. One value of having a mental-models form of representation is that it can do considerable inferential work without the person having to actually compute inferences and can also narrow the scope of possible inferences. For example, moving an object immediately changes all of its spatial relationships and makes only specific ones possible. The hypothesis that we do such inferencing via mental models gains plausibility when we consider that, as biological organisms, we have had to adapt to a changing environment. In fact, artificial intelligence researchers have run into considerable problems handling the widespread effects of even small changes in knowledge-representational systems that are represented propositionally.

To continue, in a cognitive theory of conceptual change, a scientific theory will, itself, be construed as a structure that picks out classes of models, which accords better with the semantic view of theories (van Fraassen 1980) than with the Carnapian view of a theory as linguistic framework. Thinking about and in terms of a theory necessitates the construction of mental models. While scientific concepts may be encoded propositionally, understanding them involves interpretation, i.e., the construction of a mental model of the entities or processes they represent. Thus, what philosophers have been calling "meaning" and "reference" (i.e., the interplay between words, minds, and the world), is, on this view, mediated by the construction of mental models that relate to the world in specified ways.

Like science itself, a theory of conceptual change in science needs to provide both descriptive and explanatory accounts of the phenomena. These dimensions of the theory will here be called, respectively, the "kinematics" and "dynamics" of conceptual change. Kinematics is concerned with problems of how to represent change, and dynamics with the processes through which change is created.

2.2.1. The Kinematics of Change

Any analysis of how to represent conceptual change in science must be solidly informed by the actual representational practices scientists use in developing and changing conceptual systems. Examinations of the conceptual changes that have been part of "scientific revolutions" yield the following insights. New concepts are created, such as 'spin' in quantum mechanics, and existing ones disappear, such as 'phlogiston' from chemistry. Some concepts in the new system, e.g., 'mass' and 'field' in relativity, are what can only be called "conceptual descendants" of existing ones. And, finally, some, such as 'ether', while appearing to be eliminated, have significant aspects absorbed by other concepts, in this case 'field' and 'space-time'.

If the situations of creation and disappearance were all we had, handling the problems would be far less complex. In that case, "conceptual change" could be characterized as the replacement of one concept or structure by another. *Given the reality of descendants and absorption,* though, in addition to representing change of conceptual systems, we need to be able to represent conceptual change at the level of individual concepts. Some philosophers have trouble countenancing what it could possibly mean for a concept to *change* its meaning. As I have argued in my book (Nersessian 1984), the failure of existing theories of 'meaning' and 'meaning change' even to allow for this possibility has led to many of the various conundrums associated with the so-called problem of incommensurability of meaning. If, as has been traditionally held, concepts are represented by neatly bundled and individuated units (i.e., sets of necessary and sufficient conditions), only replacement, not change, is possible. Therefore, *a* *different form of representation is needed to accommodate the data of change.*

Psychological research on categorization supports the view that in many instances people do not represent concepts by means of sets of necessary and sufficient conditions (see, e.g., Rosch and Mervis 1975; Smith and Medin 1981; and Murphy and Cohen 1984). Examination of cases from the history of science also substantiates the view that for numerous scientific concepts — or even for a concept within a single theory — it is not possible to specify a set of necessary and sufficient conditions that will take in all their historical usages. For example, I have shown in my book how there is no set of necessary and sufficient conditions defining 'electromagnetic field'. Yet there is a traceable line of descent between the Faradayan and the Einsteinian concepts and a significant overlap in salient features between successive field concepts.

The question now is: Can we assume that how scientists structure *mental* representations is reflected in their *external* representations? That is, Does what they write give a clue to how they think? Psychological studies all start from the assumption that how people represent mentally is reflected in their use of language, and there is good reason to make the same assumption here. Since I have dealt with the kinematics of conceptual change extensively, though far from exhaustively, in my book and in a number of articles (Nersessian 1985, 1986, and in press[a]), I want the focus of this essay to be its dynamics. We need to keep in mind, though, that the two problems are connected. In order to be complete a cognitive theory will have to determine how the historical data on the individual units of change — "concepts" — mesh with those from psychology and also with attempts at constructing psychologically realistic representational systems in artificial intelligence.

2.2.2. The Dynamics of Change

By what processes are new scientific representations constructed? The prevailing view among philosophers is that the discovery processes are too mysterious to be understood. This view receives support from numerous stories of discovery through flashes of insight of geniuses, such as Kekulean dreams and Archimedean eureka-experiences. What is omitted from such renderings are the periods of intense and often arduous thinking and, in some cases, experimental activity that precede such "instantaneous" discoveries. There again, the rendering of conceptual changes as "gestalt switches" reinforces the prevailing prejudice. Even Kuhn substitutes the phrase "exploitation by genius" for analysis of actual constructive processes when discussing how Galileo formed his concept of 'pendulum' (Kuhn 1965, p. 119). There is, however, no *in-*

herent conflict between the view that discovery processes are creative and the view that they are reasoned. We need to give up the notion that "creativity" is an *act* and try to fathom it as a *process*.

Historical evidence supports the conviction that conceptual change in science is at heart a problem-solving process. While pragmatist philosophers, such as Dewey, Mead, and Popper, have strongly defended this view of science, conceptual change has not been included in their analyses. Laudan (1977) does introduce "conceptual problems" into the realm of the scientist's concerns, but offers no account of the specific processes of conceptual change. I want to extend the conception of science as a problem-solving enterprise to include what has traditionally been called "conceptual change." New representations do not emerge fully grown from the heads of scientists but are constructed in response to specific problems by systematic use of heuristic procedures. Problem solving in science does differ from much of "ordinary" problem solving in that scientific problems are more complex and often less well-defined, and the solution is not known in advance to anyone. A cognitive theory of conceptual change assumes the position long advocated by Herbert Simon that "the component processes, which when assembled make the mosaic of scientific discovery, are not qualitatively distinct from the processes that have been observed in simpler problem-solving situations" (Simon, Langley, and Bradshaw 1981, p. 2). The plausibility of this assumption is not diminished by the fact that the computer "discovery" programs implemented by Simon and his co-workers to model these processes thus far have tackled only the simplest of problem-solving heuristics. While the ability to model the problem-solving techniques that have brought about major conceptual changes seems a long way off, the type of cognitive analysis advocated here is within our grasp and is also a necessary preliminary step to more realistic computer modeling.

The next section of this essay will be devoted to examination of how a selection of problem-solving heuristics create new representations in science. As in my earlier work, I will draw largely, though not exclusively, on historical data from the construction of the field representation of forces from Faraday to Einstein. These are very rich data and as yet have been far from exhausted in their fertility for our purposes. Extending my earlier work, I subject mostly novel data to fresh layers of analysis.

Throughout the history of scientific change we find recurrent use of (1) analogical reasoning, (2) imagistic reasoning, (3) thought experiment, and (4) limiting case analysis. These are all modeling activities, and although they constitute a substantial portion of scientific method, none except analogy has received more than scant attention in the philosophical literature. The main problems philosophers have had in countenancing these as methods are that they are nonalgorithmic and, even if used correctly, may lead to the wrong solution or to no solution. This very feature, however, makes them much more realistic from a historical point of view.

Limiting scientific method to the construction of inductive or deductive arguments has needlessly blocked our ability to make sense of many of the actual constructive practices of scientists. I call the particular subset of practices I will be discussing "abstraction techniques." As we will see, they are strongly implicated in the explanation of how existing conceptual structures play a role in constructing new, and sometimes radically different, structures.

3. Abstraction Techniques and Conceptual Change

3.1. Analogical and Imagistic Reasoning

3.1.1. Background

There are numerous cases where analogy has played a central role in the construction of a new scientific concept: Newton's analogy between projectiles and the moon ('universal gravitation'), Darwin's analogy between selective breeding and reproduction in nature ('natural selection'), and the Rutherford-Bohr analogy between the structure of the solar system and the configuration of subatomic particles ('atom') are among the more widely known. Also, although less well known, there are numerous cases that establish the prominence of reasoning from pictorial representations in the constructive practices of scientists who were struggling to articulate new conceptualizations. Such imagistic representations have often been used in conjunction with analogical reasoning in science.

The major problem, as was noted above, is that while amply documented, these constructive practices have received scant attention from analysts of scientific method. Analogy has received the most attention, but the thrust of those analyses has been to conceive of it as a weak form of inductive argument. Following Campbell's (1920) lead, Hesse (1966) broke some ground in stressing the importance of analogy in giving meaning to new theoretical terms and in trying to formulate how it could be an inductive method without being a *logic*, i.e., algorithmic. Sellars (1965; see also H. Brown 1986) argued that, in general, analogical reasoning creates a bridge from existing to new conceptual frameworks through the mapping of relational structures from the old to the new.

I intend to show here that the insights of those who have recognized the importance and power of analogical reasoning in concept formation and change can be furthered by cognitive-historical analysis. To do this I will go beyond my earlier analysis of the construction of the field representation of forces to discuss how it is that imagistic and analogical reasoning were used by Faraday and Maxwell in their constructive efforts. I will also draw from current cognitive theories to show how it is possible that such reasoning *generates* new representations from existing ones.

3.1.2. Case Study: Faraday, Maxwell, and the Field

This analysis will focus on Maxwell as we find him in our opening scene. Naturally, like all scientists, Maxwell was working in a context. His analysis depends heavily, among other things, on a method he claims to have "poached" from William Thomson, on specific representations he takes from Faraday, and on a certain mathematical approach to analyzing continuum phenomena being developed at the time by Cambridge mathematical physicists.

The field representation of electromagnetic forces had its origin in vague speculations that there might be physical processes in the regions surrounding bodies and charges that could account for the apparent action of one body on another at some distance from it. Faraday was the first to attempt to construct a unified representation for the continuous transmission and interconversion of electric and magnetic actions. His formulation is primarily in qualitative form, and reasoning from a specific imagistic representation figures predominantly in its construction. He constructed his field concept by reasoning from representations of the "lines of force" such as those that form when iron filings are sprinkled around a magnetic source (see Figure 1). Many linelike features are incorporated into his representation. He characterized the lines as "expanding," "collapsing," "bending," "vibrating," "being cut," and "turning corners," and attempted to devise experiments to capture the diverse motions of the lines. Thus, he transformed the static visual representation of the lines into a qualitative dynamical model for the transmission and interconversion of electric and magnetic forces, and, ultimately, for all the forces of nature and matter. As Maxwell ([1855] 1991) remarked, although this model is qualitative, it embodies within it a great deal of mathematical understanding, which Maxwell himself was able to extract from it.

In the most complete formulation of Faraday's field representation nothing exists but a "sea" of lines of force: all the forces of nature are unified and interconvertible through various motions of the lines, with matter itself being nothing but point centers of converging lines of force. The centrality of the image in his reasoning can also be seen in the only quantitative relationship he formulated explicitly: that between the number of lines cut and the intensity of the induced force. This re-

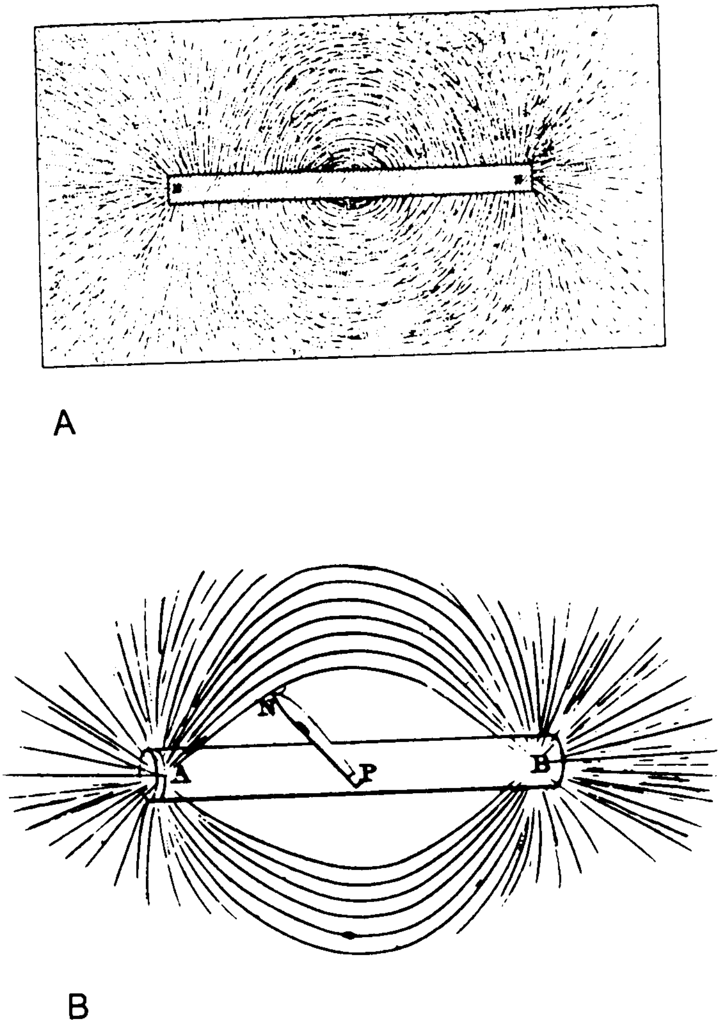

Figure 1. **A** *Actual pattern of lines of force surrounding a bar magnet (from Faraday 1839–55, vol. 3);* **B** *Schematic representation of lines of force surrounding a bar magnet (vol. 1).*

lationship is incorrect because "number of" is an integer, while "field intensity" is a continuous function. With our hindsight we can trace the "mistake" directly to the fact that lines are discrete entities and the image represents the filings as such, whereas, except in rare cases, the actual lines of force spiral indefinitely in a closed volume.

Near the end of his research Faraday introduced another image that was to play a key role in Maxwell's construction of a quantitative field representation. That image was of interlocking curves that represent the dynamical balance between electricity and magnetism (see Figure 2A) —

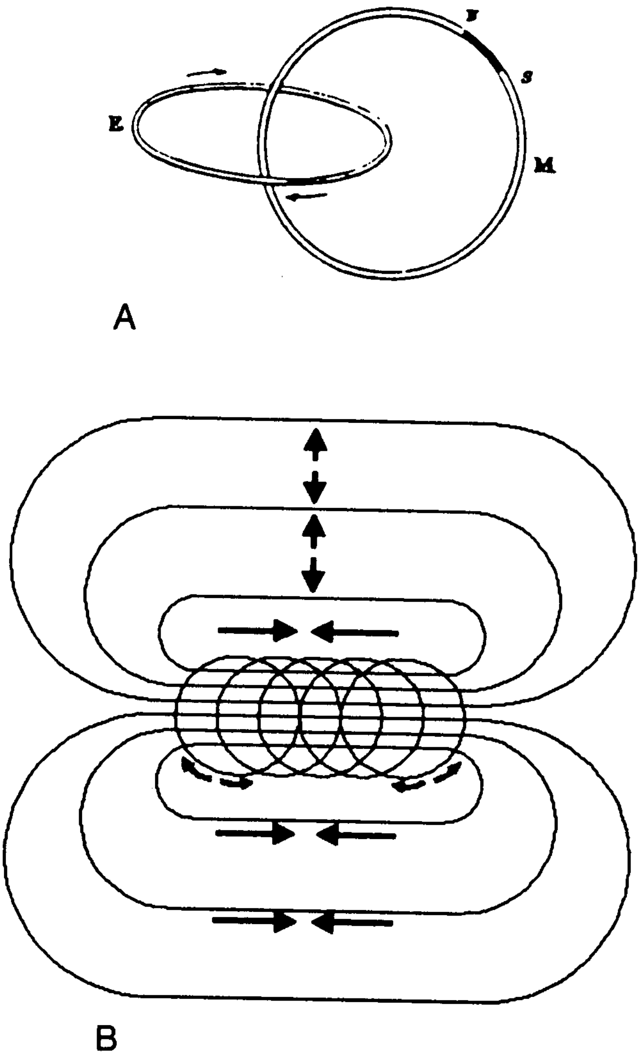

Figure 2. **A** *Faraday's representation of the interconnectedness of electric currents and magnetic force (from Faraday 1839–55, vol. 3);* **B** *My schematic representation of the reciprocal relationship between magnetic lines of force and electric current lines.*

what Faraday called "the oneness of condition of that which is apparently two powers or forms of power" (Faraday 1839–55, vol. 3, para. 3268). The image represents the structural relations between the electric and magnetic lines of force and is, thus, itself abstracted from the lines of force image. We can see this from Figure 2B. Here the outer lines represent magnetism and the inner lines, current electricity. The lateral repulsion of the magnetic lines has the same effect as a longitudinal expansion of the electric current lines, and the lateral attraction of the current lines has the same effect as a longitudinal contraction of the magnetic lines. This dynamical balance is captured in the image of interlocking curves.

Maxwell called the reciprocal dynamical relations embodied in the image "*mutually embracing* curves" (Maxwell 1890, vol. 1, p. 194n). While he does not include a drawing of this visual representation in the paper, Maxwell does describe it while referring the reader to the appropriate section of Faraday's *Experimental Researches.* Wise (1979) offers a convincing account of exactly how the image plays a role in the mathematical representation Maxwell constructed in his first paper on electromagnetism (Maxwell 1890, vol. 1, pp. 155–229) and throughout his work in his complicated use of two fields each for electric and magnetic forces: one for "intensity," a longitudinal measure of power, and one for "quantity," a lateral measure. Wise further provides a plausible argument that the image could even have provided a model for propagation of electromagnetic actions through the ether. If we expand Figure 2A into a "chain," then summations of the quantities and intensities associated with the electric and magnetic fields would be propagated link-by-link through interlocking curves.

While Maxwell constructed his full quantitative representation over the course of three papers, his central analysis is in the second, "On Physical Lines of Force" (Maxwell 1890, vol. 1, pp. 451–513). It is in this paper that he first derived the field equations, i.e., gave a unified mathematical representation of the propagation of electric and magnetic forces with a time delay, and calculated the velocity of the propagation of these actions. He achieved this by using a method he called "physical analogy" to exploit the powerful representational capabilities of continuum mechanics in his analysis. According to Maxwell, a physical analogy provides both a set of mathematical relationships and an imagistic representation of the structure of those relationships drawn from a "source" domain to be applied in analyzing a "target" domain about which there is only partial knowledge. In this case the method worked as follows.

Maxwell began by transforming the problem of analyzing the production and transmission of electric and magnetic forces into that of analyzing the potential stresses and strains in a mechanical electromagnetic medium ("target" domain) and then constructed an analogy between these and well-formulated relationships between known continuum mechanical phenomena ("source" domain). The process of application of the method of physical analogy comprised identifying the electromagnetic quantities with properties of the continuum mechanical medium; equating the forces in the electromagnetic ether with mechanical stresses and strains; abstracting what seemed to be appropriate relationships from the source domain and fitting them to the constraints of the target domain. In all this Maxwell explicitly provided imagistic representations to accompany the mathematical analysis.

Maxwell first constructed a simple representation consistent with a set of four constraints: the physical observations that (1) electric and magnetic actions are at right angles to one another and (2) the plane of polarized light is rotated by magnetic action, plus Faraday's speculative notions that (3) there is a tension along the lines of force and (4) there is a lateral repulsion between them. A mechanical analogy consistent with these constraints is a fluid medium, composed of vortices and under stress (see Figure 3B). With this form of the analogy Maxwell was able to provide a mathematical representation for various magnetic phenomena.

The problem of how to construe the relationship between electric current and magnetism led to an elaboration of the analogy. As we can see from Figure 3A, the vortices are all rotating in the same direction, which means that if they touch, they will stop. Maxwell argued that mechanical consistency requires the introduction of "idle wheels" to keep the mechanism going. He thus enhanced the source by surrounding the vortices with small spherical particles revolving in the direction opposite to them. There is a tangential pressure between the particles and the vortices, and for purposes of calculation Maxwell now had to consider the fluid vortices to be rigid pseudospheres. Maxwell's own imagistic representation of this enhanced source is seen in Figure 3B. He represented the dynamical relationships between current and magnetism mathematically by expressing them in terms of those between the particles and the vortices.

At this point Maxwell submitted the paper for publication. It took him several months to figure out the last — and most critical — piece of the representation: electrostatic actions. This is the point at which we joined him in the garden in Galloway. He found that if he made the vortices elastic and identified electrostatic polarization with elastic displacement he could calculate a wave of distortion produced by the polarization, i.e., what he called the "displacement current." He now had a unified, quantified representation of the continuous transmission of electromagnetic actions with a time delay. A testable consequence followed: electromagnetic actions are propagated at approximately the speed of light.

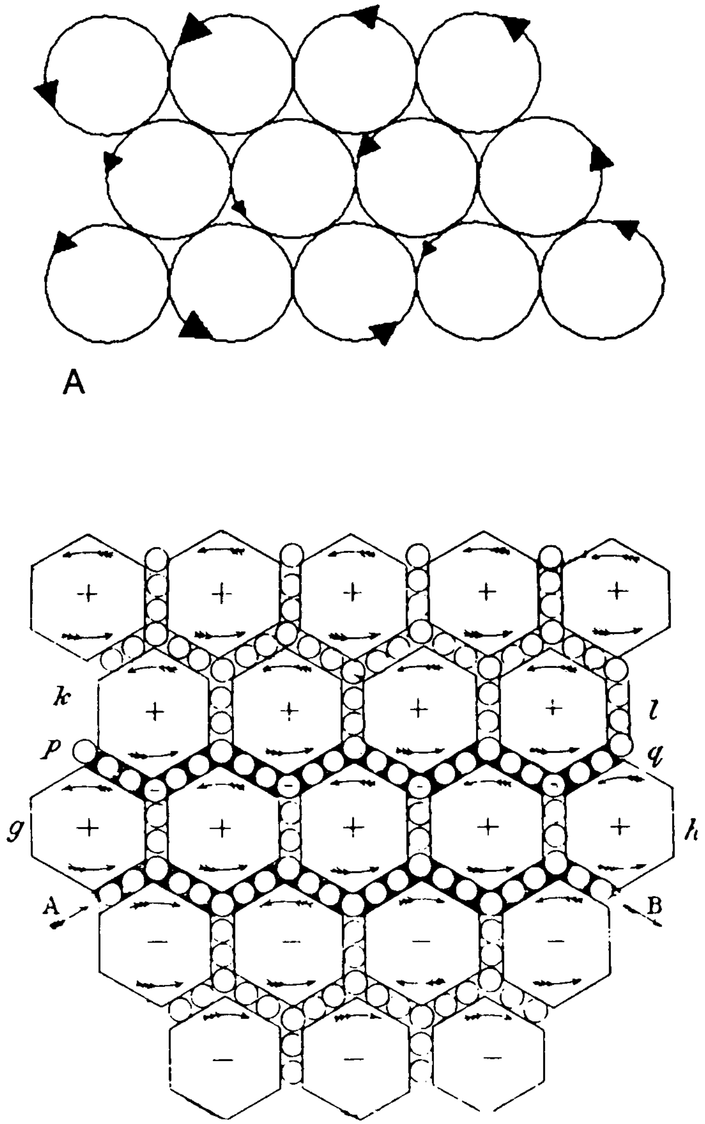

Figure 3. **A** *My schematic representation of initial crude source retrieved by Maxwell;* **B** *Maxwell's representation of his fully elaborated "physical analogy" (from Maxwell 1890, vol. 1).*

3.1.3. Cognitive Analysis: Analogical and Imagistic Reasoning

Analogical problem solving has been the subject of much recent work in cognitive psychology and artificial intelligence. It is widely recognized that analogy is a primary means through which we transfer knowledge from one domain to another. Knowing how this process works is essential for understanding learning and for developing expert systems. While still undergoing formulation, most cognitive theories of analogy agree that the creative heart of analogical reasoning is a modeling process in which relational structures from existing modes of representation and problem solutions are abstracted from a source domain and are fitted to the constraints of the new problem domain. A complete theory must give an account of the processes of retrieval, elaboration, mapping, transfer, and learning and of the syntactic, semantic, and pragmatic constraints that operate on these processes. Further, these processes need to be hooked up with other cognitive functions, such as memory. Most computational proposals are variants of three major theories of analogical problem solving: "structure mapping" (Gentner 1989), "constraint satisfaction" (Holyoak and Thagard 1989), and "derivational analogy" (Carbonell 1986).

Since my purpose is to show how we might conceive of what scientists like Faraday and Maxwell were doing with their images and analogies, I will not give detailed descriptions and evaluations of these theories. Rather, I will note some pertinent results from the empirical studies that inform them. Many psychological studies have been undertaken to understand how analogy functions in problem solving, especially in learning science. The results most germane to the issues of this essay are as follows. First, productive problem solving has the following features: (1) "structural focus": preserves relational systems; (2) "structural consistency": isomorphic mapping of objects and relationships; and (3) "systematicity": maps systems of interconnected relationships, especially causal and mathematical relationships. Second, the analogical reasoning process often creates an abstraction or "schema" common to both domains that can be used in further problem solving. Finally, in investigations of analogies used as mental models of a domain, it has been demonstrated that inferences made in problem solving depend significantly upon the specific analogy in terms of which the domain has been represented. For example, in one study where subjects constructed a mental model of electricity in terms of either an analogy with flowing water or with swarming objects, specific inferences — sometimes erroneous — could be traced directly to the analogy (Gentner and Gentner 1983). This result gives support to the view that analogies are not "merely" guides to thinking, with logical inferencing actually solving the problem, but *analogies themselves do the inferential work and generate the problem solution.*

Do these findings lend support to the interpretation I gave the case study, i.e., that analogical and imagistic reasoning are generating the respective field representations of Faraday and Maxwell? Can we model

Maxwell's use of physical analogy in cognitive terms? And what about the imagistic representations used by both him and Faraday?

While no current cognitive theory is comprehensive enough even to pretend to be able to handle all of the complexity of this case study, using what we believe we understand, cognitively, thus far, does enhance our understanding of it. It enables us, e.g., to fathom better what Maxwell was doing that summer day in Galloway and why he presented the physical analogy to his peers. Furthermore, this case points to areas of needed investigation in the cognitive sciences as well. I will first outline a cognitive analysis of Maxwell's generation of the field equations via the method of physical analogy and will then discuss the role of the imagistic dimension of that analogy along with the function of Faraday's imagistic representations.

Figure 4 provides a chart of Maxwell's modeling activities. His overall goal was to produce a unified mathematical representation of the production and continuous transmission of electromagnetic forces in a mechanical ether. The obvious source domain lay within continuum mechanics, a domain Maxwell was expert in. Continuous-action phenomena, such as fluid flow, heat, and elasticity, had all recently been given a dynamical analysis consistent with Newtonian mechanics and it was quite plausible to assume that the stresses and strains in an electromagnetic ether could be expressed in terms of continuum mechanical relationships. Using this source domain — if the analogy worked — he could presume to get: (1) assurance that the underlying forces are Newtonian; (2) continuity of transmission with the time delay necessary for a field theory; and (3) unification through finding the mathematical expression for the dynamical relations through which one action gives rise to another. He got all three, but the first was a false assurance. As we will discuss, the electromagnetic field equations represent a non-Newtonian dynamical system.

Maxwell retrieved a crude source from this domain by applying the four constraints we discussed above. He broke the overall goal down into subproblems, namely, representing magnetic induction, electricity, electromagnetic induction, and electrostatic induction. He then produced mappings between the electromagnetic quantities and mechanical properties of the fluid vortex medium and between the presumed stresses and strains and those known to take place in fluid medium under stress. The mappings are isomorphic and maintain causal interconnectedness. He reiterated the process twice, altering and enhancing the source to fit the constraints of the target domain. In the process he made "mistakes," most of which can be explained in terms of the model. For example, he takes the "displacement current" to be in the direction opposite from

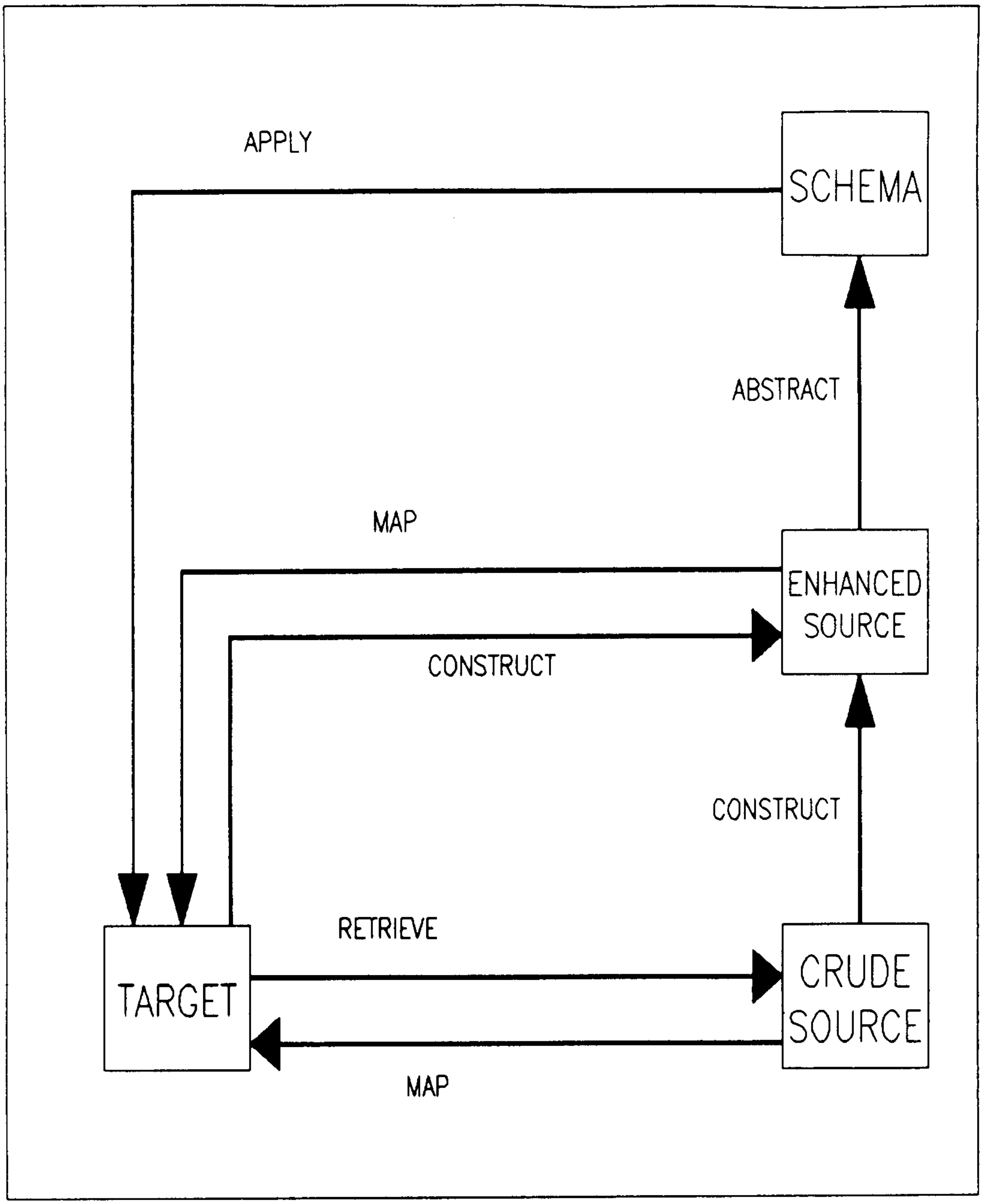

Figure 4. Maxwell's use of "the method of physical analogy."

the field intensity, and not — as is now customary and as he later held — in the same direction as the field intensity. In terms of the analogy, the equation is correct as it stands; i.e., in an elastic medium the restoring force and the displacement would have opposite orientation.

Through this process he abstracted a schema (i.e., a set of general dynamical [Lagrangian] relations) that could then be applied back to the target domain directly, without the need for any specific analogy. That is, at the end of the process he knew how to represent velocity, momentum, potential energy, and kinetic energy. With this knowledge he could re-derive the field equations using solely electromagnetic considerations, which is what he went on to do in his next paper on electrodynamics, "A Dynamical Theory of the Electromagnetic Field" (Maxwell 1890; vol. 1, pp. 526–97). But he only obtained this knowledge — the necessary general dynamical relations — *by abstracting it from the analogy he constructed with continuum mechanics.* This interpretation of what he actually did is consistent with what we have been learning about analogical reasoning in general. The physical analogy was generative and the process consistent with what are thought to be good analogical problem-solving techniques. But it also points to an area in need of investigation by cognitive theorists. None of the current theories addresses the kind of reiteration process Maxwell carried out: the modification and enhancement of the analogy to fit the constraints of the target domain. Yet one would expect such re-representation to be common in ordinary problem solving as well.

I want to underscore the point of my message by returning briefly to the historical story. In Maxwell's eyes his problem solution was never complete. To be so he would have to give an account of the actual underlying forces in the ether. The real power of analogical reasoning in creating new representations is driven home quite forcefully when we realize that he could *never* have done this: Maxwell's laws are those of a non-Newtonian system. How could it have turned out this way? That is, How could he have created a representation for a non-Newtonian system using a Newtonian system as the source of his analogy?

Maxwell did not take an existing physical system and plug the electromagnetic parameters into the equations for that system to solve the problem, as Thomson had done in finding a potential field representation for electrostatics in terms of Fourier's analysis of heat. Rather, he used various aspects of continuum mechanics and constraints from electromagnetism *to put together* a system, which *in its entirety* even he maintained did not exist — *possibly could not exist* — in nature. And *by drawing inferences from this abstracted model* he was able to extract a mathematical structure of greater generality than Newtonian mechanics. Contrary to what was believed at the time, Newtonian mechanics and general dynamics are not coextensive. This belief made Maxwell think that by giving a general dynamical formulation of electromagnetism he had shown the underlying forces to be Newtonian in nature. We now know that many different kinds of dynamical systems can be formulated in general dynamical terms, e.g., relativity and quantum mechanics.

Interpreted in the way I have been proposing, concept formation by analogical reasoning can, thus, be characterized as a process of abstraction from existing representations with increasing constraint satisfaction. This interpretation leads to a novel interpretation of continuity and commensurability between representations in cases like the one we just analyzed: these are to be found in the abstracted relationships common to the source and the target structures.

What about the imagistic representation of Maxwell's analogy and Faraday's earlier ones? While there are differences between the kinds of mappings made by Maxwell and those made by Faraday, both treated the imagistic representations as embodiments of an analogical source. I suggest that we construe this kind of imagistic reasoning as a species of analogical reasoning. In the early, qualitative phase, i.e., Faraday's work, the imagistic representation was taken as more nearly identical with what it represents than in Maxwell's quantitative analysis. That is, the schematic lines were identified with the physical lines they were to represent, and too many aspects of the specific image were incorporated into the field representation. A possible interpretation for this is that when there are insufficient constraints from the target domain to guide the process, analogies can be too generative. In the quantitative phase, i.e., in Maxwell's work, the function of the image is more abstract. It serves primarily to make certain structural relationships visualizable, and, thus, it is possible that any imagistic representation that embodied these relationships would serve the same purposes.

What further cognitive purposes might these physical embodiments serve? There is very little in the cognitive science literature on the possible role of imagistic representations like these in problem solving. I will offer some speculations by putting together recent work by Larkin and Simon (1987) on diagrams, and myself and James Greeno (1990) and Roschelle and Greeno (1987) on abstracted models in the hope of opening some avenues for research into possible roles these representations on paper play in constructing and reasoning with mental models, especially in the process of constructing a mathematical representation.

The main idea centers on the fact that perceptual inferences are easy for humans to make. By clustering connected information and making visual a chain of interconnected inferences the imagistic representations support a large number of immediate perceptual inferences. The representations on paper are presented in a form that already focuses on and abstracts specific aspects of phenomena. A great deal of mathematical information is implicit in such representations. By embodying structural relations thought to underlie phenomena they could facilitate access to the quantifiable aspects of the phenomena. As such, they provide an intermediate level of abstraction between phenomena and mathematical forms of representation (formulae). Additionally, they stabilize the image for the reasoner and make various versions accessible for direct comparison, in a way not available for internal images, and may thus take some of the load off memory in problem solving. Finally, they potentially play an important role in communicating new representations by providing a stable embodiment that is public. The imagistic representation could make it easier for others to grasp parts of the new representation than text and formulae alone. For example, Maxwell did not comprehend all the subtleties of the field concept Faraday articulated, but he did grasp the mathematical structures inherent in the lines of force representation and in the dynamical balance embodied in the interlocking curves. And Maxwell, in trying to communicate his new field representation to his colleagues, felt it necessary to provide them with the *physical* (i.e., embodied) analogy and extensive commentary on how to understand the method, in addition to leading them through the reasoning and rather than just presenting the mathematical arguments.

3.2. Thought Experiments and Limiting Case Analysis

3.2.1. Background

Another heuristic that occurs frequently in cases of major conceptual change is thought experimentation. The notion that an experiment can take place in thought seems paradoxical. Earlier scholarship presented two poles of interpretation of the role of thought experiments in creating conceptual change. Duhem dismissed them as bogus precisely because they are "not only not realized but incapable of being realized" (Duhem 1914, p. 202); i.e., they are not "experimental" in the customary sense. Koyré (1939, 1968), in contrast, argued that their logical force is so compelling that they supplant real experimentation in the construction of new representations for phenomena. The "thought" part of the experiment predominates and shows the synthetic a priori nature of scientific knowledge.

Contemporary historians and philosophers of science by and large reject both the extreme empiricism and rationalism of their forerunners. They acknowledge that thought experiment, while not eliminating the need for real experiment, is an important heuristic for creating conceptual change in science. Despite the consensus, based on ample historical documentation, there has been little theoretical analysis of just how such experiments function. As with the heuristics of analogy and imagery, I will attempt a cognitive-historical analysis of thought experiment, subsuming limiting case analysis as a species of this form of reasoning.

There have been a few recent attempts to analyze the function of thought experiments. While these are sketchy and limited, some do yield useful insights. Kuhn's analysis provides the starting point for most of these discussions. Kuhn claims that "thought experiment is one of the essential analytical tools which are deployed during crises and which then help to promote basic conceptual reform" (1964, p. 263). The importance of his analysis is that it is the first to try to come to grips with both the experimental and the thought dimensions of thought experiments. He argues that thought experiments show that there is no consistent way, *in actual practice*, of using accepted existing conceptions. That is, the thought experiment reveals that it is not possible to apply our conceptions consistently to real-world phenomena, and this practical impossibility translates into a logical requirement for conceptual reform. Gooding (1990), in his analysis of Faraday's experimental practices, picks up on Kuhn's analysis, rendering the empirical force of thought experiments in terms of their demonstration of what he calls the "impracticability of doing." Gooding is concerned to show how the real-world experimenter's knowledge of practical skill is utilized in the construction and manipulation of thought experiments.

My point will be that cognitive-historical analysis, by placing thought experiments within the framework of mental-models theory, offers the possibility of explaining *how* an experiment made in thought can have both the logical and empirical force Kuhn and Gooding argue for. In an unpublished work, Manukian (1987) has begun a cognitive-historical analysis of Galileo's thought experiments focusing on thought experiment as a species of information processing that employs mental models. He is the only one thus far to attempt to understand thought experiments as a form of "world modeling," as he calls constructing idealizations. Since his attention, however, is directed toward concerns in the field of artificial intelligence, he restricts his analysis to what I see as only a special case of thought experiment, the limiting case analysis.

Two other analyses need mentioning for the contrast in approach they provide. First, Brown (J. Brown 1986) claims that thought experiments are a species of a priori reasoning and that they get at something that cannot be derived from logical argumentation. His positive suggestion is that they provide a special window through which the mind grasps universals. The most cogent rendering I can give this claim is that Brown is trying to capture the idealizing function of thought experiments. Unlike Manukian, however, his approach through linguistic analysis does not afford the possibility of understanding their experimental nature.

Second, Norton's (1986) view that thought experiments can, in essence, be reconstructed as and replaced by arguments is the most sympathetic for philosophers who wish to restrict reasoning to logical argumentation — whether deductive or inductive. Certainly thought experiments contain an argument. As Norton himself acknowledges, however, the argument can be constructed only *after the fact*; i.e., it is not evident until after the thought experiment has been executed. By concentrating exclusively on this aspect he misses the importance of their experimental dimension. Additionally, while his claim that the presentation contains particulars irrelevant to the generality of the conclusion is correct, this emphasis reveals that he has also failed to see the *constructive function of the narrative form* in which thought experiments are customarily presented.

Both of these analyses take the traditional philosophical route of construing thought experiments in terms of propositional representations. My contention is that propositional representations cannot do the trick. *Mental simulation is required for a thought experiment to be both thought and experimental.* The original thought experiment is the construction of a mental model by the scientist who imagines a sequence of events. She or he then uses a narrative form to describe the sequence in order to communicate the experiment to others. Considerations of space will not permit here an analysis of comparable depth with that on analogical and imagistic reasoning. I do want, however, to put an analysis of a different heuristic before the reader to underscore the power of cognitive-historical analysis and to show that its utility is not restricted to what it offers for understanding the constructive practices of nineteenth-century British field theorists.

3.2.2. Case Studies: Galileo and Einstein

Although thought experiment occurs repeatedly in conceptual change, the thought experimenters who have attracted the most attention are Galileo and Einstein. I will give a brief presentation of a few of their thought experiments to convey some sense of the variety such experiments display and to elicit some common features for further analysis.

Galileo's importance as a pivotal figure in the transition from the qualitative categories of Aristotelian and medieval theories of motion to the quantitative representation of motion provided by Newton's mechanics is widely recognized. As shown in analyses by Koyré (1939) and Clavelin (1968), among others, Galileo drastically transformed the problem of how to go about constructing a mathematical representation of the phenomena of motion. That process, which Koyré called "mathematization," required constructing an idealized representation, quantifying this representation, and mapping the quantified representation back onto the real world. While it is now clear that Galileo must have performed many more real-world experiments than Koyré would have liked (see, e.g., Drake 1973; Naylor 1976; and Settle 1961), no one would deny the importance of thought experiment and limiting case analysis in the mathematization process. Take, as example, his analysis of falling bodies (Galilei 1638, pp. 62–86).

According to the Aristotelian theory, heavier bodies fall faster than lighter ones. This belief rests on a purely qualitative analysis of the concepts of 'heaviness' and 'lightness'. Galileo argued against this belief and constructed a new, quantifiable representation through a sustained analysis using several thought experiments and limiting case analyses. The outline of his use of these procedures is as follows. He calls on us to imagine we drop a heavy body and a light one, made of the same material, at the same time. We would customarily say that the heavy body falls faster and the light body more slowly. Now suppose we tie the two bodies together with a very thin — almost immaterial — string. The combined body should both fall faster and more slowly. It should fall faster because a combined body should be heavier than two separate bodies and should fall more slowly because the slower body should retard the motion of the faster one. Clearly something has gone amiss in our understanding of 'heavier' and 'lighter'. Having pinpointed the problem area, Galileo then goes on to show that it is a mistake to extrapolate from what is true at rest to what happens when bodies are in motion. That is, he has us consider that when the two bodies are at rest, the lighter will press on the heavier, and therefore the combined body is heavier. But when we imagine the two bodies are in motion, we can see the lighter does not press on the heavier and thus does not increase its weight. What Galileo has done up to this point is use the thought experiment to reveal the inconsistencies in the medieval belief, the ambiguities in the concepts, and the need to separate the heaviness of a body from its effect on speed in order to analyze free fall. He then goes on, using the methods of thought experiment and limiting case analysis in tandem, to show that the apparent difference in the speed of falling bodies is due to the effect of the medium and not to the difference in heaviness between bodies.

As Clavelin has pointed out, it is crucial for quantifying the motion of falling bodies that 'heaviness' not be the cause of the difference in speed because then we could not be sure that motion would be the same for all bodies. Galileo again used a thought experiment to demonstrate that the observed differences in speed should be understood as being caused by the unequal way media lift bodies. He asks us to suppose, e.g., that the density of air is 1; that of water, 800; of wood, 600; and of lead, 10,000. In water the wood would be deprived of 800/600ths of its weight, while lead would be deprived of 800/10,000ths. Thus, the wood would actually not fall (i.e., would float) and the lead would fall more slowly than it would in a less dense medium, such as air. If we extrapolate to a less dense medium, such as air, we see that the differential lifting effect is much less significant (e.g., 1/600 to 1/10,000 in air). The next move is to consider what would happen in the case of no medium, i.e., in extrapolating to the limiting case. With this move, Galileo says, "I came to the opinion that if one were to remove entirely the resistance of the medium, all materials would descend with equal speed" (Galileo 1638, p. 75). Having performed the extrapolation in this way we can quantify this idealized representation of the motion of a falling body and know that it is relevant to actual physical situations; we need only add back in the effects of a medium.

Galileo repeatedly used thought experiments and limiting case analyses in tandem as shown by this example both in constructing a quantifiable representation of bodies in motion and in attempting to convey this new representation to others. Later I will propose the cognitive functions of thought experiments and of limiting case analysis are much the same. Before getting to the cognitive analysis, though, I want to lay out somewhat different thought experiments used by Einstein in the development of the special and general theories of relativity.

Einstein began his paper "On the Electrodynamics of Moving Bodies" (1905) with the following thought experiment. Consider the case of a magnet and a conductor in relative motion. There are two possibilities. In the first case, the magnet is at absolute rest and the conductor moving. According to electromagnetic theory, the motion of the conductor through the magnetic field produces an electromotive force that creates a current in the conductor. In the second case, the conductor is at rest and the magnet moving. In this case, again according to electromagnetic theory, the motion of the magnet creates a changing magnetic field that induces an electric field that in turn induces a current in the conductor. With respect to the relative motions, however, it makes no difference whether it is the magnet or the conductor that is considered to be in motion. But according to the Maxwell-Lorentz electromagnetic theory, the absolute motions create a difference in how we would explain the production of a current in the conductor. Since the explanatory asymmetry could not, in principle or in practice, be accounted for by the observable phenomena — the measurable current in the conductor — Einstein argued that this supported his conclusion that "the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest" (p. 37).

Although I will not discuss them at any length, two more thought experiments figure crucially in this analysis. The second thought experiment in the paper is the famous one in which Einstein constructed an "operational" definition for the concept of simultaneity. According to Newtonian theory it is possible for distant events to occur simultaneously. Indeed this is necessary for there to be action at a distance. The thought experiment shows that we can define 'simultaneity' and thus what is meant by 'time' only for a particular reference system, and not in general. This experiment feeds into the next one, in which Einstein established the relativity of length; that is, he established the incorrectness of the Newtonian assumption that a body in motion over a specific time interval may be represented by the same body at rest in a definite location.

In a similar manner many thought experiments figured in Einstein's constructing and communicating the general theory of relativity, i.e., the field representation of gravitational action. We will just consider one he presented in various formats but claims to have first conceived in 1907. Einstein (1917, pp. 66–70) asks us to imagine that a large opaque chest, the interior of which resembles a room, is located in space far removed from all matter. Inside there is an observer with some apparatus. In this state, the observer would not experience the force of gravity and would have to tie himself with strings to keep from floating to the ceiling. Now imagine that a rope is connected to the outer lid of the chest and a "being" pulls upward with a constant force, producing uniform acceleration. The observer and any bodies inside the chest would now experience the very same effects, such as a pull toward the floor, as in a gravitational field. The experiment demonstrates that the behavior of a body in a homogeneous gravitational field and one in a uniformly accelerated frame of reference would be identical. Once we see that there is no way of distinguishing these two cases we can understand the importance of the Newtonian law that the gravitational mass of a body equals its inertial mass: these are just two manifestations of the same property of bodies. That is, we have a different interpretation for something we already knew.

3.2.3. Cognitive Analysis: Thought Experiments as Mental Modeling

Will rendering thought experiments as a species of mental modeling support the interpretation that when they are employed in conceptual change, they are "essential analytical tools" in the process? We can only speculate about what goes on in the mind of the scientist in the original thought experiment. Scientists have rarely been asked to discuss the details of how they went about setting up and running such experiments. As stated previously, it is quite possible that the original experiment involves direct construction, without recourse to language. However, reports of thought experiments are always presented in the form of narratives that call upon the reader/listener to simulate a situation in his or her imagination. Thus, drawing on what we think we know about both the process through which we imagine in general and the process through which we comprehend any narrative may help us to answer that most perplexing question about thought experiments: *How can an "experiment" carried out in thought have such powerful empirical force?* As was the case in the analysis of analogy and imagistic reasoning above, much research and development need to be done in this area, but I hope the sketch I present will persuade the reader that following this direction does offer good prospects of accounting for both the "thought" and the "experiment" aspects of thought experiments.

The most pertinent aspect of mental-models theory for this analysis is the hypothesis that understanding language involves the construction of a mental model. In the case of thought experiments we need to understand how: (1) a narrative facilitates the construction of an experimental situation in thought, and (2) thinking through the experimental situation has real-world consequences. Framed in mental-models theory, the "thought" dimension would include constructing a mental model and "running" a mental simulation of the situation depicted by the model, while the "experimental" dimension comprises the latter and the connection between the simulation and the world.

Briefly, the mental-models thesis about text comprehension is that understanding involves relating linguistic expressions to models; i.e., the relationship between words and the world is mediated by the construction of a structural analog to the situations, processes, objects, events, etc. depicted by a text (Franklin and Tversky 1990; Mani and Johnson-Laird 1982; Johnson-Laird 1983; McNamara and Sternberg 1983; Morrow et al. 1989; and Tversky 1990). What it means for a mental model to be a "structural analog" is that it embodies a representation of the spatial and temporal relationships between, and the causal structure connecting, the events and entities of the narrative. In constructing and updating a representation, readers would call upon a combination of conceptual and real-world knowledge and would employ the tacit and recursive inferencing mechanisms of their cognitive apparatus.

That the situation is represented by a mental model rather than by an argument in terms of propositions is thought to facilitate inferencing.

We can actually generate conclusions without having to carry out the extensive computations needed to process the same amount of background information propositionally. The conclusions drawn are limited to those that are directly relevant to the situation depicted. The ease with which one can make inferences in such simulative reasoning has suggested to some that mechanisms either used in — or similar to those used in — perception may be involved. As we saw in the discussion of the function of imagistic representations in problem solving, if we do employ perceptionlike mechanisms here many inferences would be immediate.

To date, most empirical investigations of the hypothesis have focused on the representation of spatial information by means of mental models. The main disagreement has been over whether the representation involves a perceptionlike image or is "spatial," i.e., allows different perspectives and differential access to locations. Although there is some research investigating the use of knowledge of causal structure in updating, on the whole there is little research investigating how knowledge and inferencing mechanisms are employed in running and revising the simulation.

Before beginning to sketch a way of understanding thought experiments and their role in conceptual change in terms of mental-models theory, we need first to glean some common features of thought experiments from the narratives presented above. While there is great variety among thought experiments, in general they do share important salient features. The Galileo and Einstein examples help us to see some of these. First, as noted, by the time a thought experiment is public it is in the form of a narrative. The narrative has the character of a simulation. It calls upon the reader/listener to imagine a dynamic scene — one that unfolds in time. The invitation is to follow through a sequence of events or processes *as one would in the real world.* That is, even if the situation may seem bizarre or fantastic, such as being in a chest in outer space, there is nothing bizarre in the unfolding: objects float as they would in the real world in the absence of gravity. The assumption is that if the experiment could be performed, the chain of events would unfold according to the way things usually take place in the real world.

Second, a thought experiment embodies specific assumptions — either explicit or tacit — of the representation under investigation. It usually exposes inconsistencies or exhibits paradoxes that arise when we try to apply certain parts of that representation to a specific situation, such as 'heavy' and 'light' to falling rocks. The paradox can take the form of a contradiction in the representation (e.g., it requires that an object be both heavy and light) or of something being not physically possible (e.g., observing the asymmetry electromagnetic theory requires).

Third, by the time a thought experiment is presented it always works and is often more compelling than most real-world experiments. We rarely, if ever, get a glimpse of failed thought experiments or avenues explored in the construction of the one presented to us.[6] Some experiments, such as Galileo's second, could potentially be carried out — at least until the analysis extrapolates to the limit. Others, such as Einstein's first, underscore that doing a real-world experiment could not provide the data the theory requires. Still others, such as Einstein's "chest" in space, are impossible to carry out in practice, either in principle or because we do not yet have the requisite level of technological achievement. Once understood, however, a thought experiment is usually so compelling in itself that even where it would be possible to carry it out, the reader feels no need to do so. The constructed situation is apprehended as pertinent to the real world either by revealing something in our experience that we did not see the import of before (e.g., the measurable current in the stationary and in the moving conductor is the same, so on what basis can we support the difference in theoretical explanation?) or by generating new data (e.g., in the case of no medium lead and wood would fall at the same speed) or by making us see the empirical consequences of something in our existing representation (e.g., the attributes called 'gravitational mass' and 'inertial mass' are the same property of bodies).

Finally, the narrative presentation has already made some abstraction from the real-world phenomena. For example, certain features of objects that would be present in a real experiment are eliminated, such as the color of the rocks and the physical characteristics of the observers. That is, there has been a prior selection of the pertinent dimensions on which to focus, which evidently derives from our experience in the world. We know, e.g., that the color of a rock does not effect its rate of fall. This feature strengthens our understanding of the depiction as that of a prototypical situation of which there could be many specific instances. In more colorful narratives there may be more irrelevant features in the exposition, but these most often serve to reinforce crucial aspects of the experiment. For example, Einstein's characterization, in one version of the chest — or "elevator" — experiment, of a physicist as being drugged and then waking up in a box served to reinforce the point that the observer could not know beforehand if he were falling in outer space or sitting in a gravitational field. In the version discussed above, the opacity of the chest is to prevent the observer from seeing if there are gravitational sources around.

We can outline the function of thought experiments in terms of mental-models theory as follows. The first performance of a thought experiment is a direct mental simulation. This hypothesis gains plausibility when we realize the likelihood that direct mental simulation precedes even real-world experiments; i.e., the scientist envisions and unfolds a sequence of steps to be carried out in the experiment.[7] The cognitive function of the narrative form of presentation of a thought experiment to others is to guide the construction of a structural analog of the prototypical situation depicted in it. Over the course of the narrative, we are led to run a simulation that unfolds the events and processes by constructing, isolating, and manipulating the pertinent dimensions of the analog phenomena. The process of constructing and running the model gives the thought experiment its applicability to the world. *The constructed situation inherits empirical force by being abstracted from both our experiences and activities in, and our knowledge, conceptualizations, and assumptions of, the world.* In running the experiment, we make use of various inferencing mechanisms, existing representations, and general world knowledge to make realistic transformations from one possible physical state to the next. In this way, the data that derive from a thought experiment, while constructed in the mind, are *empirical* consequences that at the same time pinpoint the locus of the needed representational change.

Limiting case analysis can be construed as a species of thought experiment. In this species the simulation consists of abstracting specific physical dimensions to create an idealized representation, such as of a body falling in a vacuum. The isolation of the physical system in thought allows us to manipulate variables beyond what is physically possible. Just what dimensions produce the variation and how to extrapolate from these may be something we determine initially in real-world experimentation, but the last step can only be made in the imagination. In physics, it is the idealized representation that is quantifiable. The idealized representation, however, is rooted in and relevant to the real world because it has been created by controlled extrapolation from it. We get from imagination to the real world by adding in some of the dimensions we have abstracted, again in a controlled process.

3.3. Summary: Abstraction Techniques and Conceptual Change

What we have seen in our discussion of the dynamics of conceptual change in science is the potential for acquiring a deeper understanding of the processes through which new scientific representations are constructed and communicated by joining historical analysis with our developing insights into how human beings represent, reason, and solve problems generally. By linking the conceptual and the experiential dimensions of human cognitive processing, mental-models theory offers the possibility of capturing and synthesizing theoretical, experimental, and social dimensions of scientific change. Our investigation demonstrated in some detail how cognitive-historical analysis helps us to fathom how the heuristics of analogy, imagistic reasoning, thought experiment, and limiting case analysis, of which we see recurrent use in what has been called "radical" and "revolutionary" conceptual change, could function to create genuinely novel representations by increasing abstraction from existing representations in a problem-solving process.

4. Wider Implications

To conclude I would like to underscore the potential fertility of the cognitive-historical method outlined and illustrated above by considering its wider implications for the disciplines it comprises. To do this we need to return to our title query: How do scientists think?

4.1. Implications for Philosophy of Science

1. While philosophers would be comfortable with the generality of the question, the detour through language taken by many philosophers of science has prevented its asking. Those who would ask it would prefer it transformed into: How ought scientists to think? And, quite generally, the "creative processes" are deemed by philosophers to be too mysterious to be understood. The main point of the investigations above is to show how cognitive-historical analysis opens the possibility of construing the representational and constructive practices of scientists as part of scientific method, and as such within the province of philosophical investigation. The analysis supports and elaborates upon the intuitions of those who have argued that *reasoning comprises more than algorithms* and for the generative role of such heuristics as analogy. Further, developing criteria for evaluating good and bad uses of heuristics will enable us to show why it is rational to believe inferences resulting from good heuristics are worth testing, holding conditionally if testing is not feasible, etc.

2. A major problem for historicist philosophers has been *how to go from a case study to a more general conclusion.* Those who want to use scientific knowledge practices as a basis from which to develop a historicized epistemology recognize the dangers of generalizing from one case, no matter how salient. Placing discovery processes within the framework of human representational and problem-solving abilities enables us to extend from case studies without committing the serious injustices of past reconstructive approaches.

3. As discussed in the body of this essay, cognitive-historical analysis

offers a way of recasting many of the puzzles associated with "incommensurability of meaning." By focusing on the people and practices that create and change representations of a domain, rather than on static linguistic representations that change dramatically from time to time, we open the possibility of understanding how scientists build new and sometimes radically different representations out of existing ones, and thus for explaining the continuous and noncumulative character of scientific change. And, lastly, we even have a way of making sense of that most paradoxical of all Kuhnian claims: Postrevolutionary scientists quite literally understand and experience the world in a manner incommensurable with their prerevolutionary counterparts (or selves in some cases). If we do negotiate the world by constructing mental models, prerevolutionary and postrevolutionary scientists would construct different mental models and would, thus, truly have different experiences of the world.

4.2. Implications for History of Science

1. Historians do not pose the question this way. Those who still do address such issues ask: How did my individual scientist, such as Maxwell, think? However, every historian — no matter how scrupulously he or she has tried to reconstruct the mosaic of a discovery process in a manner faithful to the historical record — must have experienced the nagging doubt: But did they *really* think this way? In the end we all face the recorded data and know that every piece is in itself a reconstruction by its author. The diaries and notebooks of a Faraday may be the closest we will ever get to "on-line" thinking, and even these cannot be taken as involving no reconstruction and as capturing all the shifts and strategies employed. If we can show, however, that what the particular scientist claims and/or seems to be doing is in line with what we know about human cognitive functioning generally, we can build a stronger case for our interpretation and fill in missing steps in a plausible manner, as I have done in the Maxwell example.

2. In claiming a generative role for heuristics such as those discussed above one often has the sense of "preaching to the converted" when talking with historians of science. *But historians do not always come down on the side of taking apparent uses of problem-solving heuristics at face value.* Witness the Maxwell case. It is still controversial as to whether or not he was reasoning through the analogical model he presented.[8] Historians who see such models as off to the side, while some other process is actually generating the representation, have at least tacitly bought the philosopher's assumption that reasoning is only by means of inductive or deductive algorithms.

Cognitive-historical analysis provides support for the idea that such heuristics are not "merely suggestive" (Heimann 1970) or an "unproductive digression" (Chalmers 1986) *but are fundamental, i.e., they constitute the essence of the reasoning process.* When they are taken in this way we get a better fit with the data and have less of a need to throw inconvenient pieces, such as Maxwell's "errors," away. Further, insights into how cognitive abilities and limitations contribute to and constrain scientific theorizing, experimentation, assessment, and choice can enrich the analyses of those who do take such heuristics seriously, irrespective of whether the scientists in question go down dead ends, contribute to "winning" science, employ different strategies to get to the same point, etc.

3. Controversies have often arisen within the history of science over such questions as: Do we find the concept of inertia in Galileo's physics? or Did Faraday have a field concept that guided his research from early on, or did he formulate it only in his last year, or never? The metatheoretical question that lies at the heart of such seemingly irresolvable disputes — *indeed at the core of historical method* — has scarcely been noticed by historians. While as historians of science we must not attribute present-day views to past scientists, even though the concepts may look quite familiar, *there is no explicit guidance on a theoretical level for how to do this.* Intuitive strategies for avoiding the problem are learned with the craft. What is missing is an explicit metatheoretical notion of what constitutes the meaning of a scientific concept.

Now why would an explicit notion yield better results than mere intuitions acquired in the craft? The answer is that underlying many of these controversies is the tacit assumption that a concept is represented by a set of necessary and sufficient conditions. As we have seen in the discussion of kinematics above, this form of representation cannot accommodate a substantial body of historical data. With it we have no criteria other than the modern for determining whether or when Faraday's concept is a field concept; no means for justifying the intuitive sense of "family resemblance" between what Galileo is discussing and what Newton called 'inertia' or for making sense of the fact that we seem able to trace out a distinct line of descent and a pattern of progress over time in a conceptual domain. *Seeing these problems as part of the wider representational problem opens a new avenue for their resolution.* In an article on Faraday's field concept, I have shown in detail how such an analysis can help to resolve standing controversies among historians (Nersessian 1985).

4. Many contemporary historians are concerned with issues about the form and rhetoric of presentation of novel ideas. These are usually

framed in terms of how scientists adopt certain modes and conventions of writing in order to persuade others of their ideas. What tends to be left out of the analysis is that in order to *persuade* one has to get one's colleagues to *comprehend* the new ideas and, again, in order to *negotiate*, one has to *comprehend* what is being proposed. Cognitive-historical analysis allows us to take the public communications of scientists presenting new representations as attempts at trying to get others to understand them. That is, we can view such communications as presented in ways that the creators find meaningful for their own construction and understanding. Success at communication does not, of course, entail success at persuasion.

Looked at in terms of the rhetoric of persuasion, Maxwell's analogies might seem utter failures. His presentation was quite out of line with the modes and conventions of his contemporaries who were publishing in *Philosophical Transactions* at that time. And even the person from whom he claims to have "poached" the technique, William Thomson, did not accept the method of analysis as transformed by Maxwell and therefore Maxwell's results. We can make better sense of what Maxwell was doing in presenting the work in that format if we assume he was trying to get his colleagues to understand his new field representation of forces by leading them through the modeling processes he used to construct the electromagnetic field concept, as well as trying to convince them of its potential.

5. Finally, we repeatedly find claims in the historical literature about the influence of wider cultural factors on a person's science; notable examples from physics are the influence of Faraday's Sandemanian religion on his belief in the unity of forces and acausality in quantum mechanics deriving from Weimar culture. Cognitive-historical analysis can capture the *locality* that is essential to historical understanding. It offers the potential for determining how it is that representational resources that are part of the scientist's local culture — whether these be construed as within a community of scientists, such as Cambridge mathematical physicists, or as within a wider Weltanschauung, such as obtained in Victorian England — are incorporated into scientific representations. *The problem becomes that of how it is that scientists, working individually or collectively, combine the cognitive abilities they have in virtue of their biology with the conceptual resources they acquire from the various facets of their lives in a wider community.*

4.3. Implications for Psychology

1. Cognitive psychologists do ask the question, but by themselves lack the means to answer it fully. "Revolutionary" science is rare and so is the

possibility of catching it "on-line." Cognitive-historical analysis greatly increases the database for psychological research. A cognitive theory of problem solving, in order to be adequate, needs to take into account the practices of scientists who have created major innovations in its formulation. This has scarcely been done to date. Combining the resources of cognitive psychologists, artificial intelligence researchers who work on modeling human cognitive processes, and historians and philosophers of science will lead to a more realistic portrayal of the complexities of scientific reasoning than current cognitive models and "discovery programs" provide.

2. Are the conceptual changes that take place in development and/or learning like those in scientific revolutions? This question is acquiring an important place in the contemporary psychological literature. A growing contingent of cognitive psychologists has been arguing that the processes of cognitive development and conceptual change (or "restructuring") in learning are indeed like those of major scientific revolutions (see, e.g., Carey 1985; Keil 1989). The main support for the psychological hypothesis comes from research that describes the initial states of children and students and compares those states with the desired final state. The kinds of changes necessary to get from one state to the other seem to resemble those that have taken place in scientific revolutions *as they have been construed by Kuhn,* whose "gestalt switch" metaphor of a scientific revolution many psychologists have uncritically adopted. And, as in history and philosophy of science, the nature of the *processes* through which conceptual change is brought about has not been explored in any depth in psychology. As demonstrated in this essay, cognitive-historical analysis points the way to a quite different understanding of the kinds of conceptual changes that take place in scientific revolutions and opens an avenue for examining the processes that bring them about.

4.4. Implications for Science Education

There is growing interest in how all three disciplines might work together on the problem of how to help students learn science. Cognitive-historical analysis opens a new area of exploration. In fact, it offers a way of fundamentally recasting the old position — proposed by Dewey, Bruner, and Conant, among others — that developing an appreciation for the historical roots of scientific ideas will facilitate learning because students will have a context in which to place them.

A cognitive theory of conceptual change views scientific "discovery" as a process in which scientists actively construct representations by employing problem-solving procedures that are on a continuum with those we employ in ordinary problem solving. With such a "constructionist" conception of discovery the cognitive activity of the scientist becomes directly relevant to learning. *The historical processes provide a model for the learning activity itself,* and, thus, have the potential to assist students in constructing representations of extant scientific theories. The history of science becomes, in this domain, a repository not of case studies, but rather of strategic knowledge of how to go about constructing, changing, and communicating scientific representations. As I have proposed elsewhere (Nersessian 1989, in press[b]), we should "mine" historical data — publications, diaries, notebooks, correspondence, etc. — for these strategies and then devise ways of integrating and transforming these more realistic exemplars of scientific problem solving into instructional procedures.

5. Return To Galloway

Coming back once more to our opening scene, what would cognitive-historical analysis have Maxwell doing that summer day? It would have him searching for a way to make electrostatic phenomena meaningful in terms of mechanical phenomena he believes he understands.

Constructing an analogical model allowed Maxwell to gain access to a systematic body of knowledge: a structure of causal and mathematical relationships with its own constraints. Maxwell generated a continuous-action representation for electromagnetism through a process of fitting the model to the constraints of the new domain. The physical embodiment facilitated his exploration of the kind of mechanical forces that could be capable of producing electromagnetic phenomena. This was done at an intermediate level of abstraction: concrete enough to give substance to the relationships he was examining and indicate possible paths to solution, and yet abstract enough to generate a novel representation — one that is dynamical but not mechanical. Once he had understood the dynamical relations through this process he was able to rederive the mathematical representation without this — or any — specific model, but just with the assumption of a general mapping between electromagnetic and dynamical variables.

The Scene

May 18, 1863, 8 Palace Garden Terrace, Kensington West.

Maxwell is sitting in a parlor sipping tea. "Aha!" he thinks, "now I know how to do it without the model."

The Problem

If we were to take this as our starting point, we could never hope to fathom the nature of conceptual change in science.

Notes

The research undertaken for this paper and its preparation were supported by NSF Scholars Award DIR 8821422. The author wishes to thank David Gooding, James Greeno, Mary Hesse, Simon Schaffer, and Robert Woolfolk for valuable discussions of the material in this paper and Floris Cohen, Richard Grandy, Larry Holmes, Philip Johnson-Laird, Paul Thagard, and Norton Wise for critical and editorial comments on the penultimate version.

1. Giere (1988) argues against a role for history of science in a cognitive theory of science. His main point is that historians of science have — by and large — not done the kind of analysis of the scientific record that a cognitive theory requires. While there are notable exceptions, such as Holmes (1985), he is correct on this point. What I am arguing here is that the historical record, itself, does contain material of great importance to a cognitive theory of science. It contains much information about the cognitive activity of scientists over the history of science.
2. I owe the name "context of development" to Richard Grandy.
3. Holmes (1985, pp. 119–20) argues in a similar vein.
4. Cohen (1990) contains an interesting discussion of the "two Kuhns" — the philosopher of science and the historian of science — and of the repercussions of the split for his analysis of the scientific revolution of the seventeenth century.
5. Notable exceptions are Clement (1983), Langley et al. (1987), McCloskey (1983), Qin and Simon (1990), and Wiser and Carey (1983).
6. An analysis of some of Faraday's explorations of thought experiments is to be found in Gooding (1990).
7. See Gooding (1990) for a fuller discussion of this point with respect to Faraday's experiments.
8. Berkson (1974), Bromberg (1968), Nersessian (1984, 1986), and Siegel (1986) present arguments in favor of the centrality of the analogical model, while Chalmers (1986), Duhem (1902, 1914), and Heimann (1970) are among those who deny its importance.

References

Berkson, W. 1974. *Fields of Force: The Development of a World View from Faraday to Einstein.* New York: John Wiley and Sons.

Bromberg, J. 1968. "Maxwell's Displacement Current and His Theory of Light." *Archive for the History of the Exact Sciences* 4:218–34.

Brown, H. 1986. "Sellars, Concepts, and Conceptual Change." *Synthese* 68:275–307.

Brown, J. R. 1986. "Thought Experiments since the Scientific Revolution." *International Studies in the Philosophy of Science* 1:1–15.

Campbell, N. R. 1920. *Physics, the Elements.* Cambridge: Cambridge University Press.

Carbonell, J. 1986. "Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition." In *Machine Learning: An Artificial Intelligence Approach*, ed. R. Michalski, J. Carbonell, and T. Mitchell, pp. 371–92. Los Altos, Calif.: Morgan Kaufmann.

Carey, S. 1985. *Conceptual Change in Childhood.* Cambridge, Mass.: MIT Press.

Chalmers, A. F. 1986. "The Heuristic Role of Maxwell's Mechanical Model of Electromagnetic Phenomena." *Studies in the History and Philosophy of Science* 17:415–27.

Clavelin, M. 1968. *The Natural Philosophy of Galileo: Essay on the Origins and Formation of Classical Mechanics.* Cambridge, Mass.: MIT Press.

Clement, J. 1983. "A Conceptual Model Discussed by Galileo and Used Intuitively by Physics Students." In *Mental Models,* ed. D. Gentner and A. Stevens, pp. 325–40. Hillsdale, N.J.: Lawrence Erlbaum.

Cohen, H. F. 1990. "The Banquet of Truth: An Historiographical Inquiry into the Nature and Causes of the 17th Century Scientific Revolution." Unpublished manuscript.

Drake, S. 1973. "Galileo's Experimental Confirmation of Horizontal Inertia: Unpublished Manuscripts." *Isis* 64:291–305.

Duhem, P. 1902. *Les théories électriques de J. Clerk Maxwell: Étude historique et critique.* Paris: A. Hermann and Cie.

———. 1914. *The Aim and Structure of Physical Theory.* Reprint, New York: Atheneum, 1962.

Einstein, A. 1905. "On the Electrodynamics of Moving Bodies." In *The Theory of Relativity.* Reprint, New York: Dover, 1952.

———. 1917. *Relativity: The Special and the General Theory.* Reprint, London: Methuen, 1977.

Faraday, M. 1839–55. *Experimental Researches in Electricity.* Reprint, New York: Dover, 1965.

Franklin, N., and B. Tversky. 1990. "Searching Imagined Environments." *Journal of Experimental Psychology* 119:63–76.

Galilei, G. 1638. *Two New Sciences.* Trans. S. Drake. Reprint, Madison: University of Wisconsin Press, 1974.

Gentner, D. 1989. "The Mechanisms of Analogical Learning." In *Similarity and Analogical Reasoning,* ed. S. Vosniadou and A. Ortony, pp. 200–241. Cambridge: Cambridge University Press.

Gentner, D., and D. R. Gentner. 1983. "Flowing Waters and Teeming Crowds: Mental Models of Electricity." In *Mental Models,* ed. D. Gentner and A. Stevens, pp. 99–130. Hillsdale, N.J.: Lawrence Erlbaum.

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

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

Gooding, D. 1990. *Experiment and the Making of Meaning: Human Agency in Scientific Observation and Experiment.* Dordrecht: Kluwer Academic Publishers.

Gorman, M. E., and W. B. Carlson. 1989. "Interpreting Invention as a Cognitive Process: The Case of Alexander Graham Bell, Thomas Edison, and the Telephone." Forthcoming in *Science, Technology and Human Values.*

Heimann, P. M. 1970. "Maxwell and the Modes of Consistent Representation." *Archive for the History of the Exact Sciences* 6:171–213.

Hesse, M. 1966. *Models and Analogies in Science.* Notre Dame, Ind.: 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.

Holyoak, K., and P. Thagard. 1989. "Analogical Mapping by Constraint Satisfaction: A Computational Theory." *Cognitive Science* 13:295–356.

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

Keil, F.C. 1989. *Concepts, Kinds, and Conceptual Development.* Cambridge, Mass.: MIT Press.

Koyré, A. 1939. *Galileo Studies.* Atlantic Highlands, N.J.: Humanities Press, 1979.

———. 1968. *Metaphysics and Measurement.* Cambridge, Mass.: Harvard University Press.

Kuhn, T.S. 1964. "A Function for Thought Experiments." In *The Essential Tension: Selected Studies in Scientific Tradition and Change.* Chicago: University of Chicago Press.

———. 1965. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

———. 1986. "Possible Worlds in History of Science: Nobel Symposium — August 1986." Unpublished manuscript.

———. 1987. "What Are Scientific Revolutions?" In *The Probabilistic Revolution,* vol. 1 of *Ideas in History,* ed. L. Kruger, L.J. Daston, and M. Heidelberger, pp. 7–22. Cambridge, Mass.: MIT Press.

Langley, P., et al. 1987. *Scientific Discovery: Computational Explorations of the Creative Processes.* Cambridge, Mass.: MIT Press.

Larkin, J.H., and H.A. Simon. 1987. "Why a Diagram Is Sometimes Worth Ten Thousand Words." *Cognitive Science* 11:65–100.

Laudan, L. 1977. *Progress and Its Problems: Towards a Theory of Scientific Growth.* Berkeley: University of California Press.

McCloskey, M. 1983. "Naive Theories of Motion." In *Mental Models,* ed. D. Gentner and A.L. Stevens, pp. 299–324. Hillsdale, N.J.: Lawrence Erlbaum.

McNamara, T.P., and R.J. Sternberg. 1983. "Mental Models of Word Meaning." *Journal of Verbal Learning and Verbal Behavior* 22:449–74.

Mani, K., and P.N. Johnson-Laird. 1982. "The Mental Representation of Spatial Descriptions." *Memory and Cognition* 10:181–87.

Manukian, E. 1987. "Galilean vs. Aristotelian Models of Free Fall and Some Modern Concerns in Artificial Intelligence." Unpublished manuscript.

Maxwell, J.C. [1855] 1991. "On Faraday's Lines of Force." In *The Scientific Papers of James Clerk Maxwell,* ed. P.M. Harman. Cambridge: Cambridge University Press.

———. 1890. *The Scientific Papers of J.C. Maxwell,* ed. W.D. Niven. Cambridge: Cambridge University Press.

Morrow, D.G., G.H. Bower, and S.L. Greenspan. 1989. "Updating Situation Models during Narrative Comprehension." *Journal of Memory and Language* 28:292–312.

Murphy, G.L., and B. Cohen. 1984. "Models of Concepts." *Cognitive Science* 8:27–58.

Naylor, R. 1976. "Galileo: Real Experiment and Didactic Demonstration." *Isis* 67:398–419.

Nersessian, N.J. 1984. *Faraday to Einstein: Constructing Meaning in Scientific Theories.* Dordrecht: Martinus Nijhoff.

———. 1985. "Faraday's Field Concept." In *Faraday Rediscovered,* ed. D. Gooding and F. James, pp. 175–87. London: Macmillan.

———. 1986. "A Cognitive-Historical Approach to Meaning in Scientific Theories." In *The Process of Science: Contemporary Philosophical Approaches to Understanding Scientific Practice,* ed. N.J. Nersessian, pp. 161–78. Dordrecht: Martinus Nijhoff.

———. 1989. "Conceptual Change in Science and in Science Education." *Synthese* 80:163–84.

———. In press(a). "Discussion: The 'Method' to Meaning: A Reply to Leplin." *Philosophy of Science.*

———. In press(b). "Constructing and Instructing: The Role of 'Abstraction Techniques' in Developing and Teaching Scientific Theories." In *Philosophy of Science, Cognitive Science, and Educational Theory and Practice*, ed. R. Duschl and R. Hamilton. Albany: SUNY Press.

Nersessian, N. J., and J. G. Greeno. 1990. "Multiple Abstracted Representations in Problem Solving and Discovery in Physics." *Proceedings of the Cognitive Science Society* 11:77–84.

Norton, J. 1986. "Thought Experiments in Einstein's Work." Unpublished manuscript.

Qin, P., and H. A. Simon. 1990. "Laboratory Replication of Scientific Discovery Processes." *Cognitive Science* 14:281–308.

Rosch, E., and C. Mervis. 1975. "Family Resemblance Studies in the Internal Structure of Categories." *Cognitive Psychology* 7:573–605.

Roschelle, J., and J. G. Greeno. 1987. "Mental Models in Expert Physics Reasoning." Technical Report No. GK-2. University of California, Berkeley.

Sellars, W. 1965. "Scientific Realism or Irenic Instrumentalism." In *Boston Studies in the Philosophy of Science 2*, ed. R. Cohen and M. Wartofsky, pp. 171–204. Dordrecht: D. Reidel.

Settle, T. 1961. "An Experiment in the History of Science." *Science* 133:19–23.

Siegel, D. 1986. "The Origin of the Displacement Current." *Historical Studies in the Physical Sciences* 17:99–145.

Simon, H. A., P. W. Langley, and G. L. Bradshaw. 1981. "Scientific Discovery as Problem Solving." *Synthese* 47:1–27.

Smith, E., and D. Medin. 1981. *Concepts and Categories.* Cambridge, Mass.: Harvard University Press.

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

Tversky, B. 1990. "Induced Pictorial Representations." Unpublished manuscript.

Tweney, R. D. 1985. "Faraday's Discovery of Induction: A Cognitive Approach." In *Faraday Rediscovered*, ed. D. Gooding and F. A. L. James, pp. 189–210. London: Macmillan.

———. 1987. "What Is Scientific Thinking?" Unpublished manuscript.

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

Wise, N. 1979. "The Mutual Embrace of Electricity and Magnetism." *Science* 203:1313–18.

Wiser, M., and S. Carey. 1983. "When Heat and Temperature Were One." In *Mental Models*, ed. D. Gentner and A. L. Stevens, pp. 267–98. Hillsdale, N.J.: Lawrence Erlbaum.