# Model-Based Reasoning in Conceptual Change

Nancy J. Nersessian
*School of Public Policy and College of Computing, Georgia Institute of Technology, Atlanta, GA, USA, nancyn@cc.gatech.edu*

Abstract: This paper addresses how specific modeling practices employed by scientists are productive methods of conceptual change in science. Within philosophy, where the identification of reasoning with argument and logic is deeply ingrained, these practices have not traditionally been considered significant forms of scientific reasoning. Embracing these modeling practices as "methods" of conceptual change in science requires expanding philosophical notions of scientific reasoning to encompass forms of creative reasoning. I focus on three forms of model-based reasoning demonstrated in my previous work as generative of conceptual change in science: analogical modeling, visual modeling, and thought experimenting. The models are intended as interpretations of target physical systems, processes, phenomena, or situations. The models are retrieved or constructed on the basis of potentially satisfying salient constraints of the target domain. In the modeling process, various forms of abstraction, such as limiting case, idealization, generalization, generic modeling, are utilized. Evaluation and adaptation take place in light of structural, causal, and/or functional constraint satisfaction. Simulation can be used to produce new states and enable evaluation of behaviors, constraint satisfaction, and other factors.

## 1. INTRODUCTION

The problem of how new concepts, in general, are formed and how they related existing concepts has vexed philosophers for centuries. The problem, with respect to science, came to the forefront of concerns in late 19$^{th}$ century, with the development of new concepts like "energy" and "field" and became critical with the radical conceptual changes in physics early in this century. On what we have come to know as "the standard view" – that associated with logical positivism – conceptual structures are treated as languages and the problem of conceptual change is that of explaining the nature of the logical and interpretive relations between the old and new conceptual structures and between concepts and the world. Possible "methods" of concept formation, as discussed, e.g. in Carl Hempel's classic monograph (1952), included setting up correlations with empirical phenomena via definition and forms of operationalization. Arguably, the spirit of this project had continued through more recent and innovative work, such as Thomas Kuhn's notion of incom-

*Model-Based Reasoning in Scientific Discovery*, edited by L. Magnani, N.J. Nersessian, and P. Thagard, Kluwer Academic/Plenum Publishers, New York, 1999

mensurability as "untranslatability" and his later approach to the problem through analysis of the lexicon of a scientific theory (1991).

In contrast to the standard approach, my approach in constructing a theory of concept formation and change in science has been to shift focus of analysis from the products of conceptual change to the processes whereby such change comes about. This shift in focus highlights the nature of the scientific practices, specifically, of the *methods* or kinds of *reasoning* through which concepts are constructed. We will focus on three such practices that have been, as claimed in my previous work (1984, 1988, 1992, 1993), generative of conceptual change in science: creating analogies, employing visual representations, and thought experimenting. On my account, what these practices have in common is that they are all forms of "model-based reasoning". What this means will be developed in the course of the paper.

Examining the historical records of major conceptual changes establishes these practices to be employed across the sciences. Since my analysis draws from physics, I will not here claim to be providing an account of how they function in other sciences. However, the hypothesis that there are significant commonalities among these practices as employed across the sciences garners support from the *cognitive* foundations indicated for them here. Specifically how they might function in other domains, though, needs to be developed by researchers with expertise in those domains. Further, although these practices are ubiquitous they are not exhaustive. Clearly, there are other modes of reasoning that generate conceptual change.

Our concern here is to determine how these specific forms of what I have called "model-based reasoning" are productive *methods* of conceptual change. From the perspective of traditional philosophy of science, the modeling practices discussed here have not been considered significant forms of scientific reasoning – in some cases, not as "reasoning" at all. Philosophical accounts of scientific reasoning have restricted the notion of reasoning primarily to deductive and inductive arguments. The practices investigated here have been understood as performing an ancillary and inessential role, with the productive reasoning being carried out by some other means. The position taken in this paper is that by developing a cognitive foundation for modeling practices as productive forms of reasoning, more general and widely applicable than the specific contexts in which they are found to be employed in science, one can mount a case for how they are extremely productive forms of reasoning in conceptual change in science.

## 2. THE TRADITIONAL VIEW OF REASONING

Embracing these modeling practices as the reasoning through which conceptual change can take place requires expanding philosophical notions of scientific reasoning to encompass forms of creative reasoning, most of which cannot be reduced to an algorithm in application, are not always productive of solutions, and can lead to incorrect solutions. To do this requires challenging one of the most sacrosanct notions in philosophy: "reasoning". In the traditional view, the identification of reasoning with argument, and thus with logic, is deeply ingrained. So, before we can develop a notion of "model-based reasoning" in science, we need first to address the question "What is reasoning?" and, specifically, "What is scientific reasoning?"

In standard philosophical accounts reasoning is employing deductive or inductive algorithms to sets of propositions. The understanding of deductive reasoning provided by classical logic stands as the model. Here the essential notion is soundness: true premises plus good reasoning yields true conclusions. A major objective of logical positivism in this century was to develop a notion of soundness for induction similar to that for deduction, to be applied in inductive justification of rationally reconstructed scientific reasoning. Loosely construed, starting from maximally probable premises and using correct inductive logic one should arrive at maximally probable conclusions.

Note that I have switched to talking about "justification" and "rationally reconstructed scientific reasoning" in discussing induction. On extending the traditional notion of reasoning to what goes on in the domain of scientific discovery, in which lies conceptual innovation, a problem arises immediately. Good reasoning, with *T* premises can lead to incorrect solutions or to no solution at all. For example, Newton's path to the concept of universal gravitational force was largely through analogy. Analogical argument is a notoriously weak form of argument and one could hold that concerns about it have been borne out in this case. According to the general theory of relativity, that conclusion is wrong. There is no gravitational force, falling bodies are just following their natural trajectory in a curved space-time. But, of course, we know this could prove to be wrong as well. The problem of unsoundness has been a factor in the contention of philosophers of various persuasions that there is no "logic of discovery". Some nontraditional philosophical accounts have allowed for the possibility of "abductive" inference, but these accounts leave mysterious the nature of the reasoning *processes* underlying abductive inference and hypothesis generation. Analyzing modeling practices provides a way of specifying the nature of some abductive reasoning processes.

A quick review establishes that, in traditional philosophical accounts, what I am calling model-based reasoning practices are considered ancillary, inessential aids to thinking. At most they have constituted – and continue to constitute – fringe topics in the literature in philosophy of science. Analogy has received the most attention and there have been some attempts to countenance it as a form of inductive reasoning. To take a classic example, in Rudolph Carnap's *Logical Foundations of Probability* (1950) analogy occupies an entire page in the Appendix. He concludes that "reasoning by analogy, although admissible, can usually yield only weak results." Even those who have taken analogy in science as a serious topic of discussion have not given made a strong case for its generative role in conceptual change. For Norman Campbell (1920) and Mary Hesse (1966) analogy is constitutive of theories in that analogies provide meaning for new theoretical terms. Thus analogy serves an explanatory function and provides some basis for hypothesis generation. Hesse's analysis of analogies treats them as propositional and focuses on the nature of the arguments enabled by specific mappings of properties and relations. Visual representation has not fared even this well, since most of the discussion from the late $19^{th}$ century until quite recently has concurred that such representations are fundamentally misleading in reasoning processes and have such a detrimental effect that it is better to eliminate them entirely, as, e.g., David Hilbert did in axiomatizing geometry.

Finally, traditional empiricist accounts of thought experiment have been weighted in favor of their being eliminable from the reasoning process. Pierre Duhem (1902), for example, argued that they can either be transformed into real experiments or are simply bogus. A modern version of this view has been developed by John Norton (1991) who argues that a thought experiment can be reformulated as and replaced by a deductive or inductive argument. Ernst Mach (1905) is one empiricist who did take them seriously and his naturalized account moves in right direction. Thomas Kuhn did see thought experiments as playing a central role in conceptual change and characterized them as "one of the essential analytical tools which are deployed during crises and which then help to promote basic conceptual reform" (1964, p. 263). The historical record does indeed show the preponderance of thought experiments in periods of conceptual change in science. But, to understand why, i.e. how thought experiment functions, requires a fundamental recasting of the problem of conceptual change that rejects both traditional and Kuhnian accounts of it.

I agree with the positivists' conclusion that there is no *logic* (classical) of scientific discovery, but disagree with equating reasoning with applying logic to sets of propositions. Rather, I propose to start by acknowledging that such uses of analogy as Newton's are instances of quite powerful *reasoning* and attempt to determine how this kind of reasoning works. The essence of my point is that the positivists were right that history and psychology are essential components to understanding the processes of conceptual change and they were right that there is no logic of discovery, but their notion of reasoning was too narrowly constrained and that led to the mistaken view there are no issues of importance to philosophy in the context of discovery. In the case at hand, the problem of the nature of the constructive processes does bear on the philosophical problem of the nature of reasoning. Further, discovery processes have implications for other traditional philosophical issues that will not be addressed here, such as, realism, objectivity, rationality, and progress.

So, the problem thus far posed is whether is it possible to articulate a notion of reasoning that includes forms that other than deductive and inductive argument and yet fruitful in that they generate potential solutions to scientific problems. Can we ascertain the kind that will come up with good results, advance new and promising solutions? Is there a way we can evaluate as good or productive, reasoning that which can lead to incorrect or no solutions? With respect to the problem of conceptual change, asking these questions shifts focus from the products, or conceptual structures, to the processes of constructing concepts, which, in turn, puts the focus on the practices and practitioners of science. Although I cannot rehearse in this paper the various arguments and justifications for philosophical naturalism, my own brand calls for the necessity of a multidisciplinary analysis of conceptual change in science: an account that is informed by the constructive practices scientist actually use and by the best scientific accounts of how human cognitive capacities and limitations produce and constrain these practices.

## 3. THE COGNITIVE SCIENCE NOTION OF REASONING BY MENTAL MODELING

The first task is to provide a cognitive foundation for taking the modeling practices employed by scientists seriously as forms of reasoning productive of conceptual change in science. In cognitive psychology there is a controversy about the nature of human reasoning that parallels the philosophical issue. This is not surprising, since philosophers of psychology who hold the traditional view have played a significant role in the shaping the contemporary debate. On the traditional psychological view, reasoning consists of applying a mental logic to propositional representations. Critics of this view contend that a purely syntactical account of reasoning cannot account for significant effects of semantic information in reasoning. An early and now classic experiment is the Wason card task (cf. Johnson-Laird, 1983, for a discussion of it and subsequent research). In the task, the subject is presented with four cards, two displaying a letter (one consonant and one vowel) and two displaying a number (one even and one odd). Such a sequence might be: A G 2 5. The subject has been told that each card has a number on one side and a letter on the other. The task is to determine how many cards have to be turned over in order to make a given generalization true, e.g. "if a card has a vowel on one side then it has an even number on the other side." Most subjects failed to make the correct inference with this form of the test. However, when meaningful information and relations were provided on the cards, the number of correct inferences increased dramatically. For example, simply substituting destinations and modes of transportation or postage stamp costs and sealed or unsealed envelopes for the letters and numbers increased people's ability to reason correctly.

This result suggested that semantic information plays a more salient role in human reasoning than the traditional view allows. It sparked numerous investigations in different domains that have led many cognitive scientists to conclude that much of human reasoning is by means of "mental modeling" rather than through a process of applying a mental logic to propositional representations. The initial hypothesis of mental modeling as a significant form of reasoning derives from a proposal made by Kenneth Craik (1943). Craik proposed that, in general, people reason by carrying out thought experiments on internal models. Since he made this proposal at the height of the behaviorist approach in psychology, the hypothesis did not receive much attention. The development of a cognitive psychology in the 1960's created a more hospitable environment for its articulation and exploration. Though not uncontroversial, the centrality of mental modeling to cognition is a hypothesis under investigation by many domains. The main impetus for the resurgence of the hypothesis is experimental outcomes, such as the Wason experiments described above, that demonstrate the inadequacy of syntactic accounts of reasoning. Mental modeling has been investigated in a wide range of phenomena, including: reasoning about causality in physical systems (cf., e.g., deKleer and Brown, 1983); the role of representations of domain knowledge in reasoning (cf., e.g., Gentner and Stevens, 1983); analogical reasoning (cf., e.g. Gentner and Gentner, 1983); deductive and inductive inferencing (cf., e.g., Johnson-Laird, 1983); and comprehending narratives (cf., e.g., Perrig and Kintsch, 1985). Because the potential range of application is so extensive, some have argued that the notion of mental models can provide a unifying framework for the study of cognition (Gilhooly, 1986). For our problem, too, the hypothesis is attractive, in part, because it opens the possibility of furnishing a unified analysis of the widespread modeling practices implicated in conceptual change.

There are several distinct theoretical aspects of mental modeling that tend to be conflated in the literature. The most significant distinction for our purposes is between those investigations that treat mental models as structures that are stored in long term memory and are then called upon in reasoning and those that treat them as temporary structures constructed in "working memory" for a specific reasoning task. This distinction itself may be artificial given that it rests on the questionable notion that long-term memory is like a storage bin and is a different structure from working memory, where both may be dynamic in nature and related processes in the brain. Our analysis of modeling practices focuses on the processes of constructing models and reasoning through manipulating them. Philip Johnson-Laird's account (1983) remains the best articulated of those analyses that focus on the temporary reasoning structure, and it informs my analysis. However, Johnson-Laird's own focus has been on mental modeling in deductive and inductive reasoning tasks and not in creative scientific reasoning. Thus, my account provides an extension of the hypothesis into this domain.

Broadly construed, for Johnson-Laird, a mental model is a structural analog of a real-world or imaginary situation, event, or process that the mind constructs in reasoning. What it means for a mental model to be a structural analog is that it embodies a representation of the spatial and temporal relations among, and the causal structures connecting the events and entities depicted and whatever other information that is relevant to the problem solving task. Most of Johnson-Laird's published work has investigated mental modeling in deductive and inductive reasoning. To accommodate scientific reasoning, we need to include models that are dynamical in nature and to expand the notion to include mental models that are functional analogs to specific dimensions and behaviors of real-world systems. Although Johnson-Laird has not investigated such cases himself, he has hypothesized that since mental models need to be causally coherent, it should be possible to carry out simulative reasoning about the behaviors of a model for those tasks that are dynamic in nature. Such models would behave in accord with constraints that need not be stated explicitly.

Advocates of mental modeling argue that the original capacity developed as a way of simulating possible ways of maneuvering within the physical environment. It would be highly adaptive to possess the ability to anticipate the environment and possible outcomes of actions, so it is likely that many organisms have the capacity for mental modeling. Given the linguistic abilities of humans, it is likely they can create models from both perception and description. This hypothesis receives support from the research in narrative comprehension noted earlier. Although the original ability to perform simulative reasoning through mental modeling may have developed as a way of anticipating possible courses of action in the world, it is highly plausible that, as human brains have developed, this ability has been extended to more esoteric reasoning contexts, such as science. Further, differences in novice and expert reasoning skill in solving scientific problems (cf., e.g., Chi, et al., 1981) can be taken to provide evidence that skill in modeling is something that develops with learning (Ippolito and Tweney, 1995; Nersessian, 1995). The nature and richness of models one can construct and one's ability to reason develops as one learns domain-specific content and techniques.

Just what *format* a mental model takes, i.e., whether it is propositional, perception-based, or non-propositional but still amodal, is a matter of some debate, as is the issue of what are the *generative processes* in the brain for creating and operating on mental models. However, these issues do not have to be resolved before it is possible to make progress on an account of model-based reasoning in science. The essential points are that a mental model can be non-linguistic in form and the mental mechanisms are such that they can satisfy the model-building and simulative constraints necessary for the activity of mental modeling. My own cognitive-historical hypothesis is that the models are analog and non-propositional (though possibly including labeling, i.e., "mixed-mode") in nature, though at present my account is agnostic between perception-based and amodal accounts. To allay possible objections it should be stressed that mental modeling, even if it were to make use of perceptual mechanisms, would not need to be identical to the process of constructing a picture in the "mind's eye". That great thought experimenters, such as Bohr, have claimed not to be able to visualize well does not undermine non-propositional forms of mental modeling. This form of mental modeling would only require the ability to reason by means of an analog model. The relationship between a mental model and what has been called "mental imagery" is something that still needs to be worked out by cognitive scientists, but is not necessary for our purposes.

To carry out an analysis of model-based reasoning in conceptual change requires only that we adopt a "minimalist" version of a mental modeling hypothesis: that in certain problem solving tasks humans reason by constructing an internal model of the situations, events and processes that in dynamic cases provide the basis for simulative reasoning. Whatever the format of the model itself, information in various formats, including linguistic, formulaic, visual, auditory, kinesthetic, can be used in its construction. Although there is still much to learn about mental modeling, this minimalist hypothesis is attractive because it provides a cognitive foundation for taking seriously the modeling practices of scientists as the reasoning through which new conceptual structures are constructed. Further, it provides a basis for constructing a unified account of the various forms of model-based reasoning we will consider. Without going into the details, Figure 1 exemplifies why a unified account is needed. The figure is taken from a paper published in *Philosophi-*

*cal Magazine* by James Clerk Maxwell, in which, I have argued (Nersessian, 1992), he was trying to get his colleagues to understand the new representation of electromagnetic forces by leading them through processes he felt were salient to his constructing the mathematical representation of the electromagnetic field concept. As I interpret it, it is a *visual* representation of an *analogical* model that is accompanied with instructions for *animating* it correctly in thought: "Let the current from left to right commence in AB. The row of vortices *gh* above AB will be set in motion in the opposite direction to a watch [...]. We shall suppose the row of vortices *kl* still at rest, then the layer of particles between these rows will be acted on by the row *gh* on their lower sides and will be at rest above. If they are free to move, they will rotate in the negative direction, and will at the same time move from right to left, or in the opposite direction from the current, and so form an *induced* electric current" (1890, v. 1, p. 477, italics in original).

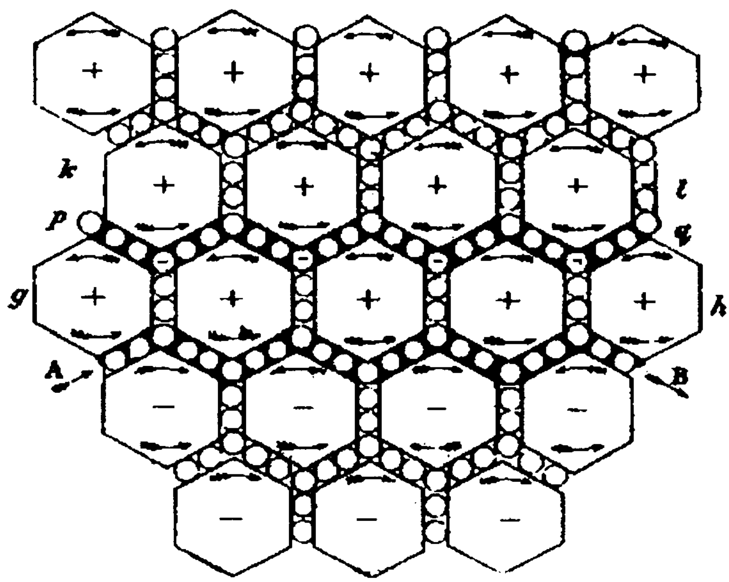


*Figure 1.* Maxwell's drawing of the vortex-idle wheel model (Maxwell 1890, Vol. I, Plate VII).

## 4. MODEL-BASED REASONING

The traditional account of reasoning has supported the belief that conceptual change cannot be the outcome of reasoned processes. By and large, philosophers consider the processes of conceptual change as mysterious and unanalyzable. Conceptual innovation is held to occur in sudden flashes of insight, with new concepts springing forth from the head of the scientist like Athena, fully grown. This does accord with retrospective accounts of some scientists, but if one examines their deeds – their papers, diaries, letters, notebooks – these records support a quite different interpretation in most cases. As I have been arguing for some years, conceptual change results from extended problem-solving processes. The records of these processes display extensive use of practices that I hypothesize constitute forms of model-based reasoning: analogical, visual, and simulative modeling. Further, the psychological theory of mental modeling provides a basis in human cognition for taking the external traces of modeling displayed in scientific practice as indicative that model-based reasoning is generative of concept formation and change.

The argument developed above in support of my hypothesis is that mental modeling is a fundamental form of human reasoning. It evolved as an efficient means of navigating the environment and solving problems in matters of significance to existence in the world. Humans have extended its use to more esoteric situations, such as constructing scientific representations. That is, the cognitive resources scientists call upon on in creative problem solving are not different in kind than those humans use in more ordinary circumstances. Mental modeling is applied by humans across a spectrum of problem solving situations and in numerous domains, ranging from solving the problem of how to get a chair through a doorway to problems traditionally classified as deductive and inductive logic problems. Additionally, there is significant experimental protocol evidence collected by cognitive psychologists to support it as a fundamental form of problem solving employed by contemporary scientists (cf. e.g., Chi, et al., 1981; Clement, 1989). These records of "think-aloud" reasoning processes provide additional support for the claim that the traces of modeling practices exhibited in the historical records of conceptual change are indicative that mental modeling played a central role in the historical process. Understood from the perspective of human cognition, these practices utilize and engage internal modeling processes that are highly effective means of problem solving and effective means of transmitting novel conceptions through a community. In the more mundane cases the reasoning performed is usually successful, e.g., one figures out how to get the chair through the door, because the models and manipulative processes embody largely correct assumptions about every-day real-world events. In the case of science where the situations are more removed from experience and the assumptions more imbued with theoretical assumptions, there is less assurance that a reasoning process, even if correct, will yield "success". In the evaluating process, a major criterion for success remains the goodness of fit to the phenomena, but success can also include such factors as enabling the construction of a viable mathematical representation.

The centrality of model-based reasoning practices in episodes of conceptual change, a creative form of problem solving, lends support to the position of several contemporary philosophers (cf., e.g., Cartwright, 1989;

Giere, 1988) that the basic units for the scientists in reasoning more generally are most often not axiom systems and propositional networks, but models. As in the theory of mental modeling, the term "model" is used here not in the logical sense of an abstract mapping of things to terms, but in the analogical sense of a structure intended as isomorphic to some aspect of a physical system. In using and constructing scientific theories, no matter how they may in principle be represented, models are the mental representations with which a scientist carries out much reasoning and by means of which she thinks and understands through the lens of a conceptual structure. In the constructive processes of conceptual change, specifically, one important lesson we should take from the historical records is that models come first, then further abstraction takes place to create formal expression in laws and axioms of theories (Nersessian, 1995, 1999). Modeling practices are employed both in experimental and in theoretical settings. The model is the mode of representation between the phenomena and expression in a language (including mathematics) and it is working with this intermediate form of representation that facilitates conceptual change. Further, there is significant evidence that much of the training of the practitioners of science takes place through learning the models of a community and developing facility with model manipulation (Clement, 1989; Giere, 1988; Nersessian, 1995). A cognitive account, then, raises the modeling practices from their traditional status of ancillary, inessential aids to reasoning to the actual forms of reasoning through which concept formation and change take place. We will now consider, briefly, how such model-based reasoning functions in conceptual change. The practices will be considered separately here, but they are all related forms of reasoning and they most often are used together in reasoning episodes I have elsewhere called "constructive modeling" (Nersessian, 1995; Griffith, et al., 1996).

## 5. THREE FORMS OF MODEL-BASED REASONING

To engage, specifically, in analogical modeling one calls on knowledge of the generative principles and constraints for physical models in a source domain. These constraints and principles may be represented mentally in different informational formats and knowledge structures that act as tacit assumptions employed in constructing and transforming models during problem solving. Inter- or intra-domain analogies may be retrieved and applied as models wholesale with suitable adaptation, but often, and especially in cases of conceptual change, no direct analogy exists and construction of an initial source model is required. In these cases the analogical domain serves as the source for constraints to be used in interaction with those provided by the target problem to create imaginary analogs. Evaluation of the analogical modeling process is in terms of how well the salient constraints of a model fit the salient constraints of a target problem.

Generic abstraction is a key reasoning process in analogical modeling, which often requires recognition of potential similarities across, and integration of information from, disparate domains. In viewing a model generically, one takes it as representing features common to a class of phenomena. This way of viewing the model can, of course, only take place in the mind. In reasoning, e.g., about a triangle, one often draws or imagines a concrete representation. However, to consider what it has in common with all triangles, one needs to imagine it as lacking specificity in the angles and the sides. That is, the reasoning context demands that the interpretation of the concrete polygon as generic. It was only through generic abstraction, e.g., that Newton could reason about the commonalities among the motions of planets and of projectiles, which enabled his formulating a unified mathematical representation of their motions.

The analogical model, understood generically, represents what is common among the members of specific classes of physical systems, viewed with respect to a problem context. Newton's inverse-square law of gravitation abstracts what a projectile and a planet have in common in the context of determining motion. The inverse-square-law model served as a generic model of action-at-a-distance forces for those who tried to bring all forces into the scope of Newtonian mechanics. In Maxwell's analogical model, represented visually in Figure 1, the dynamical relations among the idle wheels and vortices must be viewed in generic form. That is, the vortex-idle wheel system is understood to represent the class of such dynamical systems and the class includes electric and magnetic interactions on the assumptions of Maxwell's model. Although constructed for the purposes of showing how contemporary scientists use theories, I think the chart constructed by Ronald Giere (1994), shown in Figure 2, can also be interpreted as illustrating how the process of generic abstraction functioned in the process of constructing the modern system of classical mechanics. Starting from specific instances of pendula and springs, generic categories are created and are further abstracted to kinds of "harmonic motion", which then fall into generic categories of "conservative" or "non-conservative" models, and so forth.

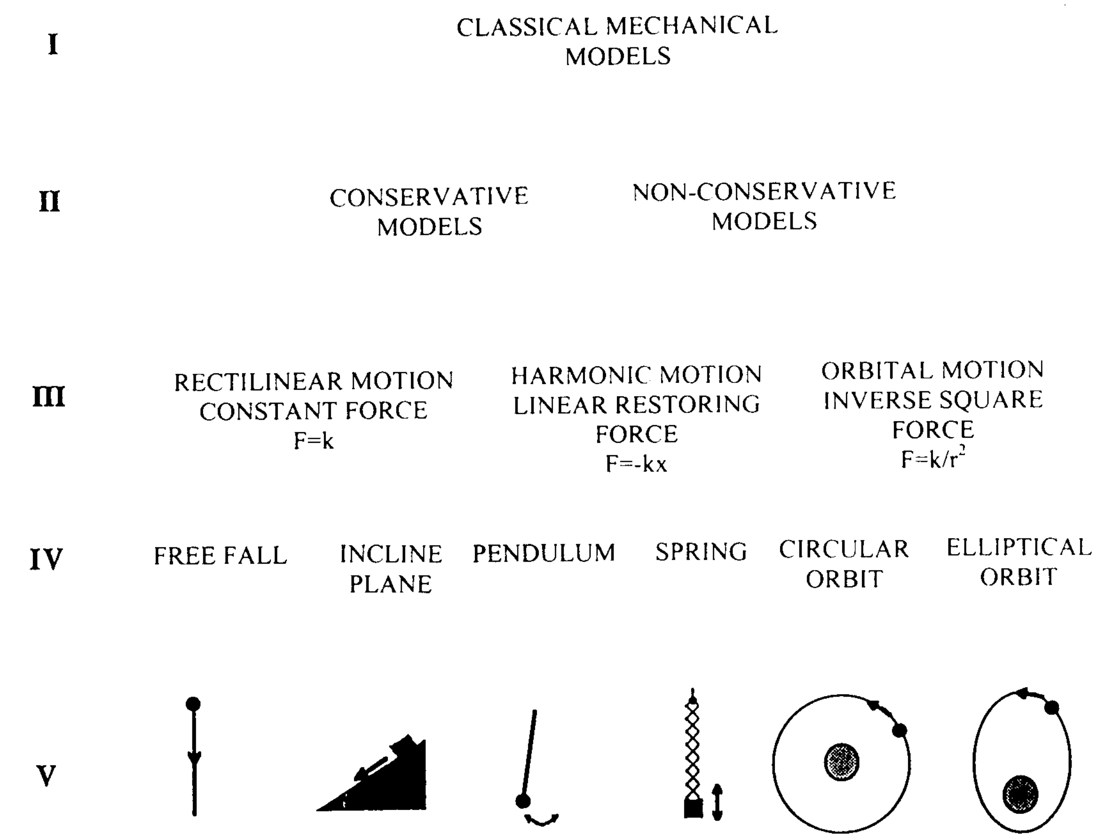


*Figure 2.* Illustration of the process of abstraction through generic modeling in classical mechanics.

Scientists use a variety of perceptual resources in modeling. These vary with the science, e.g., as Cameron Shelley (1996) has shown, an archeologist would make more use of kinesthetic information in model construction than a physicist. Both internal and external visual representations figure prominently in conceptual change across the science. A possible reason for this is that employing the visual modality may enable the reasoner to bypass specific constraints inherent in the linguistic and formulaic representations of existing conceptual structures. External visual representations provide support for the processes of constructing and reasoning with a mental model. They aid significantly in organizing cognitive activity during reasoning, such as fixing the attention of the salient aspects of a model during reasoning, enabling retrieval and storage of salient information and exhibiting salient interconnections, such as structural and causal, in appropriate co-location. Further external visual representations, such as the figure by Maxwell (Figure 1), facilitate the construction of shared mental models in a community and the transportation of a model out of the local milieu of its construction. The mental models perspective hypothesizes that the external visual representations support construction of an internal model. Internal visual representations need not be "pictorial" in format, but can be highly schematic.

Thus this modality may be operative even in the reasoning of scientists, such as Bohr, who claim not to experience imagery in reasoning.

As used in modeling in physics, external visual representations tend to be in diagrammatic form. These representations can model phenomena in several ways, including providing idealized representations of aspects of phenomena and embodying aspects of theoretical models. For example, early in Faraday's construction of a field concept the visual model represented in Figure 3 provided an idealized representation of the lines of force surrounding a magnetic bar.

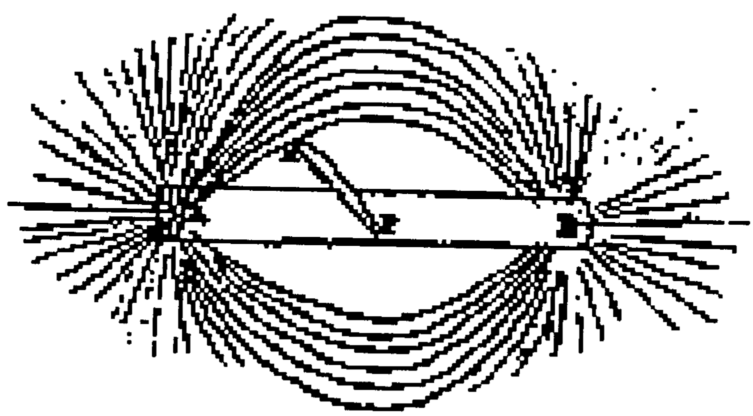

*Figure 3.* Faraday's drawing of the lines of force surrounding a bar magnet (Faraday 1839-1855, Vol I, Plate I).

Later in his problem solving, the visual model of lines of force functioned as the embodiment of a dynamical theoretical model of forces generally (Gooding, 1990; Nersessian, 1984, 1992). But, the visual model represented by Maxwell in Figure 1 is an embodiment of an imaginary system and not a theoretical model of electromagnetic actions.

There is a vast literature on mental imagery (Kosslyn, 1994) that provides evidence that humans can perform simulative imaginative combinations and transformations that mimic perceptual spatial transformation. These simulations are hypothesized to take place using internalized constraints assimilated during perception. Cognitive research also indicates that people use various kinds of knowledge of physical situations in imaginary simulations. For example, when objects are imagined as separated by a wall, the spatial transformations exhibit latency time consistent with having simulated moving around the wall rather than through it.

Thought experimenting is a specific form of the simulation that can occur in various forms of model-based reasoning. Because the thought-experimental narratives are what we have access to and because they are a central form of effecting conceptual change within a scientific community, my analysis begins with examining how these function. From that analysis one can infer that the original experiment involves a similar form of reasoning. To explicate the notion that thought experimenting is simulative model-based reasoning, we need to discuss: (1) how a narrative facilitates the construction of a model of an experimental situation in thought and (2) how one can reach conceptual and empirical conclusions by mentally simulating the experimental processes.

From a mental modeling perspective, the function of the narrative form of presentation of a thought experiment would be to guide the reader in constructing a structural analog of the situation described by it and to make inferences through simulating the events and processes depicted in it. So, extending from research on other forms of discourse models (cf., e.g., Perrig and Kintsch, 1985), the operations carried out in executing the thought experiment are performed not on propositions but on the constructed internal model. Unlike the fictional narrative, however, the context of the scientific thought experiment makes the intention clear to the reader that the situation is one that is to represent a potential real-world situation. That a thought experiment is presented in a polished form should make it an effective means of getting comparable mental models among the members of a community of scientists. The narrative has already made significant abstractions that aid in focusing attention on the salient dimensions of the model and in recognizing the situation as prototypical, so that the experimental consequences are understood to go beyond the specific situation of the thought experiment.

Although some kinds of mental modeling may employ static representations, those derived from thought-experimental narratives are inherently dynamic. The narrative delimits which are the specific transitions that govern what takes place. Constructing and conducting the experiment makes use of inferencing mechanisms, existing representations, and scientific and general world knowledge to make realistic transformations from one possible physical state to the next. Much of the information employed in these transformations is tacit. Thus, expertise and learning play a crucial role in the practice; as does what Gooding (1990) has called "embodiment". The constructed situation inherits empirical force by being abstracted both from our experiences and activities in the world and our knowledge, conceptualizations, and assumptions of it. In this way, the data that derive from thought experimenting have empirical consequences and at the same time pinpoint the locus of the needed conceptual reform. This understanding forms the basis of further problem-solving efforts to construct an empirically adequate conceptualization.

Thought experimenting plays a crucial role in conceptual change by showing that existing systems of constraints cannot be integrated into consistent models of the physical world. Thought experimenting may facilitate recognizing the undesirable consequences of a conceptualization in much the way that experimenting by computer simulation exposes undesirable consequences of the constraints of a scientific representation. By creating a simulative model that attempts to integrate specific systems of constraints, thought experimenting enables the scientist to grasp essential points of conflict and infer their consequences more readily than would reasoning through the logical consequences of a representation. Once the initial experimenter understands the implications of a thought experiment, she can guide others in the community to see them as well by crafting a description of the experiment into a narrative.

## 6. CONCLUSIONS

Assuming that analogy is at best a form of inductive argument leads to the conclusion of Carnap that "reasoning by analogy can yield only weak results" (1950, p. 589). Looking at examples of productive reasoning by analogy in science shows it to yield powerful and creative results. The way to resolve the discrepancy is to see that reasoning by analogy is not argument but model construction. So, although analogical modeling enables arguments, the heart of analogy is employing generic abstraction in the service of model construction, manipulation, and evaluation. In this, as opposed to the standard "argument by analogy," it is a powerful form of reasoning.

Visual representations may indeed have the potential to lead a reasoner astray. However, visual modeling appears to be highly developed and effective form of human reasoning in a wide variety of circumstances. Duhem was a particularly vocal opponent of visual representation in science. He went so far as to use the late $19^{th}$ Century British proclivity for using visual representation and analogy as demonstrating the inferiority of the mind of the British scientist as compared with the analytical mind of the French scientist. The great irony, of course, is that the British produced the desired unified theory of electromagnetism, and not the French. Visual representation is powerful tool for science when sufficient constraints are incorporated into the reasoning process. Although many thought experiments can often be reconstructed as arguments, their modeling function cannot be supplanted by an argument. The argument is not evident until after the thought experiment has been constructed and executed. Exhibiting the soundness of a thought experiment by reconstructing it as an argument can perform an important rhetorical function. However, real-world experimental outcomes can be recast in argument form as well, but no one would argue that the experiment can be replaced by the argument. In similar fashion, we need to differentiate between the reasoning that is done with the thought experiment and that which is done with the reconstruction of it. The thought-experimental process, by linking the conceptual and the experiential dimensions of human cognitive processing, demonstrates the undesirable real-world consequences of a representation, thereby compelling representational change.

All three forms of model-based reasoning are complex forms of reasoning that integrate various forms of information – propositions, models, and equations – into mental models. There are several key common ingredients to the various forms of model-based reasoning we have considered. They are semantic reasoning processes in that the models are intended as interpretations of a target physical system, process, phenomenon, or situation. The models are retrieved or constructed on the basis of potentially satisfying salient constraints of the target domain. In the modeling process, various forms of abstraction, such as limiting case, idealization, generalization, generic modeling, are utilized. Evaluation and adaptation take place in light of structural, causal, and/or functional constraint satisfaction and enhanced understanding of the target problem through the modeling process. Simulation can be used to produce new states and enable evaluation of behaviors, constraint satisfaction, and other factors.

## REFERENCES


Cartwright, N., 1989, *Nature's Capacities and Their Measurement*, Clarendon, Oxford.

Chi, M.T.H., Feltovich, P.J., and Glaser, R., 1981, Categorization and representation of physics problems by experts and novices, *Cognitive Science* 5:121-52.

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

Clement, J., 1989, Learning via model construction and criticism, in: *Handbook of Creativity: Assessment, Theory, and Research*, G. Glover, R. Ronning, and C. Reynolds, eds., Plenum, New York, pp. 341-81.

Craik, K., 1943, *The Nature of Explanation*, Cambridge University Press, Cambridge.

de Kleer, J., and Brown, J.S., 1983, Assumptions and ambiguities in mechanistic mental models, in: D. Gentner and A.L. Stevens, pp. 155-190.

Duhem, P., 1902, *The Aim and Structure of Physical Theory*, Atheneum, New York (reprinted 1962).

Faraday, M., 1839-1855, *Experimental Researches in Electricity*, Dover, New York (reprinted 1965).

Gentner, D., and Gentner, D.R., 1983, Flowing waters or teaming crowds: mental models of electricity, in: D. Gentner and A.L. Stevens, pp. 99-130.

Gentner, D., and Stevens, A.L., 1983, *Mental Models*, Lawrence Erlbaum, Hillsdale, NJ.

Gilhooly, K.J., 1986, Mental modeling: A framework for the study of thinking, in: *Thinking: Progress in Research and Teaching*, J. Bishop, J. Lochhead, and D. Perkins, eds., Lawrence Erlbaum, Hillsdale, NJ, pp. 19-32.

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

Giere, R.N., 1994, The cognitive structure of scientific theories, *Philosophy of Science* 61:276-296.

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

Griffith, T., Nersessian, N., and Goel, A., 1996, The role of generic models in conceptual change, in: *Proceedings of the Cognitive Science Society* 18, Erlbaum, Hillsdale, NJ, pp. 312-317.

Hempel, C.G., 1952, *Fundamentals of Concept Formation in Empirical Science*, University of Chicago Press, Chicago.

Hesse, M., 1966, *Models and Analogies in Science*, University of Notre Dame Press, Notre Dame, IN.

Ippolito, M.F., and Tweney, R.D., 1995, The inception of insight, in: *The Nature of Insight*, R.J. Sternberg and J.E. Davidson, eds., MIT Press, Cambridge, MA.

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

Kosslyn, S.M., 1994, *Image and Brain*, MIT Press, Cambridge, MA.

Kuhn, T.S., 1964, A function for thought experiments, in: *The Essential Tension: Selected Studies in Scientific Tradition and Change*, University of Chicago Press, Chicago, pp. 240-265

Kuhn, T.S., 1991, The road since structure, in: *PSA 1990*, A. Fine, M. Forbes, and L. Wessels, eds., Vol. 2, PSA, East Lansing, MI, pp. 3-16.

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

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

Nersessian, N.J., 1988, Reasoning from imagery and analogy in scientific concept formation, in: *PSA 1988*, A. Fine and J. Leplin, eds., Vol. 1, PSA, East Lansing, MI, pp. 41-48.

Nersessian, N.J., 1992, How do scientists think? Capturing the dynamics of conceptual change in science, in: *Cognitive Models of Science, Minnesota Studies in the Philosophy of Science 15*, R. Giere, ed., University of Minnesota Press, Minneapolis, pp. 3-44.

Nersessian, N.J., 1993, In the theoretician's laboratory: thought experimenting as mental modeling, in: *PSA 1992*, D. Hull, M. Forbes, and K. Okruhlik, eds., Vol. 2, PSA, East Lansing, MI, pp. 291-301.

Nersessian, N.J., 1995, Should physicists preach what they practice? Constructive modeling in doing and learning physics, *Science & Education* 4:203-226.

Nersessian, N.J., 1999, Abstraction via generic modeling in concept formation in science, in: *Correcting the Model: Idealization and Abstraction in Science*, N. Cartwright and M.R. Jones, eds., Editions Rodopi, Amsterdam, in press.

Norton, J. 1991, Thought experiments in Einstein's work, in: *Thought Experiments in Science and Philosophy*, T. Horowitz and G. Massey, eds., Rowman and Littlefield, Savage, MD, pp. 129-148

Perrig, W., and Kintsch, W., 1985, Propositional and situational representations of text, *Journal of Memory and Language* 24:503-518.

Shelley, C., 1996, Visual abductive reasoning in archaeology, *Philosophy of Science* 63:278-301.