7 The cognitive basis of model-based reasoning in science

Nancy J. Nersessian

The issue of the nature of the processes or 'mechanisms' that underlie scientific cognition is a fundamental problem for cognitive science, as is how these facilitate and constrain scientific practices for science studies. A full theory of human cognition requires understanding the nature of what is one of its most explicitly constructed, abstract and creative forms of knowing. A rich and nuanced understanding of scientific knowledge and practice must take into account how human cognitive abilities and limitations afford and constrain the practices and products of the scientific enterprise. Here I want to focus on the issue of the cognitive basis of certain model-based reasoning practices – namely, those employed in creative reasoning leading to representational change across the sciences. Investigating this issue provides insights into a central problem of creativity in science: how are genuinely novel scientific representations created, given that their construction must begin with existing representations? I will start by considering methodological issues in studying scientific cognition; then address briefly the nature of specific model-based reasoning practices employed in science; and finally provide outlines of an account of their cognitive basis, and of how they are generative of representational change.

1 How to study scientific cognition?

The project of understanding scientific cognition is inherently inter-disciplinary and collaborative. It requires a detailed knowledge of the nature of the actual cognitive practices employed by scientists; knowledge of a wide extent of existing cognitive science research pertinent to explaining those practices, such as on problem-solving, conceptual change and imagery; and employment of the customary range of cognitive science methods used in investigations of specific aspects of scientific cognition. In its approach to studying science, cognitive science has been working under the assumption made by Herbert Simon at its outset: that scientific problem-solving is just an extension of ordinary problem-solving – study the latter and you will understand the former.

This research was supported by National Science Foundation Scholar's Award SES 9810913.

Cognitive practices that take place within the context of doing scientific work have not received much scrutiny by cognitive scientists – in part because of the complexity of scientific work, and in part because the methodological practices of earlier cognitive science did not afford study of scientific cognition in the contexts in which it occurs. Computational analyses of 'scientific discovery' have tended to focus on a small range of computationally tractable reasoning practices gleaned from selective historical cases (Langley *et al.*, 1987). Psychological studies of scientific reasoning are customarily carried out in the context of studies of expert–novice problem-solving (Chi, Feltovich and Glaser, 1981) and protocol analysis of scientists or students solving science-like problems posed to them by cognitive researchers (Clement, 1989; Klahr, 1999). It is only recently, as some cognitive scientists have begun to examine scientific cognition in its own right and in the contexts within which it occurs, that they have begun to see that although it may be an extension, there are features of it that afford insight into cognition not provided by studies of mundane cognition, and that studying scientific cognition could lead to revising how we view mundane problem solving.

Study of scientific cognition has been facilitated by an important methodological shift in the field of cognitive science towards more observational studies conducted in naturalistic settings – that is, the settings in which the cognition under study naturally takes place. For example, Kevin Dunbar (1995, chapter 8 in this volume) has proposed two methods for the cognitive study of science, *in vivo* and *in vitro* studies. *In vivo* studies are those of scientific practices in 'naturalistic' – or real-world – settings, such as research laboratories. These studies employ standard protocol analysis and ethnographic methods. *In vitro* studies, on his account, employ the traditional methods of experimental psychology to investigate how subjects in studies solve authentic discovery problems in the traditional experimental settings. I will extend *in vitro* studies to encompass also 'toy' science problems given to either expert or novice subjects, and computational modelling of scientific discovery processes. Although both *in vivo* and *in vitro* studies provide valuable windows into scientific cognition, they can supply only a partial view of the nature of scientific practice. To obtain a more complete view, findings from another mode of analysis, which (following Dunbar's mode of expression) could be called *sub specie historiae* studies need to be integrated into the analysis of scientific cognition.

Sub specie historiae studies provide the perspective of how scientific practices develop and are used over periods that can extend lifetimes rather than hours and days. These practices can be examined at the level of individuals and at the level of communities. They are set in the context of training, earlier research, the knowledge base, community, collaborators, competitors and various material and socio-cultural resources. The findings derive from examining a multiplicity of sources, including notebooks, publications, correspondence and instruments.

They often involve extensive meta-cognitive reflections of scientists as they have evaluated, refined and extended representational, reasoning and communicative practices.

These studies of past science employ a *cognitive–historical* method (Nersessian, 1992a, 1995). Cognitive–historical analysis creates accounts of the nature and development of science that are informed by studies of historical and contemporary scientific practices, and cognitive science investigations of aspects of human cognition pertinent to these practices. The 'historical' dimension of the method is used to uncover the practices scientists employ, and for examining these over extended periods of time and as embedded within local communities and wider cultural contexts. The 'cognitive' dimension factors into the analysis how human cognitive capacities and limitations could produce and constrain the practices of scientists. Thus the practices uncovered are examined through a cognitive 'lens', i.e. in light of cognitive science investigations of similar practices in both ordinary and in scientific circumstances. The objectives of this line of research are to identify various cognitive practices employed in scientific cognition; to develop explanatory accounts of the generativity of the practices; and to consider, reflexively, the implications of what is learned for understanding basic cognitive processes generally. For example, my own research on conceptual change has centred on using historical case studies to identify candidate generative 'mechanisms' leading to conceptual change in science, to develop an explanatory account of how the reasoning processes employed are generative, and to use this account reflexively in addressing issues pertaining to mundane cognition, such as the nature of visual analogy.

Cognitive–historical analyses make use of the customary range of historical records for gaining access to practices and draw on and conduct cognitive science investigations into how humans reason, represent and learn. These records include notebooks, diaries, correspondence, drafts, publications and artefacts, such as instruments and physical models. The cognitive science research pertinent to the practices spans most cognitive science fields. What research is utilized and conducted depends on the issues that arise in the specific investigation. Dimensions of scientific change amenable to cognitive–historical analysis include, but are not limited to: designing and executing experiments (real-world and thought), concept formation and change, using and inventing mathematical tools, using and developing modelling tools and instruments, constructing arguments, devising ways of communicating and training practitioners.

Underlying the cognitive–historical method is a 'continuum hypothesis': the cognitive practices of scientists are extensions of the kinds of practices humans employ in coping with their physical and social environments and in problem-solving of a more ordinary kind. Scientists extend and refine basic cognitive strategies in explicit and critically reflective attempts to devise methods for understanding nature. That there is a continuum, however, does not rule out the possibility that there are salient differences between scientific and ordinary cognition. Most of the research in cognitive science has been conducted on mundane cognition in artificial contexts and on specific cognitive processes considered largely in isolation from other processes. Further, the point as argued from the perspective of situated cognition about mundane cognition (Greeno, 1998) clearly applies even more strongly to scientific cognition. Scientific 'cognition refers not only to universal patterns of information transformation that transpire inside individuals but also to transformations, the forms and functions of which are shared among individuals, social institutions and historically accumulated artefacts (tools and concepts)' (Resnick, Levine and Teasley, 1991, p. 413). To fathom scientific cognition we must examine it in a contextualized fashion.

The complex nature of scientific cognition forces integration and unification of cognitive phenomena normally treated in separate research domains such as analogy, imagery, conceptual change and decision making. In so doing, investigating scientific cognition opens the possibility that aspects of cognition previously not observed or considered will emerge, and may require enriching or even altering significantly current cognitive science understandings. Thus the cognitive–historical method needs to be reflexive in application. Cognitive theories and methods are drawn upon insofar as they help interpret the historical and contemporary practices, while at the same time cognitive theories are evaluated as to the extent to which they can be applied to scientific practices. The assumptions, methods and results from both sides are subjected to critical evaluation, with corrective insights moving in both directions. Practices uncovered in cognitive–historical investigations can provide a focal point for observational studies and for designing experiments. The point is that all three kinds of investigation are needed to develop an understanding of this complex phenomenon.

2 Model-based reasoning in conceptual change

One aspect of scientific cognition that has received significant attention in the cognitive–historical literature is conceptual change. This form of representational change has also been the focus of much research in history and philosophy of science. This research has established that conceptual innovations in 'scientific revolutions' are often the result of multiple, inter-connected, problem-solving episodes extending over long periods and even generations of scientists. The nature of the specific conceptual, analytical and material resources and constraints provided by the socio-cultural environments, both within and external to the scientific communities in which various episodes have taken place, have been examined for many episodes and sciences. What stands out from this research is that in numerous instances of 'revolutionary' conceptual change across the sciences the practices of analogy, visual representation and thought experimenting are employed. My own historical investigations centre on practices employed in physics (Nersessian, 1984a, 1984b, 1985, 1988, 1992a, 1992b, 1995, 2001a, 2001b), but studies of other sciences by philosophers, historians, and cognitive scientists establish that these practices are employed across the sciences (Rudwick, 1976; Darden, 1980, 1991; Holmes, 1981, 1985; Latour, 1986, 1987; Tweney, 1987, 1992; Giere, 1988, 1992, 1994; Griesemer and Wimsatt, 1989; Gooding, 1990; Lynch and Woolgar, 1990; Griesemer, 1991a, 1991b; Thagard, 1992; Shelley, 1996; Gentner *et al.*, 1997; Trumpler, 1997).

In historical cases, constructing new representations in science often starts with modelling, followed by the quantitative formulations found in the laws and axioms of theories. The same modelling practices often are used in communicating novel results and 'instructing' peers within the community in the new representations. They have been shown to be employed in conceptual change in science in *in vivo* (Dunbar, 1995, 1999a) and *in vitro* studies (Klahr 1999; Clement, 1989), and also in computational studies (Thagard, 1992; Gentner, 1997; Griffith, Nersessian, and Goel 1996, 2001; Griffith, 1999). Although these practices are ubiquitous and significant they are, of course, not exhaustive of the practices that generate new representational structures.

The practices of analogical modelling, visual modelling, and thought experimenting (simulative modelling) are frequently used together in a problem-solving episode. For example, figure 1 is a drawing constructed by James Clerk Maxwell in his derivation of the mathematical representation of the electromagnetic field concept, that provides a visual representation of an analogical model that is accompanied by verbal instructions for simulating it correctly in thought. Such co-occurrences underscore the significant relationships among these practices and have led me to attempt a unified account of them as forms of 'model-based reasoning'. In this chapter I take it as a given that model-based reasoning is generative of representational change in science. The project of the chapter is to determine the cognitive capacities that underlie it, and to provide an explanation of how it is generative.

Within philosophy, where the identification of reasoning with argument and logic is deeply ingrained, these practices have been looked upon quite unfavourably. Traditional accounts of scientific reasoning have restricted the notion of reasoning primarily to deductive and inductive arguments. Some philosophical accounts have proposed abduction as a form of creative reasoning, but the nature of the processes underlying abductive inference and hypothesis generation have largely been left unspecified. Conceptual change has customarily been portrayed as something inherent in conceptual systems rather than as a reasoned process, with the philosophical focus on *choice* between competing systems rather than *construction* of the alternatives. The main problem with embracing modelling practices as 'methods' of conceptual change in science is that it requires expanding philosophical notions of scientific reasoning to encompass forms of creative reasoning, many of which cannot be reduced to an

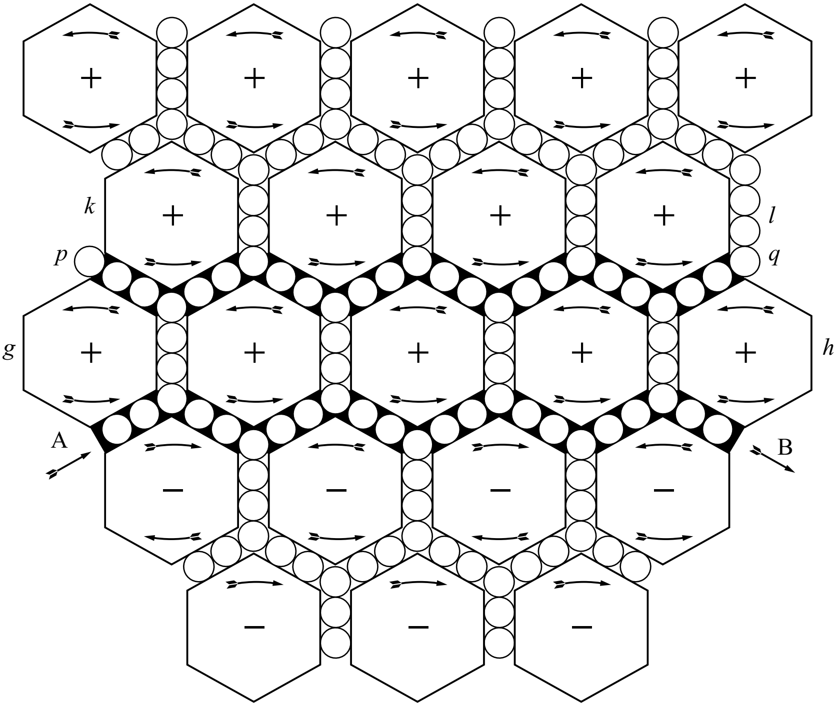

Fig. 1 Maxwell's drawing of the vortex-idle wheel model
Source: Maxwell (1890, 1, plate VII).

algorithm in application and are not always productive of solutions, and where good usage can lead to incorrect solutions. The cognitive–historical approach challenges the *a priori* philosophical conception of reasoning both with historical case studies that serve as counter-examples and as data for a richer account of scientific reasoning, and with cognitive science research that leads to a more expansive notion of reasoning.

3 The cognitive basis of model-based reasoning

Although it is not possible to go into the details in depth within the confines of this chapter, the account of model-based reasoning derives from extensive historical and cognitive research. The historical research includes my own studies – mainly of, but not limited to, nineteenth- and early twentieth-century field physicists – and pertinent research by historians and philosophers of science into other scientific domains and periods, such as noted above. As stated earlier, the nature of these scientific practices is determined by historical research and *in vivo* investigations. These provide the focal points for examining cognitive science research in search of findings that help to explain the cognitive underpinnings of the scientific practices, to formulate hypotheses about why these practices are effective and to discern ways in which the cognitive research might be challenged by the findings from examinations of scientific cognition. The cognitive science research pertinent to model-based reasoning is drawn, primarily, from the literatures on analogy, mental modelling, mental simulation, mental imagery, imagistic and diagrammatic reasoning, expert–novice problem-solving and conceptual change. In this section, a cognitive basis for model-based reasoning in science will be established by considering the representational and reasoning processes underpinning modelling practices. I will first locate my analysis of model-based reasoning within the mental modelling framework in cognitive science. I will then discuss the roles of analogy, visual representation and thought experimenting in constructing new conceptual structures.

3.1 The mental modelling framework

Akin to the traditional philosophical view, the traditional psychological view holds that the mental operations underlying reasoning consist of applying a mental logic to proposition-like representations. The work by Jean Piaget and Barbel Inhelder provides an exemplar of this position in its explicit identification of reasoning with the propositional calculus (Inhelder and Piaget, 1958). For some time critics of this view have contended that a purely syntactical account of reasoning can account neither for significant effects of semantic information exhibited in experimental studies of reasoning, nor for either the logical competence or the systematic errors displayed by people with no training in logic (Wason, 1960, 1968; Johnson-Laird, 1982, 1983; Mani and Johnson-Laird, 1982; McNamara and Sternberg, 1983; Perrig and Kintsch, 1985; Oakhill and Garnham, 1996). Instead, they propose adopting a hypothesis, first put forth by Kenneth Craik (1943), that in many instances people reason by carrying out thought experiments on internal models. In its development within contemporary cognitive science, the hypothesis of reasoning via 'mental modelling' serves as a framework for a vast body of research that examines understanding and reasoning in various domains including: reasoning about causality in physical systems (DeKleer and Brown, 1983); the role of representations of domain knowledge in reasoning (Gentner and Gentner, 1983); logical reasoning (Johnson-Laird, 1983); discourse and narrative comprehension (Johnson-Laird, 1983; Perrig and Kintsch 1985); and induction (Holland *et al.*, 1986). Additionally, there is considerable experimental protocol evidence collected by cognitive psychologists that supports claims of mental modelling as significant in the problem-solving practices of contemporary scientists (Chi, Feltovich and Glaser, 1981; Clement, 1989; Dunbar, 1995, 1999a).

Advocates of mental modelling argue that the original capacity developed as a means of simulating possible ways of manoeuvring 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 modelling from perception. Given human linguistic abilities, it should be possible to create mental models from both perception and description. This hypothesis receives support from research in narrative and discourse comprehension. It is also plausible that, as human brains developed, this ability extended to wider understanding and reasoning contexts, including science. Additionally, the differences in novice and expert reasoning skill in solving scientific problems (Chi, Feltovich and Glaser, 1981) lend support to the possibility that skill in mental modelling develops with learning (Ippolito and Tweney, 1995; Nersessian, 1995). That is, the nature and richness of models one can construct and one's ability to reason develops with learning domain-specific content and techniques. Thus facility with mental modelling is a combination of an individual's biology and learning.

The notion of understanding and reasoning via 'mental modelling' is best considered as providing an explanatory 'framework' for studying cognitive phenomena. There is not a single unitary hypothesis about the specific nature of the format of the representation of a mental model. Further, little is known about the nature of the generative processes underlying the construction and use of mental models. Given the constraints of this chapter it will not be possible to go into the details of various forms of the hypothesis invoked in explanations. Rather we will briefly consider hypotheses about the format of a mental model and discuss the reasoning processes associated with these formats.

In the first place, a mental model is a form of knowledge organization. There are two main usages of the term 'mental model' that tend to get conflated in the literature: (1) a structure in long-term memory (LTM), and (2) a temporary structure created in working memory (WM) during comprehension and reasoning processes. The first usage focuses on how the mental representation of knowledge in a domain is organized in LTM and the role it plays in supporting understanding and reasoning. Numerous studies have led to the claim that the LTM structures representing knowledge in a domain are not organized as abstract, formal structures with rules for application. Rather, it is proposed that knowledge is organized by means of qualitative models capturing salient aspects of objects, situations and processes such as the structure and causal behaviours of various systems in a domain. Mental models are schematic in that they contain selective representation of aspects of the objects, situations and processes and are thus able to be applied flexibly in many reasoning and comprehension tasks. These models are hypothesized to be generative in reasoning processes because specific inferences can be traced directly to a model, such as inferences about electrical phenomena based on a model of electricity as particulate ('teeming crowds') or as a flowing substance ('flowing water') (Gentner and Gentner, 1983), or of the operation of a thermostat based on either a valve or threshold model of its operation (Kempton, 1986). Much of the research in this area has focused on specifying the content of the mental models in a domain, with issues about the format of the mental model usually not addressed.

The second usage focuses on the nature of the structure employed in WM in a specific comprehension or reasoning task. This literature maintains that mental models are created and manipulated during narrative and discourse comprehension, deductive and inductive logical reasoning and other inferential processes such as in learning and creative reasoning. The LTM knowledge such reasoning processes draw upon need not be represented in a model – e.g. Holland *et al.* (1986) hold that LTM knowledge is organized in proposition-like schemas. Although Philip Johnson-Laird's own research focus has been on mental modelling in deductive and inductive reasoning tasks and not mental modelling in other domains, his 1983 book provides a general account of mental models as temporary reasoning structures that has had a wide influence. He holds that a mental model is a structural analogue 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 analogue is that it embodies a representation of the salient spatial and temporal relations among, and the causal structures connecting, the events and entities depicted, and whatever other information is relevant to the problem-solving task. This characterization needs expansion to include functional analogues as well.

The mental model is an analogue in that it preserves constraints inherent in what is represented. Mental models are not mental images, although in some instances an accompanying image might be employed. The representation is intended to be isomorphic to dimensions of the real-world system salient to the reasoning process. Thus, for example, in reasoning about a spring the mental model need not capture the three-dimensionality of a spring if that is not taken to be relevant to the specific problem-solving task. The nature of the representation is such as to enable simulative behaviour in which the models behave in accord with constraints that need not be stated explicitly. For example, for those tasks that are dynamic in nature, if the model captures the causal coherence of a system it should, in principle, be possible to simulate the behaviours of the system. Thus, the claim that the inferential process is one of direct manipulation of the model is central to the WM usage. The specific nature of the model-manipulation process is linked to the nature of the format of the representation.

The format issue is significant because different kinds of representations – linguistic, formulaic, imagistic, analogue – enable different kinds of operations. Operations on linguistic and formulaic representations, for example, include the familiar operations of logic and mathematics. These representations are interpreted as referring to physical objects, structures, processes, or events descriptively. Customarily, the relationship between this kind of representation and what it refers to is 'truth', and thus the representation is evaluated as being true or false. Operations on such representations are rule-based and truth-preserving if the symbols are interpreted in a consistent manner and the properties they refer to are stable in that environment. Additional operations can be defined in limited domains provided they are consistent with the constraints that hold in that domain. Manipulation, in this case, would require explicit representation of salient parameters including constraints and transition states. I will call representations with these characteristics 'propositional', following the usual philosophical usage as a language-like encoding possessing a vocabulary, grammar, and semantics (Fodor, 1975) rather than the broader usage sometimes employed in cognitive science which is co-extensive with 'symbolic'.

On the other hand, analogue models, diagrams and imagistic (perceptual) representations are interpreted as representing demonstratively. The relationship between this kind of representation and what it represents – that, following Peirce, I will call 'iconic' – is 'similarity' or 'goodness of fit'. Iconic representations are similar in degrees and aspects to what they represent, and are thus evaluated as accurate or inaccurate. Operations on iconic representations involve transformations of the representations that change their properties and relations in ways consistent with the constraints of the domain. Significantly, transformational constraints represented in iconic representations can be implicit, e.g. a person can do simple reasoning about what happens when a rod is bent without having an explicit rule, such as 'given the same force a longer rod will bend farther'. The form of representation is such as to enable simulations in which the model behaves in accord with constraints that need not be stated explicitly during this process. Mathematical expressions present an interesting case in that it's conceivable they can be represented either propositionally or iconically (Kurz and Tweney, 1998).

Dispersed throughout the cognitive science literature is another distinction pertinent to the format of mental models, concerning the nature of the symbols that constitute propositional and iconic representations – that between 'amodal' and 'modal' symbols (Barsalou, 1999). Modal symbols are analogues of the perceptual states from which they are extracted. Amodal symbols are arbitrary transductions from perceptual states, such as those associated with language. A modal symbol representing a cat would retain perceptual aspects of cats; an amodal symbol would be the strings of letters 'cat' or 'chat' or 'Katze'. Propositional representations, in the sense discussed above, are composed of amodal symbols. Iconic representations can be composed of either. For example, a representation of the situation 'the circle is to the left of the square which is to the left of the triangle' could be composed of either the perceptual correlates of the tokens, such as ●–■–△, or amodal tokens standing for these entities, such as C–S–T. One can find all possible flavours in the mental modelling literature: propositional, amodal iconic and modal iconic mental models.

Among the WM accounts of mental modelling, Holland *et al.* (1986) maintain that reasoning with a mental model is a process of applying condition–action rules to propositional representations of the specific situation, such as making inferences about a feminist bank-teller on the basis of a model constructed from knowledge of feminists and bank-tellers. Johnson-Laird's mental models are not propositional in nature, rather they are amodal iconic representations. Making a logical inference such as *modus ponens* occurs by manipulating amodal tokens in a specific array that captures the salient structural dimensions of the problem and then searching for counter-examples to the model transformation. 'Depictive mental models' (Schwartz and Black, 1996) provide an example of modal iconic mental models. In this case, manipulation is by using tacit knowledge embedded in constraints to simulate possible behaviours, such as in an analogue model of a set-up of machine gears. In both instances of iconic models operations on a mental model transform it in ways consistent with the constraints of the system it represents.

Although the issues of the nature of the LTM representations and the WM format and processes involved in reasoning with mental models need eventually to be resolved in mental models theory, these do not have to be settled before it is possible to make progress on an account of model-based reasoning in science. My analysis of model-based reasoning adopts only a 'minimalist' hypothesis: that in certain problem-solving tasks humans reason by constructing a mental model of the situations, events and processes in WM that in dynamic cases can be manipulated through simulation. The WM model is held to be iconic but leaves open the questions of the nature of the representation in long-term memory, and whether the format of the WM representation employed in reasoning is amodal or modal.

3.2 Conceptual change and generic modelling

To explain how model-based reasoning could be generative of conceptual change in science requires a fundamental revision of the understandings of concepts, conceptual structures, conceptual change and reasoning customarily employed explicitly in philosophy and at least tacitly in the other science-studies fields. Only an outline of my account will be developed here. A basic ingredient of the revision is to view the representation of a concept as providing sets of constraints for generating members of classes of models. Concept formation and change is then a process of generating new, and modifying existing, constraints. A productive strategy for accomplishing this is through iteratively constructing models embodying specific constraints until a model of the *same type* with respect to the salient constraints of the phenomena under investigation – the 'target' phenomena – is achieved.

In the model-construction process, constraints drawn from both the target and source domains are domain-specific and need to be understood in the reasoning process at a sufficient level of abstraction for retrieval, transfer and integration to occur. I call this type of abstraction 'generic'. Although the instance of a model is specific, for a model to function as a representation that is of the same kind with respect to salient dimensions of the target phenomena inferences made with it in a reasoning process need to be understood as generic. In viewing a model generically, one takes it as representing features, such as structure and behaviours, common to members of a class of phenomena. The relation between the generic model and the specific instantiation is similar to the type–token distinction used in logic. Generality in representation is achieved by interpreting the components of the representation as referring to object, property, relation, or behaviour types rather than tokens of these.

One cannot draw or imagine a 'triangle in general' but only some specific instance of a triangle. However, in considering what it has in common with all triangles, humans have the ability to view the specific triangle as lacking specificity in its angles and sides. In considering the behaviour of a physical system such as a spring, again one often draws or imagines a specific representation. However, to consider what it has in common with all springs, one needs to reason as though it as lacked specificity in length and width and number of coils; to consider what it has in common with all simple harmonic oscillators, one needs to reason as though it lacked specificity in structure and aspects of behaviour. That is, the reasoning context demands that the interpretation of the specific spring be generic.

The kind of creative reasoning employed in conceptual innovation involves not only applying generic abstractions but creating and transforming them during the reasoning process. There are many significant examples of generic abstraction in conceptual change in science. In the domain of classical mechanics, for example, Newton can be interpreted as employing generic abstraction in reasoning about the commonalities among the motions of planets and of projectiles, which enabled him to formulate a unified mathematical representation of their motions. The models he employed, understood generically, represent 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; for example, that within the context of determining motion, planets and projectiles can both be represented as point masses. After Newton, the inverse-square-law model of gravitational force 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.

My hypothesis is that analogies, visual models and thought experiments are prevalent in periods of radical conceptual change because such model-based reasoning is a highly effective means of making evident and abstracting constraints of existing representational systems and, in light of constraints provided by the target problem, effective means of integrating constraints from multiple representations such that novel representational structures result. I will now provide brief encapsulations of how this occurs.

3.3 Analogical modelling

There is a vast cognitive science literature on analogy, and a vast number of historical cases that substantiate its employment in conceptual change. This literature provides theories of the processes of retrieval, mapping, transfer, elaboration and learning employed in analogy and the syntactic, semantic and pragmatic constraints operating on these processes (Gick and Holyoak, 1980, 1983; Gentner, 1983, 1989; Holyoak and Thagard, 1989, 1996; Thagard *et al.*, 1990; Gentner *et al.*, 1997). Most of the analogy research in cognitive science examines cases in which the source analogies are provided, either implicitly or explicitly. This had led to a focus on the nature of the reasoning processes involved with the objective of finding and making the correct mappings. There is widespread agreement on criteria for good analogical reasoning, drawn from psychological studies of productive and non-productive use of analogy and formulated by Gentner (Gentner 1983, 1989): (1) 'structural focus' – preserves relational systems; (2) 'structural consistency' – isomorphic mapping of objects and relations; and (3) 'systematicity' – maps systems of interconnected relationships, especially causal and mathematical relationships. Generic abstraction can been seen to be highly significant in analogical modelling, since objects and relations often need to be understood as lacking specificity along certain dimensions in order for retrieval and transfer to occur. This is especially evident in instances that involve recognizing potential similarities across disparate domains, and abstraction and integration of information from these. Gick and Holyoak's (1983) analysis of how knowledge gained in one analogical problem-solving task is transferred to another by creating a 'schema' common to both target and source domains provides an example of its use in mundane reasoning.

As employed in model-based reasoning, I propose that analogies serve as sources of constraints for constructing models. In this use of analogy the source domain(s) provide constraints that, in interaction with constraints drawn from the target domain, lead to the construction of initial as well as subsequent models. Model construction utilizes knowledge of the generative principles and constraints for models in a known 'source' domain, selected on the basis of target constraints. The constraints and principles can be represented in different informational formats and knowledge structures that act as explicit or tacit assumptions employed in constructing and transforming models during problem-solving. Evaluation of the analogical modelling process is in terms of how well the salient constraints of a model fit the salient constraints of a target problem, with key differences playing a significant role in further model generation (Griffith, Nersessian and Goel, 1996; Griffith, 1999). Unlike the situation typically studied for mundane cognition, in science the appropriate analogy or even analogical domain is often unknown. And it even happens that no direct analogy exists once a source domain is identified, and construction of the source analogy itself is required. In the case of Maxwell's (1890) construction of a mechanical model of the electromagnetic aether, the initial source domain was continuum mechanics and the target domain electromagnetism. No direct analogy existed within continuum mechanics, the initial source domain, and Maxwell integrated constraints first from electromagnetism and continuum mechanics to create an initial model and later used constraints from machine mechanics to modify the model, creating a hybrid model consistent with the constraints of electromagnetic induction (Nersessian, 1992a, 2001a, 2001b).

The cognitive literature agrees with the position that analogies employed in conceptual change are not 'merely' guides to reasoning but are generative in the reasoning processes in which they are employed. For example, 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. One example already mentioned is the study where subjects constructed a mental model of electricity in terms of either an analogy with flowing water or with swarming objects, and then specific inferences – sometimes erroneous – could be traced directly to the analogy (Gentner and Gentner, 1983). Here the inferential work in generating the problem solution was clearly done through the analogical models.

3.4 *Visual modelling*

A variety of perceptual resources can be employed in modelling. Here I focus on the use of the visual modality since it figures prominently in cases of conceptual change across the sciences. A possible reason why is that employing the visual modality may enable the reasoner to bypass specific constraints inherent in current linguistic and formulaic representations of conceptual structures. The hypothesis that the internal representations can be imagistic does not mean that they need to be picture-like in format. The claim is that they are modal in format and employ perceptual and possibly motor mechanisms in processing. They can be highly schematic in nature. Thus the fact that some scientists such as Bohr claim not to experience mental pictures in reasoning is not pertinent to the issue of whether this kind of perceptual modelling is playing a role in the reasoning.

There is a vast cognitive science literature on mental imagery that provides evidence that humans can perform simulative imaginative combinations and transformations that mimic perceptual spatial transformation (Kosslyn, 1980; Shepard and Cooper, 1982). These simulations are hypothesized to take place using internalized constraints assimilated during perception and motor activity (Kosslyn, 1994). Other research 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. There are significant differences between spatial transformations and transformations requiring causal and other knowledge contained in scientific theories. Although the research on imagery in problem-solving is scant, cognitive scientists have recently undertaken several investigations examining the role of causal knowledge in mental simulation involving imagery – for example, experiments with problems employing gear rotation provide evidence of knowledge of causal constraints being utilized in imaginative reasoning (Hegarty, 1992; Hegarty and Just, 1994; Hegarty and Sims, 1994; Schwartz and Black, 1996).

As used in physics, for example, imagistic representations participate in modelling phenomena in several ways, including providing abstracted and idealized representations of aspects of phenomena and embodying aspects of theoretical models. Thus, early in Faraday's construction of an electromagnetic field concept, the imagistic model he constructed of the lines of force provided an idealized representation of the patterns of iron filings surrounding a magnet (see figure 2). However, cognitive–historical research substantiates the interpretation that later in his development of the field concept, the imagistic model functioned as the embodiment of a dynamical theoretical model of the transmission and inter-conversion of forces generally, through stresses and strains in, and various motions of, the lines (Gooding, 1981, 1990; Nersessian, 1984b, 1985; Tweney, 1985, 1992). But, as I have argued, the visual representation Maxwell presented of the idle wheel–vortex model was intended as an embodiment of an imaginary system, displaying a generic dynamical relational structure, and not as a representation of the theoretical model of electromagnetic field actions in the aether (figure 1).

External visual representations (including those made by gesturing and sketching) employed during a reasoning process are a significant dimension of cognitive activity in science and should be analysed as part of the cognitive system. These can be interpreted as providing support for the processes of constructing and reasoning with a mental model. In model-based reasoning

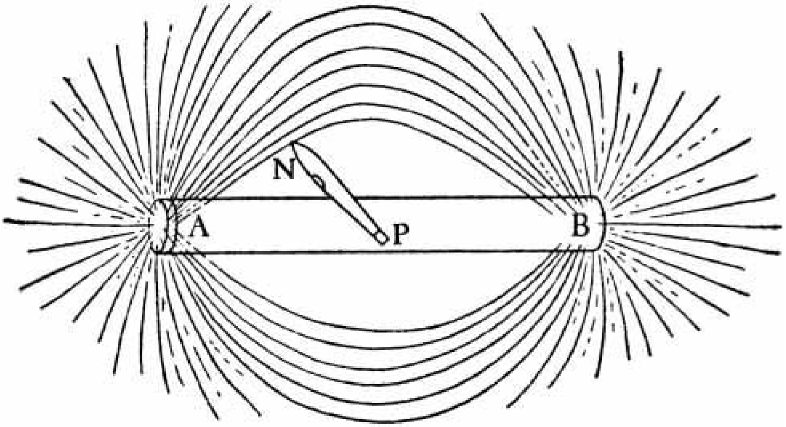

Fig. 2 Faraday's drawing of the lines of force surrounding a bar magnet. *Source*: Faraday (1839–55, vol. 1, plate 1). A, B mark the ends of the magnetic poles and P, N delineate a silver knife blade laid across the lines of force.

processes they function as much more than the external memory aids they are customarily considered to be in cognitive science. They aid significantly in organizing cognitive activity during reasoning, such as fixing attention on the salient aspects of a model, enabling retrieval and storage of salient information and exhibiting salient constraints, such as structural and causal constraints, in appropriate co-location. Further they facilitate construction of shared mental models within a community and transportation of scientific models out of the local milieu of their construction.

3.5 Simulative modelling

As a form of model-based reasoning, thought experimenting can be construed as a specific form of the simulative reasoning that can occur in conjunction with the other kinds of model-based reasoning. In simulative reasoning inferences are drawn by employing knowledge embedded in the constraints of a mental model to produce new states. Constructing a thought-experimental model requires understanding the salient constraints governing the kinds of entities or processes in the model and the possible causal, structural and functional relations among them. Conducting a simulation can employ either tacit or explicit understanding of the constraints governing how those kinds of things behave and interact and how the relations can change. A simulation creates new states of a system being modelled, which in turn creates or makes evident new constraints. Changing the conditions of a model enables inferences about differences in the way that a system can behave. Various kinds of knowledge of physical situations is employed in imaginary simulations. Because the simulation complies with the same constraints of the physical system it represents, performing a simulation with a mental model enables inferences about real-world phenomena. Note that understanding of the mathematical constraints governing a situation is one kind of knowledge that can be used in simulative reasoning by scientists.

In the case of scientific thought experiments implicated in conceptual change, the main historical traces are in the form of narrative reports, constructed after the problem-solving has taken place. These have often provided a significant means of effecting conceptual change within a scientific community. Accounting for the generative role of this form of model-based reasoning begins with examining how these thought-experimental narratives support modelling processes and then making the hypothesis that the original experiment involves a similar form of model-based reasoning. What needs to be determined is: (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 modelling perspective, the function of the narrative form of presentation of a thought experiment would be to guide the reader in constructing a mental model of the situation described by it and to make inferences through simulating the events and processes depicted in it. A thought-experimental model can be construed as a form of ‘discourse’ model studied by cognitive scientists, for which they argue that the operations and inferences are performed not on propositions but on the constructed model (Johnson-Laird, 1982, 1989; Perrig and Kintsch, 1985; Morrow, Bower and Greenspan, 1989).

Simulation is assisted in that the narrative delimits the specific transitions that govern what takes place. The thought-experimental simulation links the conceptual and the experiential dimensions of human cognitive processing (see also Gooding, 1992). Thus, the constructed situation inherits empirical force by being abstracted both from experiences and activities in the world and from 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.

Unlike a fictional narrative, however, the context of the scientific thought experiment makes the intention clear to the reader that the inferences made pertain to potential real-world situations. The narrative has already made significant abstractions, which aid in focusing attention on the salient dimensions of the model and in recognizing the situation as prototypical (generic). Thus, the experimental consequences are seen to go beyond the specific situation of the thought experiment. The thought-experimental narrative is presented in a polished form that 'works', which should make it an effective means of generating comparable mental models among the members of a community of scientists.

The processes of constructing the thought-experimental model in the original experiment would be the same as those involved in constructing any mental model in a reasoning process. In conducting the original thought experiment a scientist would make use of inferencing mechanisms, existing representations and scientific and general world knowledge to make constrained transformations from one possible physical state to the next. Simulation competence should be a function of expertise. As with real-world experiments, some experimental revision and tweaking undoubtedly goes on in conducting the original experiment, as well as in the narrative construction, although accounts of this process are rarely presented by scientists.

Finally, in mundane cases the reasoning performed via simulative mental modelling is usually successful because the models and manipulative processes embody largely correct constraints governing everyday real-world events. Think, for example, of how people often reason about how to get an awkward piece of furniture through a door. The problem is usually solved by mentally simulating turning over a geometrical structure approximating the configuration of the piece of furniture through various rotations. The task employs often implicit knowledge of constraints on such rotations, and is often easier when the physical item is in front of the reasoner acting to support the structure in imagination. In the case of science where the situations are more removed from human sensory experience and the assumptions more imbued with theory, there is less assurance that a simulative reasoning process, even if carried out correctly, will yield success. Clearly scientists create erroneous models – revision and evaluation are crucial components of model-based reasoning. In the evaluation process, a major criterion is goodness of fit to the constraints of the target phenomena, but success can also include such factors as enabling the generation of a viable mathematical representation that can push the science along while other details of representing the phenomena are still to be worked out, as Newton did with the concept of gravitation, and Maxwell with the concept of electromagnetic field.

4 Reflexivity: cognitive hypotheses

There are several key ingredients common to the various forms of model-based reasoning practices under consideration. The problem-solving processes involve constructing models that are of the *same kind* with respect to salient dimensions of target phenomena. The models are intended as interpretations of target physical systems, processes, phenomena, or situations. The modelling practices make use of both highly specific domain knowledge and knowledge of abstract general principles. Further, they employ knowledge of how to make appropriate abstractions. Initial models are retrieved or constructed on the basis of potentially satisfying salient constraints of the target domain. Where the initial model does not produce a problem solution, modifications or new models are created to satisfy constraints drawn from an enhanced understanding of the target domain and from one or more source domains (same as target domain or different). These constraints can be supplied by means of linguistic, formulaic, and imagistic (all perceptual modalities) informational formats, including equations, texts, diagrams, pictures, maps, physical models and various kinaesthetic and auditory experiences. In the modelling process, various forms of abstraction, such as limiting case, idealization, generalization and generic modelling, are utilized, with generic modelling playing a highly significant role in the generation, abstraction and integration of constraints. Evaluation and adaptation take place in light of structural, causal, and/or functional constraint satisfaction and enhanced understanding of the target problem that has been obtained through the modelling process. Simulation can be used to produce new states and enable evaluation of behaviours, constraint satisfaction, and other factors. Figure 3 illustrates the interactive nature of these construction processes.

What cognitive account of representational and reasoning processes involved in model-based reasoning might the scientific practices support? In the preceding I have attempted to carry out the analysis by remaining as neutral as possible on some contentious issues within cognitive science, in order to show that progress can be made on understanding conceptual change in science with certain significant issues unresolved. In this section I conclude by briefly noting the cognitive hypotheses made in the analysis that would bear further investigation by cognitive science. The modelling practices exhibited by scientists utilize and engage internal modelling processes that are highly effective means

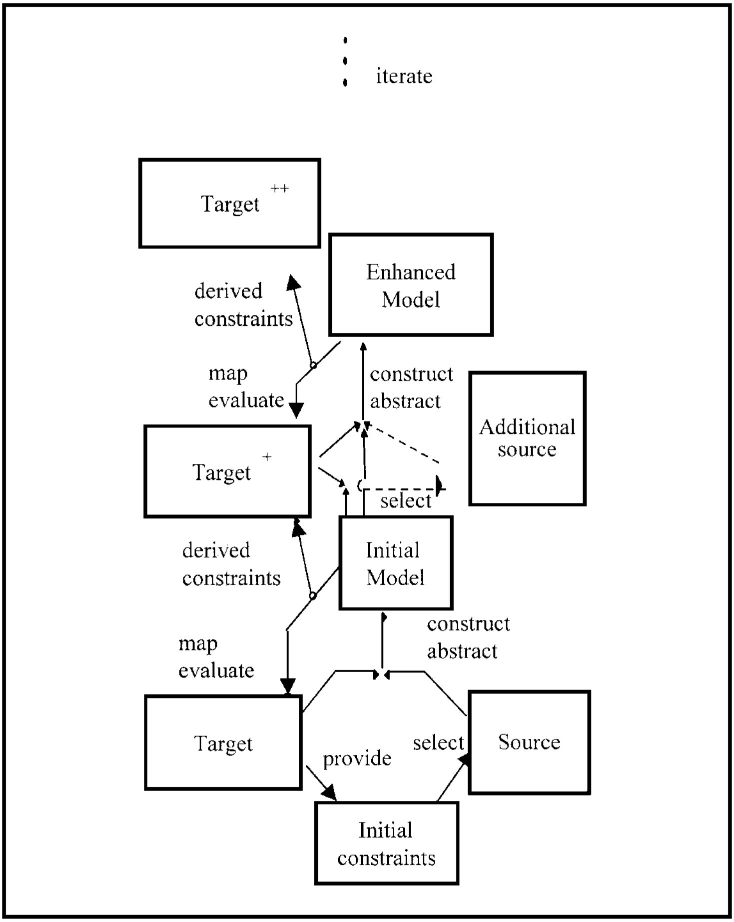

Fig. 3 Modelling processes

of generating representations and transmitting novel representations through a community. Model-based reasoning is not reducible to operations of applying mental logic to proposition-like representations. The representations are iconic in nature. The reasoning process is through model manipulation and involves processing mechanisms used in perceptual–motor activity. This implies that modal symbols are employed to some extent. Concept formation and change involve processes of generating new and modifying or replacing existing constraints. This assumes that the representations of scientific concepts provide sets of constraints for generating members of classes of models. Model-based reasoning is generative of conceptual change in science because analogical modelling, visual modelling and thought experimenting (simulative modelling) are effective means of abstracting and examining constraints of existing conceptual systems in light of constraints provided by the target problem; effective means of bypassing constraints inherent in linguistic and formulaic representations of conceptual systems; and effective means of generating and synthesizing constraints into new–revised conceptual systems. Generic abstraction enables the extraction and integration of constraints from diverse sources, leading to genuine novelty.