COGNITIVE SCIENCE

Paul Thagard
University of Waterloo

Thagard, P. (forthcoming). Cognitive science. In M. Curd & S. Psillos (Eds.), *The Routledge companion to the philosophy of science*. London: Routledge.

INTRODUCTION

Cognitive science is the interdisciplinary investigation of mind and intelligence, embracing psychology, neuroscience, anthropology, artificial intelligence, and philosophy. There are many important philosophical questions related to this investigation, but this short chapter will focus on the following three. What is the nature of the explanations and theories developed in cognitive science? What are the relations among the five disciplines that comprise cognitive science? What are the implications of cognitive science research for general issues in the philosophy of science? I will argue that cognitive theories and explanations depend on representations of mechanisms and that the relations among the five disciplines, especially psychology and neuroscience, depend on relations between kinds of mechanisms. These conclusions have implications for central problems in general philosophy of science such as the nature of theories, explanations, and reduction between theories at different levels.

THEORIES AND EXPLANATIONS: MECHANISMS

The primary goal of cognitive science is to explain the operations of the human mind, but what is an explanation? In general philosophy of science, explanations have often been discussed as deductions from general laws, or sometimes as schematic patterns that unify diverse phenomena. It is becoming increasingly clear, however, that explanations in cognitive science employ representations of mechanisms that provide causal accounts of such mental phenomena as perception, memory, problem solving, and learning. Theories are sets of hypotheses about the constituents of the explanatory mechanisms. Numerous philosophers of science have defended the mechanistic account of explanations in various fields: see Bechtel and Abrahamsen (2005) and the chapter in this volume by Glennan. I will now describe how explanations of human thinking involve mechanisms.

Cognitive science began in the mid-1950s when psychologists, linguists, and researchers in the nascent enterprise of artificial intelligence realized that ideas from the emerging field of computer science could be used to explain how minds work. The first operational computational model of mind was the Logic Theorist of Newell, Shaw, and Simon, which simulated how people do proofs in deductive logic. This model and many later ones worked with what became the fundamental analogy of cognitive science: just as computer programs run by applying algorithms to data structures, so the human mind works by applying computational procedures to mental representations.

Many competing proposals have been made about what are the most important mental representations in human thinking, ranging from rules to concepts to images to analogies to neural networks. And for each kind of mental representation there are different kinds of computational procedure; for example, rules are IF-THEN structures that work by matching the IF part and then applying the THEN part. But all of these approaches assume that thinking is like computation in that it applies algorithmic procedures to structured representations.

Explanations that employ computational ideas are clearly mechanistic. A mechanism is a system of objects related to each other in various ways including part-whole and spatial contiguity, such that the properties of the parts and the relations between them produce regular changes in the system. For example, a bicycle is a mechanism consisting of parts (e.g. the frame, wheels, and pedals) that are related to each other so that the bicycle moves when force is applied to the pedals. Similarly, according to the computational hypothesis of cognitive science, the mind is a mechanism whose parts are mental representations of various sorts that are organized such that there are computational procedures that operate on them to produce new representations. No one would disagree that computers are mechanisms built out of hardware and software that enable them to perform complex tasks, and the computer-mind analogy made it possible for the first time to see how highly complex thinking could be performed mechanically. Before the emergence of cognitive science in the 1950s, many mental mechanisms had been proposed, ranging from clockwork to association of ideas to telephone switchboards to stimulus-response connections. But only with the development of advanced hardware and software did it become possible to understand how the most sophisticated kinds of human problem solving, learning, and language could operate mechanistically.

Of course, not everyone in the constituent fields of cognitive science has been attracted to the computational approach to explaining mental phenomena. In philosophy there are still dualists who think that consciousness is not explicable in terms of physical mechanisms, but their arguments consist of thought experiments that merely serve to reinforce their own prejudices (see Churchland, 2002, for an accessible review). More usefully, a host of cognitive scientists have pointed out that we should not explain thinking solely in terms of the internal operations of the mind, but should also take into account ways in which humans have bodies that enable them to interact causally with the world. But robots can also have bodies that enable them to interact with the world and form meaningful representations of it, so the claims that cognition is embodied and situated are extensions to the computational view of mind, not replacements for it.

If cognitive explanations consist of showing how mental mechanisms can produce psychological phenomena, then psychological theories are representations of such mechanisms. Representations of mechanisms can be verbal, as when I described a bicycle in terms of its parts and their relations. But they can also be visual, for example when a bicycle is portrayed using a diagram or, even better, using a movie that shows it in operation. Similarly, psychological theories are usually presented via a combination of verbal and visual representations. For example, theories of concepts are often presented by a combination of verbal descriptions, mathematical equations that describe such procedures as spreading activation, and diagrams that portray how different concepts are related to each other. Similarly, theories about how neural networks produce psychological phenomena are presented using a combination of verbal, mathematical, and visual representations. These multimodal representations of theories may seem puzzling from the traditional view in philosophy of science that theories are universal statements in a formal language, but they make complete sense if explanation is understood, not as deduction in a formal system, but as application of mechanisms. From this perspective, the primary purpose of theories is to depict mechanisms, and visual representations are often more effective means of representing the part-whole and spatial relations of objects in a mechanism than purely verbal representations. Later in this chapter I will argue that most scientific theories, not just cognitive ones, can be understood as representations of mechanisms.

So far, I have been discussing computational mechanisms of the sort that dominated cognitive theorizing in the second half of the twentieth century. But rapid increases in knowledge about how brains work have increasingly led to psychological explanations that are based on neural mechanisms rather than abstract computational ones. In the 1980s, there was a revival of interest in computational models that employ artificial neural networks, but until recently these were so artificial that they are more aptly classified as abstract computational models rather than as neurological ones. What were called "connectionist" or "parallel distributed processing" models are giving away to more biologically realistic ones.

Neurocognitive theories are now being proposed that have three key properties that differentiate them from their much simpler predecessors. First, their main components are artificial neurons that are much more biologically realistic than connectionist neurons, which typically possess an activation value that represents their rate of firing. The new wave of neurocognitive models takes into account that neurons not only have firing rates – how often they fire in a given stretch of time – but also firing patterns. Two neurons may each fire 20 times a second, but have very different patterns of when they are firing and resting, and there are psychological and computational reasons to believe that such patterns are important. Second, whereas connectionist models typically used small numbers (often less than a hundred) of artificial neurons to model psychological phenomena, more biologically realistic neurocognitive models usually have thousands of less artificial neurons interacting with each other. Such models are still puny compared to the billions of neurons operating in the brain, but they have greater capacity to capture the representational and computational power of brains.

Third, and probably most important, the new wave of neurocognitive models takes brain anatomy seriously, organizing groups of artificial neurons in correspondence to actual brain areas. The brain is not just one big neural network, but is highly organized into functional areas that accomplish particular tasks such as vision, motor control, language and reasoning. The different areas are highly interconnected, so that there are not isolated modules operating independently, but the interconnections within a particular brain area are much denser than the connections with other brain areas. Accordingly, neurocognitive models increasingly have dedicated groups of neurons representing particular brain areas such as parts of the prefrontal cortex, the hippocampus, and the amygdala. For examples of neurocognitive models that use more biologically realistic neurons and neural organization, see Eliasmith and Anderson (2003).

It should be evident that the more biologically realistic neurocognitive models are still mechanistic and computational. They are mechanistic in that they consist of objects – neurons – organized via part-whole and spatial relations. Neurons are parts of neuronal groups that are parts of brain areas such as the prefrontal cortex. Neurons are related to each other not just by spatial contiguity, but more importantly by axons and dendrites that connect them physically via synapses, making them capable of exciting or inhibiting the firing of other neurons. Hence changes in the firing patterns of individual neurons lead to changes in the activity of entire brain areas and ultimately to changes in behavior. Thus the complexes of neurons postulated by neurocognitive models are clearly mechanisms, and theories of neural functioning are well understood as representations of mechanisms. Perusal of textbooks in neuroscience and cognitive psychology will confirm that such representations are usually multimodal, involving a combination of verbal description, mathematical equations that describe neural behavior, and diagrams that indicate spatial and temporal relations.

But are biologically realistic models still computational? The cognitive models discussed earlier are computational in a double sense, in that they not only use computers to do the complex calculations required for modeling, but also postulate that minds are actually performing a kind of computation. Contrast computer models in fields such as physics, chemistry, and weather forecasting, where no one thinks that the systems being modeled are actually doing computations. Neurocognitive models are also computational in the double sense, in that it is reasonable to postulate that brains are actually computing by encoding, decoding, and transforming information. Hence they do not involve rejection of the fundamental analogy of cognitive science that thinking is computing, only a substantial enrichment of it in terms of more biologically realistic neural processes.

Computational and neural mechanisms are not the only ones relevant to explaining human thinking. Humans are social animals, and much thinking takes place in interaction with other people. Decision making, for example, is often not just one person deciding alone, but groups of people interacting to work things out together. Social groups can also be understood as mechanisms, in which the parts are people and sub-groups and the relations are interpersonal ones such as communication. As indicated by the inclusion of anthropology as one of the disciplines of cognitive science, the field is open to including the social dimension of thinking, so that attention to social mechanisms is a natural part of cognitive science.

Similarly, cognitive science should be amenable to moving down levels of organization as well as up. Neuroscience is increasingly paying attention to molecular mechanisms that explain how neurons work. That molecular biology is mechanistic is evident from explanations of the functions and behavior of cells based on the chemical reactions of their constituents such as proteins. Explanations in molecular biology are not alternatives to psychological explanations, but complement them just as social explanations do. In the next section, I will describe how such complementation works in terms of interactions of mechanisms at different levels.

DISCIPLINARY INTERRELATIONS

Consider the highly interesting phenomenon of falling in love, for example when it happened to Shakespeare's Romeo and Juliet. A full understanding of this phenomenon needs to pay attention to at least four levels of explanation: social, psychological, neurological, and molecular. The star-crossed lovers meet at a social event, a party at Juliet's house; this social interaction occurs in the larger context of a feud between her family and Romeo's. Once they begin to interact, they have many thoughts about each other, for example when Romeo compares Juliet to the sun. Such thoughts must be understand in terms of various psychological processes including perception, analogy, and language production and comprehension. Unknown to Shakespeare, these psychological processes have corresponding neurological processes such as the firing of neuronal groups in cognitive brain areas such as the prefrontal cortex and in emotional brain areas such as the nucleus accumbens. Presumably Romeo and Juliet experienced high levels of activity in the latter brain area as they anticipated seeing each other with intense pleasure. Finally, there is evidence that neurotransmitters such as dopamine are highly relevant to explaining what happens when people fall in love, so that the molecular level of explanation must also be taken into account. What are the relations among the social, psychological, neurological, and molecular explanations of falling in love?

Philosophical answers to this question are usually reductionist, claiming that each higher level reduces to the next lower level, or antireductionist, claiming that higher levels are largely independent of lower levels. The most ruthlessly reductionist position would claim that ultimately everything must be explained in terms of the fundamental constituents of matter identified by sub-atomic physics, but it is hard to see how anything about quarks or strings is relevant to understanding how Romeo and Juliet fell in love. Similarly, although the fact that Romeo's dopamine levels spiked when he first saw Juliet is certainly relevant to understanding his falling in love with her, the molecular occurrences in his and Juliet's brain tell only part of the story about what was going on when they met at the party. Hence reductionism that claims that there is a fundamental level of explanation is implausible in cognitive science.

But antireductionism is implausible also, as it would be folly to try to give a purely sociological account of the Romeo and Juliet falling in love without also paying attention to their thoughts about each other, for example their mental representations of each other and each other's families. Hence the social explanation needs the psychological one, and there is abundant evidence from recent work in cognitive science that psychological explanations can be enriched by neurological ones that identify the brain areas and kinds of neural activity responsible for perception, inference, and emotion. These neurological explanations in terms employ molecular processes such as cascades of dopamine activity in the nucleus accumbens and other brain areas. So if both reductionism and antireductionism are implausible as accounts of the multilayer explanation of falling in love, how can philosophers of science give a plausible account of the relations among different levels of explanation?

The ideas about mechanism described in the previous section are very useful for describing the relations between different levels of explanation. Table 1 schematizes some of the mechanisms operating at the various levels. At each level, there are components consisting of objects with relations to each other, whose interactions produce changes in the whole system. The components form a part-whole hierarchy, as when the Montague family includes Romeo, and Romeo has a mind with many representations and procedures, and Romeo's body includes a brain with numerous neuronal groups, and his neurons are cells made up of various molecules such as proteins. This hierarchy supports a kind of ontological reductionism, according to which the higher level entities are nothing more than the kinds of things that make them up, for example that families are constituted by the people who make them up. But it does not support an epistemological reductionism which is concerned with how explanations are actually carried out. A full blown reductionism of this kind would require that the changes at each level would have to be explained by the changes at the subordinate level, with all changes ultimately being explained at some lower level. But there are at least two reasons why understanding of mechanisms does not work that way.

First, we often have a good understanding of how a mechanism works without being able to say how it arises from subordinate mechanisms. For example, there are many social mechanisms such as verbal and nonverbal communication that can be described in detail without knowing all the psychological mechanisms that make them possible. Similarly, there are currently good computational theories of inference and problem solving that work well at the psychological level even though the specific neural mechanisms that support them are little understood. Given the enormous complexity of social, psychological, and neural mechanisms, it is unlikely that we will ever be able to fill them out fully at the molecular level, let alone the subatomic physics level.

Mechanisms	*Components*	*Relations*	*Interactions*	*Changes*
Social	Persons and social groups	Association, membership	Communication	Influence, group decisions
Psychological	Mental representations such as concepts	Constituents, associations, implication	Computational processes	Inferences
Neural	Neurons, neural groups	Synaptic connections	Excitation, inhibition	Brain activity
Molecular	Molecules such as neurotransmitters and proteins	Constituents, physical connection	Biochemical reactions	Transformation of molecules

Table 1

Constituents of mental mechanisms.

Second, the interactions between levels are not always upward, from molecular to neural to psychological to social. For example, the best explanation of why Romeo has molecules of cortisol circulating in his bloodstream at a particular time may not operate purely at the molecular level, but should also take into account the social fact that Romeo has encountered members of the opposing clan, the psychological fact that he believes them to be hostile, the neural fact that his amygdala has neurons firing rapidly in a fear response, as well as the molecular fact that amygdala activity had activated his glands to pump out more cortisol, a hormone influenced by stress. Hence a social mechanism – interaction of conflicting groups – is a key part of the explanation of what happens to the molecular mechanism of cortisol production. Intervening between these two levels are the other two, because the social interaction produces mental representations comprised by neural activity that causes changes in cortisol levels. Hence explanation of why Romeo and Juliet fell in love operates best at multiple linked levels, invoking all the relevant mechanisms.

These two reasons show why we should not expect there to be a purely neurochemical theory of falling in love. The neurochemistry should not be ignored, as dopamine activity in the brain's reward areas are undoubtedly part of the process by which two people become romantically attached to each other. But all the other levels are highly relevant as well, including the social level concerning the kinds of group-based interactions that Romeo and Juliet had, the psychological level concerning the mental representations that they have of each other and their situation, and the neural level concerning how their brains process information about each other. We are unlikely ever to have enough knowledge of all the relevant mechanisms to be able to reduce the social to the molecular, and even if we did we would have to appreciate that the explanations do not all proceed from lower to higher levels. For example, if we want to understand why Romeo and Juliet both have high dopamine levels we would have to cite the relevant social fact that they are gazing into each other eyes, the relevant psychological fact that they have mental representations about each other, and the relevant neural fact that neurons are spiking rapidly in their brain's reward areas.

Figure 1 illustrates a multilevel, multidisciplinary explanation of why Romeo fell in love with Juliet, including social, psychological, neurological, and molecular factors. Table 1 provided a more specific view of what the components, interactions, and changes are at each level. The resulting picture is partly reductionist in that it shows how components at each level can be constituted by components at the lower level, for example when social groups are understood as consisting of individual thinkers. It is also reductionist in that the interactions at each level are to be understood at least in part by interactions at lower levels, for example when people communicate with each other by virtue of psychological processes of language production and comprehension. But it is emphatically not reductionist in that the characterization of components, interactions, and changes at each level does not have to be fully specified in lower level terms. Moreover, the bidirectional arrows allow changes at a higher level to causally produce changes at the lower levels, as in my examples of social conflict increasing cortisol and lovers' gazes increasing dopamine.

Thus the relations among different disciplines in cognitive science involves representations of mechanisms operating at different levels. Anthropology, psychology, and neuroscience illustrate interactions among the social, psychological, neural, and molecular levels of explanation. Linguistics cuts across these levels, as the use of language is clearly a social and psychological phenomenon that is carried out in specifiable brain areas governed by molecular processes such as genetics. Because cognitive science supports the materialist view that mental changes can be wholly explained naturalistically in physical terms, the philosophical position defended here can be called *Multilevel Mechanistic Materialism*.

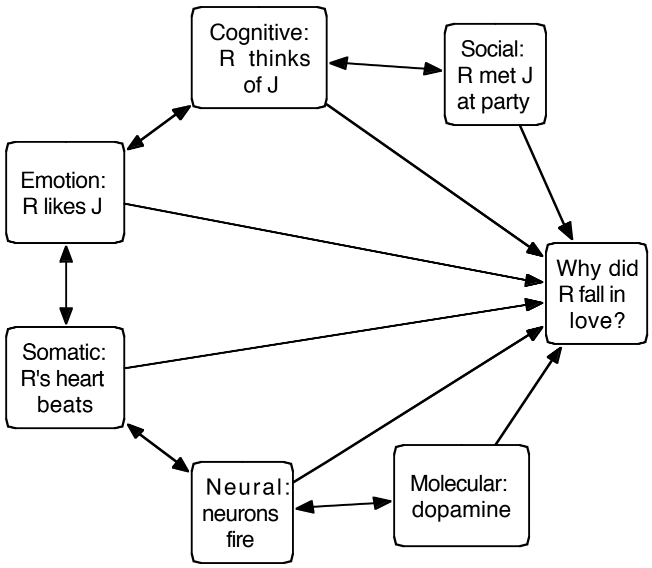

Figure 1. Sketch of a multilevel mechanistic explanation of why Romeo fell in love. A full causal picture would have more arrows.

Where does philosophy fit in cognitive science? Some philosophers see themselves as standing above the sciences, using a priori reflection to critique the conceptual confusions that arise there. Others see philosophy as providing under-laborers to clear away some of the rubbish that lies in the way of the development of scientific knowledge. My own view is that the interconnection between science and philosophy is much tighter than either of these more common views reflects. Philosophy of mind and cognitive science are tightly intertwined, with philosophical reflections ideally going in hand in hand with scientific developments in fields such as anthropology, psychology, neuroscience, and molecular biology. Philosophy differs from the sciences in two main ways, in its concerns with very general matters and with normative issues. Philosophy has greater generality than particular sciences that concern themselves with a narrower range of phenomena, as psychologists, for example, seek explanations of processes such as perception, memory, and problem solving. Such topics are of great philosophical interest, but they are only part of a more general concern with the nature of knowledge and existence. The generality of philosophy makes it of great importance to an interdisciplinary field such as cognitive science, because philosophy can attend to the full range of phenomena concerning the mind studied by people in different fields, and help to provide some of the theoretical glue that holds them all together.

The second way that philosophy differs from specific cognitive sciences is that is concerned not only with how thinking works but also with how it can work better. Epistemology and ethics are both fields that are essentially normative, the former concerned with how people ought to think if their thinking is to constitute knowledge, and the latter concerned with how people ought to treat each other. Theories about how people ought to think and how they ought to act should be connected with scientific theories about how people do think and act, but the connections are not so simple that the normative concerns of epistemology and ethics can be dispensed with in favor of purely descriptive matters. For description of how the normative issues of philosophy can cohere with empirical matters, see Thagard (2000).

This completes my picture of how the different disciplines of cognitive science are related to each other. Philosophical reflection on the nature of theories, explanations, and mechanisms provides a way of seeing how disparate disciplines can cooperate to promote understanding of the nature of mind and intelligence. Now I want to describe how this view of the nature of scientific activity has important implications for philosophy of science in general.

GENERAL PHILOSOPHICAL IMPLICATIONS

Cognitive science is not only a subject for discussion in the philosophy of science, like other special sciences. It is also a source of new ways of thinking about the structure and growth of scientific knowledge, with implications for general questions about the nature of theories, explanations, justification, and discovery. This section will review some of the general contributions that cognitive science can make to the philosophy of science.

Much of twentieth-century philosophy of science was dominated by the philosophical views of the logical positivists, who understood scientific theories as formalized statements in logical systems and explanations as deductions in such systems. Many problems were identified with logical analyses of scientific theories and explanations, but it was difficult to see what might be an alternative to giving a rigorous and insightful account of scientific knowledge. Some philosophers turned to other formal methods such as set theory and probability theory to attempt to provide richer accounts of scientific practice, but mapping them onto actual scientific theories and reasoning has been problematic. Other philosophers of science have taken a more historical approach, but have had to resort to vague notions like the paradigms of Thomas Kuhn and the research programmes of Imre Lakatos to describe the structure and development of scientific knowledge.

Cognitive science provides a whole new set of intellectual tools for addressing issues in the philosophy of science, and cognitive accounts have been proposed by such philosophers as Lindley Darden (2006), Ronald Giere (2002), Nancy Nersessian (2002), and myself (1992, 1999). On my version of the cognitive approach, we should think of a scientific theory as a complex mental representation, a structure in human brains that contributes to various mental processes. The nature of these mental representations varies with different sciences, and not all sciences seem to work with theories that are mental representations of mechanisms of the sort I discussed above as appropriate for theories in cognitive science. Some do: biological theories such as genetics and evolution by natural selection can naturally be understood as representations of mechanisms, and so can many theories in chemistry and many areas of physics. But mathematical theories at the quantum level or qualitative theories in sociology may need to be understood as representations of a different sort.

If theories are mental representations, then explanations are mental processes that apply the theories to mental representations of phenomena to be explained. How this works is best understood by means of computer programs such as the one described in Thagard (1988). It is possible to develop computational models of scientific thinking that have just as much rigor as models relying on formal logic, set theory, and probability theory, but with much greater applicability to actual scientific theories and applications. The mental representations that constitute theories are usually verbal and mathematical, but they can also be visual as we saw with the representations of mechanisms discussed earlier.

Many philosophers such as Frege have thought that the sort of naturalistic, psychologistic account of reasoning that cognitive science offers is incompatible with rationality and objectivity. On the contrary, an approach to the theory of knowledge based on cognitive science can avoid the sheer irrelevance that models based on formal logic and probability theory have to actual scientific practice. Computational models of scientific reasoning can be intended not merely as descriptive of how scientists think, but also as normative of scientific thinking at its best. For example, my theory of explanatory coherence, which has been used to model many important episodes of scientific reasoning including major scientific revolutions, is both descriptive and normative (Thagard, 1992, 1999, 2000). It enables us to see how theory evaluation is both a process that occur in actual human minds and a process that can be thoroughly rational when done right. Because the theory has a direct connection with human psychology, it can also tie in with explanations of cases where rationality fails, for example where personal motivations lead scientists to ignore evidence and alternative theories in ways that make their coherence-based inferences less than rational (Thagard, 2006).

Within logic-based approaches to the philosophy of science, it is difficult to say much about the nature of discovery, one of the most exciting aspects of scientific practice. But if theories are mental representations, then their construction can be explained by specifying mental processes that generate new hypotheses, such as analogy and abductive inference. Claims about processes that are supposed to be sufficient to generate discoveries can be evaluated by building computer programs to see if the processes are computationally feasible and sufficiently powerful to produce the desired discoveries. For example, cognitive scientists have developed computational models of how analogies can be used to generate scientific discoveries.

Hence, just as computational modeling has provided a powerful set of tools for understanding psychological and neurological processes, it can also be used to address central issues in the philosophy of science concerning epistemological processes.

Philosophers do not typically have these tools, but they can be acquired by developing familiarity with the relevant theories and methods from cognitive psychology, neuroscience, and artificial intelligence. A new direction for work in philosophy of science from a cognitive science perspective will develop models of how the brains of human scientists function to understand complex phenomena. For example, Thagard and Litt (forthcoming) have developed a computational model of how thousands of neurons can operate to generate explanations of surprising phenomena. Another promising area of general philosophical research might be to apply the mechanism-based account of interdisciplinary relations that I gave for cognitive science to other combinations of fields, producing a more general theory of reductionism and its limits.

SUGGESTED READINGS

For an accessible interdisciplinary introduction to cognitive science, see Thagard (2005). Boden (2006) provides a two-volume history of cognitive science. On the philosophy of cognitive science, introductions include Goldman (1993) and Clark (2001).

REFERENCES

Bechtel, W., & Abrahamsen, A. A. (2005). Explanation: A mechanistic alternative. *Studies in History and Philosophy of Biology and Biomedical Sciences, 36*, 421-441.

Boden, M. (2006). *Mind as machine: A history of cognitive science*. Oxford: Oxford University Press.

Churchland, P. S. (2002). *Brain-Wise: Studies in neurophilosophy*. Cambridge, MA: MIT Press.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science.* New York: Oxford University Press.

Darden, L. (2006). *Reasoning in biological discoveries*. Cambridge: Cambridge University Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Giere, R. (2002). Scientific cognition as distributed cognition. In P. Carruthers, S. Stich & M. Seigal (Eds.), *The cognitive basis of science* (pp. 285-299). Cambridge: Cambridge University Press.

Goldman, A. I. (1993). *Philosophical applications of cognitive science*. Boulder, CO: Westview.

Nersessian, N. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich & M. Siegal (Eds.), *The cognitive basis of science* (pp. 133-153). Cambridge: Cambridge University Press.

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

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

Thagard, P. (1999). *How scientists explain disease*. Princeton: Princeton University Press.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Thagard, P. (2005). *Mind: Introduction to cognitive science* (2nd ed.). Cambridge, MA: MIT Press.

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.

Thagard, P., & Litt, A. (forthcoming). Models of scientific explanation. In R. Sun (Ed.), *The Cambridge handbook of computational cognitive modeling*. Cambridge: Cambridge University Press.