

Paul Thagard

The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change.

Cambridge, MA: The MIT Press 2012.

xii + 376 pages

\$40.00 (cloth ISBN 978-0-262-01728-2)

The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change is Paul Thagard's (along with some co-authors of some chapters) excellent overview of his many contributions to the field of the cognitive science of science. Thagard is indeed one of the pioneers of the field and thus this book is a very good and well-referenced overview of the cognitive science of science. It is split into four sections. The first section provides a summary of what the cognitive science of science amounts to. Namely, it is the endeavour to understand scientific development and its underlying mental processes from the perspective of cognitive science. Thagard stresses the interdisciplinary nature that the cognitive science of science should have and argues that philosophy, history, and sociology of science can all benefit from the ideas in this book.

The second section of the book (chapters 2 to 6) is concerned with explanation and justification. Thagard presents 'a picture in which explanation is a key aspect of justifying the acceptance of hypotheses and theories' (22). But what is an explanation? Thagard argues that we should think of explanations in terms of the *mechanista* view of scientific method. On this view, science aims for the discovery of underlying mechanisms rather than laws, where a mechanism is a system of interacting parts that produce regular changes. Thagard believes that in the history of science this method has produced deep explanations and has been applied in biology, medicine, psychology, neuroscience, and the social sciences. Thagard then moves on to computational models of scientific explanation, for he argues that the provision, generation, and evaluation of explanations can all be modelled computationally.

Thagard doubts the psychological relevance of computational models that use Bayesian networks and instead argues for his own ECHO system. ECHO 'shows how a very high-level kind of cognition, evaluating complex theories, can be performed by a simple neural network performing parallel constraint satisfaction' (43). Thus, thanks to ECHO 'it is now possible to model the evaluation of competing explanations using more biologically realistic neural networks' (*Ibid.*). Chapter 4 provides a neural account of these mental models by attempting to describe some of the brain mechanisms that produce them. It also attempts to describe how a 'pattern of activation in the brain constitutes a representation of something' (51). This is achieved, according to Thagard, 'when there is a stable causal correlation between the firing of neurons in a population and thing that is represented, such as an object or group of objects in the world' (*Ibid.*). Thagard claims that this is a 'radical departure' from cognitive psychology, but I do not see why that is the case. There are different levels at play here: the computational/representational and the implementational. Of course the mental representations involved are implemented as patterns of firing in neural populations. If the representations are in the brain, then what else can they be? Thagard claims that 'Neural populations can acquire the ability to encode features of the world as their firing activity becomes causally correlated with

those features' (52). However, the notion of causal correlation of mind and world is at best problematic and has been criticised for decades in the philosophical literature, only a small section of which Thagard cites or engages with.

The next chapter shows how Thagard's theory of belief revision applies to the issue of global warming by explaining how scientists come to accept – or not accept in the case of a few scientists and a large number of people in business and politics – the conclusions of climate science. This is a very interesting and thought-provoking discussion that attempts to show that belief revision about global warming can be modelled by Thagard's theory of explanatory coherence, which he has applied to many other cases of scientific belief change. In brief, according to the theory of explanatory coherence, belief revision proceeds by the evaluation of all relevant hypotheses with respect to the evidence. Belief revision then takes place 'when a new proposition has sufficient coherence with the entire set of propositions that it becomes accepted and some proposition previously accepted becomes rejected' (66).

Thagard argues in favour of what he sees is a highly effective and computationally efficient way of modelling belief revision. But the reader is left wondering what the relation is between such models and the cognitive and social factors that are involved in actual scientific practice. Merely stating that such models are implemented in the brain is not enough because other competing theories can also make the same claim. Thagard does cite other works of his and others that discuss this issue, but adding such a discussion to *this* book would have made his case much more convincing.

The third section of the book (chapters 7 to 11) discusses discovery and creativity in science. It shows how cognitive science can contribute to answering questions such as 'How do scientists make discoveries?' or 'How did Newton discover his theory of gravitation?' Thagard also stresses the importance to science education of understanding how scientific discovery occurs. There is a need to motivate students to acquire new concepts, theories, and scientific methods. 'Motivation should be increased if students are not simply force-fed a stock of information to acquire, but can also get some sense of the thrill of figuring out things for themselves' (103).

Chapter 8 attempts to address the question 'What neural processes underlie the wonderful *Aha!* Experiences that creative people sometimes enjoy?' (107). Thagard argues that human creativity requires 'the combination of previously unconnected mental representations constituted by patterns of neural activity', and that 'creative thinking is a matter of combining neural patterns into ones that are both novel and useful' (*Ibid.*). It is noteworthy, however, that Thagard does not cite any linguists in this discussion; indeed, barely any are mentioned in the book. This is surprising, given the amount of work linguists have done on concepts and mental representations and their meaning.

Thagard claims that many scientific discoveries can be understood as instances of the combination of concepts to create a novel concept. However, the 'famous examples' he gives are problematic. One of the examples is 'the wave theory of sound, which required development of the novel concept of a sound wave' (109). The concepts of *sound* and *wave* were supposedly put together to create the novel representation of *sound wave*. But is this really a cognitive science

explanation? It strikes me as more akin to a folk psychological explanation. It says nothing of the underlying mechanisms that cognitive science, according to Thagard himself, is supposed to uncover.

The philosophical controversies about the meaning and content of concepts are not discussed much in this book. Instead, Thagard points to some of his other work and provides a sketch based on Chris Eliasmith's 'neurosemantics' thesis. That said, however, Thagard is correct in wishing to avoid the term *content* 'because it misleadingly suggests that the meaning of a representation is some kind of thing rather than a multi-faceted relational process' (131).

The fourth section of the book (chapters 12 to 16) is concerned with conceptual change and case studies. Chapter 13 shows how resistance to conceptual change 'derives both from (1) cognitive difficulties in grasping the superiority of mechanistic explanations ... and (2) from emotional difficulties in accepting the personal implications of the mechanistic worldview' (200). Chapter 14, which provides an especially good discussion, argues that the cognitive obstacles to adopting evolution by natural selection include 'conceptual difficulties, methodological issues, and coherence problems that derive from the intuitiveness of alternative theories' (219). Thagard discusses the implications that such difficulties have for science education. He concludes that the best strategy is to engage with alternative views and show why the scientific view is to be preferred. Because of this, 'political attempts to keep creationism out of the schools may undermine cognitive strategies that help students appreciate the strengths of Darwinism by contrasting it with theological approaches' (233).

Chapter 15 deals with the dramatic differences between Western and traditional Chinese medicine and argues that there are no insurmountable barriers to a rational evaluation of acupuncture. Chapter 16 deals with the history of explanations and the conceptual changes of mental illness.

The fifth and last section of the book (chapters 17 and 18) offers new directions in the cognitive science of science. Chapter 18 provides a technical account of the content of scientific concepts. It gives good details of several issues that were glossed over in previous chapters but again the crucial controversial issues in, for example, philosophy of mind, are not addressed. In chapter 17 Thagard argues that there has been a general neglect of the values that are part of scientific practice and that there should be a debate about whether there should be a role for social and ethical values in the assessment of scientific theories. Thagard gives an account of values as emotionally valenced mental representations and provides a technique for displaying concepts and their emotional interconnections. He then shows how values can distort scientific deliberations in ways that diminish rationality, but also, positively, how values can 'legitimately affect scientific developments at all three stages of research: pursuit of discoveries, evaluation of results, and practical applications' (284). Values, says Thagard, are inextricable from science because they are irremovable from the minds of scientists.

Eran Asoulin
Independent Scholar