AI Magazine Volume 19 Number 1 (1998) (© AAAI)

Mind: Introduction to Cognitive Science

A Review

Bonnie Holte Bennett, Dwight Nelson, Russell Pannier, Thomas Sullivan, and Gregory Robinson-Riegler

Understanding the mind is one of the great "holy grails" of twentieth-century research. Regardless of training, most people who come in contact with the field of AI are at least partially motivated by the glimmer of hope that they will get a better understanding of the mind. This quest, of course, is a rich and complex one. It is easy to get mired in minutiae along the way, be they the optimization of an algorithm, the details of a mental model, or the intricacies of a logical argument. Thagard's book attempts to call us back to the larger picture and to draw in new devotees—and, in general, he succeeds.

This book begins, "Cognitive science is the interdisciplinary study of mind and intelligence…" (p. ix); so, we assembled a cross-disciplinary review team that included researchers from the fields of AI, cognitive science, neuroscience, and philosophy. This multidisciplinary approach seemed appropriate because this book attempts to be a *bridge book,* written for a wide audience covering these areas and more.

The book is divided into two major sections: (1) Approaches to Cognitive Science and (2) Challenges to Cognitive Science. The first section is a survey of major trends in the research. The second section delineates and analyzes open problems and research issues.

Approaches to Cognitive Science

This section is a broad review of the literature and theories put forth to date. Thagard summarizes results of research to date: "the central hypothesis of cognitive science: Thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures" (p. 10). He uses a shorthand notation for this approach: the computational-representational understanding of mind (CRUM). One wonders at this point just how pessimistic his view will be of research to date. Is CRUM just a nice, pronounceable acronym, or does he intend to include all the overtones associated with the homonymous *crumb* (that is, crummy, crumbly, crumbling)? Section 2 shows that it is somewhere in between.

Paul Thagard, Mind: Introduction to Cognitive Science, *Cambridge, Mass.: The MIT Press, 199,. 212 pp., $22.50, ISBN 0-262-20106-2.*

The chapters in this section are "Representation and Computation," "Logic," "Rules," "Concepts," "Analogies," "Images," "Connections," and "Review and Evaluation." From an AI perspective, these chapters adequately cover the research. Representation and computation are generally recognized as the two major divisions or perspectives on research. Logic, rules, and concepts (meaning frames, semantic networks, and so on) are widely used subdivisions. It is arguable whether the analogies section is at the same level of generality as the previous chapters, but there is a sufficient quantity of work in this area to constitute a separate chapter. The images chapter discusses visual images and processing. This discussion is necessarily at a high level and from the expected cognitive science perspective. Thus, it would be different from what an image-processing researcher might expect. However, the discussion is appropriate given the goal of the book and the perspective of the intended (general) audience. The connections chapter introduces connectionist and parallel distributed processing research in a general way.

Reading this text as a teacher of cognitive psychology, you can spot one of the book's strengths. Our cognitive psychologist writes, "Previously, I had been hesitant about covering cognitive science at any length in my cognitive psychology class, much less offering an entire course in cognitive science. I had yet to see an overview that was truly accessible to novices in the field." Cognitive science textbooks tend to lose the reader in a morass of labyrinthine detail. Thagard's style is clear and readable but still conveys the complexity and breadth of the issues involved. The organization is appealing; it is arranged around the type of mental representation (for example, logic, rules, concepts, images) instead of disciplinary line. This organization helps to make the text more accessible to everyone, regardless of his/her particular area of interest.

One aspect of the literature review that Thagard adds is an analysis of all the approaches (logic, rules, concepts, analogies, images, and connections) in terms of their representational method, problem-solving capabilities, learning approaches, and the types of language that have developed to facilitate these tasks. Thus, you have a neat and useful framework to understand a variety of aspects of these approaches.

In summary, this discussion covers the field and does so in a way that, although unremarkable to a seasoned AI researcher, is accessible to a general audience. In short, this section of the book could be useful in a class, seminar, or discussion group with a diverse audience.

Challenges to Cognitive Science

In this second section, which constitutes the last third of the book, Thagard presents chapters entitled "Emotions and Consciousness," "Physical and Social Environments," "Dynamic Systems and Mathematical Knowledge," and "The Future of Cognitive Science." This area is where the real technical "meat" lies for the seasoned reader. Each chapter presents several interesting challenges to CRUM and explores solutions that include deny, expand CRUM, supplement CRUM, and abandon CRUM. The chapter on emotions and consciousness describes the mind-body problem and explores how emotions and consciousness challenge CRUM. The chapter on physical and social environments discusses the effects of the world and "being in it," with implications for robotics, situated action, the body and direct perception, and intentionality. Thagard also relates these concerns to the social context of knowledge. In the chapter on dynamic systems and mathematical knowledge, Thagard brings in chaos theory and the complexities of large dynamic systems. Mathematical knowledge turns out to be a discussion of Gödel's incompleteness theorem and Penrose's (1994, 1989) version of Gödel's argument against computational views of the mind. In the final chapter, on the future of cognitive science, Thagard reviews the vast array of open questions and extends an invitation to interested parties to join the quest.

From an AI perspective, these chapters seem to do a good job of delineating the major objections and concerns about the field. Many AI texts touch on one or several of these objections, but Thagard clearly categorizes them and provides instructive delineations. His analysis of the possible responses to the objections is rigorous in its repeated analysis of alternatives: deny, expand, supplement, and abandon. Although, occasionally, the distinction between expand and supplement is lost on the reader.

From a cognitive psychology perspective, Thagard does a nice job of incorporating cognitive psychology theory and research into his general overview of cognitive science. His discussion manages to include both traditional research and more recent developments that might be placed under the rubric of ecological approaches to the study of cognition. Ecological approaches emphasize the study of cognition within its everyday context. Relevant topics discussed by Thagard include the relationship between cognition and emotion and the important role that social context plays in cognition.

From a philosophy perspective, Thagard's attempt to deal with some challenges to CRUM has the virtue of bravely accepting the difficult facts. Defenders of CRUM sometimes deny the existence of consciousness and intentionality, but Thagard wants no part of the "eliminativist's" way out. He instead makes broad proposals about expanding and supplementing CRUM. Perhaps a richer account of representation and computation will help with consciousness, and he theorizes, biological considerations might have to be invoked to help explain intentionality if we are to avoid dualism or any other anticomputational view.

Gödel's incompleteness theorem: One of the challenges to CRUM taken up is based on Gödel's incompleteness theorem. The discussion focuses on a particular version of the challenge formulated by Roger Penrose. In general, although Thagard has some interesting things to say, his discussion is somewhat difficult to follow. This is unfortunate given that the text purports to be an introduction and the likelihood that at least some readers will presumably come to the book lacking a background in the theory of computable functions.

Here is Thagard's formulation of Penrose's argument:

First, anything a computer can do, a Turing machine can do; so, any task that no Turing machine can do is a task that no computer can do.

Second, for any Turing machine, we can devise a task that it cannot do, namely, define a computation such that the Turing machine will not be able to tell whether it stops.

Third, human mathematicians, if they know that the computation is sound (consistent, error free), can tell that the Turing machine does not stop.

Fourth, humans can do something that no Turing machine and, hence, no computer can do, namely, recognize mathematical truth.

One of the difficulties with the formulation is that the second and fourth steps themselves express inferences. Of course, this problem is not serious, but it does force the careful reader to break the steps apart.

Another difficulty is that the steps are not formulated as precisely as one might want. For example, what exactly is the quantificational structure of step 2: "For any *x,* there exists a *y*…," or "there exists a *y* such that for any *x*…"? Of course, these are not equivalent forms. Literally read, step 2 appears to use the first form, but it's not clear that the first is strong enough for the intended conclusion.

Another ambiguity in step 2 concerns the phrase "a task that it cannot do." What exactly is this task—defining a computation such that an arbitrarily selected Turing machine will not be able to tell whether it stops or having an arbitrarily selected Turing machine determine whether it will stop?

A more serious problem with the formulation is that the argument does not appear to be formally valid as it stands. Just one indication is the fact that the conclusion contains a predicate ("recognize mathematical truth") that does not appear in any of the premises.

In any case, it's unclear exactly how Thagard's response to Penrose's argument relates to its deductive structure. His apparent rejection of the conclusion ("Penrose has not shown something that a human mathematician can do that no computer could…" [p. 178]) logically commits him to challenging either the argument's validity, its soundness, or both, which is where things get a bit murky. He says that even if a full cognitive model (computer-aided machine [CAM]) of a human mathematician could be constructed (an unlikely accomplishment according to the author), it is improbable that any human mathematician would ever be able to construct the Turing machine equivalent of a CAM or prove that a CAM is sound. He closed the discussion with, "Since CAM is not using a knowably sound algorithm, it is not obviously different from a human mathematician, characterized in Penrose's ***G*** as also not using a knowably sound algorithm" (p. 178). How precisely does this relate to the deductive structure of Penrose's argument? Does he maintain that it is formally invalid? If so, what does he think are the missing premises? However, does he maintain that a premise is false? If so, which one? In this regard, it is not obvious that anything Thagard says explicitly denies any of the steps he specifies as premises (1, 2, and 3).

Conclusions

Thagard presents a broad introduction aimed at a general audience and generally succeeds very well. His organization is appealing and accessible to a variety of disciplines. Both experts and novices will find it a rewarding read, with much fodder for discussion. The small section on Gödel's incompleteness theorem raised more questions for us than it answered, but this, too, can be a virtue.

References

Penrose, R. 1994. Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford, U.K.: Oxford University Press.

Penrose, R. 1989. The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics. Oxford, U.K.: Oxford University Press.

Bonnie Holte Bennett is director of the Artificial Intelligence Laboratory and an associate professor in the Graduate Programs in Software at the University of St. Thomas. Her e-mail address is bhbennett@stthomas.edu.

Dwight Nelson is a professor in the Biology/Neuroscience Department at the University of St. Thomas.

Russell Pannier is a professor at the William Mitchell College of Law.

Thomas Sullivan is a professor in the Department of Philosophy at the University of St. Thomas.

Gregory Robinson-Riegler is a professor in the Department of Psychology at the University of St. Thomas.