AI Magazine Volume 7 Number 1 (1986) (© AAAI)

BOOK REVIEWS

Artificial Intelligence: The Very Idea. John Haugeland. Cambridge, Massachusetts; The MIT Press, 1985. 287 pp.

In his introduction, the author states three goals for his book: "to explain, clearly and with an open mind, what AI is really all about; second, to exhibit the philosophical and scientific credentials behind its enormous appeal; and finally, to take a look at what actually has and has not been accomplished." Readers who are willing to accept the author's definition of AI will find that these three goals have been met quite well. AI is not viewed in this book as a particular collection of tools and techniques, nor is it seen as an effort to make machines efficiently perform certain tasks which are done well at present only by people. The author ignores definitions of the field which do not promote the assumption that one is grappling with fundamental questions having vast implications. He defined artificial intelligence as the attempt to create "machines with minds, in the full and literal sense." Fortunately, this most dramatic of definitions is not used as the excuse to weave an intricate tangle of abstruse speculations from which few ideas and fewer readers would emerge. To the contrary, the author's discussion is generally very clear, concrete, accurate, and interesting. My main regret is that the author does not take (or was not given by his editors) the space to more fully explore certain ideas.

In the first chapter, AI (as defined by the author), is placed in historical perspective as the most recent of investigations into the relations which hold between our mental and physical universes. Overviews of the work of Copernicus, Galileo, Hobbes, Descartes, and Hume take the reader from the ancient commonplace that things are not always what they seem to the more modern view that there is no intrinsic connection between thoughts and their alleged objects. Thus, we are faced with two questions which lie at the heart of AI: What makes a notation suitable for symbolizing some subject matter? What makes a suitable notation actually symbolize that subject matter? The author mainly addresses the second question, by discussing the meaning of meaning at various points throughout the book.

The next three chapters exhibit some of AI's credentials by discussing "interpreted automatic formal systems," one instance of which is purposefully programmed computers. The only programs discussed at any length are SHRDLU and GPS. As implied by the author's choice of goals for the book, AI has many credentials apart from its repertoire of working programs. Instead of focusing closely on existing programs, the discussion of AI's credibility revolves around the plausibility of certain assumptions. Perhaps the most obvious of these is "medium independence," the assumption that "essentially the same formal system can be materialized in any number of different media, with no formally significant difference whatsoever." The media presently of concern to AI researchers, of course, are neurons and transistors. The author discusses other assumptions as well, including the very interesting one that "just as our smooth visual experience is somehow based in a 'grainy' retina, perhaps our own easy, flexible good sense is ultimately based in (sufficiently fine grained) rules and stereotypes." By the way, the author is neither strongly "pro-AI" nor "anti-AI"; rather, he takes the stand that AI is based on some very good ideas which may or may not be correct.

I was disappointed to find no references to Schank or any other natural language understanding researcher in the chapter "Semantics." I also wish the ideas of a "semantic division of labor" and reinterpretation of symbols had been more fully explored. On the other hand, the overview of Babbage's, Turing's, Von Neumann's, McCarthy's, and Newell's virtual machines in the chapter "Computer Architecture" could not be better. This chapter concludes that even though, for reasons of convenience, most AI programs are written in LISP, "the mind could have a computational architecture all its own. In other words, from the perspective of AI, mental architecture itself becomes a new theoretical "variable," to be investigated and spelled out by actual cognitive science research." This conclusion is a natural lead-in to a chapter discussing current work on knowledge representation, but there is no such chapter. Instead, the author turns to his third goal of looking at what has actually been accomplished in AI. He discusses early work on machine translation of natural languages, and describes the behavior of the programs GPS and SHRDLU. Here again, the reader could have benefited from more discussion of current AI work, although some references are given in the footnotes. Knowledge representation is discussed in a very general way. Schank's scripts, Minsky's frames, Bartlett's schemata, and Husserl's noemata are all thrown together in the single footnote relating to actual AI work; these are referred to collectively in the text as "linked stereotypes." There is, however, an interesting look at what the author calls the "frame problem." The example given is the task of correctly updating a knowledge base which represents a simple real-world physical situation after one of the objects being represented is moved.

In the final chapter, the author examines some of the fundamental differences between people and programs. The chapter is full of little gems. For example, the author points out that people follow an "ascription schema." That is, we ascribe beliefs, goals, and faculties so as to maximize a system's overall manifest competence. If someone says "Careful! That chair is hot!" and then sits on the chair himself, we will conclude that he lied to get the chair for himself, that he enjoys hot seats, or make some similar conclusion which maximizes our opinion of his competence. The author points out that “the ascription schema constrains mental ascriptions once a system is specified; but it puts no limit on which systems should have mental states ascribed to them.” He postulates a “Supertrap” which strikes matches in the presence of gas-soaked mice, topples dictionaries on mice, and, of course, snaps shut whenever mice nibble its bait “These habits betray a common malevolent thread, which is generalizable by (and only by) ascribing a persistent goal: dead mice.” When we see other Supertrap behaviors, such as failure to harm cats that reek of gasoline, we become involved in a “semantic intrigue,” an effort to understand how mental ascriptions cohere and interact. Whimsical examples aside, ascription is important for AI because it provides one more way to detect patterns that might otherwise go unnoticed. The ascription schema is proposed during the author’s discussion of people’s pragmatic sense. The final chapter also examines other fundamental differences between people and programs: our use of mental images, feelings, and ego involvement. Even if this chapter were not as thought-provoking and enjoyable as it is, it would be worth reading simply to remind oneself how extremely difficult problems in AI can (should?) be.

John W. L. Ogilvie
Modula Corporation
Provo, Utah

Heuristics: Intelligent Search Strategies for Computer Problem Solving. Judea Pearl. Addison-Wesley Publishing. 1984. 382 pp.

The view of AI science offered by Judea Pearl is thoroughly traditional and standard, and therein lie both its strengths and its weaknesses as a monograph, a reference, or a textbook in its field. As a graph-theoretic analysis of search strategy that clearly conforms to well-established AI methods and techniques, it expands upon these to incorporate probabilistic performance analysis principles, thus providing a (partial) formal framework of search strategy, evaluation criteria, and decision methods that are all firmly grounded in operations research.

To those readers for whom mathematical logic and probability calculus represent the most promising theoretical foundations of AI science, especially if understood in terms of graph theory and standard probabilistic models, this book will be quite useful and illuminating for the purposes of a textbook and as a reference. Pearl’s survey of search strategies with respect to various probabilistic features of “heuristic information” provides valuable insights for general readers, students, and practicing researchers alike. From this perspective, the strength and value of Pearl’s work will not be questioned here. For the purposes of teaching and promoting the general aspects of that theoretical approach, his book is clearly worthwhile and even innovative. Granting all of this, the only complaint that might be raised is altogether excusable, if not also entirely minor, *i.e.,* that the material presented might not be so easily grasped by the “casual reader” as the author supposes.

Discursively considered, however, and especially for the purposes of AI research, these very same strengths can be seen as weaknesses from the viewpoint of at least two alternative approaches: (1) nonformalist or antiformalist theories, which completely reject standard mathematical logic and traditional probability theory; or perhaps, (2) nonstandard or alternative formal theories, which can displace those views as prevailing paradigms. Now it clearly was not Pearl’s aim to forestall alternative theories or to justify his own approach in contrast to other views. The comments that follow are not being offered as criticisms *per se.* They should instead be regarded as advice for those who may wish to pursue such alternative approaches, but who could benefit from a survey of precisely that direction in AI science they might ultimately choose to oppose, for reasons of their own.

Advocates of nonformalism and antiformalism in AI science tend to regard “heuristics” as their last line of defense, so to speak, against formal encroachment upon their research territory, as Pentland and Fischler (1983) or Bierre (1985) stand opposed to Nilsson (1983), for example, or as the notorious “Great Debate” runs, in general. Pearl’s functional analysis of heuristics as the (somewhat arbitrary) catalyst for algorithmic procedures does not yield “heuristics” at all on this view, it seems, since these are “formally ineffable” by virtue of being exactly that which algorithms are not. The objection that Pearl’s analysis is pervasively algorithmic, however, has some merit after all, if the “algorithmic properties” of “heuristic methods” (*i.e.,* those of completeness,” admissibility,” “dominance,” and “optimality” in Chapter 3) are just the kinds of properties that “heuristics,” by definition, cannot have. But it should come as no surprise to any antiformalist that these are the kinds of properties any formalist would seek to identify and establish, even under the name of “heuristics.” Yet this does not count against the analysis itself, nor does it diminish the usefulness (in its particular domain) of the search strategies, evaluation criteria, and decision methods provided by Pearl’s account.

Pearl’s conception of heuristics as rules of thumb, intuitive judgments, educated guesses, and common sense hints at their subjective character as inferential guidelines that are defeasible in light of new information. In particular, he defines these techniques as “strategies using readily accessible though loosely applicable information to control problem-solving processes in human beings and machine(s)[1] (p. vii). As such, Pearl’s heuristics that

[1] On the notion of a scientific paradigm, see Kuhn (1970)