Book Reviews

Automated Reasoning: Thirty-Three Basic Research Problems

Ulrich Wendl

To read the book *Automated Reasoning: Thirty-Three Basic Research Problems* (Prentice Hall, Englewood Cliffs, N.J., 1987, 300 pp., $11.00) by Larry Wos "it is not necessary to be an expert in mathematics or logic or computer science" (from the preface). However, even if you are such an expert, you will read it with interest, and likely, with enjoyment.

The book is outstanding for its presentation of the theme. Following the introductory chapter, Wos discusses some obstacles to the automation of reasoning in Chapter 2. In Chapter 3, he lists the research problems (with short descriptions) in nine groups: six problems on strategy, five on inference rules, six on demodulation, one on subsumption, three on knowledge representation, two on global approach, one on logic programming, two on self-analysis, and six on other areas. After a short review of automated reasoning (AR) in Chapter 4, these problems are discussed in detail in Chapter 5. Chapter 6 gives some sets of test problems, all concerning a mathematical discipline. An appendix as interesting as the bibliography follows. Last but not least there is an excellent index.

The discussion of the obstacles in Chapter 2 is relaxing; although some repetition exists in the book, the author spares you a puffy pseudophilosophical treatise on AI in general and AR specifically. After reading this chapter, you are convinced there is no general problem solver, and you have a pleasant introduction to the belly of the beast.

A non-expert might have some problems in mapping the eight obstacles described to the problems and problem areas cited earlier. However, the detailed discussion that forms the heart of the book provides the reader with ample reward. It is clear that the

author writes with a freshness and an obvious love for logic.

Between the listing of the problems and their detailed discussion is a review of AR. It seems to be too sketchy for the novice, but a companion book by the same author (Wos, L.; Overbeek, R.; Lusk, E.; and Boyle, J. 1984. *Automated Reasoning: Introduction and Applications*. Englewood Cliffs, N.J.: Prentice-Hall.) gives far more details.

Chapter 5, the in-depth discussion, is remarkable for its style. It is fascinating to immerse oneself into the details guided more by questions from the author than answers; indeed, a whole paragraph consists of nothing but questions! You will enjoy this chapter and the whole book, especially if you dislike the 'definition-theorem-proof' style. The author's presentation lets you forget that completely solving one of the problems is considered equivalent to finishing a Ph.D.!

A non-expert will have the most difficulty with the given test problems primarily because they are all out of the field of mathematics. Again, it is not trivial to see a correspondence to the problems or the obstacles. At this point, it is likely the author's opinion—"zeal, interest, and curiosity" suffice as prerequisites—is too optimistic. However, this book is also a workbook, and the appendix shows the reader how to get the appropriate software. Finally, besides a machine such as a VAX, you only need one thing: a lot of time. Perhaps the best thing to say about the book is that it tempted me to wait for a copy of the companion software.

The book as a whole is a rarity in that it successfully serves several audiences at the same time. The layperson gets a real background knowledge of one of the main disciplines of AI, and the theorist gets a good occasion to put the theory to practice. Most astonishingly, both can enjoy it.

One last point is worth mentioning. It seems to be a law that the

price of a (computer science) book is inversely proportional to its content. This book confirms this law: It is delightfully low priced. As Georg Christoph Lichtenberg, a German physicist of the eighteenth century said: "If you have two trousers, sell one, and buy this book."

Ulrich Wendl is at the SCS Information-stechnik GmbH, Hoerselbergstr. 3, 8000 Munchen 80.

Logical Foundations of Artificial Intelligence

Drew McDermott

Knowledge is important to intelligent programs: Just about everyone in AI would agree, but the agreement doesn't extend much further. There is a weak sense of "know" in which a computer program knows P if its correctness depends on P. For instance, an airline reservation system might "know" every passenger has a 0.9 probability of showing up for a reserved flight in the sense that it books 11 percent too many passengers. However, more interesting possibilities exist. A program might have a notation that facts can be expressed, and it might consult a database of such facts to move through problems. In this case, we have a more explicit concept that a program knows P if P is in the database (or can be derived from it when required). This concept is the idea of *declarative knowledge;* the more mundane concept is called *procedural knowledge.* The book *Logical Foundations of Artificial Intelligence* (Morgan Kaufmann, Los Altos, Calif., 1987, 406 pp., $48.95) by Michael Genesereth and Nils Nilsson is about the declarative version. (Declarative knowledge could be false, and we would do better to call it belief, as I often do in this review.)

Declarative knowledge requires a notation, often called a knowledge representation system. Those who

study declarative knowledge representations are divided about whether such notations ought to be thought of as variants of the predicate calculus and related systems of mathematical logic invented by logicians, mathematicians, and philosophers in this century. Lately, those who think they ought to be so regarded seem to be winning. Genesereth and Nilsson have no doubt and take it for granted that the semantic tools of mathematical logic are indispensable for analyzing knowledge representations. They adopt the label *logicism* for this doctrine.

In Chapter 2, they develop these tools from the point of view of AI, which is rather different from the logician's point of view. They introduce the idea of *conceptualization,* which is roughly what a philosopher calls the intended model of a formal theory. A formal theory provides predicates and functions for talking about (some part of) the world and axioms involving these symbols. Formal semantics specifies a mapping from the symbols to the entities in the world. It is one of the key insights (and disturbing insights) of mathematical logic that unique mappings are rare. Any given theory has many interpretations that make its axioms true, that is, several models. Mathematicians cheerfully study them all, but philosophers worry more about how the correct model can be picked out. Fortunately, in AI, we can ignore these problems and treat ontological commitment as an engineering decision. It doesn't matter that the robot we build cannot intend a model; its builders can. Genesereth and Nilsson give a lucid explanation of this topic.

When they reach the topic of inference, however, they seem to lose their way. Chapters 3 through 5 are concerned with this topic, but it is central to the entire book. Inference can be considered as the deriving of probably useful, probably correct conclusions from a set of beliefs. This characterization is vague, but the vagueness is inevitable. Almost any algorithm can be thought of as doing inference in some sense, and if we arrange that its premises and conclusions are expressed in a logical notation (not a difficult requirement), then it can be thought of as doing inference on a declarative knowledge representation. The authors give several examples throughout the book. Chapter 7, for instance, is devoted to

concept learning, in which new rules are inferred from facts that would follow from them. The algorithms described—based on Tom Mitchell's idea of version spaces—are specific to the domain of the concept of learning. Similarly, Chapter 8 describes Bayesian inference, from a priori probabilities to a posteriori probabilities given new evidence.

In competition with this diversity is the idea of a unified model of inference. The desire for such a model is strong among those who study declarative representations, and Genesereth and Nilsson are no exception. As are most of their colleagues, they are drawn to the model of inference as the derivation of conclusions that are entailed by a set of beliefs. They wander from this idea in a few places but not for long. It is not hard to see why: Deduction is one of the fews kinds of inference for which we have an interesting general theory.

The authors made a conscious decision not to talk much about search processes. Unfortunately, a deductive process is always confronted with the problem of what to deduce next; it is impossible to correctly choose with any confidence, so programs that deduce must try various options. Sometimes, they can generate many useless inferences before finding useful ones. Chapters 4 and 5 are devoted to the topic of implementing deductive processes using the resolution method with various refinement strategies. However, throughout the rest of the book, little focus is given to the actual computational consequences of relying on these methods. Usually, proofs are given with a passing warning that finding the proof might be expensive.

It's clear what the authors are thinking. They are studying foundations; so, what's important is what should be inferred in a situation, not what actually can be practically inferred (compared, say, to theoretical mechanics). However, as I argued earlier, what should be inferred is probably different from what is entailed by current beliefs because not all that is entailed is useful and not all that is probably correct is entailed. The authors are forced to acknowledge these discrepancies in their detours to alternative inference techniques in Chapters 6, 7, and 8, but their heart isn't in it. When it comes to actual applications, they always revert to classical deduction. The unsophisticated reader should be

warned that the foundations being explored are not exactly the foundations of AI.

The authors' viewpoint causes distortions throughout the book. The chapter on planning (Chapter 12) is typical. The planning problem is defined as finding a constructive proof that a series of actions will bring about a desired state of the world. It is assumed that the right way to find this proof is to tailor a general-purpose theorem prover with a few specialized strategies. This description of the planning problem is quite remote from any description that active researchers in the field would produce. Perhaps Genesereth and Nilsson feel that techniques for solving planning problems will come and go, but the foundations can be secure. If so, they are too casual about the implicit claim.

I admit to bias on these questions. My skepticism should be balanced by the agreement of many perfectly reasonable people with Genesereth and Nilsson's view. However, it bothers me that questions about the scope of their enterprise are treated so superficially in this book.

Let me put such doubts aside and pretend from now on that deduction is the foundation of AI. Within this perspective, the book has some strengths and some weaknesses. Chapter 6 is an excellent discussion of *nonmonotonic logic,* a family of logic extensions that allow defeasible conclusions, which can be blocked by knowing more. Most real-world inference is nonmonotonic because knowing more usually causes one to change one's mind in some way. For example, if told that person P is an adult living in suburbia, you would probably infer P owns a car. Now, suppose the further information that P is blind. It is a problem with the logicist approach to infer that logic lacks this ability; if a proposition is entailed by a set of beliefs, it is entailed by any superset. Various nonmonotonic variants of logic have been proposed. They are all covered in Chapter 6, especially John McCarthy's circumscription, which augments a standard theory with a special axiom schema that can change nonmonotonically as new beliefs are added. This description of circumscription is the best I have seen; no one could ask for more from a textbook.

Chapter 7, on induction, is too short. *Induction* is defined as the

problem of finding a hypothesis that entails observed data. An excellent description follows of *version spaces,* a useful framework for thinking about sets of hypotheses that have not been ruled out by data observed so far. However, no mention is made of Ehud Shapiro's or Gordon Plotkin's generalizations of the idea to a logical framework, which is an amazing omission. It would have been nice to see a discussion of *explanation-based generalization,* a more recent idea of Mitchell's in which a problem solution, expressed as a proof, is generalized and stored as a new lemma. Perhaps, this idea is too recent, but it would fit the authors' world view well.

Chapter 8, on probabilities, is weaker. It deals with two topics, with little hint about how they are related or how much of the problem of reasoning under uncertainty is covered. The first topic is Bayesian inference, which underlies much work on expert systems. The discussion is clear and helpful. The second topic is Nilsson's probabilistic logic, an attempt to generalize logical entailment so that the probability of a conclusion can be found given the probabilities of some premises. It is not clear what the theoretical or practical significance of probabilistic logic is. The maximum-entropy method is briefly mentioned here; I wish it had been covered in more depth because it is of interest in its own right.

Chapter 9 is a description of the logic of knowledge and belief using two different conceptualizations: Kurt Konolige's syntactic approach and J.K.K. Hintikka's possible worlds approach. This discussion is lucid and thorough. In the syntactic approach, belief is a predicate on formulas in an agent's database. In the possible worlds approach, an agent does not believe P if for all the agent knows, P might be false; and this statement is formalized as "There exists a world, possible as far as the agent believes, in which P is false." In either approach, we can make inferences such as "If Fred knows that Mary's phone number is the same as John's, and he knows Mary's phone number, then he knows John's" without knowing what the phone number is.

Chapter 10 is about *metareasoning,* or reasoning about reasoning. This concept mesmerizes many in AI, probably because it seems intimately connected with our ability to consciously introspect and observe ourselves thinking. However, except for

this connection, the value of metareasoning as a programming or representation technique has seldom been demonstrated. Chapter 10 is mainly concerned with pointing out various alternative architectures for metareasoning. Little discussion is given of what it is for. One possibility is for reasoning about the reasoning of other agents, which connects to the syntactic belief calculus of Chapter 9. However, the possibility is not really explored, except for a baffling detour into an alternative formalization of belief that is never related to the approaches of Chapter 9.

Chapters 11 and 12 are about temporal reasoning and planning. They are out of date, based on the situation calculus devised by John McCarthy and Patrick Hayes in the late 1960s. A lot of work has been done in the last 20 years in the logicist tradition on formalizing temporal reasoning to handle continuous time and alternative possible worlds. It is mentioned only in the bibliographic notes.

Chapter 13 is entitled Intelligent-Agent Architecture, but it is nothing of the sort. It is hard to say what it is about. I think it was meant to be a nod toward robotics.

In summary, this book has some excellent parts, notably its treatments of formal semantics, nonmonotonic logic, and logics of knowledge and belief. However, it omits detailed discussion of the work of many logicists, including James Allen, Alan Bundy, Ernie Davis, Patrick Hayes, Robert Kowalski, Ray Reiter, and Udi Shapiro. Opportunities were missed for tracing a consistent thread through the topics covered. Nonmonotonic reasoning is rarely mentioned after Chapter 6, even though it is relevant in several subsequent chapters. Reasoning about reasoning is covered in Chapters 9, 10, and 13, each time from a different perspective. Finally, the book is unself-conscious about the importance of logic in AI, which might be less than the authors believe.

Drew V. McDermott is a professor at the Computer Science Department, Yale University, P.O. Box 2158 Yale Station, New Haven, CT 06520.

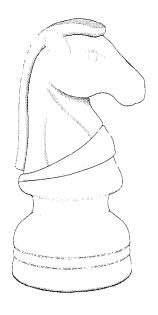# Response to Drew McDermott's Review of *Logical Foundations of Artificial Intelligence*

Nils Nilsson

McDermott makes some valid points in his review. I acknowledge some of these in this response. It's too bad that he chose to embed the helpful comments in the context of his by-now-tiresome doubts about the value of logic in AI. (The reader who wants to be saturated with the arguments surrounding these doubts should see McDermott's 1977 A Critique of Pure Reason and the accompanying commentary in *Computational Intelligence,* 3(3).

McDermott accurately summarizes our book as being about the representation and use of declarative knowledge in AI. He duly notes that declarative knowledge requires a notation and that some AI researchers think most of the notational schemes being used are variants of the predicate calculus. He even says, "Lately those who think they ought to be so regarded seem to be winning." Under these circumstances, it does seem odd for McDermott to devote much space to complaining about the logical basis of a book whose very title proclaims it is about logical foundations. In any case, given such a title, it wouldn't seem necessary that readers "should be warned that the foundations being explored are not exactly the foundations of [all of] AI."

It seems one of McDermott's main complaints about the logical approach is based on his view that "logicists" think of inference as being simply sound deductive inference. Actually, we happen to agree with McDermott's statement that "almost any algorithm can be thought of as doing inference in some sense, and if we arrange that its premises and conclusions are expressed in a logical notation (not a difficult requirement), then it can be thought of as doing inference on a declarative knowledge representation." He inappropriately and unfairly characterizes our treatment of various nondeductive inference techniques (nonmonotonic reasoning, inductive reasoning, and probabilistic reasoning) as mere detours from what he would like to regard as our commitment to deductive inference. We had hoped it would be obvious that our commitment regarding inference isn't exclusively

to soundness but to an understanding of the underlying theoretical properties of various inference methods. Soundness is just one property that an inference method might or might not have; minimal-model entailment and version-space properties are others We are curious about what subtleties McDermott found in Chapters 6, 7, and 8 to make him think that "our heart wasn't in it."

McDermott's specific criticisms are better than his general ones. He makes some good points about our chapters on planning. Our goal in these chapters was to present the fundamental conceptualization of the state-based approach and the situation calculus based on it. No foundations text could omit these basic and classical ideas. We agree with McDermott that it would have been good to include more material on practical planning (for example, hierarchical planning, opportunistic planning, and constraint-based planning). It is hard, though, to decide which of this additional material is foundational. A

chapter or more on temporal reasoning and how time-based (rather than state-based) techniques might be used in planning would also have been useful.

McDermott appropriately observes that our chapter on induction was too short.

The last chapter in the book, Chapter 13 on agent architectures, is admittedly speculative and strayed somewhat close to research frontiers (as did some of the other chapters, perhaps more successfully). We think that the material in Chapter 13 presents an interesting way to think about an agent which is embedded in a world that it must sense and in which it must act. We like to think of it more as a foundation for thinking about robots than as "a nod toward robotics."

Nils J. Nilsson is a professor and chairman of the Computer Science Department, Stanford University, Stanford, CA 94305.