Mind Design II
Philosophy, Psychology, and Artificial Intelligence

by John Haugeland, editor

A Bradford Book
The MIT Press
Cambridge, MA, 1997
476 pages
ISBN 0-262-58153-1 (paper) $19.50
ISBN 0-262-08259-4 (cloth) $40.00

Reviewed by Varol Akman

Department of Computer Engineering and Information Science
Bilkent University, Bilkent, Ankara 06533, Turkey
akman@cs.bilkent.edu.tr

all things contain the seeds
of their own completion.
all seeds contain the things
of their own destruction.

from "Feasts of Death, Feasts of Love"
-Stuart Z. Perkoff (1957)

John Haugeland is a professor of philosophy at the University of Pittsburgh. He is well known in AI circles. With interests--among others--in the philosophy of mind and philosophical psychology, he has given numerous thoughtful works in the past and continues to do so. The title under review, *Mind Design II*, is the second, revised version of his 1981 book *Mind Design* (MD in the sequel). His other books include (Haugeland 1985), and the most recent (Haugeland 1997) which this reviewer has not yet seen.

Mind Design II (MD2 in the sequel) continues in the tradition of quality one naturally associates with its editor. It is a book worthy of careful study and is heartily recommended to the readers of this *Bulletin*, particularly to those with an interest in the philosophical aspects of AI. Many of the essays included in MD2 assume little specialized background, tend to be self-contained, and can be profitably read by someone with a general appreciation of the main ideas of AI. MD2 is a sizable yet cheap book (at

least the paperback version) and can be an excellent choice for an "AI & Philosophy" sort of course, especially at the graduate level.

With these somewhat standard but genuinely felt remarks out of the way, let me now take a look at MD2 in finer detail. In the "Acknowledgments" section of MD (Haugeland 1981, p. vii) Haugeland said:

> This book is conceived as a sequel to Alan Ross Anderson's *Minds and Machines* (1964), augmenting it and bringing it up to date. All of the essays included here were written after that work appeared, and I daresay most of them were directly or indirectly influenced by it. There is, of course, no duplication of contents, and there could be no replacing the classic articles in that little volume; I can only hope to have chosen as well from the burgeoning subsequent literature.

It can safely be said that Haugeland's choices were most fitting; MD quickly became a classic and found a secure place for itself in philosophy and AI circles. It also served to introduce important contemporary philosophers of mind to the AI community. The revised and enlarged successor of MD consists of the following chapters:

1. What is Mind Design? (by John Haugeland)
2. Computing Machinery and Intelligence (by A. M. Turing)
3. True Believers: The Intentional Strategy and Why It Works (by Daniel C. Dennett)
4. Computer Science as Empirical Inquiry: Symbols and Search (by Allen Newell and Herbert A. Simon)
5. A Framework for Representing Knowledge (by Marvin Minsky)
6. From Micro-Worlds to Knowledge Representation: AI at an Impasse (by Hubert L. Dreyfus)
7. Minds, Brains, and Programs (by John R. Searle)
8. The Architecture of Mind: A Connectionist Approach (by David E. Rumelhart)
9. Connectionist Modeling: Neural Computation / Mental Connections (by Paul Smolensky)
10. On the Nature of Theories: A Neurocomputational Perspective (by Paul M. Churchland)
11. Connectionism and Cognition (by Jay F. Rosenberg)
12. Connectionism and Cognitive Architecture: A Critical Analysis (by Jerry A. Fodor and Zenon W. Pylyshyn)
13. Connectionism, Eliminativism, and the Future of Folk Psychology (by William Ramsey, Stephen Stich, and Joseph Garon)
14. The Presence of a Symbol (by Andy Clark)

15. Intelligence without Representation (by Rodney A. Brooks)
16. Dynamics and Cognition (by Timothy van Gelder)

Of these, Chapters 4, 5, 6, and 7 survive from MD. The following chapters from MD have disappeared:

2. Complexity and the Study of Artificial and Human Intelligence (by Zenon Pylyshyn)
4. Artificial Intelligence--A Personal View (by David Marr)
5. Artificial Intelligence Meets Natural Stupidity (by Drew McDermott)
7. Reductionism and the Nature of Psychology (by Hilary Putnam)
8. Intentional Systems (by Daniel C. Dennett)
9. The Nature and Plausibility of Cognitivism (by John Haugeland)
11. Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology (by Jerry A. Fodor)
12. The Material Mind (by Donald Davidson)

It is fair to note that of these Pylyshyn, Dennett, Fodor, and Haugeland appear in MD2 with their newer contributions. So, in some sense, only Marr, McDermott, Putnam, and Davidson are truly missing. I would love to see at least the last two of these contribute an essay to MD2. They are both remarkable philosophers, with exciting things to say. In case of Putnam, his (now recanted) work on functionalism, or better yet, his newer insights in the post-functionalist tradition would make lots of sense (Guttenplan 1994, p. 513). In case of Davidson, an essay explicating his doctrine of anomalous monism would be apt. Davidson's thesis is that the nature of mental phenomena does not permit law-like regularities connecting the mental phenomena with physical events in the brain. In his own words (Guttenplan 1994, p. 231): "There are no such things as minds, but people have mental properties, which is to say that certain psychological predicates are true of them. These properties are constantly changing and such changes are mental events." (ASIDE: In a similar vein, Searle once said that weather concepts such as "partly cloudy in London" are not systematically related to the concepts of physics.)

Haugeland's introduction to MD ("Semantic Engines: An Introduction to Mind Design") is replaced by Chapter 1 of MD2, serving an analogous role. Here, Haugeland defines mind design as the endeavor of understanding mind in terms of its design. Accordingly, AI, with its historical mission to construct intelligent artifacts, makes up the core of mind design. There is also a cognitive psychology flavor to all this; in the memorable phrase of Haugeland mind design can be conceived as "psychology by reverse engineering."

A fundamental and, in fact, long overdue addition to the new volume is Alan Turing's landmark paper--contributed to *Mind* 59: 433-460 (1950)--which, by the way, is available on-line (see the recent Turing Test Homepage at

http://www.cs.bilkent.edu.tr/~psaygin/ttest.html). It is one of most beautiful pieces in philosophical AI, with a forever fresh, optimistic insight, and prophetic vision.

John McCarthy likes to suggest (perhaps somewhat whimsically) that even devices as simple as thermostats have beliefs. A simple thermostat has, at any given time, one of the following beliefs: it's hot in here, it's chilly in here, and it's just comfortable in here. But what is a belief anyway? How does it acquire the content it has? Dennett studies these questions and more in Chapter 3. His essay, written in the exceptional, inimitable style of the author, comes from his (Dennett 1987). In it he describes a strategy which he calls adopting the intentional stance. In a nutshell, Dennett's criterion to be a true believer is to be an intentional system, a system whose behavior can be reliably predicted by way of the intentional strategy.

Newell and Simon's contribution (Chapter 4) is their 1975 Turing Award Lecture, published in *Communications of the Association for Computing Machinery* 19: 113-126 (1976). While somewhat simpleminded in its philosophical outlook, this paper introduces another famous idea, viz. the physical symbol system hypothesis. The hypothesis claims that a physical symbol system has the necessary and sufficient means for general intelligent action. It is one contribution that could have been safely left out, for its contents (PSS and heuristic search) now appear, with commensurate economy, in all modern AI textbooks. Obviously, this judgment has nothing to do with the quality of the contribution which is first-rate.

The ever popular "frames" paper of Minsky (Chapter 5) is edited in such a way that every part of the original MIT AI Laboratory Memo #306 appears either in this chapter or in (Winston 1975). While I realize that this shows admirable decorum on the part of Minsky for sole publication, it would be profitable to have the complete memo published in MD2. After all the Winston volume is long out-of-print, available only in libraries.

The Dreyfus essay (Chapter 6) comes, with minor revisions, from the introduction of (Dreyfus 1979), one of my favorite books. It is one of the most informed and intelligent critiques of AI as it has developed in the seventies. Although its tone remains skeptical, Dreyfus's discussion of AI advances such as Terry Winograd's SHRDLU, David Waltz's vision program, Patrick H. Winston's learning program, Minsky's frames, Roger C. Schank's scripts, and Winograd and Daniel G. Bobrow's KRL is both eye-opening and objective.

Searle's classic essay (Chapter 7) probably needs no words of introduction. First published in *Behavioral and Brain Sciences* 1: 417-424 (1980) and widely anthologized, it is an exemplar of philosophical clarity and purity. In this paper, Searle devises his now famous Chinese Room--a thought experiment directed to the claims of what he calls "strong" AI. Very briefly, the latter is the general viewpoint that given the right programs,

computers can be literally said to understand (have a mind). AI needs more philosophers of Searle's caliber to challenge its philosophical foundations.

Starting with Chapter 8, MD2 shows clear signs of the editor's decision to place heavy--some readers may think unwarranted--emphasis on the connectionist approach to AI. Thus, Rumelhart's essay (which first appeared in (Posner 1989)) offers a no-frills account of brain-style computation; it is interesting to notice that most of his account can now be found in the appropriate chapters of AI textbooks. In a slightly revised version of his contribution to (Nadel et al. 1989), Smolensky studies, in Chapter 9, questions of the following sort: What is the relation between connectionist systems and the brain? How does the connectionist approach compare with the traditional symbolic approach? Can connectionism help us understand the nature of processing characterizing the mind?

Chapter 10 finds Churchland surveying the problems afflicting the classical (viz. "sentential" or expressible in FOL) approach. His paper, slightly abridged from its counterpart in (Churchland 1989), proposes that a neurocomputational perspective should replace the sentential paradigm, and explores the consequences of this recommendation for assorted issues in the philosophy of science. Everybody knows that the Churchlands have been extremely productive and always find a lot of imaginative things to say--as a dozen or so references in the bibliography of MD2 manifest. Well, this essay is no exception.

Rosenberg's contribution in Chapter 11 is a slightly revised version of his paper in *Acta Analytica* 6: 33-46 (1990). While he is sympathetic to the insights and accomplishments of the connectionist networks, in this essay he plays the devil's advocate and voices doubts about just what the connectionist paradigm equips us to understand. His arguments are directed towards Churchland's eagerness to advocate the connectionist paradigm as an alternative to the so-called sentential epistemologies (cf. Chapter 10 of MD2). It is good to know that this brilliant rebuttal has spared itself a more visible place in the literature.

The Fodor and Pylyshyn paper (Chapter 12) is a drastically shortened version and loses more than one third of the initial text which appeared in *Cognition* 28: 3-71 (1988). Similar in flavor to the Rosenberg article, it is a philosophical criticism of the connectionist models, probably more careful in its attendance to technicalities. Fodor and Pylyshyn explain what makes connectionist and classical theories of mental structure incompatible. In the authors' view, the connectionist decision to reject the classical picture led connectionists to put forward implausible theses re logic and semantics. Again, this is a superb essay; it is unfortunate that it could not be printed in its entirety in MD2.

The essay by Ramsey et al. (Chapter 13) is also philosophical in character (having

appeared in (Ramsey et al. 1990)) but does not really argue about the truth or falsity of connectionism. The authors' goal is just to indicate how connectionism and eliminativism are related. They believe that certain kinds of connectionist models bring a sweeping change since the models are seriously eliminativist about the central notions of folk psychology, e.g., belief. Ramsey et al.'s claim is that if the authors are right then we might witness what they call an "ontologically radical theory change," one that will conclude that propositional attitudes such as belief do not exist (and are thus eliminated).

Clark's work (Chapter 14) first appeared in *Connection Science* 4: 193-205 (1992). It is a rather unique contribution in that it argues against both Fodor and Pylyshyn (Chapter 12), and Ramsey et al. (Chapter 13). Its major thesis is that the presence of a symbol is always processor relative. Thus, when we are asked whether an agent explicitly represents a given content we can only answer relative to an environment in which the agent is situated.

The Brooks paper (Chapter 15) is another, substantially different version of his original article, published in *Artificial Intelligence* 47: 139-159 (1991). It includes the photos of two robots he built, named Allen and Herbert, in a tongue-in-cheek style. Brooks hypothesizes that the world itself should serve as its own model and that representation is the wrong unit of abstraction. The right approach to constructing intelligent agents must be through an understanding the essence of acting and reacting. He notes that this has conceptual similarities to the work of Heidegger, the famed German philosopher. However, he is quick to put a disclaimer: "[My work] is based purely on engineering considerations. That does not preclude it from being used in philosophical debate as an example on any side of any fence, however." (ASIDE: The reader must be told that Haugeland is finishing a book on Heidegger.)

Van Gelder's Chapter 16 conclude MD2. Previously unpublished, its central claim is that cognitive systems are dynamical systems and not computers. Cognitive processes correspond to state-space evolution rather than computation. Van Gelder's is a brazen trial to establish yet another research program, in addition to the traditional computational approach and attempts such as connectionism, neurocomputing, embedded computation, and the like. I want to believe that it turns out to be a barren program, just witness his chilling assertion vis-a-vis the Watt centrifugal governor for controlling the speed of a steam engine: "[It] is preferable to the Turing machine as an archetype for models of cognition."

It is time for a general outlook. Aaron Sloman and John McCarthy, two major proponents of fundamental research in AI, demonstrated that realistic AI requires arming a program with a philosophy. Their *IJCAI '95* papers ("A philosophical encounter," available at http://www.cs.bham.ac.uk/~axs/cog_affect/ijcai95.text, and "What has AI in common witl philosophy?" available at http://www-formal.stanford.edu/jmc/aiphil/aiphil.html,

respectively) consolidated the need for deeper inquiries into foundations by suggesting numerous pathways for growth. In the words of McCarthy, ‘‘AI needs many ideas that have hitherto been studied only by philosophers.’’ But McCarthy also notes that ‘‘many philosophical problems take new forms when thought about in terms of how to design a robot.’’ He then issues a warning: ‘‘Some approaches to philosophy are helpful and others are not.’’ As I have already hinted at in the preceding presentation, MD2 places heavy--some would think unusual, even biased--emphasis on connectionism, neurocomputing, situated AI, and yes, dynamic systems theory. In the words of Margaret A. Boden (from a blurb appearing in the cover of MD2), "which of these is the most stable, and whether the foundations need to be re-worked, are questions readers will be eager to explore." Haugeland uses the umbrella term NFAI (New-Fangled AI) to mean scientific mind design that is not GOFAI (Good Old-Fashioned AI, that is, thinking as internal symbol manipulation). Connectionism directly falls under NFAI, and several other possibilities such as the embodied mind design present themselves as worthy candidates. (ASIDE: I had to look "fangle" up from *The Dictionary of Phrase and Fable* by E. Cobham Brewer (1894): A new-fangled notion is one just started or entertained. The word springs from Saxon "fengan," to begin.)

Finally, a few observations on matters not concerning substance. There are 22 pages of references in MD2 compared to 14 pages in MD. I found them to be well-done (i.e., complete) in general. I’ve also liked the fact that each reference has pointers indicating which chapters cite it. All the figures in MD2 have been redrawn, rendering it a more harmonious look. Overall, there are quite a few pestering typos, but they are not of the confusing kind. A valid charge may be that there are no name or subject indices; these would indeed come very handy in a book of this size. And as a considerable blemish, while the cover illustration is fine, the color and the quality of the cover are both of inferior quality.

I would like to end this review by explaining the relevance of the opening stanza of Perkoff as follows. MD contained the seeds of a second edition and MD2 shows that at least in this particular case the destruction didn’t take place; what we have is as accomplished as MD. And clearly in MD2 there are the seeds of a possible MD3. I, for one, will look forward to reading that third edition too.

References

Churchland, Paul M. 1989. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA: MIT Press.

Dennett, Daniel C. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Dreyfus, Hubert L. 1979. *What Computers Can’t Do: A Critique of Artificial Intelligence*.

New York: Harper and Row. Second edition.

Guttenplan, Samuel, ed. 1994. *A Companion to the Philosophy of Mind*. Oxford: Blackwell.

Haugeland, John, ed. 1981. *Mind Design*. Cambridge, MA: MIT Press. First edition. [Abbreviated as MD in this review.]

Haugeland, John. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.

Haugeland, John. 1997. *Having Thought*. Cambridge, MA: Harvard University Press.

Nadel, Lynn, Cooper, Lynn A., Culicover, Peter, and Harnish, R. Michael, eds. 1989. *Neural Connections, Mental Computation*. Cambridge, MA: MIT Press.

Posner, Michael I., ed. 1989. *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Ramsey, William, Stich, Stephen P., and Rumelhart, David E., eds. 1991. *Philosophy and Connectionist Theory*. Hillsdale, NJ: Lawrence Erlbaum.

Winston, Patrick H., ed. 1975. *The Psychology of Computer Vision*. New York: McGraw Hill.