Mind Design II
Philosophy, Psychology, and Artificial Intelligence

John Haugeland, editor

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

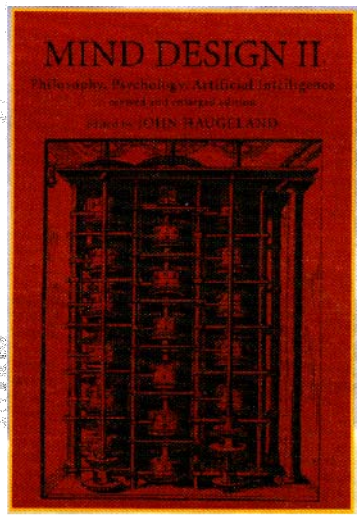
Reviewed by: Varol Akman
Department of Computer Engineering and Information Science
Bilkent University
Bilkent, Ankara 06533, Turkey
akman@cs.bilkent.edu.tr

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, professor of philosophy at the University of Pittsburgh, is well known in the Artificial Intelligence community. With interests — among others — in the philosophy of mind and philosophical psychology, he has produced numerous thoughtful works in the past and continues to do so. The title under review, *Mind Design II*, is the revised version of his 1981 book *Mind Design*. Other Haugeland books include *Artificial Intelligence: The Very Idea* [Haugeland 1985] and *Having Thought* [Haugeland 1997].

Mind Design II continues in the tradition of the 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 *Mind Design II* 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. *Mind Design II* is a substantial yet inexpensive book and would be an excellent choice for an "AI & Philosophy" course, particularly at the graduate level.

Let me now take a look at *Mind Design II* in detail. In the "Acknowledgments" section of *Mind Design* [Haugeland 1981, p. vii] Haugeland states:



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; *Mind Design* quickly became a classic and found a secure place for itself in philosophical and AI circles. It also served to introduce important contemporary philosophers of mind to the AI community.

The revised and enlarged successor of *Mind Design* 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 are taken from *Mind Design*. The following chapters of the original book are not included in *Mind Design II*:

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)

Of these, more recent contributions of Pylyshyn, Dennett, Fodor, and Haugeland appear in *Mind Design II*. So, in some sense, only Marr, McDermott, Putnam, and Davidson have been deleted. I would love to have seen at least the last two of these represented in *Mind Design II*. They are both remarkable philosophers, with exciting things to say. In the case of Putnam, his latest insights into the post-functionalist tradition would have been appropriate [Guttenplan 1994, p. 513]. In the 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. Davidson states [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." (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 *Mind Design* ("Semantic Engines: An Introduction to Mind Design") is replaced by Chapter 1 of *Mind Design II*, serving an analogous role. Here, Haugeland defines mind design as the endeavor of understanding the 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 is available on-line (*http://www.cs.bilkent.edu.tr/~psaygin/ttest.html*). It is one of most beautiful pieces in philosophical AI, with a fresh, optimistic insight and prophetic vision.

John McCarthy suggests (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 *The Intentional Stance* [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 unsophisticated 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. This is one contribution that could have been easily omitted, for its contents (PSS and heuristic search) now appear, with commensurate economy, in all modern AI textbooks. Obviously, this opinion 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 useful to have the complete memo published in *Mind Design II* since 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 *What Computers Can't Do: A Critique of Artificial Intelligence* [Dreyfus 1979], one of my favorite books. It is one of the most informed and intelligent critiques of AI as it developed in the '70s. 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, *Mind Design II* 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 *Mind Design II* manifest. 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 toward Churchland's eagerness to advocate the connectionist paradigm as an alternative to the so-called sentential epistemologies (cf. Chapter 10 of *Mind Design II*). 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 significantly 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, somewhat 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 concerning logic and semantics. Again, this is a superb essay; it is unfortunate that it could not be printed in its entirety in *Mind Design II*.

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 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, 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 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 a substantially altered version of his original article, published in *Artificial Intelligence* 47: 139-159 (1991). It includes the photos of two of his robots, Allen and Herbert. 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 of 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."

Van Gelder's Chapter 16 concludes *Mind Design II*. 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 bold attempt to establish yet another research approach, different from the traditional computational approach as well as from connectionism, neurocomputing, embedded computation, and such. Consider 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." I would like to believe that this turns out to be a barren program.

To summarize: 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 with 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 stated, *Mind Design II* 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 the back cover of *Mind Design II*), "which of these is the most stable, and whether the foundations need to be reworked, 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.

A few observations on the structure of the book: There are 22 pages of references in *Mind Design II* compared to 14 pages in *Mind Design*. I found them to be quite complete. I also liked the fact that each reference has pointers indicating in which chapters it was cited. All the figures in *Mind Design II* have been redrawn, rendering a more harmonious look. There are quite a few typos, but they are not of the confusing kind. A valid complaint could be made that there are no name or subject indices; these would indeed be very useful in a book of this size. Finally, while the cover illustration is fine, both the color and the quality of the cover could be improved.

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

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.

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

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

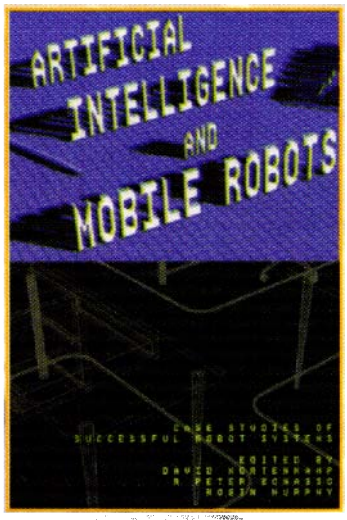
Nadel, Lynn, Cooper, Lynn A., Culicover, Peter, and Harnish, R. Michael, eds. 1989. *Neural Connections, Mental Computation.* Cambridge, MA: MIT Press.

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.

* * * * * * * * *



Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems

D. Kortenkamp, R.P. Bonasso, and R. Murphy

The MIT Press
Cambridge, MA, 1998
390 pp.
$37.50
ISBN 0-262-61137-6

Reviewed by: John Clarke
Instituto de Sistemas e Robotica
Universidade de Coimbra
Polo II
Coimbra 3030 Portugal

Artificial Intelligence and Mobile Robots is a compilation of 13 chapters by well respected robotics researchers describing some of their most recent work. The book aims to provide a "how-to" guide to building the control system of a mobile robot. These case studies are definitely worth studying: seven of the 13 have placed highly in various robotics competitions. In the introduction, the editors give a good, yet brief, history of