Comparing Artificial and Natural Intelligence

Rainer Born (Ed.)
Artificial Intelligence: The Case Against
London: Croom Helm, 1987. 220 pp.
ISBN 0-7099-5923-0 (paperback).
$15.95 paperback; $21.00 (Canada) hardcover

Review by
Thomas R. Shultz

Rainer Born is lecturer of philosophy of science at Johannes Kepler University of Linz (Austria). ■ *Thomas R. Shultz is professor of psychology and director of the McGill Cognitive Science Centre at McGill University (Montreal).*

The rapid growth of the field of artificial intelligence (AI) in recent years has spawned a revival of interest in mental philosophy. Much of this philosophical analysis attacks AI for its extravagant claims about the psychological power of AI computer programs. There is apparently something unnerving to many philosophers of mind about the prospect of thinking machines. Despite claims on the book jacket to offer constructive criticism and to advance AI research, this book is clearly in the anti-AI mold. The negative tone of the book is unsubtly conveyed by both its subtitle and the rather bizarre reference in the first paragraph of the introduction to false alarms in computerized antimissile warning systems.

Most of the contributors to the eight chapters in this book attempt to draw sharp contrasts between the artificial intelligence of various computer programs and the natural intelligence exhibited by humans. Unfortunately, most of these comparisons fail to shed much light on intelligence of either the artificial or the natural type.

Implications for AI

I find it useful to examine separately the implications of these philosophical analyses for the fields of AI and of cognitive psychology. From the standpoint of AI, much of the analysis seems trite and anthropocentric. The recurrent claim is that AI programs do not, and indeed cannot, exhibit intelligence. We are told that silicon microchips are the wrong sort of material in which to implement intentional states (Searle), that cognition and behavior are determined not by facts and features but by a sense of the situation (Dreyfus), that AI programs fail to generalize beyond the restricted domains for which they were designed (Dreyfus), and that it is unintelligible to even speak of a machine following a rule or learning (Shanker).

Searle's well-known argument is analogous to claiming that airplanes don't fly because they don't fly like birds (i.e., with bird bodies). The psychological facts are that we know very little about either intentionality or about how it might be implemented in the brain. If intentional and other mental states are not implemented by the representation and transmission of information, then it would be helpful to know how Searle thinks it is done. To claim that intentionality can only be manifested in neuroprotein is thus premature.

Searle's argument that AI programs are noncomprehending procedures for symbol manipulation fails to appreciate the nature of much psychological explanation. It is precisely because scientists want to explain psychological processes that they posit underlying processes that do not have their capacities. Explanations that merely assume what they attempt to explain are circular and thus devoid of explanatory power. Searle's chapter appeared a full 9 years ago with extensive commentary (Searle, 1980) and has been reprinted elsewhere (Haugeland, 1981). Its reappearance in this volume is more historical than newsworthy.

Dreyfus, whose arguments are likewise well known (Dreyfus, 1979; Dreyfus & Dreyfus, 1986), fails to indicate how he would explicate situational knowledge. His comment that connectionist models could better represent human intelligence neglects the fact that the definitions of at least *input* and *output units* would need to be mapped to the definitions of those same terms in the human users in much the same manner for which he criticizes classical AI models. Dreyfus's argument about the lack of generalizability of AI programs neglects research with SOAR, a learning and problem-solving program that has been successfully applied to a wide range of phenomena (Steier et al., 1987).

Shanker's assertion that the very notion of a thinking machine is unintelligible conflicts with the widely held view that humans are, in fact, machines and reveals an essentially anthropocentric view that seems more common among philosophers of mind than among psychologists or AI researchers. One of the most important contributions of AI has been to fit human learning and reasoning into a broader than human perspective.

Much of the philosophical criticism of AI programs derives from the assumption that such programs are pure syntax with no semantic content. However, this assumption has been questioned by those who argue that programs can embody a causal semantics in which the meaning of a symbol is provided by its causal links to other phenomena. Programs can implement causes of the construction or activation of symbols as well as implement effects of symbol activation (Boden, 1988; Sloman, 1986; Smith, in press).

Another probable source of the recent discord between the philosophy of mind and AI is that they have pursued somewhat distinct interests. Whereas AI researchers have been concerned with how to represent knowledge to facilitate reasoning and learning, philosophers of mind have been concerned with the problem of reference (i.e., the mapping of symbols to what they represent).

Implications for psychology

From the standpoint of cognitive psychology and cognitive science, I can be much briefer because this book is largely irrelevant to these fields. This is because the chief focus of interest in these fields is the use of computational models to develop and test theories rather than the building of thinking machines. Whether these programs exhibit the intelligence that they are designed to simulate is beside the point. They are valuable tools in the service of scientific discovery and have been essential in demonstrating the degree to which psychological theories are computationally sufficient. (Usually, psychological theories are far from computationally sufficient, but this shortcoming is not often noticed by noncomputational psychologists and is ignored in the present book.)

Dreyfus's psychological analysis of the development of expertise is much less compelling than previous computational accounts emphasizing the compilation and composition of knowledge with practice (Anderson, 1983).

Kobsa is undoubtedly correct in his argument that computational models are not theories about mechanisms but rather the mechanisms themselves. More abstract theories at the computational (Marr, 1977) or task analysis (Newell & Simon, 1972) level are required to explain the mechanisms implemented in actual programs. I would add that the implemented mechanisms do explain the behavior they generate, thus adding a critical level to the hierarchy of explanation.

Other matters

The editor's rather baroque, diagrammatic organizational scheme, presented in the introduction and the final chapter, is wholly inadequate to characterize the contents of this book. One of the chapters, Klix's psychological study of cognitive efficiency, seems not to fit at all; it is an unsophisticated model of analogy recognition, lacking consideration of the essential aspects of analogical reasoning: search, mapping, tweaking, and application (Holland, Holyoak, Nisbett, & Thagard, 1986).

The book could have benefitted a great deal from the inclusion of other relevant philosophers such as Dennett, Fodor, Haugeland, and Pylyshyn. Dennett (1978) has espoused an instrumentalist view of intentionality. He argues that an intentional system's behavior can be explained and predicted by ascribing intentional states to it. In his analysis, computer programs could be considered intentional.

One of the more interesting chapters is Neumaier's Wittgensteinian view of AI. Although Wittgenstein died just before the birth of AI, he had stressed the explanatory value of theory and the need for outward verification of inner states, both of which are quite consistent with the practices of AI and of cognitive science. However, because AI programs do not participate in the "bustle of human life," they would not be likely Wittgensteinian exhibitors of intelligence. Neumaier further notes that in AI it is not intelligence that is artificial but the mechanism by which intelligence is realized.

Putnam's chapter, reprinted here from an unnamed source, identifies what seems to be a deep problem for both mental philosophy and computational models, whether of the AI or psychological type. The problem examined by Putnam concerns the inherent difficulty of establishing sameness of meaning in two or more different mental representations. He argues that this can be done only by examining the extensions of the representations as they exist in the collective minds of the participants' communities. The implications of this problem for AI, cognitive science, and psychology may be quite serious.

This book is somewhat useful because it has brought this material together. However, it is not a particularly good introduction to AI or computational modeling for psychologists or other nonspecialists. (A much better introduction is provided in Boden [1988].) Exclusion of some of the present chapters and inclusion of others would have yielded a stronger book.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Boden, M. A. (1988). *Computer models of mind: Computational approaches in theoretical psychology.* New York: Cambridge University Press.

Dennett, D. C. (1978). *Brainstorms: Philosophical essays on mind and psychology.* Cambridge, MA: MIT Press.

Dreyfus, H. (1979). *What computers can't do.* New York: Harper & Row.

Dreyfus, H., & Dreyfus, S. (1986). *Mind over machine.* New York: Macmillan.

Haugeland, J. (Ed.). (1981). *Mind design: Philosophy, psychology, and artificial intelligence.* Cambridge, MA: MIT Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.

Marr, D. (1977). Artificial intelligence: A personal view. *Artificial Intelligence, 9,* 37–48.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3,* 417–457.

Sloman, A. (1986). What sorts of machines can understand the symbols they use? *Proceedings of the Aristotelian Society, 60,* 61–80.

Smith, B. C. (in press). *Is computation formal?* Cambridge, MA: MIT Press.

Steier, D. E., Laird, J. E., Newell, A., Rosenbloom, P. S., Flynn, R. A., Golding, A., Polk, T. A., Shivers, O. G., Unruh, A., & Yost, G. R. (1987). Varieties of learning in SOAR. *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 300–311). Los Altos, CA: Morgan Kaufman.

A Psychiatric Perspective on Affective Disorders

J. John Mann (Ed.)
Phenomenology of Depressive Illness
New York: Human Sciences Press, 1988. 263 pp. ISBN 0-89885-369-9. $34.95

Review by
Michael L. Raulin

J. John Mann is professor of psychiatry at Cornell University Medical College (New York City). ■ *Michael L. Raulin is clinical associate professor and director of the Psychological Services Center at the State University of New York at Buffalo. He is coauthor, with A. M. Graziano, of* Research Methods: A Process of Inquiry.

The occasional feeling of depression is a nearly ubiquitous experience, but, fortunately, the experience of severe depressive illness is considerably less common. This volume seeks to describe the variants of severe affective disturbance and the currently accepted medical treatments for these disorders.

Phenomenology of Depressive Illness is apparently the first in a series of volumes on depressive illness. As such, its purpose appears to be to describe the various types of affective disorders encountered in psychiatric clinics and hospitals. The description of these disorders is very close to atheoretical, with one exception: The assumption that affective disorders represent a disease process with a primarily biological etiology forms the basis for the entire volume. Many psychologists may find this perspective unnerving or at least annoying. The causal chain of events from biological risk factor to onset of symptoms must certainly in-