"Emperor AI, Where Is Your New Mind?"

Wai Kiang Yeap

■ Any theory of the mind must explain how the mind works, and an AI theory is no exception. Many critics have correctly argued that AI researchers have failed to produce such a theory. However, their discussion has focused mainly on what current computer (or particular programs) can or cannot do. Few have examined whether the field itself provides a foundation for producing a theory of the mind. If it does, what has been learned, and what do we need to do next? This article is an attempt to show how AI research has progressed in its quest for a theory of the mind. The "emperor's new mind" is not here yet, but it is argued that it is in the process of being developed.

I s AI dead as a science for investigating the nature of intelligence, or as suggested by Penrose (1989), is the emperor's mind actually hollow? Should AI be used simply to superficially describe those studies that produce intelligent programs? Several indicators suggest that the answer to these questions might be yes. These indicators include beliefs that (1) the field has made limited or little progress to date, and worse, interests in these questions among AI researchers are on the wane or nonexisting; (2) the serious criticisms from those who champion the study of human mind and behavior have cast a strong doubt on the possibility of an AI theory of the mind; and (3) the continuing challenges of new AI approaches such as connectionism have shown that the research itself is in turmoil.

This essay argues that the answer should be an emphatic no. Amid the chaos in the past history of AI research, some serious lessons have been learned, and good progress has been made. However, given that AI has not yet produced a theory of the mind, its progress as a science is best discussed in terms of the foundation it has built and the important questions it is trying to solve. The appropriateness and relevance of these questions will determine whether AI has anything to contribute. Thus, this article is neither a survey of the different kinds of AI research carried out in the past (Bobrow and Hayes 1985; Boden 1994), nor is it an attempt to provide a unified view with a synthesis of past research (Simon 1995). It is only concerned with what AI has to offer to mind research.

I claim that AI provides a coherent three-tier strategy for unraveling the mystery of the mind: (1) develop the technology for advanced computations, (2) develop computational theories about how the mind-body complex works, and (3) develop machines that can compute with original intent (Haugeland 1980). The first two tiers are necessary for developing an AI theory of the mind simply because without the right technology, one just could not implement such a theory. However, as demonstrated in past AI research and as I argue later, they alone are not sufficient; the machines created might be powerful, but without the ability to compute with original intent, they are not intelligent machines. Note that although the first two tiers have been mentioned frequently in past reviews of AI research, they are seldom perceived as the parts of AI that contribute to the development of a theory of the mind. In fact, the first tier is often described as its main alternative. The second tier consists of various computational studies of individual perceptual processes (such as vision and natural language), and these are treated as a subject in their own right. I show why these views persisted and why they should not be the case. An overview of the problems to be discussed is given here.

The first tier appeared in the early days when AI researchers were dreaming of a machine that could reason. The results were machines that manipulate symbols, and their existence

From the difficulty faced in defining the term intelligence in past AI research, an immediate lesson seems to be that intelligence is a quality *that one ascribes to the behavior of a given system.*

raised serious questions about the nature of intelligence. What do we think of these machines? Are they intelligent? These machines were hailed as a technological breakthrough, and their emergence led many to believe that the goal of AI research was to develop a new technology rather than a theory of the mind. This possible alternative goal, and a very attractive one, caused a lot of confusion among AI researchers and still does. The first section explains this dual nature of AI research, both as a technology and as a science. It shows where the important link between them lies and why both are essential to the development of an AI theory of the mind.

The early symbolic manipulator machines definitely did not think, and there was a subsequent rush to create one that did. Many different approaches were experimented with, and chief among them were incorporating more knowledge and allowing the machines to interact with their environment. Interactive machines involved work on building integrated robots or systems embedded in their environment and work on developing models of various perceptual processes. This work was significant because it highlighted the intricate processing that is being carried out in these processes prior to the information reaching the brain. For example, where does vision end and the mind begin?

It is unlikely that one could understand how the mind works without knowing first what information is being extracted from the environment. In other words, the construction of a mind begins very much with the input itself; understanding what is involved in existing mind-body complexes would help to identify the necessary conditions for creating a machine with original intentionality. What exactly is the role of a neural brain, fingered hands, languages, and other features of humans in developing the human mind? One problem with past AI research in this area is that it seldom went beyond investigating the mechanics of these particular processes, and thus, individuals became, for example, vision researchers rather than AI researchers. The second section discusses how and why we need AI theories of the mind-body complex.

The third tier is concerned with building machines that can compute with original intent. A working definition for such a machine is one that can reason about its world based only on information computed by its sensors. A robot built with sensors to avoid objects is not such a machine if all it does is run around the environment and avoid objects. However, a similar robot, which, after

interacting with the environment, discovers the notion of avoidance from its action, is. After lengthy debate about Searle's (1980) Chinese room experiment, AI researchers are slowly beginning to realize that a serious gap between AI reasoning programs and the real mind is that AI programs do not reason with original intentionality (see also Rey [1986] and Sayre [1986]). A popular AI solution is to ground the symbols used in the program to the outside world, and because of the recent development of more powerful AI methods, this problem is arousing renewed interest. The third section discusses why this symbol-grounding solution is inappropriate and instead suggests that AI researchers should focus on how the input from the sensors are interpreted or, more appropriately, symbolized and not grounded. In particular, Harnad's (1990) solution, whereby a connectionist network is used to ground the symbols to the outside world, is reviewed and refuted.

The concluding section shows how the three-tier AI strategy could produce a theory of the mind. We should stop arguing about whether existing machines are intelligent or not. These machines are not and will never be if we fail to provide an answer to these questions. The foundation is well laid for us to build on.

Understanding AI: Ai and aI

From the difficulty faced in defining the term *intelligence* in past AI research, an immediate lesson seems to be that intelligence is a quality that one ascribes to the behavior of a given system. If so, it would be incorrect to make the goal of AI the study of the nature of intelligence without referring to the specific system under consideration. From this observation, it is no surprise that AI has a dual role: to study human intelligence and to study machine intelligence. The human system is a natural choice, being the most intelligent system known to us. The question to ask is, "How and why has the human system achieved its level of intelligence compared with other living systems?" The computer system is a popular choice because it is the only physical system with which one can implement one's ideas of intelligence. In this case, the research seeks to discover how to make a computer that behaves intelligently. Note that there is no single universal set of criteria for judging whether what has been implemented is intelligent.

It is important to realize that the study of human intelligence implies the use of the computer as a tool, whereas the study of

machine intelligence treats the computer as the subject of study. To claim more, as in Searle's (1980) notion of strong AI and the idea that the mind is a physical symbol system (Newell 1980), is to fantasize, at least for the moment. Highlighting this different emphasis, I refer to the study of human intelligence as artificial *intelligence* (henceforth aI, or "little AI") and the study of machine intelligence as *artificial* intelligence (henceforth Ai, or "big AI"). Ai emphasizes building artificial systems, whereas aI stresses the study of the nature of intelligence, humans and beyond. When no such distinction is made and especially when reference is made to earlier work, the usual notation, AI, is used.

The question for AI now becomes, "How can a given system be intelligent?" rather than, "What is intelligence?" If the system in question is a computer, one can assume that it has a zero level of intelligence, and the Ai question is how to make a computer more intelligent. If the system is a human being, one can assume that it has a high level of intelligence. The aI question then is, "How does the human system become intelligent?"

For Ai research, one should be free to choose what constitutes an intelligent task, and the goal is to make the computer perform the task according to some selected criteria. Hence, Ai is about writing intelligent programs in the sense that the programs solve a predefined intelligent task. Much of the past research in AI is about Ai. For past summaries and achievements, see Nilsson (1983), Waltz (1983), and Reddy (1988). Not surprisingly then, Ai has mushroomed into many different areas of study. Not surprisingly, too, these works show that one of the hallmarks of Ai research is to discover the necessary mechanisms to build intelligent machines. Therefore, Ai is concerned with the search for efficient and universal problem solutions on a computer, and issues such as scalability, parallel versus serial, and distributed versus central organization of memory are rightly important. Bundy (1981) suggested some useful criteria for assessing Ai research.

Why then does the confusion between Ai and aI research arise? The problem lies in that Ai researchers not only try to make a computer perform a task intelligently but also to perform one that is typically done by humans and at the same level of competence. Achieving this goal apparently gives one the illusion that he/she has contributed a theory about how humans might have solved the problem. The reason why these theories are not adequate lies in the manner in which these programs are developed. Ai researchers pay little attention to the complex ways that these problems are solved by humans, ranging from how the brain works to how things are perceived and used. This lack of attention is evident in many of the past critical reviews, such as Dresher and Hornstein (1976) on understanding natural languages and Reeke and Edelman (1988) on neural networks. Because of this failure, it is more appropriate to treat these results as computer solutions to a similar problem rather than as explanations of how the mind might work.

I claim that this "confusion" between Ai and aI research is the important link that cements the two together. Without this link, their distinction might as well be disregarded; one might as well, as some researchers did, refer to Ai as something else and AI strictly as aI, or vice versa. It is only by studying problems based on, but not constrained by, how humans performed a particular task that Ai researchers were able to investigate all kinds of weird ideas (McDermott 1981) and produce interesting solutions. Such interesting solutions are typically suggestions about how a particular solution to a problem is physically realized. They present a different but important perspective for aI researchers when considering these problems: "How could I possibly implement a solution to the problem?" For example, artificial neural networks show how computing with weights in a physical system is possible, and aI researchers now have some idea of how neural computation could be realized physically. Brady and Hu (1994) recently presented some insightful comments on the nature of intelligence from the perspective of building robotic systems.

Many have criticized the ad hoc approach in Ai research and argued that Ai researchers should raise their "psychological assumptions from the level of ad hoc intuitions to the level of systematic empirical observations" (Ringle 1983, p. 238). Doing so would be a pity and is unnecessary. It is a pity because many AI problems really begin by simply asking how a computer can perform such a task. Sometimes, the only way forward is to freely imagine different kinds of possible solutions. It is unnecessary because such systematic studies naturally follow from initial exploratory work. What is needed is the clear distinction between the different methods (Ai and aI) afforded by AI research. In particular, it is important not to think that every AI solution has to immediately be a part of a theory of the mind (or, for that matter, that it has to incorporate some fascinating and complex algorithms).

Why then does the confusion between Ai and aI research arise? The problem lies in that Ai researchers not only try to make a computer perform a task intelligently but also to perform one that is typically done by humans and at the same level of competence.

Understanding the Mind-Body Complex

What kind of information is made explicit in each (body) process prior to higher-level reasoning? How does the mind relate the different pieces of information from the different processes so that they can be perceived as a combined whole? From what is perceived, how does reasoning with original intentionality emerge? These are the central questions that aI researchers should ask when investigating how the mind-body complexes work.

Early AI studies simply mimicked these processes, but when more sophisticated systems were developed, it became clear that one could and should develop theories to explain these processes, focusing on the flow of information in them. Marr's (1982) theory of vision clearly showed how such a theory could be developed, and since then, many studies have adopted Marr's computational approach (for example, see Richards [1988] and Cosmides and Tooby [1994]). There is even a growing acceptance in the more traditional sciences, such as psychology (Boden 1988) and neuroscience (Sejnowski, Koch, and Churchland 1988), to adopt this approach, which has helped to more rapidly advance the understanding of these processes.

Understanding the flow of information in and between each process is what is needed for aI research. However, it is important to stress that for aI, the interests lie less in knowing exactly how these individual processes work and more in understanding what their general nature is. In particular, how will the information extracted be used for higher reasoning? For example, consider the first two stages in vision. Although the gap between Marr's primal sketch (a description of what is implicit in the intensity images) and the two-and-one-half-dimensional (2-1/2D) sketch (a viewer-centered description of surfaces computed from a single view) is significant, what aI researchers need to know most is what is available at the 2-1/2D–sketch level and how the information is then processed by humans to understand their environments. Therefore, given what is made explicit in the 2-1/2D sketch, it is important that aI researchers ask how the concepts of impediment, permanence, three dimensionality, and others are understood by the system itself. The numerous processes and phenomena observed at the level below the 2-1/2D sketch might be interesting, but unless they lead to theories about how one could understand the environment, they would not be useful for aI.

> *The aI goal is not to duplicate the human's system, just learn from it.*

The aI goal is not to duplicate the human's system, just learn from it. aI researchers must focus on theories that help solve the puzzle as a whole rather than those that get bogged down by explaining the workings of individual pieces. To do so, it is important to ensure that these models satisfy the following two constraints: The first constraint (C1) is that the input must correspond, at least in information terms, to that used by humans when they are solving a similar problem. The second constraint (C2) is that the output must be useful for solving the next task, which, in turn, must be identified as a relevant task to be solved at the particular level in the human system. C1 ensures that one investigates a problem similar to the one presented to humans, and C2 ensures that the solution developed contributes to one's understanding of the process as a whole and not just a particular subprocess.

The need to satisfy the first requirement shows why past AI work on perceptual problems has tended to be most relevant to aI research. Researchers working on these problems are likely to begin with the right kind of input. However, there is a stronger interpretation of C1, namely, that aI is best understood by working on the perceptual problems first, in a bottom-up fashion. This stronger interpretation is not strictly necessary because what is important is that one pay attention to what could be computed as input for the task at hand. Doing so encourages one to think seriously about perceptual problems, but this problem is not the same as studying them, at least not in the sense of requiring the development of complete models. Thus, one might call this a *think-bottom-up approach*.

C2 has often been neglected in past AI research. Marr (1982, p. 272) first highlighted its significance when he argued why one should abandon trying to segment an image using specialized knowledge about the nature of the scenes:

> Most early visual processes extract information about the visible surface directly, without particular regard to whether they happen to be part of a horse or a man or a tree. It is these surfaces—their shape and disposition relative to the viewer—and their intrinsic reflectances that need to be made explicit at this point in the processing.

In short, what is computed directly from an intensity image is not a three-dimensional (3D) object but what Marr (1982) called a 2-1/2D sketch.

However, Marr failed to ask similar questions when developing his theory for the next

stage of the visual process: A 3D model using generalized cones (Marr and Nishihara 1978) was suggested to represent shape information. A strong argument for suggesting that a 3D model is necessary is that the information in the 2-1/2D sketch is not stable enough for the 3D object-recognition task. This is not unreasonable provided that the focus is on understanding how the 3D model itself is derived from the 2-1/2D sketch, but this question was not asked. Instead, the research focused on the design of a suitable representation for recognizing 3D objects in the 2-1/2D sketch. Marr thus fell into a common trap: producing application programs as theories of the mind. Because little is known about the relationship between our knowledge of 3D objects and the surface information that we perceive, any representation suggested is, at best, a wild speculation about what is actually needed. The gap between the 2-1/2D sketch and the 3D sketch remains unbridged, even though many sophisticated object-recognition systems have been built (Hurlbert and Poggio 1988). Hoffman and Richards (1984) later and rightly questioned the use of any specialized representation (such as the use of generalized cones in the 3D model) for recognition purposes. Yeap, Jefferies, and Naylor (1991) argued that a more immediate—and, hence, more appropriate—module that takes the 2-1/2D sketch as input is the raw cognitive map (see also Yeap [1988] and Yeap and Handley [1991]).

In summary, the task here is to understand how the mind-body complex works. When doing so, it is clear that aI research must pay close attention to what the human system has to offer and not be tempted to develop a better solution. It is not the solution as such that is of interest. The emphasis should be on the construction of the entire inner process based (initially) on the bits and pieces that psychologists and philosophers have discovered to this point. The prize will be a computational theory of the mind-body complex.

Understanding Computing with Original Intentionality

AI researchers have enjoyed theorizing about the mind from a variety of platforms. Examples of these include (1) simulation studies based on psychological findings (for example, Anderson's [1983] ACT and Laird, Newell, and Rosenbloom's [1987] SOAR), (2) ideas based on one's own experience of developing intelligent machines (for example, Minsky's [1985] society of the minds and Newell's [1980] physical symbol systems), and (3) extensive tests of a

particular method of developing intelligent machines (for example, the CYC Project [Lenat et al. 1990] and the Japanese Fifth-Generation Project [Feigenbaum and McCorduck 1983]). However, as long as AI researchers fail to address the problem of how the system itself creates its own meanings for each symbol, that is, has original intentionality (Haugeland 1980), such theorizing will remain useful only for discussing what the mind might be but not for demonstrating how it is. The research, as before, will simply produce yet another theory of the mind.

The problem of creating a machine with original intentionality is, however, seriously misunderstood by many in AI research. The current popular belief is that the problem is equivalent to that of symbol grounding (Harnad 1990). That is, once the symbols that the machine uses are somehow grounded to the outside world, the machine will have original intent. I disagree, but first I briefly review the symbol-grounding problem.

The symbol-grounding problem has gained much attention in recent years because of the emergence in AI of methods that show how the symbols might be grounded. One example is Brooks's (1991) subsumption architecture for building robots. It emphasizes building simple robots that interact with the world as a prerequisite to gradually increasing their competence. Thus, it is ensured that the robots developed have their symbols grounded to the real world. When higher-level cognitive components are added, these robots, they argue, continue to act with complete understanding and original intent (for example, Malcolm and Smithers [1989], Cliff [1990], and Schnepf [1991]). For an interesting variation of the same theme in constructing robots, see Wilson's (1991) "animat approach." Another example, and an even more popular one, is the connectionist approach to modeling the mind. The much-publicized learning ability of such networks and their supposed brainlike machinery have led many to believe that such networks, perhaps combined with traditional AI systems, would truly be intelligent (for example, Hendler [1989] and Harnad [1990]).

Let's take a closer look at Harnad's solution to the problem. Harnad singled out the processes of discrimination and identification as the two most important initial steps in human understanding of their environment. He then argued that one uses iconic representations for discriminating input and categorical representations for identification. *Iconic representations* are internal analog transforms of the projections of distal objects onto one's

The problem of creating a machine with original intentionality is … seriously misunderstood by many in AI research.

sensory surfaces, whereas *categorical representations* are formed by the grouping of icons with common invariant features. Harnad argued that once categories are formed and used as a basis for reasoning in a symbolic system, the meanings in the system are grounded. Thus, he believes that a solution to the symbol-grounding problem lies in knowing how categories could be derived. He proposed using a connectionist network to identify icons. The network operates by dynamically adjusting the weights on the features and the feature combinations that are reliably associated with a particular category.

Although Harnad pointed out that icons are internal analog transforms of the projections of distal objects onto our sensory surfaces, his example only discussed how the category *horse* could be learned from icons of horses. However, in the real world, there are no individual icons of horses available as input. It is more likely that icons from more complex scenes, such as one with a man wearing a hat, galloping down a hill, and chasing a wagon pulled by several horses, are provided (which would be the case if you were watching one of the all time favorite John Wayne movies). If so, how could the system, human or otherwise, pick up the icon of a horse, or anything else for that matter, from the input and why? Knowing why is important because it is the intent of the system. Why pay attention to icons of horses? Where does the notion of a horse come from? Why interpret the scene as having different objects? Where does the notion of object come from? How does one interpret any movement perceived, and how does one interpret when things move out of sight? How could the system derive the notion of motion from observing things that move in its view? Thus, much of one's understanding is far from identifying invariant features; it has more to do with the ability to form one's own concepts about the world. Concepts are different from invariant features in that they are based on one's reasoning or belief of what is happening out there.

Some of the early concepts that children learn about their world include object permanence, impediment, and other problems concerned with moving in a 3D world. aI researchers should pay attention to developing a machine that allows it to discover such basic concepts on its own. In this sense, the symbol-grounding problem continues to miss its target. Many of the solutions suggested to date still presuppose the a priori presence of categories in the world (Reeke and Edelman 1988); the difference between the traditional symbolic approach and the new approaches lies main-

ly in the more sophisticated mechanisms for entering the different categories into the system. For example, in Brooks's approach, the low-level activities are prebuilt into the robots, and in Harnad's approach, the neural network has to be trained with predefined categories. However, for the Harnad approach, even if the neural network does not need to be trained with predefined categories, by, say, using more biologically realistic networks such as Reeke, Sporns, and Edelman's (1990) DARWIN systems, one still has to find the algorithm that enables the system itself to realize the significance of these categories. The fact that many animals have a complex neural brain—but only one (that is, humans) has truly demonstrated the capability to perform symbolic reasoning—might well let us argue that having a neural brain does not in itself guarantee a solution to the problem.

In summary, to compute with original intentionality, symbol grounding is the wrong problem to solve. It fails to provide a mechanism for allowing the individual to create its own symbol. A symbol is different from the uninterpreted tokens-categories computed from one's sensory input. It represents one's intention about the world, and it is created after one has formed certain concepts-hypotheses about his/her world. The initial concept need not be correct because it could be modified later. The symbols created will be used to develop other concepts.

Creating such symbols from (initially) observing icons literally dancing in our eyes is the key to constructing a machine with original intentionality. This problem is cognate with an age-old problem in concept formation, namely, the way that an infant comes to understand the world in the same way that it is perceived by adults. With all the sounds that an infant picks up and all the images that appear on his/her retina, how does the baby know that the word *mama* is associated with his/her mother? Also, how does he/she know what a mother is? By the time such an association is made, what else does the infant know about the word *mama*? Psychologists have been puzzled by these questions for decades and have provided rich sources of observations (Streri 1993; Premack 1990) and ideas (Gergely et al. 1995; Spelke 1994; Astington, Harris, and Olson 1988). aI research must draw on these studies to resolve the problem.

Conclusion

AI is about building intelligent machines, and naturally, we must work toward producing the
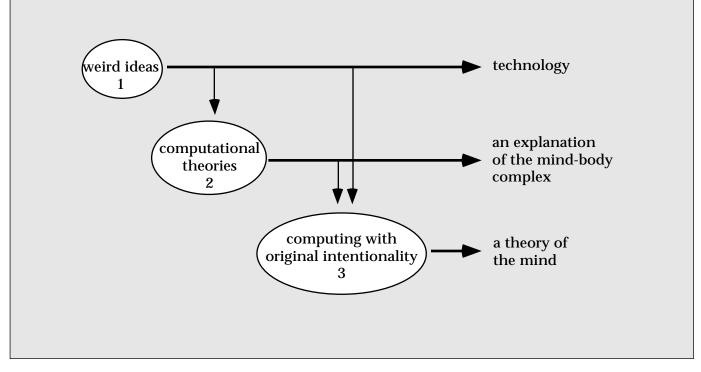
Figure 1. Steps toward an AI Theory of the Mind.

underlying theory and the necessary technology. Figure 1 shows how the three-tier AI strategy outlined previously will contribute to this success.

The figure shows how initial questions about the nature of intelligence are explored, giving rise to a demonstration of how the idea might be realized physically. Computational studies, investigating similar problems being solved in a mind-body complex, help develop a more rigorous model and, more importantly, provide an understanding of the flow of information in and between these processes. These results combine to help aI researchers to develop a machine with original intentionality: given only the input from the sensors, a machine that will formulate its own symbols describing the world and developing higher-level reasoning similar to that exhibited by humans. Isn't this an important goal for all researchers interested in discerning the nature of intelligence?

Acknowledgments

References

Anderson, J. R. 1983. *The Architecture of Cognition.* Cambridge, Mass.: Harvard University Press.

Astington, J. W.; Harris P. L.; and Olson, D. R. 1988. *Developing Theories of Mind.* Cambridge, U.K.: Cambridge University Press.

Bobrow, D. G., and Hayes, P. J. 1985. Artificial Intelligence—Where Are We? *Artificial Intelligence* 25:375–415.

Boden, M. A. 1994. New Breakthroughs or Dead-Ends? *Philosophical Transactions of The Royal Society of London A* 349:1–13.

Boden, M. A. 1988. *Computer Models of Mind.* Cambridge, U.K.: Cambridge University Press.

Brady, M., and Hu, H. 1994. The Mind of a Robot. *Philosophical Transactions of The Royal Society of London A* 349:15–28.

Brooks, R. A. 1991. Intelligence without Representation. *Artificial Intelligence* 47:139–159.

Bundy, A. 1981. Some Suggested Criteria for Assessing AI Research. *AISB Quarterly*, 26–28.

Cliff, D. T. 1990. Computational Neuroethology: A Provisional Manifesto, Technical Report, CSRP 162, School of Cognitive and Computing Sciences, University of Sussex.

Cosmides, L., and Tooby, J. 1994. Beyond Intuition and Instinct Blindness: Toward an Evolutionarily Rigorous Cognitive Science. *Cognition* 50:41–77.

Dresher, B. E., and Hornstein, N. 1976. On Some Supposed Contributions of Artificial Intelligence to the Scientific Study of Language. *Cognition* 4:321–398.

Feigenbaum, E. A., and McCorduck, P. 1983. *The Fifth Generation: AI and Japan's Computer Challenge to the World.* Reading, Mass.: Addison-Wesley.

Gergely, G.; Nadasdy, Z.; Csibra, G.; and Biro S. 1995. Taking the Intentional Stance at Twelve Months of Age. *Cognition* 56:165–193.

Harnad, S. 1990. The Symbol-Grounding Problem. *Physica D* 42:335–346.

Haugeland, J. 1980. Programs, Causal Powers, and Intentionality. *The Behavioral and Brain Sciences* 3:432–433.

Hendler, J. A. 1989. Problem Solving and Reasoning: A Connectionist Perspective. In *Connectionism in Perspective,* eds. R. Pfeifer, Z. Schreter, F. Fogelman-Soulie, and L. Steels, 229–243. New York: Elsevier Science.

Hoffman, D. D., and Richards, W. A. 1984. Parts of Recognition. *Cognition* 18:65–96.

Hurlbert, A., and Poggio, T. 1988. Making Machines (and Artificial Intelligence) See. In *The AI Debate: False Starts, Real Foundations,* ed. S. R. Graubard, 213–239. Cambridge, Mass.: MIT Press.

Laird, J. E.; Newell, A.; and Rosenbloom, P. S. 1987. SOAR: An Architecture for General Intelligence. *Artificial Intelligence* 33:1–64.

Lenat, D. B.; Guha, R. V.; Pittman, K.; Pratt, D.; and Shepherd, M. 1990. CYC: Toward Programs with Common Sense. *Communications of the ACM* 33(8): 30–49.

McDermott, D. 1981. Artificial Intelligence Meets Natural Stupidity. In *Mind Design,* ed. J. Haugeland, 143–160. Cambridge, Mass.: MIT Press.

Malcolm, C., and Smithers, T. 1989. Symbol Grounding via a Hybrid Architecture in an Autonomous Assembly System, Research Paper, 420, Department of Artificial Intelligence, University of Edinburgh.

Marr, D. 1982 . *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco, Calif.: Freeman.

Marr, D., and Nishihara, H. K. 1978. Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes. *Proceedings of the Royal Society* B200:269–294.

Minsky, M. 1985. *The Society of Mind.* New York: Simon and Schuster.

Newell, A. 1980. Physical Symbol Systems. *Cognitive Science* 4:135–183.

Nilsson, N. J. 1983. Artificial Intelligence Prepares for 2001. *AI Magazine* 4(4): 7–14.

Penrose, R. 1989. *The Emperor's New Mind: Concerning Computers, Minds, and the Law of Physics.* Oxford: Oxford University Press.

Premack, D. 1990. The Infant's Theory of Self-Propelled Objects. *Cognition* 36:1–16.

Reddy, R. 1988. Foundations and Grand Challenges of Artificial Intelligence. *AI Magazine* 9(4): 9–21.

Reeke Jr., G. N., and Edelman, G. M. 1988. Real Brains and Artificial Intelligence. In *The AI Debate: False Starts, Real Foundations*, ed. S. R. Graubard, 143–173. Cambridge, Mass.: MIT Press.

Reeke Jr., G. N.; Sporns, O.; and Edelman, G. M. 1990. Synthetic Neural Modeling: The "Darwin" Series of Recognition Automata. *Proceedings of the IEEE* 78(9): 1498–1530.

Rey, G. 1986. What's Really Going on in Searle's "Chinese Room." *Philosophical Studies* 50:169–185.

Richards, W. 1988. *Natural Computation.* Cambridge, Mass.: MIT Press.

Ringle, M. 1983. Psychological Studies and Artificial Intelligence. *AI Magazine* 4(1): 237–243.

Sayre, K. M. 1986. Intentionality and Information Processing: An Alternative Model for Cognitive Science. *The Behavioral and Brain Sciences* 9:121–166.

Schnepf, U. 1991. Robot Ethology: A Proposal for the Research into Intelligent Autonomous Systems. In *From Animals to Animats,* eds. J. A. Meyer and S. W. Wilson, 465–474. Cambridge, U.K.: Cambridge University Press.

Searle, J. R. 1980. Minds, Brains, and Programs. *The Behavioral and Brain Sciences* 3:417–424.

Sejnowski, T. J.; Koch, C.; and Churchland, P. S. 1988. Computational Neuroscience. *Science* 241:1299–1306.

Simon, H. A. 1995. Artificial Intelligence: An Empirical Science. *Artificial Intelligence* 77:95–127.

Spelke, E. 1994. Initial Knowledge: Six Suggestions. *Cognition* 50:431–445.

Streri, A. 1993. *Seeing, Reaching, Touching: Relations between Vision and Touch in Infancy*. London: Harvester.

Waltz, D. L. 1983. Artificial Intelligence: An Assessment of the State of the Art and Recommendation for Future Directions. *AI Magazine* 4(3): 55–67.

Wilson, S. W. 1991. The Animat Path to AI. In *From Animals to Animats,* eds. J. A. Meyer and S. W. Wilson, 15–21. Cambridge, U.K.: Cambridge University Press.

Yeap, W. K. 1988. Toward a Computational Theory of Cognitive Maps. *Artificial Intelligence* 34:297–360.

Yeap, W. K., and Handley, C. C. 1991. Four Important Issues in Cognitive Mapping. In *From Animals to Animats,* eds. J. A. Meyer and S. W. Wilson, 176–183. Cambridge, U.K.: Cambridge University Press.

Yeap, W. K.; Jefferies, M. E.; and Naylor, P. S. 1991. An MFIS for Computing a Raw Cognitive Map. In Proceedings of the Twelfth International Joint Conference on Artificial Intelligence. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.

Wai Kiang Yeap is a director of the AI Research Laboratory at the University of Otago, New Zealand. His primary research interest is to develop an aI theory of the mind. To date, he has developed a computational theory of cognitive maps and is currently working on a computational theory of natural language understanding. His e-mail address is yeap_wk@cs.otago.ac.nz.