AI Magazine Volume 12 Number 1 (1991) (© AAAI)

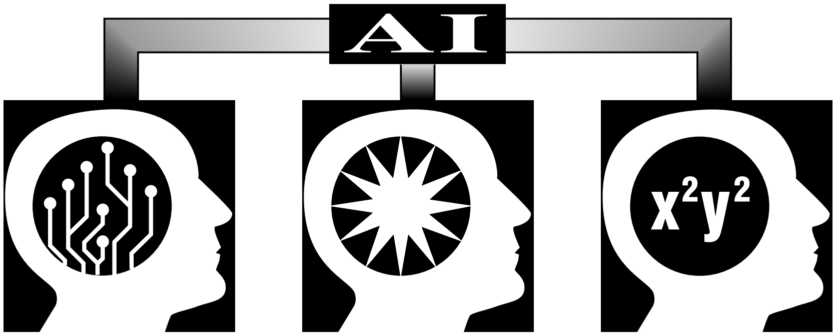
The Computational Metaphor and Artificial Intelligence: A Reflective Examination of a Theoretical Falsework

David M. West and Larry E. Travis

Advocates and critics of AI have long engaged in a debate that has generated a great deal of heat but little light. Whatever the merits of specific contributions to this ongoing debate, the fact that it continues points to the need for a reflective examination of the foundations of AI by its active practitioners. Following the lead of Earl MacCormac, we hope to advance such a reflective examination by considering questions of metaphor in science and the computational metaphor in AI. Specifically, we address three issues: the role of metaphor in science and AI, an examination of the computational metaphor, and an introduction to the possibility and potential value of using alternative metaphors as a foundation for AI theory.

Since its inception at Dartmouth in 1956, AI has strongly been championed by its advocates (H. Simon, A. Newell, E. Feigenbaum, P. McCorduck, and so on) and strongly challenged by its critics (J. Weizenbaum, J. Searle, H. Dreyfus and S. Dreyfus, T. Roszack, and so on). Despite (perhaps because of) the significant amount of highly charged debate between defenders and debunkers, little merit has been afforded those who would engage in a reflective analysis of the underlying assumptions, paradigms, metaphors,and models of AI.

Just how little can be illustrated by the reaction to Winograd and Flores's (1986) recent book *Understanding Computers and Cognition*. In personal comments, the book and its authors have been savaged. (Winograd, in particular, has been described as a turncoat because the book constitutes such a radical departure from his early work.) Published comments are, of course, more temperate (Vellino et al. 1987) but still reveal the hypersensitivity of the AI community in general to any challenge or criticism. Similar reactions to



0738-4602/91/$4.00 ©1991 AAAI

Penrose's (1989) even more recent book *The Emperor's New Mind* have been observed.

Winograd and Flores concluded that the premises supporting mainstream AI (and the tradition from which they derived) are fundamentally flawed. Most of the AI community reacted by categorizing their book as just another polemic to be dismissed out of hand. Lost in the rush to dismiss was the real value that might have been obtained by the open, reflective discussion of the underlying assumptions and presuppositions of AI and how they have affected (positively and negatively) our research programs. Winograd and Flores pointed out the need for a reflective examination of the perspective from which AI problems are traditionally approached.

Like Suchman (1987) and Clancey (1987), we feel that insights of significant value are to be gained from an objective consideration of traditional and alternative perspectives. Some efforts in this direction are evident (Haugeland [1985], Hill [1989], and Born [1987], for example), but the issue requires additional and ongoing attention.

The Importance and Role of Metaphor

One starting point for such an objective and reflective consideration concerns the use of metaphor or, more specifically, the role of the computational metaphor in establishing and maintaining the traditional perspective of AI research.

Why metaphor? One answer begins with Quine's (1979) observation:

> Along the philosophical fringes of science we may find reasons to question basic conceptual structures and to grope for ways to refashion them. Old idioms are bound to fail us here, and only metaphor can begin to limn the new order. If the venture succeeds, the old metaphor may die and be embalmed in a newly literalistic idiom accommodating the changed perspective. (p. 159)

Although AI has deep philosophical (R. Descartes and T. Hobbes) and allegorical roots,[1] its formalization as a discipline is usually traced to the 1956 Dartmouth conference organized by John McCarthy and Marvin Minsky. Because of its relative youth, AI still inhabits Quine's fringe of science where metaphor plays a major (if not the central) role, shaping theory, guiding research, and defining concepts.

A second answer to our question is provided by MacCormac's (1985) cogent discussion of the computational metaphor; its central role in AI; and its potential for obscuring, as well as illuminating, research in this area. (We make frequent references to MacCormac's work in the following pages.) Failure to fully consider metaphors and their attendant presuppositions or premature attempts to kill the metaphor by taking it literally (as advocated by Pylyshyn [1985]) is to risk the creation of myth or dogma instead of fostering the appropriate climate of open investigation.

Our reflective discussion of the role of metaphor in AI proceeds in three parts: (1) a brief review of the role metaphor plays in scientific thought in general, (2) a discussion of the computational metaphor and its role as the base metaphor of AI (particular attention is given to its association with a broader philosophical tradition and the reasons that it continues to operate despite factors that might ordinarily be expected to weaken it), and (3) a consideration of some alternatives to the standard computational metaphor that focuses on the potential theoretical insights that might be gained by adopting (or including) the perspectives suggested by these alternative metaphors.

Our discussion deals with theoretical and philosophical issues, so it is important to register our awareness of the distinction between what Winograd (1987) calls the dream of AI (a unified—if ill-defined—goal for duplicating human intelligence in its entirety) and its technical program (a fairly coherent body of techniques that distinguish the field from others in computer science). Similarly, because we are extensively dealing with metaphor, we need to also keep firmly in mind Shanker's (1987) distinction between prose and technical descriptions used in the cause of scientific theory. Failure to make the distinction leads to arguing against the technical accomplishments of AI because they fail to embody the fullness of its prose. Noting this failure goes a long way toward putting the attacks of AI critics such as J. Searle, H. Dreyfus, and S. Dreyfus in their proper place.

Throughout the discussion, the reader is admonished to remember that our intent is to conduct a reflective examination of AI, not an attack on it.

Metaphor in Scientific Discourse

> Explanations without metaphor would be difficult if not impossible, for in order to describe the unknown, we must resort to concepts that we know and understand, and that is the essence of metaphor—an unusual juxtaposition of the familiar and the unfamiliar. (MacCormac 1985, p. 9)

Metaphors provide… significant heuristic value as guides to further investigation.

That metaphor is used in scientific discourse or even the how or why of its use is no longer questioned. However, one important aspect of why metaphor is used should be noted. Metaphors provide not only the basis for explaining the unknown but also significant heuristic value as guides to further investigation.

A metaphor can, in fact, be evaluated in terms of its fruitfulness—the ability to suggest (frequently through subsidiary metaphors) possible lines of exploration. A primary criterion for the continued use of a metaphor is its fruitfulness, a criterion, it should be emphasized, that does not preclude the use of multiple metaphors applied to a particular domain. (We return to this issue later.)

AI is a theoretical domain that makes liberal use of metaphor at every level of inquiry. As such, it provides a rich subject for students of metaphoric language use and the consequences of such use.

MacCormac addressed metaphor in science and AI in some detail but with a focus on the theory of metaphor rather than the theory of AI. We want to use his work as a foundation for our own analysis, so it is appropriate to review those issues raised by MacCormac that are most germane to our primary concerns. Three issues have direct bearing:

First is the variance in the strength of a metaphor: How effectively does a metaphor express a similarity relationship between two objects that would otherwise be considered dissimilar and unrelated?

Second is the manner in which a metaphor evolves from a highly poetic expression to a lexical assertion of truth and the conditions that support this evolution.

Third is the direction of a metaphor, the attempt to define or explicate a strange entity in terms of a known entity.

Metaphoric direction is the simplest issue and is addressed first. That a metaphor has direction is implicit in the definition of *metaphor* as a linguistic device that attempts to provide insight into the nature of an unknown entity in terms of a known entity. A well-known example is the use of a solar-planetary system as a metaphor for atomic structure. The direction in this metaphor is from the known planetary system toward the (at the time the metaphor was first coined) unknown atomic structure.

Once established, the direction of a metaphor would not be expected to reverse except and unless the unknown entity came to be better understood than the known entity, and it was felt that understanding of the formerly known (now strange) entity would be enhanced by a comparison with the formerly strange (now known) entity.

Examples of this kind of directional reversal are rarely found in the realm of scientific metaphors. However, AI and its computational metaphor seems to be one such rare example. Whether it is and what the implications are is central to the discussion in the next section.

The strength or expressiveness of a metaphor is our second issue of interest. When a metaphor is established, it is an assertion of the similarity of two entities. The assertion is not, however, the similarity of whole entity to whole entity but the similarity of an attribute (what MacCormac labels a *referent*) of one entity to an attribute of the other. Each entity has a number of attributes or referents, and a metaphor is really a statement of analogy between paired members of the two sets of referents.

Not all paired referents are similar, and it is the balance between the number of analogous and the number of disanalogous referents that determines the expressive power of a metaphor. The strength or expressiveness of a metaphor can be thought of as the ratio of similar to dissimilar pairs of referents.

A corollary observation is that the expressiveness of a metaphor can vary as a function of the complexity (number of perceived referents) of each of the objects it relates. For example, the Apache metaphor “whitemen are carrion beetles” can be considered highly expressive if the set of attributes for both “whiteman” and “carrion beetle” are reduced to a set of particular behavioral referents.

At this general level (low complexity, few referents), there is a great deal of similarity (at least in the eyes of the Apache) between the two entities, and the metaphor can be considered highly expressive. The more complexity that is allowed in the definition of "white-man" and "carrion beetle," the greater the chance that paired referents will be found dissimilar, which weakens the metaphor.

Thus, we would expect that the strongest metaphors are those that relate objects about which relatively little is known. As the actual nature of metaphorically related entities is studied, and the dissimilarity of their constituent referents revealed, it is not at all unusual for the metaphor to become strained and disappear from scientific discourse. One excellent example comes from physics. The metaphor that presented particles as tiny spheres behaving like planetary bodies dissolved as the scientific understanding of atomic structures increased.

MacCormac divides metaphors into two types based on a measure of their expressiveness. *Diaphors* are primarily suggestive and poetic; they are the metaphors where either the referents are few in number, or there is a low ratio of similar-dissimilar referents. *Epiphors* are more descriptive and expressive because they have a higher similarity-dissimilarity referent ratio and, usually, because the total number of paired referents is also greater. Poetry makes much greater use (and is appreciated on the basis) of diaphor, but descriptive prose, especially scientific prose, strives toward epiphoric use. Although the scientist is not prohibited from creative expression (witness the names used in particle physics to describe quanta attributes), it is the poet who is generally allowed greater leeway in coining tenuous metaphors. This basis for measuring the expressiveness of metaphor leads to the third issue we want to raise: the tendency, over time, for a metaphor to evolve from diaphor to epiphor or vice versa.

When a metaphor is introduced to aid some scientific understanding, it can be highly suggestive or speculative in nature, for example, Kenneth Johnson's (1979) introduction of colored quarks. Only one, specific attribute of color (blending characteristics) is diaphorically related to one specific aspect of quark behavior (composition phenomena).

A diaphor of this type is unlikely to evolve to either epiphor or lexical status because of the limited number of analogous referents. If, however, further research into quarks revealed additional similarities between attributes of quarks and color, the number of analogous referents would increase as would the chances of the diaphor shifting to the more concrete status of epiphor.

The shifting of a metaphor's status between diaphor and epiphor is, therefore, to be expected. As understanding of each of the metaphorically related entities increases, the number of referents should increase, and the similarity-to-dissimilarity ratio among referent pairs should increase or decrease. If they decrease, then it is likely that the metaphor will evaporate. If they increase, the diaphor will solidify into an epiphor and perhaps even become a commonplace lexical statement. Either outcome is nonproblematic if the change in status is the result of empirical observation and increased understanding.

However, MacCormac notes, there is another manner in which a diaphor can assume epiphoric or even lexical status. Popular and repeated use of the metaphor is sufficient. A metaphor that captures the popular imagination and is often repeated because of its clever appeal comes to be accepted as a literal expression even when the level of analogy among referents can severely be questioned. (A well-known example of this phenomenon is the planetary model of atomic structure and its continued widespread use in the popular culture in the United States during the 1950s.)

It is useful to make a further distinction in MacCormac's classification of metaphors to account for metaphors that change status through common use rather than empirical confirmation. We propose the term *paraphor* for this class of metaphors.

Paraphor is derived from the term paradigm (as it is used by Kuhn [1970[) and the term metaphor. It is used to denote a metaphor whose use is so common (sometimes because it is so useful) that its metaphoric nature is commonly ignored, even in scientific discourse, unless there is a direct and explicit challenge made to its literal status. A paraphor acts as a kind of first principle whose use is pervasive and convenient to the point that it becomes a standard reference point for any expansion of theory in the realm in which it is used.

Paraphors corrupt the manner in which metaphor is commonly used in science. To illustrate, consider the way that metaphors suggest auxiliary metaphors. A primary metaphor,

a computer is a kind-of-mind,

yields secondary metaphors,

computing is a kind-of-thinking,

and

pattern matching is a kind-of-seeing

This use of metaphor has its drawbacks, noted by MacCormac (1985):

> Metaphors can be dangerous not only in bewitching us into thinking that what they suggest really does exist but also in leading us to believe that the attributes normally possessed by any of

Metaphoric use continues to be central to the understanding and teaching of AI *theory…*

> the referents in the metaphor are possessed by the others. (p. 17)

Secondary, derivative metaphors should be regarded with greater caution than primary metaphors, and when the primary metaphor is diaphoric or epiphoric in status, this caution is usually acted on. However, if the primary metaphor is a paraphor, the secondaries are often accepted without sufficient critical examination simply because they are consistent with the primary. In essence, the metaphoric nature of the primary is ignored or forgotten.

> By forgetting that theories presuppose basic metaphors and thereby by taking theories literally, both scientists and theologians create myths. (MacCormac 1985, p. 17)

From myth, it is but a short step to dogma and the situation where theoreticians come to be classified as true believers or heretics. (The terms true believer and heretic are themselves metaphors, ones that illustrate the transfer of a polemic function from one domain to another.)

Ordinarily, some set of first principles, axioms, presuppositions, or articles of faith must exist to support dogmatic (hegemonic) assertions. A paraphor can, however, subsume first principles, draw attention away from axioms, and become a convenient shorthand expression of the articles of faith. The paraphor becomes the symbol that is publicly exchanged to communicate the background assumptions that are assumed to be shared by all practitioners in a domain of science.

Paraphors can also assume the status of lexical assertion in common use, hiding behind the label of metaphor when directly challenged. This status represents another contrast between paraphors and epiphors that come to be regarded as lexical terms. Lexical status for an epiphor is achieved when empirical observation increasingly confirms similarities between two entities. Paraphors need not pass this rigorous test to achieve lexical or quasi-lexical status. This statement opens the question about when and under what circumstances it is appropriate to allow a metaphor (whether epiphor or paraphor) to assume lexical status.

In some cases, taking a metaphor literally is an obvious error—animism being a simple example. At other times, the literal acceptance of a metaphor can create a completely false but essentially benign understanding of a scientific phenomenon, as in the case of the planetary model of atoms, which is wrong in detail but nevertheless retains usefulness as an illustration at elementary levels of science. Occasionally, the argument is made that a science cannot advance unless the metaphor is granted lexical status.

> Given that computation and cognition can be viewed in these common abstract terms, there is no reason why computation ought to be treated as merely a metaphor for cognition, as opposed to a hypothesis about the literal nature of cognition. In spite of the widespread use of computational terminology, much of this usage has had at least some metaphoric content. There has been a reluctance to take computation as a literal description of mental activity, as opposed to being a mere heuristic metaphor. In my view this failure to take computation literally has licensed a wide range of activity under the rubric of information processing theory, some of it representing a significant departure from what I see as the core ideas of a computational theory of mind. Accepting a system as a literal account of reality enables scientists to see that certain further observations are possible and others are not. It goes beyond merely asserting that certain things happen as if some unseen events were taking place. In addition, however, it imposes severe restrictions on a theory-builder, because he is no longer free to appeal to the existence of unspecified similarities between his theoretical account and the phenomena he is addressing—as he is when speaking metaphoricly. (Pylyshyn 1980, p. 115)

To a degree, Pylyshyn's points are well taken. Metaphor does allow a wide range of sometimes wildly speculative theory, but such is the purpose of metaphor—to illuminate the unknown with the spotlight on "as if." It is equally true that a science is not advanced by certain metaphor-based speculations. Science fiction, however entertaining, is not science. There is also nothing wrong, in principle, with a requirement that theorizing be constrained by a certain analytic rigor.

Pylyshyn's arguments are far less convincing, however, in a situation where a prevailing metaphor has attained its prominence simply from overwhelming use rather than an increasing body of corroborative similarities in the referents of the metaphor. In the former case, to take the metaphor literally could be a major mistake. There are many other aspects of metaphor theory and use, but the three preceding issues are sufficient background for introducing the proper status and role of the computational metaphor in AI.

The Computational Metaphor and Artificial Intelligence

Defining the computational metaphor is simultaneously a simple and a difficult task. It is simple in that it can be abridged as a simple statement:

mind (brain) is a computer is a mind (brain)

It is difficult for at least two reasons. The first is nuance. The simple form, such as the evolutionary statement "man is descended from apes," is carelessly stated. Just as the evolutionary statement should be "man and apes share a common evolutionary ancestor," so it should be stated that minds and machines are both instances of a single abstract entity, perhaps a physical symbol system. The second reason is that, precisely speaking, the metaphor is not a single metaphor. Rather, it is a family of closely related metaphors, each of which attributes an aspect of similarity between referents of the objects we call computers and those we call minds (as embodied in brains).

Compounding the difficulty is the extent to which and the speed with which various aspects of the metaphor have essentially invaded every realm of scientific investigation as well as infused the popular culture and vernacular, a process noted by Turkle (1984) and Bolter (1984):

> By promising (or threatening) to replace man, the computer is giving us a new definition of man, as an information processor, and of nature, as information to be processed. (Bolter 1984, p. 13)

Bolter also provides a historical context for the computer metaphor by showing that it is, in fact, only the latest example in a long line of technological metaphors used to explain man and nature. The Greeks, for example, explained the functioning of the universe and the passage of human lives in terms of the spinning wheel.[2]

The complex of metaphoric use that is commonly subsumed under the label of the computational metaphor has been accepted and defended by AI practitioners from the inception of the discipline.

> To ascribe certain beliefs, knowledge, freewill, intentions, consciousness, abilities, or wants to a machine or computer program is legitimate when such an ascription expresses the same information about the machine that it expresses about a person. It is useful when the ascription helps us understand the structure of the machine, its past or future behavior, or how to repair or improve it. (McCarthy 1979, p. 147)

Metaphoric use continues to be central to the understanding and teaching of AI theory, even in introductory textbooks:

> The fundamental working assumption, or central dogma of AI is this: what the brain does may be thought of at some level as a kind of computation. (Charniak and McDermott 1985, p. 6)

Note that McCarthy ascribes mental attributes to computers, and Charniak ascribes computer attributes to minds. This issue is important later.

Given that the computational metaphor is the base metaphor of AI, what status should it be accorded? Is it diaphoric, epiphoric, paraphoric, dogmatic, or lexical? Should it be dissolved or radically redefined in some manner?

Arguments for dissolving the metaphor arise when we consider that the theory of metaphor predicts that any metaphor dissolves when ongoing investigation increases the number of referents for the related entities, and the dissimilarity among pairs of these referents increases. This situation seems to be the case when considering the mind-is-a-computer metaphor.

Minds are organically based, not programmed, inconsistent, only generally persistent, poor at math, great at pattern recognition, and so on, but computers are the exact opposite in essentially every way. Computers incorporate transistors, consist of clearly defined and limited-function electric circuits, operate on discrete packets of data, maintain memory in specific locations, are generally nonadaptive, are completely deterministic, are brittle (in the sense that power interruptions or magnetic fields render them useless), cannot deal with ambiguity in a satisfactory manner, take immense amounts of time to perform certain classes of operations, and are amazingly accurate. Minds, however, exhibit none of these properties.

Why then has the metaphor not evaporated? Is it the case that the computational

…technological success… is not sufficient to account for the persistence of the metaphor.

metaphor, like the planetary metaphor of atomic theory, is generally acknowledged as incorrect in most details but still useful in situations requiring only an elementary and superficial explanation? Few AI practitioners would agree that this case is true. How then do we account for the persistence of the metaphor?

Three reasons might be given: (1) technological success, (2) metaphoric conflation, and (3) expressive power. Each of these reasons is briefly discussed.

Technological Success

Technological success is a simple, pragmatic reason that acknowledges that the metaphor has been enormously useful in the development of AI as a technology. Even the strongest critics generally agree that AI is responsible for major technological accomplishments. Most agree that systems built in accordance with the computational metaphor can usefully simulate human performance even if they will never be able to fully emulate such performance. AI advocates further argue that the approach has been sufficiently fruitful that it merits continued pursuit.

Although technological success is a strong motivation for perpetuating the metaphor, there have been sufficient setbacks and cogent criticisms of AI's theoretical approach that such success is not sufficient to account for the persistence of the metaphor.

Metaphoric Conflation

Metaphoric conflation arises from the fact that the computational metaphor is operationally bidirectional rather than unidirectional.

To fully see the bidirectional nature of the computational metaphor, it is necessary to briefly adopt a historical perspective. Although it is impossible to reconstruct exactly how and why the first metaphors were employed in the computing arena, it is not surprising that mental images were used. (It should be noted, however, that Babbage used metaphors of mills and stores derived from the mechanistic technology of his era rather than mental metaphors.)

The first computers were unknown, speculative entities. Metaphor was required to explain what the computer was and what it was intended to do. The computer was the strange object in the metaphoric relationship, and the known object was mind. Mind was, of course, known only in the common introspective sense where people think they are generally familiar with how the mind works or, at least, have a vocabulary for vaguely describing how it works.

When computers were connected to peripheral devices such as tape drives and check sorters that were controlled by signals generated without immediately obvious human action, it was natural that the computer was seen as the controlling (volitional) agent. This attribute was added to others, such as the computer being able to read input data and write output information. Periods of processing that did not exhibit any outward sign of action came to be described as thinking. Architecturally, the computer had both a memory and a brain.

A subtle but significant distinction can be made between the early computer-as-mind metaphors and those like colored quarks. In the latter case, the known object (color) was, in fact, an objectively understood phenomenon. An accepted theory of color combination existed to supply referents for the color portion of the metaphor. This case is not true in the computer-as-mind example. No objectively understood and accepted theory of mind existed (nor exists) to fill the position of a known object in the metaphoric construct. What does exist to fill this position is a set of commonsense (experience-derived) terms for mental states.

When initially proposed, mental metaphors of computer function were clearly diaphoric. Even today, when they are applied to a particular incarnation of a computer, they are less than epiphoric even if more than diaphoric—still, clearly metaphors. They are more than diaphoric because we sufficiently understand the workings of computers that we find it difficult to think of our personal computer as really thinking or talking. However, this difficulty seems to dissipate when the computer in question is an abstract entity (a symbol-processing system). Mentation-derived metaphors applied to the operation of computer systems became so commonplace that sheer use advanced their status from clear diaphor to strong paraphor.

By the late 1950s, computers were no longer strange and exotic. They were well understood in theoretic, as well as engineering, terms. Computers had even become a fixture in the popular culture of the Western world as sources of humor, anxiety, and frustration. At this point, a second set of metaphors emerged that related the computer as a known entity to the mind as an unknown entity.

The commonly adopted but scientifically erroneous notion that the mind was something we understood well enough to use as a metaphor for the unknown computer faded even though the mental metaphors applied to computers persisted. Instead, the computer became the metaphor of choice for a renewed research effort directed toward understanding the mind.

These new metaphors were not tentatively proposed as mere diaphors but blossomed with full epiphoric status. In large part, this status came about because the distance between computers and minds as dissimilar entities had already been bridged and significantly reduced by the first set of metaphors that related minds to computers. Those hearing the new metaphors were predisposed to accept them.

Computer-derived metaphors for the mind were found to be less satisfactory than mind-derived metaphors for computers. Resistance to their glib application was encountered. Several reasons can explain the different levels of acceptance for this second class of metaphors. One explanation is that insights generated by these metaphors did not result in a consumer product analogous to the computer, so they did not capture the popular imagination to the same extent. Another explanation is that people resisted the notion that they were simply meat-based computer systems. In large part, however, they were less successful because the dissimilarities between computers and minds assumed greater theoretical significance.

A situation existed that should have resulted in the dissolution of the metaphor, in accordance with the theory outlined by MacCormac. Instead, the metaphors were preserved; however, they were held to apply between minds and computers in the abstract rather than the physical manifestations of these abstractions. Abstract computers (for example, Turing machines) and abstract function (algorithms, programs, pattern matching, and so on) supplied abstract entities for mind-as-computer metaphors.

Abstractions, however, do not provide the well-known entities that are capable of supplying referents for correlation with referents associated with the unknown entity—the metaphoric object. Tacit acknowledgment of this state of affairs can be seen in the shift from attempts that metaphorically relate minds and computers to the assertion that both minds and computers are examples of a single abstract class, perhaps physical symbol systems.

We now have two reasons to expect that the computational metaphor would have long since disappeared: dissolution by dissimilarity and replacement by a definition. Instead, both classes of metaphor persist and have, in fact, merged into a single, persistent supermetaphor. This situation results in part from the blurred distinction between any two metaphorically related entities.

> In an interaction metaphor both parts of the metaphor are altered. When we claim metaphorically that computers think, not only do machines take on the attributes of human beings who think... but thinkers (human beings) take on the attributes of computers. And that is exactly what has happened in the case of the computational metaphor: the mind of a human being is described in terms of the attributes of a computer. We talk about the neuronal states of the brain as if they were like the internal states of a computer; we talk of the mental processes of thinking as if they were algorithmic. (MacCormac 1985, p. 10)

Expressive Power

The third reason for the persistence of the metaphor derives from the seeming expressive power of the metaphor. We noted in the previous section that the expressive power of a metaphor is a function of the number of paired referents of each entity that are found to be similar. Expressiveness is also a measure of the heuristic value of the metaphor to suggest new lines of research.

Expressiveness is essentially determined by the ratio of similar referent pairs to total referent pairs. This situation allows a kind of pseudoexpressiveness in those situations where the total number of referents is small, and several of them are similar.

The computational metaphor is considered

Thus… the computational metaphor should no longer be considered a metaphor in the normal sense of the term.

highly expressive, but it might be more properly considered pseudoexpressive for at least two reasons. First, the process of abstraction, as previously described, reduces the total number of referents available for pairing. A Turing machine, for example, has significantly fewer referents than an actual, physically realized digital computer. Those referents that are available are virtually assured to be similar to referents of other abstractions, but this situation is mainly a matter of the commonalities in the definitions of the abstractions, for example, the Turing machine and the symbol-processing system.

The second reason comes into play when metaphorically relating real computers with physical minds and derives from the bidirectionality. Because common terms are used to label the referents of both entities in the metaphors, most of these referents can be paired with similar referents in the opposing entity. For example, the term memory is used (and similarly defined) in both the computer context and the mind context. Referents paired in this manner are not, however, discovered referents; they are defined. Their definition presupposes the metaphor, and hence, a circular feedback loop is created.

As a consequence, the metaphor can be considered highly expressive in a technical sense, even though this expressiveness measure is inconsistent with the spirit of the theory of metaphor expressed by MacCormac. The notion of referent similarity as a result of empirical observation has mostly been lost.

Thus, we can conclude that the computational metaphor should no longer be considered a metaphor in the normal sense of the term. Should it be considered a lexical term (as Pylyshyn advocates)? Should we consider it to be a member of the class paraphor? We believe that the latter is more tenable, in part because of the strong association of the computational metaphor with the philosophic and scientific tradition that is often labeled formalist.

One of the basic assumptions behind this approach, sometimes referred to as information processing, is that cognitive processes can be understood in terms of formal operations carried out on symbol structures. Thus, it represents a formalist approach to theoretical explanation (Pylyshyn 1980).

Pylyshyn is an outspoken representative of a significant majority of ai theorists that, as a group, can be seen as inheritors of a longstanding philosophic tradition. Labels for this tradition (depending on which of its many themes is emphasized) include rationalism, dualism, formalism, and mechanism. In terms of modern Western philosophy, this tradition began with Hobbes, Descartes, Liebniz, and Locke and motivated the work of Russell, Whitehead, and Carnap. The work of Babbage, Turing, von Neumann, and Weiner is solidly grounded in the tradition, as is the theory of the founders of AI—McCarthy,. Minsky, Simon, and Newell.

A detailed discussion of this philosophic tradition is obviously outside the scope of this article. Instead, several topical points are briefly discussed: representation, formal operations, and mechanism. Pylyshyn (1980) raises two of these topics: formal operations and symbol structures.

Symbol structures are representations, maps that exist in the mind and that stand in place of the cruder, sensory objects that populate the real world. The notion of representation is relatively new in epistemology and is usually attributed to Descartes and Locke. Without this concept, however, the idea that computers could think would likely not have occurred.

Scholastic philosophers held that to know something you had to assimilate some portion of that thing's form:

> A thing's form is what makes it the kind of thing that it is, so that in knowing it the knower must in some sense become the same sort of thing as the object known. To know a horse is in some sense to become a horse, or perhaps to become horsey, to know God is in some sense to become divine. (Pratt 1987, p. 14)

Descartes (and successors) insisted on dissociating mind from matter, establishing the need for an intermediary between the mind and the world it perceived. This intermediary consisted of representations (ideas, symbol structures).

> Ideas are mental entities, the only items with which the mind can deal directly, but they stand for non-mental things about which the thinker has occasion to think. (Pratt 1987, p. 18)

Gardner notes that Descartes's notions of representation and dualism remain central to cognitive science and AI:[3]

> …the cognitive scientist rests his discipline on the assumption that, for scientific purposes, human cognitive activity must be described in terms of symbols, schemas, images, ideas and other forms of mental representation. (Gardner 1985, p. 39)

Genesereth and Nilsson (1987) also profess Descartes's theory:

> Note that in talking about the behavior of an intelligent entity in its environment, we have implicitly divided the world into two parts. We have placed an envelope around the entity, separating it from its environment, and we have chosen to focus on the transactions across that envelope. (p. 2)

These suppositions establish a position that requires the entire environment in which a thinking entity exists be recreated, in some sense, as knowledge before the thinking entity can deal with it. Thus, "A spider, for example, must use quite a bit of knowledge about materials and structures in spinning a web" (Genesereth and Nilsson 1987, p. 3).

Metaphysical Cartesian dualism, which deals with mind and body as separate substances, need not concern us, but methodological Cartesian dualism is a necessary precondition if entities such as computers are to receive serious consideration as thinkers.

Commonsense notions of thinking, however, retain vestiges of the premodern nondualistic concept of interactive knowledge. It is precisely this contrast between dualistic and nondualistic perspectives that is at the root of many of the debates about the ability of computers to really think, feel, and know. It is not our intent to discuss the merits of the two positions but, merely, to note that the concept of representations and Cartesian dualism are at the heart of the disagreement. This same dualistic notion is central to Keller's (1985) discussion of the differences between B. McClintock's genetics (allowing interaction) and mainstream genetics (Cartesian dualist) and is a focal point for Winograd's theoretical reevaluations.

The second Pylyshyn topic involves the formal operations that are applied to the representations. In modern philosophy, the notion of a set of formal operations that would encompass all human thought can also be traced to Descartes and his project to codify what he called the "laws of thought." Liebniz (another precursor) dreamed of "a universal algebra by which all knowledge, including moral and metaphysical truths, can some day be brought within a single deductive system" (Genesereth and Nilsson 1987, p. 4).

Despite a long and distinguished pedigree that includes Boole, Frege, Russell, Whitehead, Chomsky, Fodor, and many others, the idea of a set of formal operations that encompass all thought is far more problematic an assumption than abstract representation.

Descartes, for example, abandoned his grand project of codifying all such formal operations (although his *Discourse on Method* was to be a foundation for his ultimate vision). Central assumptions in the work of Frege, Russell, and Whitehead were disputed by Gödel. For everyone who advocates the position that a formal system of operations must exist, there is an opponent who maintains that the whole of human knowledge and understanding exceeds the limits of any formal system.

Regardless of the merits of this debate, why, beyond a desire for order and prediction, is a formal system a prerequisite to an adequate understanding of mentation? The answer is that a formal system might not be required to describe and understand mental operations, but one is certainly required if we are to build a machine capable of replicating these operations. It is not surprising, therefore, that many of the strongest advocates of formal systems (for example, Leibniz, Babbage, Turing, and von Neumann) were also actively engaged in the construction of machinery that would embody their systems.[4]

Preoccupation with the idea of building an autonomous machine capable of manipulating symbols according to rules of thought or logic necessarily limited formal representations to that subset that were also mechanical and constructivist. With the possible exception of some connectionist machines (derived from an alternative metaphor discussed in the next section), all attempts to build thinking machines are in the tradition of the earliest calculating clocks of Wilhelm Schickard and Blaise Pascal.

Following the conviction of Giambattista Vico that "one is certain of only what one builds" (Genesereth and Nilsson 1987, p. 1), AI researchers want to build machines (write computer programs) and are intensely distrustful of ideas that cannot be expressed in this material fashion. AI is so strongly steeped in this philosophic tradition of formalism and mechanism that the mind-as-computer and computer-as-mind metaphors found an immediate, accepting, and enthusiastic audience. This widespread acceptance, combined with the metaphoric conflation, effectively literalized the computational metaphor.

What remains is a phrase, "the computational metaphor," that does not denote a metaphor at all (except, perhaps, as a paraphor) but is a kind of shorthand expression for a philosophic point of view. Using the phrase is a declaration that the user is a follower of the modern dualistic tradition (beginning with Descartes, Leibniz, and Locke) that holds that (1) mind and nature are absolutely separate, and the scholastic

Critics can be given some credit for turning our attention to important problems…

concept of interactive knowledge is nonsense; (2) the mind manipulates abstract representations of the environment, never the environment directly, and it is this manipulation that constitutes thinking; and (3) the manipulations that make up mental functions can be expressed in a formal language; and (4) this formal language is basically mechanical and deterministic, at least sufficiently so that it can be embodied in a machine (computer) where it can function as a replicant of the functioning of the human mind.

As a form of shorthand, computational metaphor is useful. It is easier to establish one's perspective with a two-word phrase than an enumeration of the full list of basic assumptions behind this perspective. It should not be forgotten, however, that because the use of the phrase computational metaphor arose from the use of a true metaphor (actually a series of metaphors beginning with the clockwork metaphor of the sixteenth and seventeenth centuries) and because the status of these metaphors has been either abrogated or forgotten, a myth (MacCormac's sense) has been created.

Recognition of the myth summarized in the computational metaphor does not imply a judgment, either of the myth itself or of the usefulness of the work done by those adopting it. Arguing a myth is like arguing a religion; the argument is essentially pointless. (Especially a myth that at one time was thought to be divinely inspired: The Angel of Truth visited Descartes in a dream on 10 November 1619.)

Awareness of the myth is, however, important for several reasons: It explains, in part, the tone of debates between strong advocates and vehement critics of AI in general; it should act as a restraint on the sometimes shameless hyperbole that has come to be associated with AI; and it provides a bridge for communication between AI advocates and critics.

Why do we advocate a dialogue with the critics of AI and the consideration of alternative metaphors or perspectives for undertaking AI research? Whatever one thinks of the critics in general, they must be given credit for reminding us of the importance of solving the kinds of problems and questions that are at the focus of their critiques. (Most of these problems are already familiar and acknowledged as difficult by researchers in the field, but in the absence of the criticism, they are too often shelved for later consideration.)

Critics can be given some credit for turning our attention to important problems and creating the context in which we can question whether these problems are intrinsic to the AI enterprise or are artifacts of the specific perspective and the computational metaphor that has been the primary influence on AI research and development to date. If they are artifacts, then a consideration of the alternatives is not only warranted but necessary.

In the final section of this article, we review the case that the core problems of AI are artifacts of the formalist perspective that is incorporated in the computational metaphor; recall some of the major, contemporary, alternative metaphors for mind; and pose the question of how any of these alternatives (or a combination of them) might offer some promise.

Metaphoric Consequences and Points of Departure, or How to Learn from Your Critics without Surrendering to Them

Perhaps the most significant consequence of the formalist perspective as embodied in the computational metaphor is the mandate to recreate, in symbolic form, the totality of the environment (everything external to the mind) in which a thinker operates—to develop a mental simulacrum of the thinker's external world.

This requirement can be deferred in some instances when we want to model a simple mental operation or a severely restricted application domain. It is only deferred, not obviated, as shown by the scaling problem so often encountered when attempts are made to generalize an AI system.

Not only is the creation of a simulacrum a formidable task in and of itself, but this simulacrum must also be amenable to processing. It must be accessible as required. Two interrelated approaches to providing access are central to all AI theory: efficient search and representation.

First, consider the search problem. Although they can simply be stated, search models rapidly assume complex dimensions. The complexity arises from the potential size of the space that needs to be searched to solve a given problem. The space is large because the required symbolic simulacrum of the environment is large.

Human behaviors are not amenable to simple search procedures. The number of state sequences that need to be considered explodes in a combinatorial manner and is certainly beyond the capabilities of computer hardware to process in limited time. (*Limited* is imprecisely defined as before the user loses interest or dies, the computer wears out, or the universe comes to an end, whichever comes first.)

Significant amounts of effort have been expended in attempts to create faster hardware (very large system integration, parallel processors, and so on), overcome the search problem by brute force, and devise more efficient search strategies (minmax, means end, truncation) to take better advantage of the hardware that is available.

In contrast, the indirect method of addressing the search problem is to organize the space to be searched. This organization is provided by knowledge representation schemes. Frames, semantic nets, and logic programming can be regarded as methods for organizing knowledge spaces so that search is limited and constrained.

These observations should not be taken as an assertion that search optimization is the only—or even the primary—objective of logic programming, frames, and semantic nets. Each of these approaches can be seen, however, as a tacit or explicit recognition of the scope and complexity of the search space required to solve nontrivial problems and the inadequacy of general search techniques to deal with such a space without the help of some abstractive organization of the base-level search space.

Representation and search are intertwined and intimately related concepts: The need for success in one area is directly proportional to the inadequacy of the other. Weak representations require a greater reliance on search mechanisms that must be highly efficient. Historically, it was the weakness in search capabilities that surfaced first and propelled interest in representation.

It is generally overlooked, however, that search and representation issues do not need to arise and would not if we were using certain alternative paradigms, in particular those that do not incorporate Cartesian dualism. If the mind is allowed interaction with the environment and if knowledge of the environment involves some degree of merger with it (the scholastic epistemology noted earlier), then the environment could directly evoke a response as a function of posing a problem. Hard-core believers in B. F. Skinner's theories would argue that the stimulus-response paradigm in psychology (Skinner 1957) does not need to bother with search and representation issues for exactly this reason.

Logic systems are exemplars of the formalist perspective in the sense that every valid expression in a system can be generated from, or resolved to, a small set of axioms and a small set of combination rules. Although logic systems seem to be the perfect exemplars of the formalist perspective, in fact, they are seldom applied in areas requiring a detailed simulacrum of the real world. Instead, they deal with abstract conceptualizations, with a priori declarations of a universe of interest.

> …a conceptualization is a triple consisting of a universe of discourse, a functional basis set for that universe of discourse, and a relationship basis set.
> (Genesereth and Nilsson 1987, p. 12)

Each portion of the triple consists of only those objects, functions, and relations that the builder of the system considers important to a problem at hand.

At this level, a logic system is a pragmatically useful tool for building applications if the system to which these applications apply can be bounded in the real world. This utility is lost, however, when a logic system attempts to address a domain with no bounds or fuzzy bounds. Only if the conceptualization of a logic system attempts to incorporate a larger part of the real world will it be a true test of the feasibility of realizing a formalist model (certainly, a part larger by orders of magnitude than anything achieved to date).

Several other areas of AI research could be discussed, but we only mention the special case of vision research. Vision is one of the primary areas acknowledged within AI, and its complete absence from this discussion would be curious.

First-layer vision systems are primarily concerned with the transduction of physical phenomena into internal representations (Charniak and McDermott 1985). Here, we have more of a problem with emulating a sensory mechanism than emulating the mind that uses this sensory input to think.

AI concerns with vision are not limited, of course, to first-layer (sense-emulation) problems. Sensation becomes perception as that which is sensed is recognized and understood, a transition that succeeds only by considering situation and context. Transduced sensory input must be related with representations already stored. Mechanisms involved in this area are duplicative of those already discussed, for example, logic systems.

…there is nevertheless value in simply considering and exploring alternative metaphors.

One observation that needs to be made about ai vision research is the vast difference in richness between natural and artificial vision systems. Computer perception is severely limited compared to human perception (immediate recognition of a multitude of objects within a range of vision and instant evocation of detailed knowledge about these objects and their relationship with each other and the viewer), yet visual perception seems to be the most important channel for connecting with the outside world. In addition, it is difficult to imagine truly intelligent machines until human visual capabilities can more completely be emulated.

In summary, the main consequence of the computational metaphor and formalist perspective is the need to recreate, inside the mind and inside the computer, a symbolic simulacrum of an extensive portion of the external world in such a manner that it is amenable to processing in pragmatically finite time. This need is not a consequence of the objective to emulate the mind but the perspective from which this problem was approached.

Other perspectives exist, and other metaphors have been used as a basis for understanding the mind and might be usable in attempts to model and emulate the mind. At this point, we want to introduce some of these alternatives.[5]

Hampden-Turner (1981) offers a catalog of alternative metaphors. Most of the entries in his catalog enjoyed some measure of success in limited application realms. Whether the relative success of any one of them in its primary application domain is equaled by the computer metaphor in AI is open to question.

There is also a major, philosophy-derived alternative to the formalist perspective that might be investigated. We speak of the hermeneutical or interpretive perspective that is adopted by critics such as the Dreyfuses and, more recently, Winograd and Flores.

Hermeneutical philosophy enjoys as long a tradition as formalist philosophy and includes the works of W. Dilthey, H.-G. Gadamer, M. Heidegger, E. Husserl, P. Ricouer, M. Merleau-Ponty, and L. Vygotsky. Winograd and Flores acknowledge Gadamer and Heidegger as principle influences on their work, but H. Dreyfus is more closely associated with Husserl and Merleau-Ponty. Leaf (1979) documents how our definitions of, and theories concerning, man have alternated between the formalist and the hermeneutical (or interpretive) poles. Gardner (1985) also notes some aspects of the tension and alternating periods of ascendancy between these two perspectives.

The central point of divergence between the two positions concerns a context that is also a process, which is missing from (its existence denied by) formalist models and is considered critical by advocates of hermeneutics. It is missing from formalist models for two reasons: the context component because of its scope, the process component because it is particularistic and ephemeral.

Several contemporary alternative metaphors appear to provide the same kind of supportive vehicle for realizing the hermeneutical perspective as the computer provided for formalism.

Connectionism is the most obvious alternative metaphor, one based on the electrophysical architecture of the human brain. The operation of connectionist systems is explained by replacing the calculating clock metaphor with the landscape metaphor. J. Hopfield said

> …neural nets have contours like the hills and valleys in a countryside; they also have stable states. (Allman 1986, p. 26)

Connectionism has the potential to eliminate one of the two drawbacks of formalism noted earlier—the need for a formal language of thought. It does not, however, directly address the dualist aspect of Cartesian formalism. Connectionist systems still seek to recreate a simulacrum of the environment in the system, albeit in distributed, rather than discrete, form.

Others, such as Bergland (1985), have also looked to the human brain for metaphoric inspiration. Where connectionists such as Hopfield see electric circuits, however, Bergland sees a gland and a hormonal soup.

Bergland criticizes the prevailing electric perspective of what the brain is and how electric activity constitutes the basis of thought. For him, the critical juncture is the *synapse*, the point at which any electric circuit is closed and, therefore, established. This closure is a function of brain chemistry in physical proximity to the synapse more than the strength of currents flowing along dendrites. This chemistry, in turn, is a function of hormonal production throughout the human body as it interacts with the physical and psychological environment outside the body.

> …the primary mechanisms of intelligent thought must be viewed differently. The mind is made pattern dependent and comes to share in the ubiquitous secret of evolutionary survival: pattern recognition. The mechanisms of mind are thus released from the conceptual confines of the reductionistic left brain. The mechanisms that drive thought are found all over the body and, wherever they live, they function at their highest level by recognizing the molecular patterns of the combination of hormones that modulate thought. (Bergland 1985, p. 109)

Bergland's metaphors are congruent with two other metaphors: Conrad's tactilizing processors and Maturana's and Varela's evolutionary-adaptive processor model.

Conrad (1987a) derives his metaphor from the operation of protein enzymes and their folded shapes:

> Any computational function that can be implemented using conventional switching elements can be implemented using tactilizing processors, and, in general, much more efficiently. All conventional switches do is recognize simple patterns (such as 11 or 10). Recognizing complex patterns requires networks of many simple switches, whereas tactilizing processors are capable of recognizing complex spatio-temporal patterns by themselves. (p. 13)

Maturana and Varela (1987) are a major source of inspiration for the arguments presented by Winograd and Flores:

> …for the operation of the nervous system, there is no inside or outside, but only the maintenance of correlations that continuously change… The nervous system … is the result of a phylogenetic drift of unities centered on their own dynamics of states. What is necessary, therefore, is to recognize the nervous system as a unity defined by its internal relations in which interactions come into play only by modulating its structural dynamics, i.e., as a unity with operational closure. (p. 169)

This metaphor is consistent with both Bergland's gland metaphor and Conrad's tactile processor metaphor, pointing, perhaps, to the possibility of a hybrid organismic metaphor for the mind.

It should be noted that Maturana and Varela and Winograd and Flores primarily use this metaphor as a basis for establishing an ethics of mind and of language-based communication rather than as a foundation for a theory or a model of the mind.

Other alternative metaphors that deserve consideration are Minsky's (1986) society of mind and Pribram's (1971) holographic model.

Whether any of these alternatives is likely to be more promising in the long run than the computational metaphor is a question that will not be answered for decades. It will never be answered if alternative metaphors and philosophical perspectives are used only as redoubts from which to lob polemic missiles at formalists.

However this question is ultimately decided, there is nevertheless value in simply considering and exploring alternative metaphors. When such consideration arises from a reflective effort to expand our understanding and awareness of our science, we will derive benefits, in the form of either new approaches to our hard problems or new insights into existing approaches.

Instead of dismissing critics of AI and those that propose alternative metaphors, we should use their criticisms to direct our attention to the presuppositions, paradigms, and metaphors at the heart of our discipline. Only after using these elements are we in a position to answer such questions as, Have we selected the best metaphors? Do our often-used metaphors serve our best research interests, or are we being misled by them?

Bibliography

Allman, W. F. 1986. Mindworks. *Science 86* 7(4): 22–31.

Bergland, R. 1985. *The Fabric of Mind.* New York: Viking.

Bolter, J. D. 1984. *Turing's Man: Western Culture in the Computer Age.* Chapel Hill, N.C.: University of North Carolina Press.

Born, R., ed. 1987. *AI: The Case Against.* New York: St. Martin's.

Charniak, E., and McDermott, D. 1985. *Introduction to Artificial Intelligence.* Reading, Mass.: Addison-Wesley.

Clancey, W. J. 1987. Book Review of Winograd's and Flores' Understanding Computers and Cognition: A New Foundation for Design. *Artificial Intelligence* 31:233–251.

Conrad, M. 1987a. Biomolecular Information Processing. *IEEE Potentials,* October, 12–15.

Conrad, M. 1987b. Molecular Computer Design: A Synthetic Approach to Brain Theory. In *Real Brains, Artificial Minds,* eds. J. L. Casti and A. Karlqvist. New York: North-Holland.

Dreyfus, H. L.; Dreyfus, S. E.; and Athanasiou, T. 1985. *Mind over Machine*. New York: The Free Press.

Feigenbaum, E. A., and McCorduck, P. 1983. *The Fifth Generation*. Reading, Mass.: Addison-Wesley.

Gadamer, H.-G. (trans., ed. D. E. Linge). 1976. *Philosophical Hermeneutics*. Berkeley: University of California Press.

Gardner, H. 1985. *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic.

Genesereth, M. R., and Nilsson, N. J. 1987. *Logical Foundations of Artificial Intelligence*. San Mateo, Calif.: Morgan Kaufmann.

Hampden-Turner, C. 1981. *Maps of the Mind: Charts and Concepts of the Mind and Its Labyrinths*. New York: Collier.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press.

Haugeland, J., ed. 1981. *Mind Design*. Cambridge, Mass.: MIT Press.

Hill, W. C. 1989. The Mind at AI: Horseless Carriage to Clock. *AI Magazine* 10(2): 28–42.

Hopfield, J. J. 1982. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. In Proceedings of the National Academy of Sciences 79, 2554–2558. Washington, D.C.: National Academy of Sciences.

Johnson, K. A. 1979. The Bag Model of Quark Confinement. *Scientific American* 241:112–121.

Keller, E. F. 1985. *Reflections on Gender and Science*. New Haven, Conn.: Yale University Press.

Kuhn, T. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Leaf, M. 1979. *Man, Mind, and Science*. New York: Columbia University Press.

McCarthy, J. 1979. Ascribing Mental Qualities to Machines. In *Philosophical Perspectives in Artificial Intelligence,* ed. M. Ringle, 161–195. New York: Humanities Press.

McCorduck, P. 1985. *The Universal Machine*. New York: McGraw-Hill.

MacCormac, E. R. 1985. *A Cognitive Theory of Metaphor*. Cambridge, Mass.: MIT Press.

MacCormac, E. R. 1983. Scientific Metaphors as Necessary Conceptual Limitations of Science. In *The Limits of Lawfulness*, ed. N. Rescher, 185–203. Pittsburgh: University of Pittsburgh Center for the Philosophy of Science.

Maturana, H. R., and Varela, F. J. 1987. *The Tree of Knowledge: The Biological Roots of Human Understanding.* Boston: New Science Library.

Minsky, M. 1986. *Society of Mind*. New York: Simon and Schuster.

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

Pratt, V. 1987. *Thinking Machines: The Evolution of Artificial Intelligence*. Oxford: Basil Blackwell.

Pribram, K. 1971. *Languages of the Brain: Experimental Paradoxes and Principles in Neuropsychology.* Englewood Cliffs, N.J.: Prentice Hall.

Pylyshyn, Z. W. 1985. *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, Mass.: MIT Press.

Pylyshyn, Z. W. 1980. Computation and Cognition: Issues in the Foundation of Cognitive Science. *The Behavioral and Brain Sciences* 3:111–132.

Quine, W. V. 1979. A Postcript on Metaphor. In *On Metaphor*, ed. S. Sacks, 159–160. Chicago: University of Chicago Press.

Searle, J. 1984. *Minds, Brains, and Science*. Cambridge, Mass.: Harvard University Press.

Shanker, S. G. 1987. The Decline and Fall of the Mechanist Metaphor. In *AI: The Case Against*, ed. R. Born. New York: St. Martin's.

Skinner, B. F. 1957. *Verbal Behavior*. New York: Appleton-Century-Crofts.

Suchman, L. A. 1987. Book Review of Winograd's and Flores' Understanding Computers and Cognition: A New Foundation for Design. *Artificial Intelligence* 31:227–233.

Turkle, S. 1984. *The Second Self: Computers and the Human Spirit*. New York: Simon and Schuster.

Vellino, A.; Stefik, M. J., and Bobrow, D. G.; Suchman, L. A.; Clancey, W. J.; and Winograd, T., and Flores, F. 1987. Four Reviews of Understanding Computers and Cognition: A New Foundation for Design. *Artificial Intelligence* 31(2): 231–261.

Weizenbaum, J. 1976. *Computer Power and Human Reason*. San Francisco: Freeman.

Winograd, T. 1987. Thinking Machines: Can There Be; Are We? Presented at Stanford University Centennial Conference, 23–27 April, Stanford, Calif.

Winograd, T., and Flores, F. 1986. *Understanding Computers and Cognition: A New Foundation for Design*. Norwood, N.J.: Ablex.

Notes

1. Paracelsus possessed a golden head that was capable of both thought and speech. The Pygmalion myth captures an early attempt to create artificial life from inorganic parts (in contrast to Dr. Frankenstein's efforts).

2. The awareness of other examples, historical and contemporary, of technology-based metaphor provides an important, cautionary perspective.

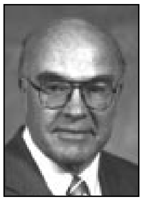
The Hill (1989) article is especially interesting and valuable from this point of view.

3. Not only are these notions central to AI, they are at the root of a number of interesting problems that AI has not adequately addressed. For example, how does one relate internal representations with representations that exist in the world external to the organism, for example, symbol structures on a blackboard or books in a library?

4. One element that formalism contributes to this task is a sense of permanency, but mechanism changes as our concept of machine evolves. Computers of 1990, for example, are machines but machines whose nature is significantly different from Babbage's calculating engine.

5. In an article to appear in the summer 1991 issue of *AI Magazine*, each of these issues will be developed in greater detail and discussed in terms of the criteria for a useful metaphor, what the alternatives might offer theoretical AI, and why they have not (to date) supplanted the computational metaphor.

David West is an assistant professor with a joint appointment in the Graduate School of Technology and the Department of Quantitative Methods and Computer Science at the University of St. Thomas. He received a Ph.D. from the University of Wisconsin at Madison in the area of cognitive anthropology and AI. His general research interests center on nonrepresentational paradigms for AI but also involve applied neural networks and object-oriented system development.



Larry Travis holds a Ph.D. in philosophy from the University of California at Los Angeles. He has been a member of the computer science faculty at the University of Wisconsin at Madison since 1964. He leads seminars and does research in the field of AI, ranging from specific applications (for example, in the areas of formalizing geographic and genetic knowledge) to the general, philosophical foundations of the field.