From Society to Landscape: Alternative Metaphors for Artificial Intelligence

David M. West and Larry E. Travis

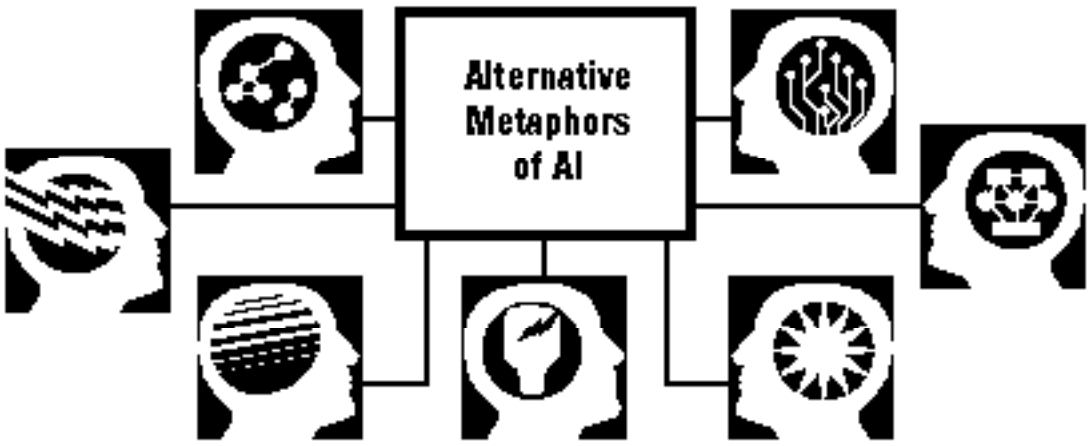
This article picks up the call for a reflective examination of the prevailing computational metaphor of AI *(and philosophical presuppositions behind it) by sketching alternatives that might serve as seeds for discussion—specifically, the seven alternatives introduced in our previous article (see* AI Magazine*, spring 1991). The relative strengths and weaknesses of the alternatives are contrasted with those of the computational metaphor.*

Previous examination of the computational metaphor exposed behavior inconsistent with that expected of metaphors in general. Specifically, despite demonstrated dissimilarity in the referents of brains (minds) and computers, the metaphor persists, not dissolves.

Seeking an explanation of this behavior led to the conclusion that the computational metaphor is not truly a metaphor at all. Instead, it is a kind of shorthand expression, a label, for a set of philosophical presuppositions. These presuppositions generate a particular perspective from which the problem of how to model a mind has been approached —a perspective that is intrinsically formalistic, mechanistic, and dependent on the methodological dualism that results from a reliance on internal representation.

Further discussion showed that many of the hard problems encountered by AI researchers derive from constraints imposed by the perspective of attack and might be avoidable attributes of the mind-modeling problem. Similarly, most of the points attacked by critics of AI are more properly directed toward one or more of the mechanist, formalistic, or representational presuppositions rather than the "grand objective" of AI itself.

It seems desirable, therefore, to determine if alternative perspectives might yield alternative methods of attack. One fruitful starting point in the quest to find alternative (or complementary) perspectives should be an inventory of metaphors that have been or are being



0738-4602/91/$4.00 ©1991 AAAI

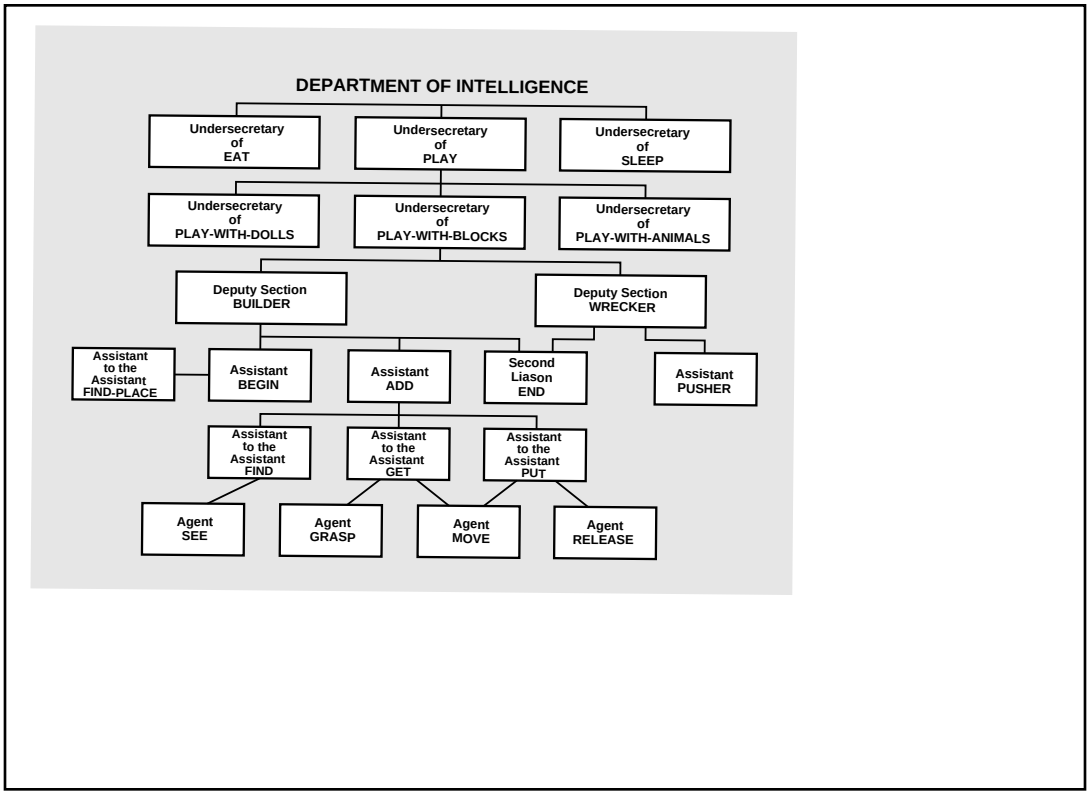


Figure 1. Minsky's Agents in a Bureaucracy.

The familiar form of an organizational chart (with bureaucratic titles) is used to depict a few of the hierarchically arranged agents noted by Minsky.

employed to explain aspects of the mind, the brain, or computation.

If Not a Computer, What?

The number of potential metaphors is large. We limit our discussion to seven metaphors that either are currently being used within some subset of the AI community or that seem to be particularly relevant to the AI problem domain. In each case, we make some attempt to ascertain the extent to which the alternative offers hope of a viable perspective or to point out why the alternative is not likely to supplant the computational metaphor.

Although such an examination might seem to be a straightforward and uncomplicated endeavor, it is not—primarily because of the situation aptly captured and summarized by Johnson (1990):

> The idea of the brain as an information processor—a machine made from matter manipulating blips of energy according to fathomable rules—has come to dominate neuroscience... . So far the main alternative to accepting the brain as a kind of computer has been to embrace holism, the belief that there is some sort of ethereal, irreducible mental stuff that is fundamentally different from any possible algorithm. (p. 45)

Given the polarization noted by Johnson, there are those who will object to any investigation of alternatives as an implicit surrender to AI's critics (the holist camp) at worst or a waste of time at best.

We disagree. We approach alternative metaphors as potential sources of insight into research problems resulting from blind spots caused by excessive dependence on the computational metaphor by AI researchers. Such a reflective examination might, of course, expose weaknesses or even fundamental conceptual errors, leading to modification or abandonment of the computational metaphor and its presuppositions. However, such an outcome is certainly not our a priori intention.

We suspect that most of the prominent critics of AI, for example, J. Searle, do have alternative metaphors in mind. Sometimes

they even deign to tell us what they are. The Dreyfus brothers, for example, state that they see potential in holographic metaphors (Dreyfus, Dreyfus, and Athanasiou 1985) and neural networks (Dreyfus 1986) for overcoming their objections to the computational metaphor.

Because the mind is both mysterious and fascinating, it is to be expected that numerous metaphoric attempts to understand it have been made. Hampden-Turner (1981) catalogs 60 metaphoric maps of the mind, surely not an exhaustive list.[1] Exploring the entire catalog would be a daunting task, one that we can avoid, at least for a while, by focusing on a much smaller set of metaphors that have attracted the attention of some subset of AI researchers or one or more of AI's critics, specifically (1) Minsky's (1987) society of mind, (2) Pribram's (1971) hologram, (3) Bergland's (1985) gland, (4) Conrad's (1987a, 1987b) enzyme substrate, (5) Kupper's (1990) self-organizing system coupled with Maturana and Varela's (1987) autopoietic organism, (6) the architectural (neurode) and process-explanatory (landscape) metaphors often used in connectionist writings, and (7) an evolutionary metaphor common to several of the other metaphors presented.

Society

Minsky proposes a metaphor of the mind as a society. In doing so, he severely underuses the common sense and anthropological notion of society, reducing it to nothing more than a cooperative aggregation of autonomous subintelligent agents. Ignored (or cursorily treated) are all referents usually associated with the common sense of a society in other (especially human) contexts, such as common language, institutions, convention, and culture.[2]

In Minsky's society of mind, aggregates of agents interact in a nonintelligence-presupposing manner, such that aggregates exhibit behavior of greater complexity than any of their parts. Aggregates of aggregates in a hierarchical structure eventually exhibit all the properties normally associated with human minds, including intelligence.[3]

Figure 1 shows a partial hierarchy of "agents in a bureaucracy" (Minsky 1987, pp. 25, 32), the beginning of a structure that would eventually accumulate to a society capable of exhibiting intelligence on par with humans. At each level of the hierarchy, the individual agent is capable only of simple nonintelligent behavior—a simple task, decision, or (de)activation of lower-level agents. As you ascend the hierarchy, the aggregation of simple tasks results in complex performance and, eventually, intelligence.

Many of the fundamentals of this metaphor seem to be (but probably are not) derivative. For example, the relationship between wholes and parts is reminiscent of Koestler's "holons" existing in a "holarchy" (Hampden-Turner 1981, pp. 162–165). Also, Minsky's layered mind (world—"A" brain—"B" brain) and illustrations bring to mind Korzybski (1958). Most telling, however, is the often-heard comment that "Minsky has renamed and repackaged object-oriented programming" because the societal metaphor, as Minsky develops it, does resemble the structure and associated paradigm of an object-oriented language such as Smalltalk.[4]

Minsky proposes a metaphor of the mind as a society.

The society approach might enhance or extend the traditional computational approach by emphasizing decomposition, delegation, and decentralization. The immense complexity of a human mind-brain is decomposed into a set of semiautonomous, highly specialized agents. To each of these agents is delegated some small and simple portion of the mentation task, a direct subperception or a simple subdecision, for example. A hierarchical relationship among base agents is presumed to ensure cooperative aggregation and provide the basis whereby aggregates of agents can exhibit behavior of greater complexity than any individual agent.

Minsky's society does address at least one of the problems encountered by the formalist computational approach—search and concomitant representational complexity—without being untrue to its other formalistic, mechanistic, and dualistic aspects. Each agent deals with a limited search space because it has a limited function. Accordingly, the representations used by each agent can be simple.

The metaphor is also compatible in many ways with some of the other metaphors of interest here, particularly the neural network architectural metaphor (despite Minsky and his colleague Papert often being blamed at one point for a two-decade neural network research drought), Conrad's tactilizing processors, and even the autopoietic organisms of Maturana and Varela.

Not adequately explained, however, is precisely how the complex behavior of a large aggregate accrues from the simple behaviors of each involved entity. Winograd (1987) suggests that Minsky engages in "sleight of hand by changing from 'dumb' agents to 'intelligent' homunculi communicating in natural language at the point of Wrecker versus Builder in a Child." An alternative to sleight of hand would be the idea (almost surely

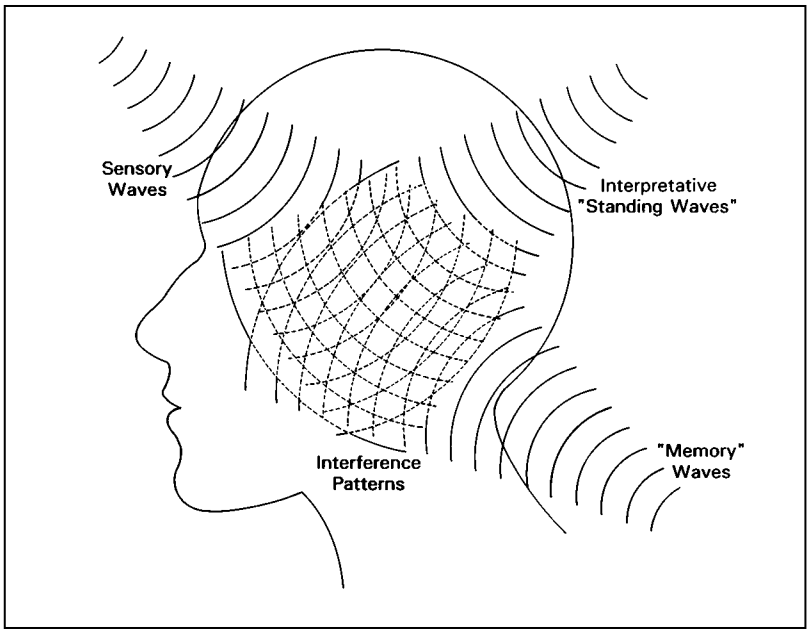


Figure 2. Pribram's Holographic Representations.

According to Pribram, memory is the result of sensory stimuli being converted into a type of brain wave, which interacts with memory waves to create a storable interference pattern. Recall occurs when semipermanent standing waves illuminate the stored interference pattern.

anathema to Minsky) that the complex behavior was an emergent property of the collective ("emergent" as opposed to a property intended and predicted on the basis of well-understood mechanisms).

The society metaphor does offer the possibility of a testable model. In so far as object-oriented programming is consistent with the society metaphor, it is reasonable to expect that the construction of large, complex, object-oriented systems might yield insights that confirm or question its fundamental soundness.

Hologram

Pribram's research began with cats but shifted to monkeys, and most of his experimental findings are based on data from the latter. The key metaphor for Pribram is the *brain wave*, a pattern of electric activation that resembles a wave front as observed in a disturbed liquid. Brain waves are a kind of surface phenomenon, moving to and fro across the brain. This area is in distinct contrast with, for example, neural network research, where the electric behavior of the brain is metaphorically spoken of as pulses traveling along neural circuits.

Sensory input (Pribram concentrated on visual input) is transformed into a brain wave. This wave travels across the brain to an area that interprets its meaning. The interpretation is a product of various kinds of persistent or standing *memory waves*. Multiple waves travel across the brain simultaneously and interfere with each other. An interference of particular relevance to Pribram's investigations is that between a memory wave and a visual sensing wave. The observed waveform phenomena seemed to Pribram suggestive of those characteristic of holograms, hence the holographic metaphor that pervades much of his work.

Figure 2 shows the interaction of sensory, memory, and standing interpretive waves within the brain. The interference patterns generated by interaction of the memory and sensory waves are storable in some sense and become a basis for the generation of subsequent memory waves. The standing interpretive waves interact with the interference pattern to recall memories stored in the interference pattern.

This holographic metaphor has exhibited an appeal for a wide audience. At the more conservative end of a continuum are the Dreyfus (1985) brothers who advance the holographic model as a possible alternative to the computational model. At the other end are a number of metaphysically inclined physicists and new age philosophers who have extended the metaphor to include the notion of a "holographic mind interpreting a holographic universe." (See Wilber [1982], Capra [1975], Bohm [1980], and Comfort [1984].)

Some particular characteristics of the metaphor that make it interesting in the context of AI include (1) an explanation for the apparent distribution of memory and its persistence even when large amounts (to 90 percent) of relevant brain tissue are removed or inhibited from operation, (2) a mechanism for explaining ubiquitously observed associational memory phenomena, (3) a mechanism whereby immense amounts of information can be encoded and stored in a limited volume, and (4) an argument for conceiving of the mind in a nondualistic fashion as an open cybernetic system of organism plus environment.

It is interesting to recall that this particular metaphor was at least as popular at one time as the neural network metaphor is today. For those immersed in the neural net paradigm, it is especially instructive because holograms seem to offer precisely the same advantages that are touted for neural networks.

Given its popularity and what many researchers find to be a highly expressive power, why did holograms not supplant the computational metaphor? One reason is simply technical. As for Babbage, those who advocated the metaphor often found themselves without the necessary technology to implement their ideas, for example, an optical computer. Another technology that was not immediately available was the ability to produce a hologram without using coherent waveforms—a limitation because the brain has no apparent source of coherent waveforms and could not function as a hologram unless holograms could be generated with incoherent or natural waveforms.

A final and somewhat ironic reason for its failure to gain a serious foothold in AI (or neuroscience in general) is the hyperbole associated with the theory.[5] This overselling did not originate with Pribram but with those that adopted his metaphor. The holographic metaphor was combined with quantum potential wave metaphors and Buddhist metaphysical metaphors to generate radical theories of the relationships between consciousness, brains, and the world at large. These theories were more enthusiastic than rigorous and generated a backlash against the central metaphor involved.

A component of such theories, however, deserves continuing attention: One valuable test of a metaphor is its ability to account for a broad range of phenomena as opposed to a narrow subset. This ability will be a continuing dynamic when evaluating formal computational metaphors and theories that are weak when confronted with pattern-recognition problems and alternative metaphors that are weak when confronted with logic problems. The human mind exhibits wide-ranging performance characteristics, and all of them need to be accounted for.

Gland

Bergland (1985) looks at the brain and sees a gland: "It produces hormones, it has hormone receptors, it is bathed in hormones, hormones run up and down the fibres of individual nerves, and every activity that the brain is engaged in involves hormones" (Preface). He then concludes that to the extent that thought is the result of brain activity, it is the product of brain chemistry.

Bergland sees his approach as a necessary antidote to an overdose of "brain as electrical circuit" metaphors that originated with L. Galvani's demonstrations and T. Schwann's microscopic observations of neural wiring and were reinforced in contemporary times by the electronic digital computer.

Bergland see his approach as a necessary antidote to an overdose of "brain as electrical circuit" metaphors…

Synapses ("clasping paws") and the interneuron networks that they potentially connect are key to Bergland's arguments. Despite the fact that they were discovered, by Ramon y Cajal in 1900, and shown to be points at which a circuit was physically broken, synapses were effectively dismissed or only grudgingly acknowledged. This state of affairs, according to Bergland, continues to this day.

What is important for Bergland is the manner in which the physical barrier created by the synapse is either maintained or overcome. Simple variation in electric potential traveling the dendritic and axonic "wires" is not sufficient explanation, he argues, even though, in some cases, it does result in a "spark" jumping across the synaptic barrier. The major explanation is the presence of specific brain hormones. Some hormones enhance transmission, and others inhibit or block transmission. "The new hormone based paradigm for the mind acknowledges that electricity does flow from nerve to nerve and can be measured on the surface of the brain or on the membranes of individual nerves. But these superficial signals are little more than the dry echoes of deeper molecular events going on within the cell" (Bergland 1985, p. 108).

The major function of the brain, therefore, is to create, transport, and use a complex hormonal soup that, in turn, determines in large part what kinds of electric circuits are established and, thereby, what kinds of mental activities can and do take place.

Synaptic closure relies on the existence of receptor sites in the synapse and the presence or absence of the hormone that can bind to the available receptor sites. This arrangement is analogous to a lock-and-key mechanism—a secondary metaphor and one that will be echoed in the following discussion of Conrad's tactilizing processors.

Figures 3 and 4 illustrate Bergland's conception of the joint roles of chemistry and circuitry in determining brain function and,

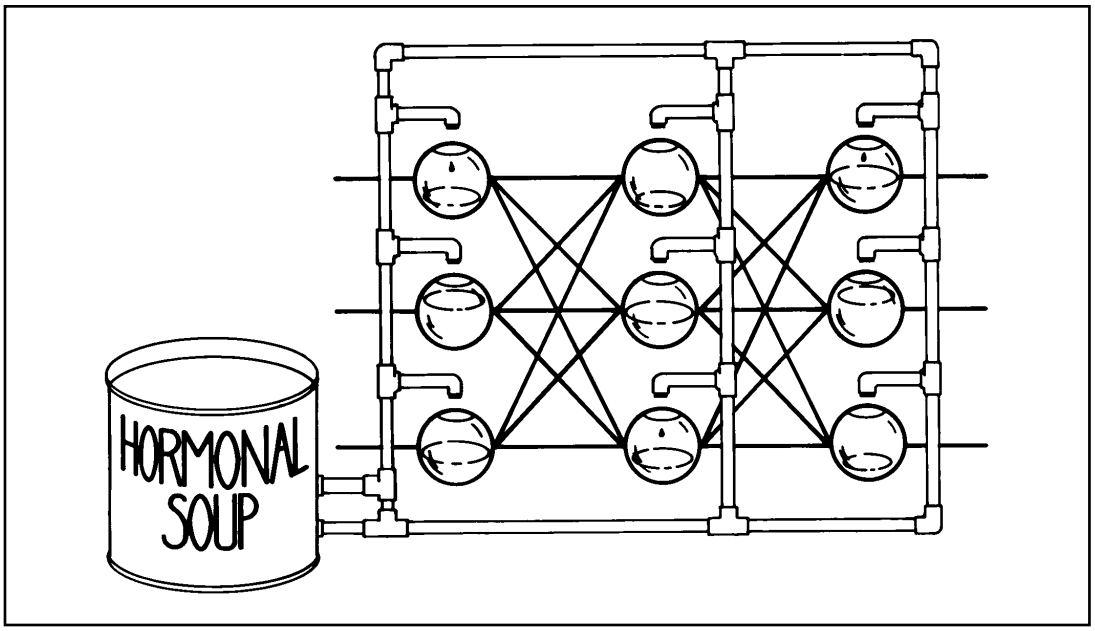


Figure 3. Bergland's Whimsical Version.

A common abstract representation (cross-connected circles) of a neural network is overlaid with a set of pipes used to transport "hormonal soup" to synapse sites, where it plays a critical role in determining synaptic firing. This picture is a visual metaphor of Bergland's idea that "the hollow brain" is a gland that secretes and transports a variety of hormones and chemicals.

ultimately, thought. Figure 3 is a fanciful depiction of the role of "brain as gland " to transport appropriate hormones and chemicals to synaptic sites in the brain. Figure 4 illustrates a synapse as a "clasping paw" and shows how any electric circuit is broken at the synapse. The break in the circuit allows the synapse to function as a switch. The switch can be closed when enough current builds on one side that it arcs to the other side or when a hormone binds to a receptor site and reduces the resistance to allow even weak currents to flow across the gap. Other hormones can bind to receptor sites that increase the resistance and effectively prevent the switch from closing as long as the hormone is present. Noting that hormones, even those that as far as we know are solely used in the brain, are produced in several sites in the body, Bergland goes on to argue the possibility that the organism as a whole is responsible in some measure for thought.

The gland metaphor presents various challenges to AI theorists, especially connectionists. The empirical evidence for hormonal activity in the brain on which Bergland focuses cannot be dismissed or ignored, and it is not easy to see how to accommodate it in other models of brain operation. In a standard computational model, it adds one more level to already intractable levels of complexity. Connectionist models have such a simple architecture that there is no obvious place to insert hormonal influences.

Tactilizing Processor

Conrad draws his inspiration from the ability of an enzyme to combine with a substrate on the basis of the physical congruency of their respective shapes (topography). This is a generalized version of the lock-and-key mechanism as the hormone-receptor matching discussed by Bergland. When the topographic shape of an enzyme (hormone) matches that of a substrate (receptor), a simple recognize-by-touch mechanism (like two pieces of a puzzle fitting together) allows a simple decision, binary state change, or process to take place, hence the label "tactilizing processor."

Like Minsky's agents, each tactilizing processor is a special-purpose entity capable only of simple, nonintelligent operations. Matching patterns, distinguishing among a variety of potential input, and making subtle distinctions of signal strength are among the tasks that such processors should be particularly adept at, especially in contrast with an algorithmic approach to such tasks.

Although each tactilizing processor func-

tions as a special-purpose pattern-recognition mechanism—capable of recognizing one specific but arbitrarily complex pattern—collections of such processors enable the construction of sophisticated mechanisms. These devices could then either function independently (Conrad [1987a] outlines a 12-step "progressively futuristic design process for a molecular-computing device" [p. 14]) or as input devices linked to conventional von Neumann machines. Extensions to Conrad's basic notion and details on the physiology of both natural and artificial biomolecular computers can be found in Hameroff (1987). (For a popular account of this area, see Drexler [1986].)

Figure 5 uses simple geometric shapes to show how an input might stimulate the generation of a tactilizing processor capable of finding and combining with other processors to form complex entities that could, in turn, be recognized by receptors linked to a variety of output. It should be noted that this figure shows only the simplest level of Conrad's conception of how a sophisticated, arbitrarily powerful computing device might be constructed.

Although Conrad and Bergland share the base key-lock metaphor of a tactilizing processor, they take the metaphor in widely divergent directions. Bergland stresses the role of tactilizing processors in the brain as a foundation for thought, and Conrad (and Hameroff) focuses on the potential for constructing biological computational entities.

Conrad is less interested in AI than in *artificial life,* an embryonic discipline that subsumes AI, robotics, bioengineering, recombinant DNA, and *nanotechnology* (the engineering of machines and computing devices at nanometer scale). Conrad and others involved in the artificial life effort such as Moravec (1988) and Rucker (1989), expect to generate AI or, in a manner reminiscent of the golden spike connecting the transcontinental railroads, connect with a partial AI developed by current approaches. Their approach might be thought of as a biologically grounded attempt to construct agents of the sort postulated by Minsky, which would then be subject to directed evolution and adaptation until they evolved and aggregated to form an intelligent entity (perhaps a Minskian society of mind).

As a metaphor, the tactilizing processor is intriguing, but as an approach, two limitations immediately come to mind. The first is simply the same kind of technology limitations that continue to impede the holographic metaphor. Biomolecular computing devices exist only at the periphery, where science and science fiction begin to overlap.

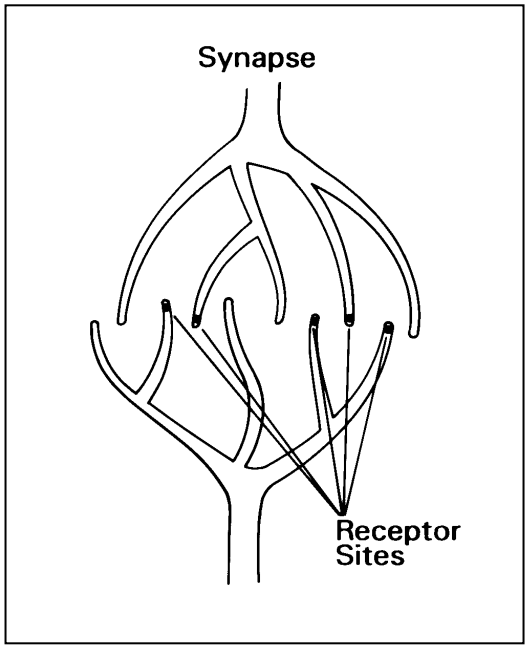


Figure 4. Bergland's Synapse and Receptors.
Brain circuits are broken at the synapse. Bergland focuses on how brain chemistry, using the binding of hormones at receptor sites on the synapse, influences or determines whether the circuit can be closed.

More important is that the same problem can be noted for Minsky—how to get from nonintelligent agents to intelligent aggregates without using a sleight of hand. Including the Bergland hormone-receptor aspect of tactilizing processors does provide a possible bridge but one that does not eliminate the possibility that intelligence is an emergent property rather than a designed property of the aggregate of tactilizing processors. Given the strong biological and evolutionary perspective adopted by proponents of this metaphor (discussed further later), the emergence of intelligence as a result of evolutionary adaptation is probably an acceptable notion for them.

A related obstacle is the lack of any clear depiction of how a complex computational architecture might be assembled. This problem is analogous to demonstrating how a complex von Neumann computer can be constructed using only basic computational primitives such as Nand or Nor. Simply showing that a Turing machine can be constructed from such primitives and, therefore, that in principle, it is possible to construct the tactilizing processor equivalent of a Cray supercomputer does not address the issue of feasibility; therefore, for many, it fails to be convincing.

Like Minsky and Conrad, Maturana and Varela… want to construct complex wholes from simple parts.

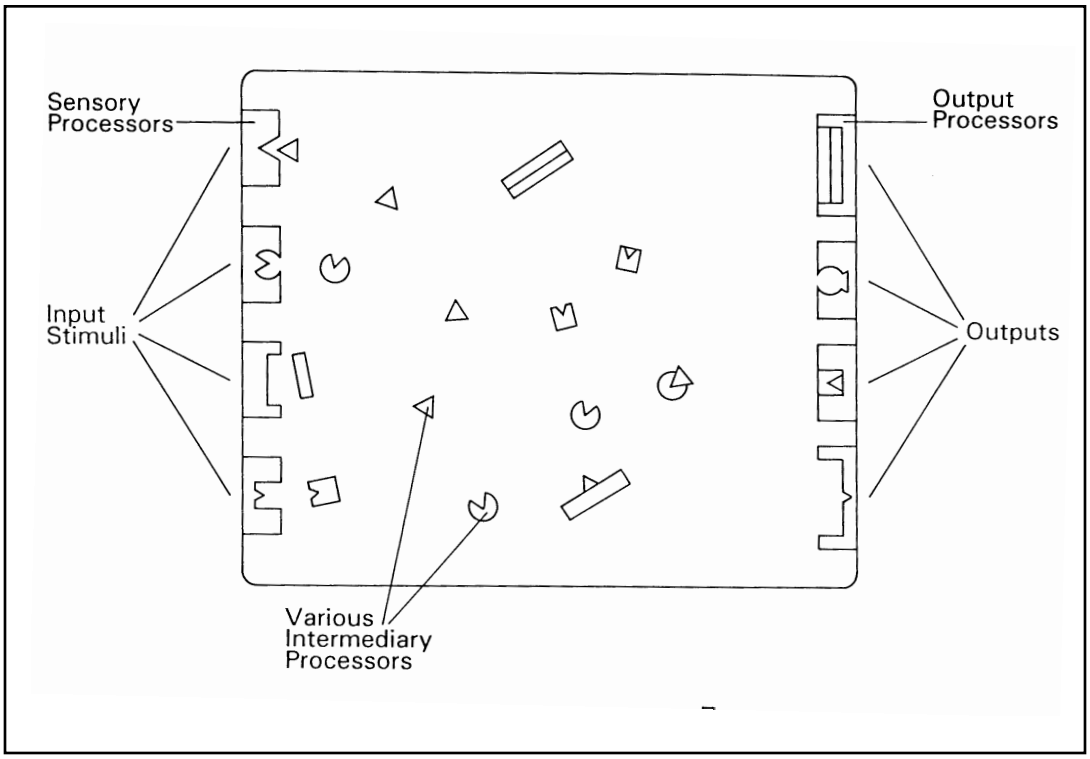


Figure 5. Conrad's Tactilizing Processors.

Conrad's notion of a molecular computing device uses tactilizing processors (illustrated with simple geometric shapes). Input-sensitive processors produce free-floating processors that can combine with others (based on their physical configurations) and subsequently be sensed by output processors that generate output signals. At each stage of the process, all interactions are based on the congruency of physical shapes.

Autopoiesis

Like Minsky and Conrad, Maturana and Varela (1987) want to construct complex wholes from simple parts. In their case, the basic part is a self-organizing, or *autopoietic*, unity, consisting of a set of internal processes or dynamics and a membrane that separates it from the environment as a whole.[6] The most common example is a biological cell, and it is with the cell as a conceptual base that Maturana and Varela begin to construct their more complex, metacellular wholes.

Although they do offer minimal explanation of how biological cells come into existence requiring nothing more than the operation of basic laws of physics and chemistry (and certainly without the need for conscious design), a more complete explanation is presented by Kuppers (1990). In both instances, the argument presented is that the macromolecules and eventually the cells on which all life is built derive from the iterative application of simple rules—the same rules basic to all chemistry and physics.

> The molecular theory of evolution has been a result of rapid progress in biology and physics in the last two decades. It is based on the one hand upon the discovery in molecular biology that all basic phenomena of life such as metabolism and heredity can be traced back to regular interactions between biological macromolecules, and thereby to the laws of physics and chemistry, and on the other hand upon the discovery in physics of open systems that, far from equilibrium, can spontaneously assemble states of material order (so called dissipative structures) that are also characteristic of living systems. (Kuppers 1990, p. 168)

Once established, autopoietic unities are subject to ongoing interactions with their environment, many of which result in internal structural changes:

> This ongoing structural change occurs in the unity from moment to moment, either as a change triggered by interactions coming from the environment in which it exists or as a result of its internal dynamics. As regards its continuous interactions with the environ-

ment, the cell unity classifies them and sees them in accordance with its structure at every instance. That structure, in turn, continuously changes because of its internal dynamics. (Maturana and Varela 1987, p. 74)

Larger, metacellular organisms arise when two or more autopoietic unities (starting with cells) share interactions that recur more consistently than interactions with the environment at large. Maturana and Varela label this process *structural coupling*, "a history of recurrent interactions leading to the structural congruence between two or more systems" (p. 75).

Figure 6 reproduces the highly abstract symbology used by Maturana and Varela to illustrate the autopoietic organism (sphere and defining circular arrow as an implicit membrane), a nervous system that arises from the self-organizing processes within the organism's membrane (also a circular arrow), and structural coupling (pairs of directional arrows) through interactions with the environment in the large (wavy line) and other autopoeitic organisms. The dashed unidirectional lines between the environment and the structural coupling between the two autopoeitic organisms—not present in Maturana and Varela's symbology—depict the common reaction of such a coupled structure with individually recognized environmental stimulus (so called third-order structural coupling).

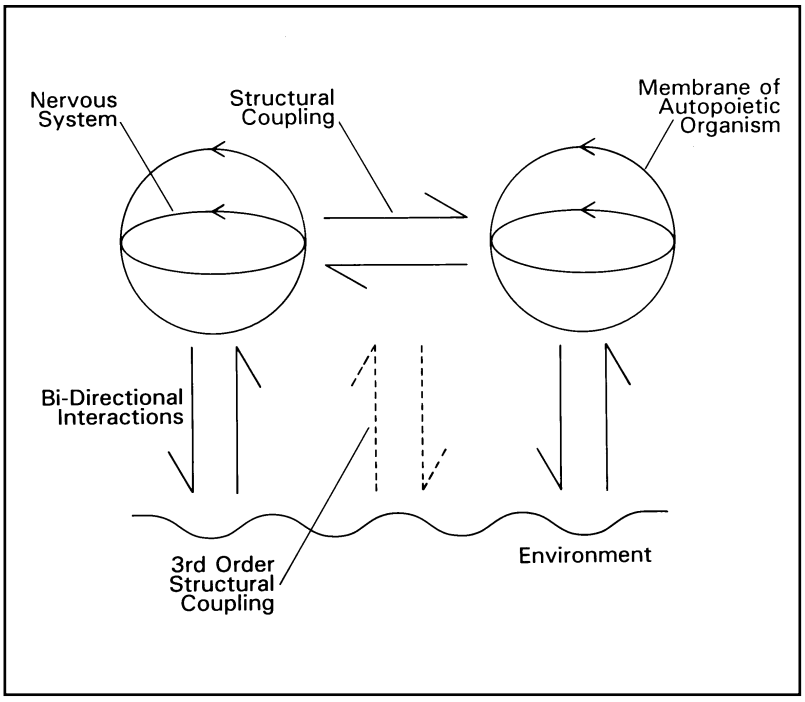


Figure 6. Maturana and Varela's Symbolic Depiction.

Maturana and Varela use a highly abstract symbology to depict their ideas of autopoeitic organisms and structural coupling. Circles with arrows depict the dynamic self-organization processes that give rise to a membrane separating the organism from its environment and, eventually, to a nervous system within the organism. Organisms are structurally coupled to the environment and each other in that persistent stimuli from outside the membrane yield persistent states within the membrane. Complex organisms arise when structural coupling between organisms cause them to act in concert to shared stimuli from the general environment (third-order coupling).

Kuppers and Maturana and Varela lay out a continuous process, beginning with inert elements; moving through the generation of macromolecules, organic cells, and cellular organisms; and culminating in the development of consciousness, language, and intelligence. At every stage of the process, nothing is used that is not consistent with the laws of physics and chemistry coupled with Darwinian evolution. This metaphor and approach is consistent with that of Conrad and Hameroff. (See Hameroff [1987] for a discussion of how several of the metaphors discussed here are subsumed in the biomolecular computing model.)

Although there is some surface similarity, two aspects of this perspective keep it from being essentially equivalent to Minsky's and also generate theses that make it antagonistic to formalistic AI in general: (1) the nature of dissipative systems and (2) the intimate integration of an autopoietic system and its environment.

Dissipative systems are self-organizing and nondeterministic, and they operate far from equilibrium conditions.[7] They exhibit complex behavior that is highly sensitive to initial conditions in that widely divergent outcomes result from minute changes in the initial value of a single variable. They have recently received a lot of attention under the label "chaotic systems." They are consistent with the notions at the heart of cellular automata theory. Such automata are similarly capable of generating complex behavior and organization from the iterative application of simple rules.

Dissipative systems provide an explanatory mechanism that directly addresses the problems, including the quantum questions (see Pool [1989] for a discussion of quantum chaos) raised by Penrose (1989). They do so, however, in a manner that is unlikely to please advocates of traditional, formalistic AI.

> On the basis of natural law, therefore, it is possible to predict that biological structures exist, but not what biological structures exist. The structures that are found reflect the historical uniqueness of living systems, and the details of their origin are in principle inaccessible for

…neural networks offer buildable models, active research issues, and even pragmatic commercial applications.

> description in terms of natural law. This means: the origin of biological information can indeed be explained as a general phenomenon, but the concrete content of biological information can not be deduced from the laws of physics and chemistry. (Kuppers 1990, p. 172)

There is little doubt that Minsky, for example, intends his society of mind to be a deterministic system where, in principle, the operation and behavior of the whole could be determined (predicted) simply from knowledge of the state of its parts. Dissipative systems are not consistent with this kind of deterministic motivation.

Kuppers echoes Maturana and Varela's conclusions when he notes that the outcome of the process he describes is both particularistic and historical (that is, development can be traced backward in time but not predicted forward in time):

> There are basically two reasons for the indeterminacy… . One is that the "direction" taken by the optimization process depends on genetic variation and this in turn is the result of fundamentally indeterminate genetic mutations. The other is that the structure… depends on the individuals taking part in the evolutionary process. (Kuppers 1990, p. 177)

Maturana and Varela generalize this conclusion and use the metaphor of M. C. Escher's self-referencing artwork to illustrate the intimate connection of the autopoietic organism and its environment. They conclude, in fact, that it is meaningless to try to speak of one apart from the other, and therefore, any scheme for modeling a mind that depends on dualism and representation (such as the computational metaphor) is doomed to failure.

Advocates of this position overstate their case against strong AI. They too assume that the only means to such an objective is the one implicit in the computational metaphor—to build (engineer) AI like one builds a machine. None of their premises or conclusions precludes the possibility of obtaining AI by other than engineering means, for example, by growing one. The notion of growing AI is not alien to our community. Moravec, for example, uses a metaphor of growing.

Growing does not preclude engineering; it merely moves it to a different level. One cannot build a tree, but it might be possible to engineer a seed (using inorganic chemicals) from which the tree can subsequently be grown. Conrad, Hameroff, and others seem to be following (without explicitly stating so) this developmental model. Engineering might come into play again at the macrolevel in a manner analogous to the way in which one plans and lays out a garden.

Even if one gets past the misdirection of criticism, adoption of this kind of metaphoric perspective is likely to be resisted. It would, after all, require a reorientation away from physics and engineering to the natural sciences, such as biology, anthropology, and neurophysiology. Despite the daunting nature of the required perspective shift, there is some movement in this direction, again most notably among those enamored with the neural network metaphor.

Neurodes and Landscapes

One of the most profound differences in referents between computers and brains is architectural. Computers have central processing units, random-access memory, disks, cathode ray tubes, and so on, and brains have neurons, synapses, dendrites, and axons (in addition, of course, to a lot of hormones). Using brain architecture as a metaphor leads one to constructing devices that consist of neurodes, interconnects, and synapses. (Terminology here is used by Caudill and Butler [1990].) The recent resurgence of interest in devices of this sort is well known.

Neural networks do contravene some of the philosophical tenets behind the computational metaphor, the most prominent example being that they do not operate through the formal manipulation of symbolic tokens. So different is the functioning of a neural network that an additional complementary metaphor, that of a landscape, has been introduced to aid in an explanation: "Neural nets have contours like the hills and valleys in a countryside" (Allman 1986, p. 24).

Input to a network are likened to rain falling on a landscape. Water flows downhill until it reaches a point where the terrain per-

mits no further descent. Similarly, the energy representing the input seeks a point of stability. These points, analogous in many cases to seas or lakes in a landscape, represent output from the network. It is the landscape surface of the network that channels input to correct output in a manner analogous to rain falling in Montana being channeled into the gulf of Mexico.

Although it is possible to preconfigure a network landscape (also its topology, technically referring to the arrangement of neurodes and interconnects), self-configuration as a result of training or learning is more important. As a result of the iterative application of an input plus some sort of feedback, the network adjusts internal parameters (connection weights) and thereby alters its landscape to the point that a presented input is appropriately channeled to the desired output.

Figure 7 illustrates connectionist architectural and performance metaphors. The underlying structure of neurodes, interconnects, feedback lines, and connection weights provides a topological architecture where memory and other capabilities of the brain reside, that is, distributed among the nodes, connections, and connection weights. Superimposed on this architectural metaphor is the processing metaphor of the landscape whose peaks, valleys, and channels arise from, and correspond to, energy contours generated by variance in connection weights distributed across the underlying network.

Unlike the other metaphors presented so far, neural networks offer buildable models, active research issues, and even pragmatic commercial applications. They have also captured the imagination of researchers in a wide-ranging, interdisciplinary community just as the computational metaphor did 30 years ago.

Despite this popularity and the success of the research agenda based on this metaphor, it has not replaced the computational metaphor or the computational perspective. At least three factors account for this situation: First (as presented so far), it is only an architectural metaphor. Accordingly, neural networks firmly remain within the class of devices known as Turing machines, and in fact, most neural network research is performed (simulated) on conventional von Neumann computers.

Second, the capabilities of such devices are complementary to, and not replacements for, the capabilities of conventional computers. The situation is analogous to the difference between using a mathematics coprocessor and software simulation of these same mathematical functions. It is true that the capabilities of a neural network are more exciting (and much more difficult) than those in the coprocessor analogy, but the principle is the same. In this regard, the neural network architectural metaphor actually strengthens the computational metaphor.

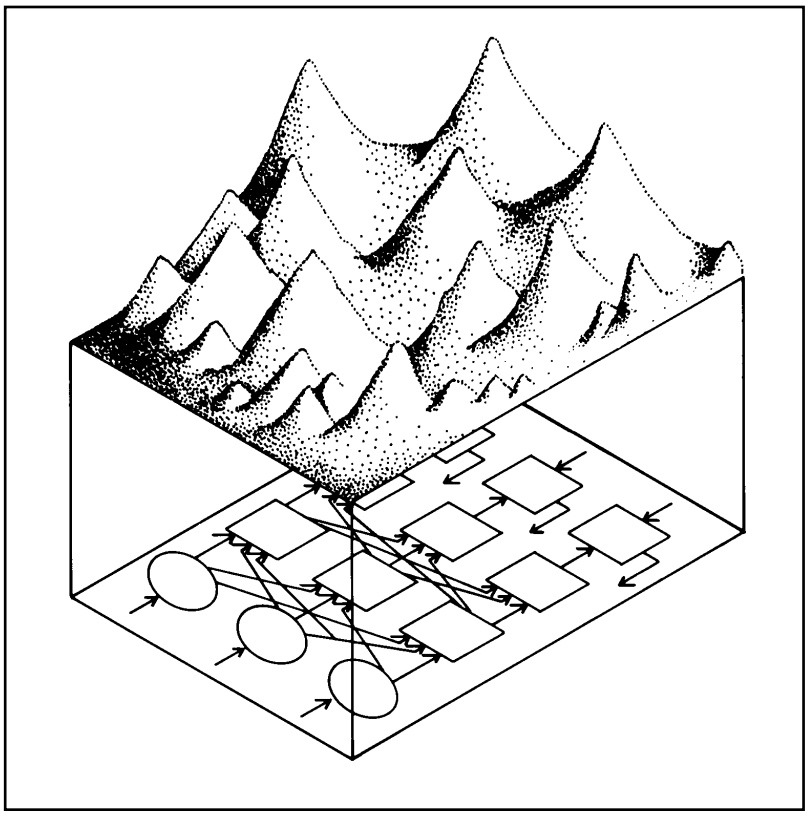


Figure 7. "Landscape" Generated by Neural Net.

Connection weights stored in the trained neural net (abstractly represented at the base of the figure) can be thought of in terms of energy contours that channel the flow of subsequently received input to appropriate output. These contours can be thought of in terms of a landscape (depicted at the top of the figure) that are capable of channeling rainfall to the ocean.

Third, critics of the computational approach to AI do not find counters in connectionism to all the concerns they direct against the computational metaphor. Two examples illustrate this point: (1) Neural networks still embody methodological dualism in that they require representation, albeit distributed, of the external realm inside the machine (in the form of connection weights). (2) Artificial neural networks are minute in comparison to those that occur naturally. The scaling factor encountered in traditional AI is potentially minor compared to the analogous problem with neural networks. Therefore, it is not a completely satisfactory alternative for

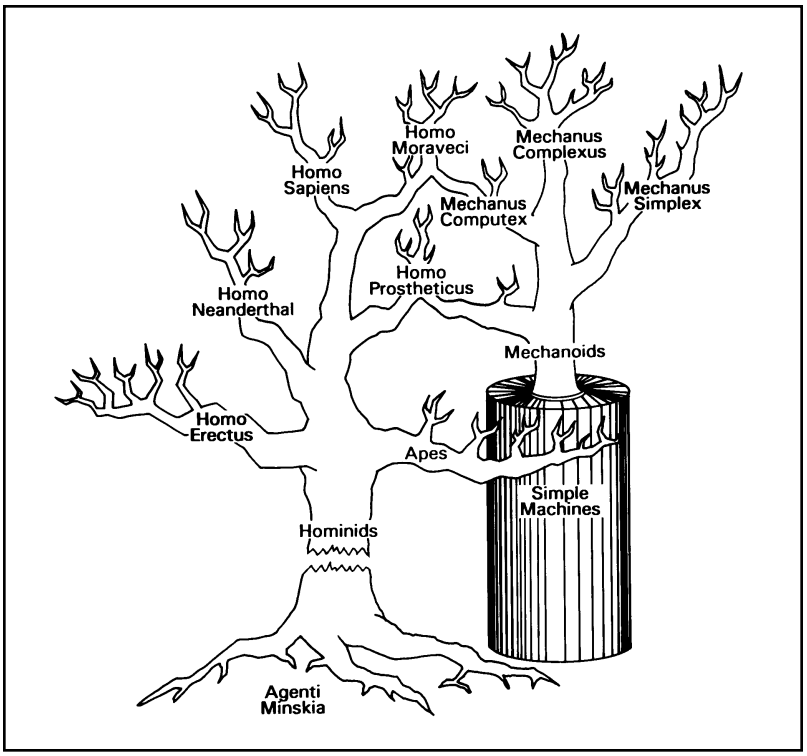


Figure 8. Evolution.

Two, incomplete, evolutionary trees of the form familiar to readers of basic biology texts show how complex forms arise from simpler forms over time, including the (current) fanciful possibility of a merged evolution between humans and machines.

many AI critics and, hence, one that only a few, such as the Dreyfuses, will even cautiously endorse.

Evolution

Growing, maturing, learning, and adapting are secondary metaphors common to a number of the positions presented so far. Their use by Kuppers, Maturana and Varela, Conrad, Bergland, and, to a lesser extent, some connectionists indicates either an implicit or explicit adoption of still another metaphor, that of biological, Darwinian evolution.

All too frequently the four terms (growing, maturing, learning, and adapting) are used in a manner that obscures some important distinctions among them. Growth and maturation, for example, refer to modifications in an organism that realize an implicit, species-shared, potential, but learning and adaptation refer to modifications reflective of individual experience. All four concepts apply to individual organisms and, therefore, are distinct from evolution, which refers to modifications of species, not individuals.

In figure 8, a central (and perhaps most commonly recognized) theme of evolution—the progression from simple to increasingly complex organisms—is illustrated. It is possible to influence this progression (demonstrated by practitioners of animal husbandry), which gives rise to the possibility of *intelligence breeders* creating and manipulating an analogous evolutionary progression from simple nonintelligences to full AI or, perhaps, hybridizing natural and artificial intelligences.

Use of these metaphors is not limited to those holding an alternative perspective on how to achieve AI. Learning, for example, is an important metaphor in conventional AI and is an explicit recognition that intelligent computers might have to obtain their programming in a manner analogous to the way that a human obtains an education.[8]

Evolution is also an important metaphor for conventional, as well as alternative, approaches to AI. The fact that human intelligence is the product of evolution is almost universally accepted. Therefore, it is reasonable to expect that an understanding of this process is a necessary prerequisite to understanding how the human brain-mind operates.

> In the case of the human brain . . . we will have to understand how brain cells work... . Then we'll have to understand how the cells of each type interact with the other types of cells to which they connect... . Then, finally comes the hardest part: we'll also have to understand how our billions of brain cells are organized into societies... . The more we can find out about how our brains evolved from those of simpler animals, the easier that task will be. (Minsky 1987, p. 25)

For Kuppers and others, the importance of the evolutionary metaphor extends beyond its use as an explanation for aspects of the human mind. For these researchers, it is also a statement of prerequisites for realizing an artificial mind. Whether the basic mechanisms are Conrad's tactilizing processors, Maturana and Varela's autopoietic cells, or one of the other alternatives discussed, these mechanisms must organize and develop through a process that includes growth, adaptation or learning, and evolution across generations. Just what details of the evolution metaphor might eventually prove useful—for example, the mutual exclusion of traits learned within the lifetimes of individuals and traits inherited across generations—remains, of course, an open, empirical, and interesting question.

Adoption of the evolutionary metaphor, even in its strongest form, does not eliminate

the possibility of human intercession in the evolutionary process. Evolution might artificially be accelerated, the basic mechanisms undergoing evolution might artificially be engineered, nondirected evolution might be replaced by directed breeding, or other possibilities.

Although evolution is accepted as an integral part of the explanation for the generation of human intelligence, it has been resisted or effectively ignored in most conventional AI research. Where it is used, it is generally misused, for example, referring to a successor computer model as having "evolved" from predecessor models. Reasons for the reluctance to adopt the evolution metaphor for AI can be deduced simply by looking at the change in the metaphoric description of the AI scientist's task. Instead of determining how to fabricate or engineer artificial intelligences, the scientist will have to determine how to breed or nurture such artifacts.

Summary and Conclusion

Our intent to this point has been to briefly sketch some promising alternative metaphors and raise at least some of the factors that prevent their widespread substitution for the computational metaphor.

Individually, each metaphor satisfies a number of criteria that would argue in favor of its use: First is *suggestiveness,* the generation of a large set of referents on both sides of the metaphoric relationship, referents that can lead to concept operationalizing and can be used to confirm or dispute the validity of the original metaphor. Second is *concreteness,* the generation of practical avenues of research or testable hypotheses. Third is *consistency,* both internally and, perhaps more importantly, in relation to what is known or believed about the mind and brain in other disciplines. Just as the computational metaphor is congruent with a major philosophical tradition and theories in other related disciplines, so, too, are the metaphors previously presented—different traditions and different theories perhaps, but the links are there.

Collectively, the metaphors represent a body of research and a set of perspectives of sufficient significance that it compels at least a reevaluation and a reconfirmation of AI's basic metaphor and associated philosophical perspective; however, with the exception of isolated examples, this reworking has not happened. At least two explanations might account for this situation.

First, most of the alternatives surveyed here were articulated as frequently polemic criticisms of AI and AI research. In this context, it is not surprising that those criticized have spent more time in defense than in evaluation.

Second and more important, despite significant overlap among the alternative metaphors and a common source of inspiration (the organism), they lack the kind of unification enjoyed by the computational metaphor. By *unification*, we mean a commonality of perspective that can be brought to bear on the myriad aspects of the large and complex problem of creating AI. The computational metaphor (which is itself a family of metaphors) derives its unification from its alliance with the formalistic philosophic tradition.

The Hindu fable of the blind men and the elephant illustrates the problem confronted by advocates of an alternative metaphor. Each metaphor presented here describes a particular view of mind. This particularism creates a situation where an alternative metaphor contrasts a characterization of an aspect of the brain or mind with the computational characterization of this same aspect plus the rest of the computational "elephant." A part is compared to a whole, and the part loses. The collection of parts is diminished to simply being reminders that alternative views exist.

Therefore, it seems that a serious challenge to the computational metaphor and perspective will depend on the ability to articulate a unified alternative. In turn, this ability seems to require an alliance with an alternative philosophic tradition. The loosely labeled hermeneutic tradition in which some of AI's most implacable critics (such as Hubert Dreyfus, Maturana, and the recently converted Winograd) have roots is an obvious candidate.

If successful, such an effort would obviously intensify the theoretical debate between the computationalists and the holists. Perhaps the debate could attain the same kind of epic status that was accorded the "big bang versus steady state" debate in cosmology or the controversy between quantum and relativistic theories in physics.

Debate between intractable adherents of polarized positions is not, however, a particularly desirable outcome. Insight-generating discussion and stimulation of investigation are the real objectives, and they might be better served if the attempt to unify alternative metaphors and the hermeneutic philosophic tradition were extended to include the prevailing computational position as well. Just as physicists seek to find a grand unified theory that will reconcile relativistic and quantum theories, AI theorists might find it valuable to pursue a comparable goal.

Whether a grand unified theory for AI is possible (or even ultimately desirable), the attempt to characterize one would in itself stimulate discussion, reflection, and experimentation. More importantly, it would open potential avenues of cooperation among those who currently find themselves at philosophical odds. We hope that someone will attempt to formulate and present such a bridging metaphor.

Having completed our survey of various metaphors, we recognize that many might say that all of them suffer in comparison to the computational metaphor because they cannot as easily be realized. The computational metaphor endures and gains much of its hold on us, it might be argued, because it can be articulated and put to the test through the building of computer-based models—something relatively easy to do with modern programming and circuit-construction tools.

However, to argue thus is to fail to understand the true universality of modern computing machines. They are capable of being designed or programmed to simulate any of the metaphors we discussed. Thus, they provide a medium for articulating and testing any metaphor of the mind whatsoever, and they provide no special advantage to the computational metaphor. There are indeed problems of computational tractability, but as is well known from painful experience, the computational metaphor would appear just as prone to this problem as any of the others.

This point is subtle but can be summed up with a dictum: Don't confuse the medium with the metaphor.

Bibliography

Allman, W. F. 1986. Mindworks. *Science86* 7(4): 22–31.

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

Bohm, D. 1980. *Wholeness and the Implicate Order*. London: Routledge Kegan Paul.

Capra, F. 1975. *The Tao of Physics*. New York: Random House.

Caudill, M. and Butler, C. 1990. *Naturally Intelligent Systems*. Cambridge, Mass.: MIT Press.

Comfort, A. 1984. *Reality & Empathy: Physics, Mind, and Science in the 21st Century*. Albany, N.Y.: State University of New York Press.

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, 197–226. New York: North-Holland.

Dreyfus, H. L., and Dreyfus, S. E. 1986. Why Computers May Never Think Like People. *Technology Review* 89(1): 42–61.

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

Drexler, E. 1986. *Engines of Creation*. New York: Harper and Row.

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.

Hameroff, S. R. 1987. *Ultimate Computing: Biomolecular Consciousness and NanoTechnology.* New York: North-Holland.

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): 28–41.

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, G. 1990. New Mind, No Clothes. *The Sciences,* July-August, 44–49.

Korzybski, A. 1958. *Science and Sanity.* Lakeville, Conn.: The Institute of General Semantics.

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

Kuppers, B.-O. 1990. *Information and the Origin of Life*. Cambridge, Mass.: MIT Press.

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

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

Minsky, M. 1987. *Society of Mind*. New York: Simon and Schuster.

Moravec, H. 1988. *Mind's Children: The Future of Robotic and Human Intelligence*. Cambridge, Mass.: Harvard University Press.

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

Pool, R. 1989. Quantum Chaos: Enigma Wrapped in a Mystery. *Science* 2:893–895.

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 Cogni-

tion: Issues in the Foundation of Cognitive Science. T*he 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.

Rucker, R. 1989. Why Cellular Automata? Rudy Rucker's Answer. In *Rudy Rucker's Cellular Automata Laboratory User Manual*, 15–20. Sausalito, Calif.: Autodesk.

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, 72–131. 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* 34:227–233.

Wilber, K., ed. 1982. *The Holographic Paradigm and Other Paradoxes*. Boulder, Colo.: Shambala.

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. Some of the more intriguing metaphors noted in Hampden-Turner include a wide range of organismic metaphors dating from Greek times, Freud's steam engines and other kinds of hydraulic mechanisms, Bertalanffy's general systems, cybernetics, and the binary oppositions of Levy-Strauss.

2. Minsky's use of the concept of society reflects, at best, its application to the behavior of social insects such as termites, ants, and bees. His society of mind reminds one of the popular science fiction theme of a *hive mind,* whereby an intelligence or superintelligence emerges from the social interaction of subintelligent, insectlike components of some kind.

3. Minsky actually states that his agents are organized in a heterarchical, rather than hierarchical, manner. His illustrations and frequent use of the terms hierarchy and bureaucracy, however, seem to point to hierarchical relationships. Heterarchy, a concept from anthropology and the science of human organizations, has many technical aspects that Minsky does not appear to carry over to his target domain.

4. As one of the originators of object-oriented programming and knowledge representation with his seminal concept of computational frames, Minsky certainly has every right to poeticize his earlier work as freely as he wants.

5. Why almost all metaphors of mind, including AI's defining computational metaphor, are so vulnerable to hype is a subject worthy of a separate essay in the sociology of science.

6. Despite the fact that autopoiesis is found in relatively few dictionaries, it is not a coined word. Its Greek roots indicate that Maturana and Varela's ideas have been around for some 2000 years.

7. There are two types of fundamental indeterminancy: epistemological and metaphysical. The two tend to be used interchangeably in discussions of dissipative systems. The latter is usually based on quantum indeterminancy and is used as a fallback position when confronting philosophical determinists who argue that in principle, nothing is indeterminate. Kuppers and Maturana and Varela follow the common tendency to intermix and confuse the two types with abandon.

8. One particular part of the modern theory of species evolution postulates mechanisms for genetic mutation, persistence, and adaptation. This part of the theory has been used as a specific metaphor for so-called genetic algorithms that enable software constructs to learn and survive in a manner analogous to biological organisms subject to the forces of natural selection. The use of metaphor to understand computational puzzles (in this case, one kind of learning), which, in turn, might be part of the problem of building AI is an important part of the story of metaphor in AI. Problems of scope prevent our addressing such piecewise use of metaphor in favor of our main theme of metaphor as the basis for establishing a research paradigm.

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 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 area of AI, ranging from specific applications (for example, in the area of formalizing geographic and genetic knowledge) to the general, philosophical foundations of the field.