.op
.po 13
.mt 7
.mb 9

Chapter II

The Computational Metaphor

Defining the computational metaphor is simultaneously a simple and a difficult task. Simple, in that it can be abbreviated as a simple equation,[1]

"mind (brain) = computer = mind (brain)."

Difficult, because the computational metaphor is not a single entity, rather it is the label for a collection 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).

1 Even this simple form has nuances. Just as an evolutionist is careful to correct the statement, "man is descended from apes," to "man and apes share a common ancestor," a proponent of the computational metaphor would correct the "mind=computer" statement to, "minds and computers are both instances of a single abstract entity, a physical symbol system."

Compounding the difficulty is the extent to which, and the speed with which, various aspects of the metaphor have invaded essentially every realm of scientific investigation as well as infusing the popular culture and vernacular. Documentation of the influence of the computational metaphor is provided by numerous books and articles. Perhaps the three best known are Bolter's Turing's Man, Turkle's The Second Self, and McCorduck's The Universal Machine. [Bolter 84, Turkle 84, McCorduck 85]

Exploring the consequences (good and bad) of the seemingly universal acceptance of the computational metaphor is the common thread linking the cited works. Bolter labels those accepting the metaphor as "Turing's men."

> "I call those who accept this view of man and nature Turing's Men. ... When the cognitive psychologist begins to study the mind's 'algorithm for searching long-term memory,' he has become Turing's Man. So has the economist who draws up input-output diagrams of the nation's business, the sociologist who engages in 'quantitative history,' and the humanist who prepares a 'keyword-in-context' concordance." [Bolter84:13]

Additions to Bolter's list would include anthropologists who speak of "culture as a system of knowledge," [Keesing 74:77] "binary oppositions," [Levi-Strauss 75]

"representation of knowledge" [Quinn 85: 292] or "cultural frames" [Colby 85].

In the realm of psychology, cognitive science, and AI the computational metaphor has become so central it is difficult to imagine conducting serious research that is not affected, in some respect, by the metaphor. Other humanistic disciplines (anthropology, sociology, economics, history, philosophy, etc.) show increasing evidence of the influence of the metaphor in their realms as well. [See Gardner 85]

More interesting than the fact that the computational metaphor is fundamental to so many areas of modern thought and life is WHY the metaphor has come to be so influential. This chapter will discuss two contributing forces that account, at least partially, for the pervasiveness of the computational metaphor. The first force involves a kind of metaphorical conflation and the second a synergy between the computational metaphor and the scientific (philosophical) tradition of rationalism (formalism).

Technological (Computer) Animism

Projective anthropocentrism (animism, personification) has is evident from the earliest periods of human existence

and self-awareness. Ascribing characteristics to our external environment that we have introspectively seen in ourselves has never been unusual. Examples can be found almost anywhere we choose to look. Nor is it unusual for technological objects to be the object of animism.

Our automobiles have "moods," "personalities" and even volition. Our appliances have sufficient "awareness" and "will" that they can burn our food or otherwise disrupt our lives. So it is not at all surprising that computers are also endowed with humanlike characteristics.

This simple animism was augmented by early attempts to explain what computers were for, what functions they were capable of, and how they performed those functions. It is impossible to reconstruct exactly how and why the first metaphors were selected to explain aspects of computers and computing,[2] but it is reasonable to assume that since even the simplest computer function - adding two numbers together - was akin to those tasks that humans (in the Western world,

2 Consider the case of the "bug" metaphor for a computer malfunction. The anecdote attributed to Grace Hopper about finding a moth short circuiting an early incarnation of a computer, although widely accepted as true, was branded as apocryphal by Hopper in recent years.

at least) learned in a school setting that mentally derived metaphors immediately suggested themselves.[3]

When computers were connected to peripheral devices like tape drives and check sorters that were controlled by signals generated without an immediately obvious human action it was almost inevitable that the computer was seen as a controlling (volitional) agent. These attributes were added to descriptions of the computer as being able to "read" input data and "write" output information. Periods of processing that did not exhibit any outward sign of action were "thinking." Architecturally the computer had both a "memory" and a "brain."

It is interesting to note that as the computer = mind metaphor gained widespread acceptance the description of the CPU as the "brain" of the computer came to be replaced with the notion that the CPU was the "heart" of the computer. "Braininess" came to be attributed more to the collection of

3 Tracing the transition from "calculating engine," the metaphor employed, roughly, from the days of Lady Lovelace and Babbage to Turing as it became, currently, simply "calculator" would provide both insight into, and documentation of, the increasing degree to which human characteristics have been ascribed to computers - and the change from simple personification to actual animism.

components that made up a computer rather than a single component or a small subset of components.

The fact that these metaphors (and the animism and/or personification behind many of them) were employed is not surprising nor particularly interesting except as yet another example of what appears to be a universally human trait.

> "Metaphorical personification, which has probably existed since the advent of human speech, has become extensive in computer science. Primitive cultures often personify natural objects by giving them a divine status; perhaps we have shifted the deification from nature to technology. [MacCormac 85:17]

However they are very significant for another reason. The awesomely powerful bolt of lightning is preceded by an all but invisible "guide bolt" that pre-traces the path of the lightning and provides the critical weakening of the resistance along that path that allows the main bolt to discharge. Animism, personification and mentation derived metaphors for computer functions can be interpreted as the initial "guide bolt" for the "lightning strike" that was to be the full computational metaphor.

When initially proposed mental metaphors of computer functionality 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 - and they are still clearly metaphors. The latter is true because we understand the workings of computers sufficiently that we find it difficult to think of our PC as "really thinking" or "really talking." When the computer in question is an abstract entity, however, the status of the metaphor approaches that of paraphor.

Mentation derived metaphors have been so commonly employed with respect to computers that they have inevitably advanced from their purely diaphoric nature to epiphoric status and perhaps a step beyond. This lessened the dynamic tension of the metaphor to such a degree that when the computer first came to be used metaphorically to describe the workings of the brain-mind, these new metaphors were accepted much more readily than they would have been had the previous metaphors not been so widely used. Hearers of the new metaphors were predisposed to accept them. Instead of being born as diaphors they blossomed as full epiphors.

Computer Metaphors for Mind

When the first computers were being built they 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 metaphorical relationship and the "known" object was our "mind."

There is a subtle but significant distinction between early mind-computer metaphors and a metaphor like "colored" quarks. In the latter case the "known" metaphor 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 is not the case in the mind-computer example. No objectively understood and accepted theory of mind existed (nor as yet exists) to fill the position of a known object in the metaphoric construct. What does exist to fill that position is a set of common sense (experience derived) terms for mental states.

The function of the metaphor - to illuminate the unknown in terms of the known - is preserved even though one of its objects is "known" in a different sense than in a metaphor like the color -> quark example. However, the distinction being made is important for two reasons.

First, it created an ambiguity that contributed to the metaphorical conflation discussed in the next section. Second, as noted in the previous chapter, the strongest metaphors sometimes arise from relating two objects where

very little is known about either one. Both of the metaphorically related objects present referents that are difficult if not impossible to relate accurately and therefore are not subject to empirical refutation. The primary means by which a metaphor evolves or is dissolved - confirming or refuting similarity among referents of the metaphor objects - is essentially inoperative. Evolution of the diaphor to epiphor and lexical status can still occur via the avenue of widespread repetitive use, but dissolution of the metaphor is more difficult.

By the 1960s computers were no longer strange and exotic entities. They were well understood theoretically as well as in engineering terms.4 Computers had even become a fixture in the popular culture of the Western world as sources of humor, anxiety, and frustration. It was at this point that mind-as-computer metaphors become less important, in terms of cognitive theory, than computer-as-mind metaphors.

At this point the pretense that the mind (embodied in the human brain) was something we understood well enough to

4 This is not to say that the theory of computation is complete; the continued existence of computer science departments attests to the need for continued investigations in the area.

employ as a metaphor for the unknown computer was dropped. The mind-in-brain became the major unknown that was amenable to explanation with metaphors derived from the now well understood computer.

Computers, however, were not very satisfactory sources of metaphor for mental operations. The referents provided by a computer and those provided by a brain or mind were obviously dissimilar in almost every instance. Brains did not consist of transistors, memory in the mind was not a specific location of a specific size that was erased everytime there was an interruption in the power supply, etc.

Demonstrable dissimilarities in metaphor referents did not cause the metaphor to dissolve, although that is what might have been expected based on the theory of metaphor presented earlier. Instead, the metaphor shifted from the physical computer to the abstract computer, to computation. Referents of the physical computer (Turing-von Neumann machine, algorithm, programs, pattern matchers, etc.) supplied the "terms" used in computer-as-mind metaphors but in themselves those terms acted as virtual pointers to abstract concepts of which they were exemplars.

Again the classical form of metaphor is violated. The "computer" in computer-as-mind metaphors is only a pointer

to an abstraction, an abstraction that is not a well known entity and which cannot provide referents to be correlated with referents provided by the mind object of the metaphor. Tacit acknowledgement of this state of affairs can be seen in the shift from a straight mind-is-a-computer assertion to its variant, "the mind and the computer are both examples of computation."

(Because the physical computer still supplies the terms used in metaphorical expression there is a tendency to limit the abstract entity (computation) to those aspects that are in fact realized in the physical entity. Although computation is not necessarily restricted to linear, algorithmic, stepwise processes, the metaphoric expressions of the computation are so constrained.)

What is asserted by this type of metaphor is not a relationship between the named objects but between two abstract entities, of which the named objects are examples.

Computation is to computers what mind is to our common sense (introspectively derived) understanding of mentation. The consequences of this relationship will be discussed in the next section.

Metaphorical Conflation

The preceding sections discussed two separate instances of metaphors used to relate computers and minds. Although they are historically distinct, to some extent, and are diametric opposites in terms of relating a "known" object to an "unknown" object, they have tended to merge together into a single entity. Two factors have contributed to this merger.

First, in any metaphor there is an interaction that tends to blur the distinction between the related objects.

> "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 85: 10]

When a metaphor is first proposed there is a "distance" between its objects that is reduced because of the mutual alteration described by MacCormac. The computational metaphor experiences this alteration in a double dose

because not only has it been metaphorically asserted that "computers think," but also that "thinkers compute."

The second influence is simple metaphorical conflation. Two distinct types of metaphor have been merged into a single entity with the combined label of the "computational metaphor." The two-way application of metaphor between computers and human minds has created a "super-metaphor" that obscures the real differences that do exist between the entities being compared.

The consequences of this metaphorical conflation are profound and provide the core issue in debates concerning ethics. The danger is succinctly stated by MacCormac:

> "If humans and computers possess memories or beliefs, then we may be seduced by the metaphorical usage to assume that the properties of human memory can be found in the computer or that the notion of belief in humans should be limited to dispositions to act since they are so limited in computers. [MacCormac 85: 17]

Weizenbaum, Hubert and Stuart Dreyfus, Rozack, and many others criticize strong application of the computational metaphor on precisely this ground - that advocates have totally confused, and obliterated, any and all distinctions between mentation and computation. That which was

originally a metaphor has become a literal statement of equivalency, an equivalency that many people find demeaning to humans and less than ethical in its consequences.

If the "computational metaphor" is not a metaphor, then what kind of expression is it? In part, the answer to this question can be found by considering the other factor that accounts for the pervasive influence of the computation-equals-mentation expression - its association with the philosophic and scientific tradition of "formalism."

Computers and the Formalist Tradition

> "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. It thus represents a formalist approach to theoretical explanation. [Pylyshyn 80:111]

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, often labeled "rationalism" or "formalism." In terms of modern Western philosophy this tradition began with Hobbes, Descartes, Leibniz and Locke and motivated the work of Russell, the early Whitehead, and Carnap.

A detailed discussion of this philosophic tradition is, obviously, outside the scope of this work. Several thematic points need to be elucidated, however, because of their particular relevance to the topic at hand. Two of the themes are introduced by Pylyshyn, i.e. "formal operations carried out on symbol structures." [ibid]

"Symbol structures" are representations, maps that exist only 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.

Pre-modern Aristotelean 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 87: 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 was the concept 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 87:18]

Gardner echoes Descartes' dualism:

> "...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 85:38-39]

as do Genesereth and Nilsson:

> "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." [Genesereth 87: 2]

Although the latter two claim to be dealing with "transactions ACROSS that envelope," (my emphasis) they are in reality discussing transactions between environmental objects represented UPON that envelope and the structure of the entity within the envelope that interacts with those representations.

Their dualism is so strong that the entire environment must 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 87: 3]

Although Descartes' severance of the "mind" from the "world" is a necessary precondition if entities like "computers" are to receive serious consideration as being "thinkers," common sense notions of "thinking" retain vestiges of the pre-modern concept of "interaction" between the mind and the world. It is precisely this contrast that is at the root of many of the debates about the ability of computers to "really think, feel, and know."

For instance, when experiencing pain or when expressing undying affection for another person, humans tend to feel as if the pain and the perceiver of pain become merged (I hurt), the lover loves the object, not the representation of the object (I love you). Humans seem to feel as if they live in both a dualistic and monistic universe, but computers would seem forever relegated to a universe of Cartesian dualism.

It is not my intent to discuss the merits of the two positions, merely to note that the concept of representation and Cartesian dualism are at the heart of the disagreement. This same dualistic notion is central to Keller's discussion of the differences between McClintock's genetics (allowing

interaction) and "mainstream" genetics (Cartesian dualist). [Keller 83] This issue will be raised again in Chapter IV.

The second theme derived from Pylyshyn 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 of human thought can also be traced to Descartes and his project to codify the "laws of thought." Leibniz 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 87: 5]

Despite a long pedigree that includes Boole, Frege, Russell, the early Whitehead, Chomsky, Fodor and many others, the idea that there is a set of formal operations could encompass all thought is far more problematic an assumption than that of abstract representation. Descartes abandoned his grand project (although his Discourse on Method was to be a foundation for his ultimate vision). Central assumptions in the work of Frege, Russell and Whitehead have been disputed by Godel. For every advocacy of the position that a formal system of operations must exist, like:

> "... scientific and engineering progress in a discipline requires the invention and use of appropriate mathematical

> apparatus with which to express and unify good ideas. ...(and) symbolic logic forms a most important part of the mathematics of AI." [Genesereth 87: vii]

there is an opponent (Searle, Dreyfus, McDermott, Geertz, Turner, Singer, etc.) who will maintain that the "whole" of human knowledge and understanding exceeds the limits of any formal system.

Here too, it is not the merits of the debate that are our concern with the exception of one point: Why, beyond a desire for order and prediction, is a formal system a pre-requisite to an adequate understanding of mentation? The answer: 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 those operations. It is not surprising, therefore, that many of the strongest advocates of formal systems (e.g. Pascal, Leibniz, Babbage, Turing, and von Neumann) were also actively engaged in the construction of machinery that would embody their systems.

Pre-occupation 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 potential exception of some

connectionist machines, all attempts to build "thinking machines" are derivative from the earliest "calculating clocks" of Wilhelm Schickard and Blaise Pascal.

Following the conviction of Giambattista Vico, "one is certain of only what one builds," [quoted in Genesereth 87: 1] formalist researchers still want to "build machines" (write computer programs in terms of a set of primitive operations and manipulation rules for those primitives) and are intensely distrustful of ideas which cannot be expressed in this material fashion.

So strong is this formalist, mechanist philosophic tradition that the mind-is-computer and computer-is-mind metaphors found an immediate, accepting and enthusiastic audience. This widespread acceptance, combined with the metaphoric conflation discussed previously, literalized the "computational metaphor."

What remains is a phrase, "the computational metaphor," which does not denote a metaphor at all but is a kind of shorthand expression for a philosophic and scientific "point-of-view." Employing the phrase is a declaration that the user is a follower of the modern dualistic tradition beginning with Descartes, Leibniz, and Locke. Lakoff summarizes the essential points of this tradition (which he identifies with the label "objectivist") as follows:[48]

- "Thought is the mechanical manipulation of abstract symbols.
- The mind is an abstract machine, manipulating symbols essentially in the way that a computer does, that is, by algorithmic computation.
- Symbols (e.g., words and mental representations) get their meaning via correspondence to things in the external world. All meaning is of this character.
- Symbols that correspond to the external world are internal representations of external reality.
- Abstract symbols may stand in correspondence to things in the world independent of the peculiar properties of any organisms.
- Since the human mind makes use of internal representations of external reality, the mind is a mirror of nature, and correct reason

mirrors the logic of the external world.

- It is thus incidental to the nature of meaningful concepts and reason that human beings have the bodies they have and function in their environment in the way they do. Human bodies may play a role in choosing which concepts and which modes of transcendental reason human beings actually employ, but they play no essential role in characterizing what constitutes a concept and what constitutes reason.
- Thought is abstract and disembodied, since it is independent of any limitations of the human body, the human perceptual system, and the human nervous system.
- Machines that do no more than mechanically manipulate symbols that correspond to things in the world are capable of meaningful thought and reason.
- Thought is atomistic, in that it can be completely broken down into simple "building blocks" - the symbols used in thought - which are combined into complexes and manipulated by rule.
- Thought is logical in the narrow technical sense used by philosophical logicians; that is, it can be modeled accurately by systems of the sort used in mathematical logic.

These are abstract symbol systems defined by general principles of symbol manipulation and mechanisms for interpreting such symbols in terms of 'models of the world.' " [Lakoff 87: xiii]

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 that perspective. However, because the use of the phrase "computational metaphor" arose out of 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 those metaphors has been either abrogated or forgotten, a "myth" (in MacCormac's sense) has been created.

Recognition of the myth summarized in the computational metaphor does not imply a judgement, either of the myth itself or on the usefulness of the work done by those adopting it. Utility is ultimately determined on the basis of applications (the subject of Chapters III and IV). Arguing for or against use of a myth is in the same category as arguing religion[5] i.e., essentially a fruitless exercise.

Awareness of the mythical underpinnings of the computational metaphor is, however, important for several reasons. Among them:

- It contributes to an understanding of the tone of the debates that have occurred between strong advocates and vehement critics.
- Claims of "scientific hegemony" [Keller 84, Harding 86] can be understood in terms of the (usually) non-conscious push for adoption of the prevailing "myth" as a standard framework for all research activities.
- The evolution of the computer = mind / mind = computer metaphors to the status of paraphor and then to lexical shorthand illustrates the establishment and maintenance of a formalist paradigm that constrains research.
- It provides a perspective (or framework) providing

either focus for additional criticism (of the philosophical presuppositions implicit in the computer metaphors) or, more importantly, points of departure for proposing alternative models.

5 A "religion" that was "divinely" inspired" when the Angel of Truth visited Descartes in the midst of a series of dreams the night of November 10, 1619?

In many cases the influence of the computer has been that of "arbiter" between two, or more, existing theories. In psychology, for example, the computer metaphor has provided considerable support to opponents of both Gestaltist and Freudian positions. [Gardner 85: Chap. 5] In the case of the Freudian position especially, this is interesting because it contrasts two positions based on technological metaphors - the computer vs. the steam engine.

> "Freud's writings on the brain still speak the old language of energy and power. His approach is reminiscent of the early days of thermodynamics, when new laws of nature evolved out of the study of steam engines and when the chief purpose of such investigations was to obtain more work for less fuel. The Freudian lexicon, full of such terms as drives, repressions, discharges, sources, and sinks, belongs to the nineteenth century world of steam power and the mystique of railways. The brain was seen as an engine waiting to have its boilers stoked." [Campbell 82:193]

Prevailing (but increasingly under attack) positions in anthropology, economics, sociology, and many other disciplines have been similarly influenced by the computer as metaphor.

Model Conflation

Conflation of distinct models occurs just as it does with metaphors. This is of special concern when the models

are used across disciplinary boundaries where the metaphors and pre-suppositions behind those models are not a part of the background understanding of those outside the discipline where the models are developed. Model conflation occurs when researchers in one discipline employ a model developed in another: "Researchers in AI have shown that minds compute; therefore we can analyze this human action in terms of a computational process." Conflation is minimized when it is noted that, "AI researchers have hypothesized, or metaphorically related, mental activities as computational algorithms and if we adopt this same metaphor then ... ." Failing to make this kind of distinction results in perpetuation of the metaphorical conflation examined previously. It also provides ammunition for those challenging the metaphor, the model, or the pre-suppositions, as will be seen in Chapter IV.