Mind, Self, Society, and Computer: Artificial Intelligence and the Sociology of Mind[1]

Alan Wolfe
New School for Social Research

Recent developments in artificial intelligence (AI) raise the possibility that humans may not, as classical sociological theory held, be unique among species. A Meadian distinction between brain as a neurological phenomenon and mind as a sociological one is introduced to examine such a claim. Software approaches to AI avoid the problem of modeling the human brain, but, because they require thorough and unambiguous instructions, they cannot model how human brains understand external reality. Hardware approaches to AI, such as parallel data-processing models, do attempt to model the brain but only in an engineering sense: in substituting procedures for meaning, they again fail to account for how human brains, let alone human minds, work. The hypothesis of human distinctiveness, consequently, is not rejected, but expanded and elaborated. The actual results of work in AI support interpretative trends in sociological theorizing rather than system-oriented ones.

I. NATURE AND ARTIFICE

Sociology developed toward the end of the 19th century in a specific intellectual milieu dominated by Darwinian evolutionary biology. While this impetus sometimes had the effect of modeling society on the homeostatic functioning of the human organism, especially in Durkheim, it also gave rise to a dualistic view of nature and society. Most thinkers in the sociological tradition shared a philosophical anthropology that understood the special character of being human as the ability to transform and control nature (Honneth and Joas 1988). Weber, for example, viewed culture as "man's emancipation from the organically prescribed cycle of natural life" (Gerth and Mills 1958, p. 356), while, for Durkheim, "it is civilization that has made man what he is: it is what distinguishes him from the animal: man is man only because he is civilized" (1973, p. 149).

[1] I wish to thank Charles W. Smith, Dean Savage, Melvin Reichler, Eugene Rochberg-Halton, and Jeff Coulter for criticism, as well as anonymous reviewers for *AJS*. Requests for reprints should be directed to Alan Wolfe, New School for Social Research, 65 Fifth Avenue, New York, New York 10003.

0002-9602/91/9605-0001$01.50

Civilization was generally viewed as artificial, a development sometimes lamented but more often than not appreciated as a precondition for the development of more purposive forms of action. (For a quintessential statement, see Elias [1978].) In contemporary sociology as well, an emphasis on the importance of the artificial—or, in the discipline's current parlance, the socially constructed—aspects of reality over the biological has been applied to such once-allegedly "natural" phenomena as sex roles (Epstein 1988), homosexuality (Greenberg 1988), and, most relevant for this article, mind (Coulter 1979).

Just as Darwinian theory stimulated sociology by posing the question of whether there were capacities that—in contrast to animal behavior—were uniquely and specifically human, recent research in artificial intelligence (AI), cognitive science, and neurobiology raises exactly the same question—only this time in contrast to machine behavior. Artificial intelligence can be viewed as a *Gedankenexperiment,* an effort to pose a series of interrelated "what if" questions. The fascination with AI is surely due to the philosophical issues raised by the possibility that machines can carry out activities once thought to be exclusively human. While many of those issues—such as how minds represent reality, how language works, whether a little person sits in the brain giving it instructions, and whether we can gain access to the thoughts of others—have been addressed in great detail in the literature spawned by AI, one important thought experiment has received little attention, especially from sociologists. What if the duality between the natural and the artificial that has shaped sociological thought is wrong?

That possibility has been raised in AI work on two fronts. On the one hand, Simon (1969) calls efforts to understand language, problem solving, rule following, and cognition based on an analogy with machines the "sciences of the artificial." From such a perspective, human activities such as producing language are viewed as "natural," thereby *reversing* the dualism at the heart of sociological theory, a contrast captured well in Boden's (1981) title, *Artificial Intelligence and Natural Man.* On the other hand, more recent work in AI posits *eliminating* that duality. This is the implication of the notion of a unified theory of cognition, that there may be similar ways of processing information that link DNA reproductive codes to animal cognition to human decision making and finally to machine calculation (Waldrop 1988*a,* 1988*b;* Churchland and Sejnowski 1988; however, see Edelman [1988]). If such is the case, then at some point sociology would disappear, its laws reduced to biochemistry. In either case, the proper study of man, as Simon puts it, would no longer be man but "the science of decision" (1969, p. 83).

Turkle (1984, p. 13) describes the computer as an "evocative . . . object that fascinates, disturbs equanimity, and precipitates thought."

Carried away by the extraordinary ability of machines to process information—especially in contrast to the sloppy, irrational, and trial-and-error methods used by humans—some adherents to what Searle (1981) has called "the strong case for AI" envision the possibility of a "postbiological" world in which robots, because their purely cognitive powers have, in some areas, already surpassed ours, will carry out most of the work that exists in society (Moravec 1988, p. 5). If indeed "these artificial intelligences will help run society and relieve mankind of the burden of being the leading species" (McCorduck 1981, p. 210), then human society can be viewed as a passing phase in the evolutionary cycle. "And in all humility, we really must ask: How smart are the humans who've taught these machines? On the evolutionary time scale, thinking animals are relatively recent arrivals. Evolution hasn't had a great deal of time to work on the perfection of human cognition" (Feigenbaum and McCorduck 1983, p. 41). With the development of AI, we are back to where sociology began, reflecting on the implications of Darwin.

This article reflects an effort to participate in the thought experiment, stimulated by work in AI, concerning the duality between nature and artifice. In the spirit of such experiments, the following "what if" questions can be asked: What would AI have to demonstrate to confirm the hypothesis that machines can be a substitute for human forms of intelligence? How far has actual work in, as opposed to euphoric claims about, AI come to meeting that goal?

These questions are of great importance for sociological theory. One of the reasons sociological theorists such as Weber (Gerth and Mills 1958) and Mead (1934) compared the human species to other forms of biological life was to argue the importance of the meaning-producing capacities of the human species. To possess the ability to attribute meaning is not simply to follow precoded rules but to bring to situations an awareness of context that allows flexibility in the application of rules. The same ground exists for comparing human intelligence with AI. If it can be shown that humans possess meaning-generating capabilities that machines do not, it follows that efforts to theorize about human behavior based on essentially algorithmic, automatically functioning, rule-following models—which have become increasingly popular in sociology (see, e.g., the work of Niklas Luhmann [1982, 1989])—are not as appropriate to the study of human endeavors as those emphasizing ambiguity, multiple realities, and the social construction of meaning.

II. MIND AND BRAIN

Researchers in AI have a test of machine intelligence called the Turing test—not, it ought to be added, in the original form in which Turing

(1950) proposed it but in the form in which most contemporary work in AI uses it (see Karelis [1986] and Shannon [1989] on this point). To determine whether a machine is intelligent, the Turing test suggests that we imagine a person being given instructions on what to do by both a machine and another person. When the person can no longer tell which of them is giving the instructions, intelligence has been modeled by the machine. How, in the same spirit, do we know we are in the presence of human, rather than machine, intelligence? To answer that question, we can turn to the one sociological theorist who most addressed questions of mind: George Herbert Mead.

Mead's (1934) argument is that the difference between human and nonhuman species involves two further distinctions: all animal species have brains, but only the human has a mind; all other species have bodies, whereas only the human has a self. In the first distinction, brains are physiological entities, organs composed of material properties and represented by what in Mead's day was called the central nervous system. (More recent biological insights into the brain utilize a different terminology; for important work in the area, see Edelman [1987].) But unlike the study of the brain, Mead wrote, "it is absurd to look at the mind simply from the standpoint of the individual human organism." This is because "we must regard mind . . . as arising and developing within the social process." Human forms of cognition are characterized by a process in which the social mind compliments the biological brain: "The subjective experience of the individual must be brought into relation with the natural, sociobiological activities of the brain in order to rend an acceptable account of mind possible at all; and this can be done only if the social nature of mind is recognized" (Mead 1934, p. 133). Mind, therefore, presupposes at least two brains. Mind supplements brain to the degree that an individual incorporates into his or her actions the point of view of another.

Can communication between a human and a machine therefore be considered mindful? Humans can, of course, put themselves in the place of a machine and identify with it, as was the case with Joseph Weizenbaum's (1976) ELIZA. Designed as a purely self-referential system, ELIZA is capable of asking questions of a subject—in the form of a therapeutic dialogue—simply by transforming the word order or form of the questions asked of it. Nonetheless, many individuals were moved by their therapeutic encounters with ELIZA to reflect and grow, which indicates something approximating Mead's (1934) triadic relationship between subject and object. Mindful interaction between a person and a computer therefore seemed to have taken place.

Even more interesting, however, are efforts by machines to take the position of individuals. Intelligent tutoring systems (ITS) are designed to flag seemingly inappropriate questions from student programmers and then to check whether the programmer really want to ask them (Sleeman and Brown 1982; Sleeman 1983). It may be the case, for example, that the machine has assumed too much knowledge (or too little) on the part of the student programmer. Or it may be that the options presented to the programmer are too restricted. Under such conditions, ITS systems are capable of spotting errors by, in a sense, substituting themselves for the questioner. GUIDON, for example, was designed to supplement the medical diagnosis program MYCIN. By comparing a student's questions to those asked by MYCIN, the program can determine when the diagnosis being followed by the student is off track. GUIDON is also capable of analyzing the discourse patterns of the questions posed to it to see whether they are consistent with earlier questions (Clancey 1984). And, as the limits of GUIDON are reached, other programs—such as NEOMYCIN, HERCULES, IMAGE, and ODYSSEUS—have been developed to further refine ITS possibilities. (For an overview, see Wenger [1987], pp. 271–83.) It is even possible to design a program, such as TALUS, that can debug other programs such as LISP (Murray 1988). If the Meadian concept of mind is based solely on the ability to take the point of view of another, including another program, these programs would also appear to possess mind.

But there is more to the Meadian analysis than reflection by putting oneself in the place of another. The second Meadian distinction is between the body and the self. What enables a physical body to become a social self is the possibility for an interaction with another social self. Since "selves can only exist in definite relationships to other selves" (Mead 1934, p. 164), qualities of mind exist instead when a gesture "has the same effect on the individual making it that it has on the individual to whom it is addressed" (Mead 1934, p. 46). No individual can therefore possess reflective intelligence—that is, be viewed as having a mind—without another individual also possessing a mind. Mead's formulation is thus the converse of the Turing test: the other must itself be a self before a self can communicate with it. Human cognition, because it requires that we filter our thoughts through the way we anticipate that other human beings will receive them, is therefore distinct from any other kind of cognition.

The distinction between a social mind and a biological brain is somewhat arbitrary. There are neurobiologists who argue for the existence of a "social brain," in the sense that many activities studied by sociologists, such as religious belief, can be explained by pure neurological functioning (Garrazzinga 1985). On the other hand, there are theorists in cognitive science who argue that we have a "cognitive mind," that thought has its own language, so that qualities of mind do not lie, as Mead argued, in things external to it (Fodor 1975, 1981*a;* for a sociological critique, see Coulter [1983, 1985]). Some work in AI even suggests that there is no distinction between mind and brain at all, but only something called "mind-brain" (Churchland 1986). Still, the distinction between mind and brain as suggested by Mead can help formulate some of the issues at stake in the claim that AI can model human intelligence.

The possibility that machine intelligence could replicate the human brain is primarily a technological question, a matter of engineering. Since the brain is composed of a series of neural nets, it ought, in principle, to be possible to develop a machine capable of equaling, or perhaps someday surpassing, the information capacity of human brains. We could thus conclude that if a machine (or series of machines) were capable of processing information faster and more efficiently than a human, the hypothesis that the brains of humans are superior to those of all other species would have to be rejected.

But the possibility of a machine replicating the human mind raises a different and more complicated set of issues. If we use Mead's distinction between mind and brain, the possibility of a machine replicating the human mind is foreclosed by definition, since at least two human selves have to be involved in an interaction, according to Mead, to define it as mindful. In that sense, our test is "rigged." There are good grounds nonetheless for using Mead's distinction to examine the kind of mindful intelligence a machine would have to at least approximate for its intelligence to be compared with human intelligence. For if human selves make sense of the world by sharing impressions with other human selves—and therefore interpret the rules by which people govern themselves—it becomes possible to understand why humans are capable of accomplishing certain tasks even if their brains may have less pure computational power than the artificial brains existing in machines. The question, in this sense, is not whether machine brains are superior to human minds or vice versa. Rather, the biological brain and the social mind work in radically different ways: one seeks information as complete and precise as possible; the other does not need hardwired and programmed instructions—or even trial-and-error learning through the strengthening of neural nets—because it can make sense out of ambiguity and context.

III. SOFTWARE APPROACHES TO AI

Work in AI is generally divided into two approaches: software and hardware. A software approach is one that tries to reproduce what the brain does without entering into the question of how the brain does it. Hardware approaches to AI reject an analogy with the computer and try to model intelligence and learning in machines by imitating the neurological structure of the brain.

Early work in AI proceeded along both approaches, but it was not long before the hardware efforts represented by Frank Rosenblatt (1962) and his ideas about "perceptrons" gave way to software approaches that seemed to offer promising payoffs (Minsky and Papert 1969). These approaches assumed that the brain possessed a central processing unit (CPU) that stored information in the form of memory. The von Neumann architecture of a computing machine could therefore attempt to replicate this structure, with software providing access to the CPU in the way it was assumed that events in the real world stimulated the memory recall operations of human brains. The beauty of this approach was that it was irrelevant whether the computer actually did model the human brain. If a program could be written that could represent reality, then intelligence lay in the application of instructions or algorithms (Minsky 1981). Only one assumption was necessary for this procedure to work, and that was that a complete set of instructions could be provided. As von Neumann put it, "Anything that can be exhaustively and unambiguously described . . . is . . . realizable by a suitable finite neural network" (cited in Gardner 1985, p. 18). Many researchers in AI are convinced that even the world of everyday life—of metaphor and ambiguity, for example—can be programmed into formal rules that a machine can understand, if only we specify them in all their complexity (Hobbs and Moore 1985).

If memory were infinite, as it could be imagined in the purely abstract theories of Alan Turing (1950), machines would simply take as long as is needed to find the information relevant to an instruction, even if the time to do so were, say, the equivalent of three human lifetimes. But in the real world, software approaches to AI rapidly ran up against the obstacle that the reality outside the computer was more complicated than anyone had realized. Some indirect way to represent reality, therefore, became the objective. Some attempted to create "expert systems," in which programs were based on a model of the decision-making process used by experts in various areas (Newell and Simon 1972; Feigenbaum and Feldman 1963). So long as they confined themselves to relatively limited domains of rule application, expert systems were a major success, especially with the creation of DENDRAL and MYCIN for medical diagnosis. They can even, as we have already seen, be expanded to guide the questions asked of them by programmers. Such successes have led researchers within this tradition back to efforts at more general programs, such as Allen Newell's attempt to create a unified theory of cognition (Waldrop 1988*b*), but, although such efforts have generated a good deal of excitement, their potential lies in the future.

Another way to get around the problem of representing reality was to take certain shortcuts, to assume that although reality outside the machine was enormously complicated, it could be broken down into smaller categories that were less complicated and then such smaller units could be combined. Minsky (1981), for example, argued for "frames" (a term sociologists recognize from Goffman [1974]), while Roger Schank talked about the existence of scripts (Shank and Abelson 1977). No one was willing to reject the CPU metaphor, but they did recognize that holding information about everything that ever happened throughout a lifetime would require an unwieldy human CPU, so memory was assumed to be stored by humans in the form of episodes. Each set of generalized episodes could be called a script. There was, for example, Schank's famous example of a restaurant script, a set of associations in the brain about what happens to an individual on entering a restaurant, within which any particular experience in a real-world restaurant can be framed (Schank and Abelson 1977).

Schank's ideas about scripts were subjected to one of the most searing critiques in the AI literature: Searle's (1981) effort to prove that machines cannot "understand" the instructions given them. The implication of Searle's "Chinese room" critique was that, while human intelligence might be replicated, there could not be any artificial, functional equivalent of human understanding. In the absence of this human capacity to understand, not even frames and scripts could avoid the problem that, to know anything, the machine first had to know everything (Rosenfield 1988, p. 113). No wonder that many of the early criticisms of work in the area were directed primarily against software approaches (Dreyfus 1979; Weizenbaum 1976). Even AI researchers themselves began to look for another approach, one based more directly on the "hardware" of the human brain.

As these examples illustrate, the development of AI raised new and important questions not only about machines but also about humans. If machines have trouble representing reality outside themselves, how do humans do it? Humans, like machines, can also be given rules that they are expected to follow. But since our memories are imperfect, it would be difficult to conclude that our rules can also be programmed in detailed specificity if the same technique did not work for computers. Work in AI stimulated neuroscientists to take a closer look at the human brain, and what some of them discovered was that the whole idea of a human CPU had to be rejected (Edelman 1987; Reeke and Edelman 1988; Rosenfield 1988). This position—associated with those who reject neurological theories (dominant since Broca) that different local sectors of the brain are responsible for different functions—offers instead what Edelman calls a nonalgorithmic understanding of the brain (1987, p. 44) or, more accurately, of brains, for Edelman's argument is that different brains develop differently in the form of a selective system, just as certain species are understood by Darwinian theory to evolve in response to new challenges. Humans have, in other words, what Edelman called a "mindful brain" (Edelman and Mountcastle 1978), which software approaches, dependent purely on algorithms, cannot have.

But if human memory is not stored somewhere, how do we, in any particular circumstance, know what to do? Putting the issue in another form, if it is true, as one neuroscientist argues, that "we are probably much better at recognition than we are at recollection" (Rosenfield 1988, p. 158), what we need to understand is not how human memory is stored but how it is activated. The answer may very well lie in the Meadian distinction between mind and brain. Humans have minds that are capable of interpreting rules and instructions: we fill in the frames or interpret the scripts, not just search through memories to match a representation to a reality, because our minds recognize the external reality that our brains cannot.

If this line of reasoning is correct, it would follow that the human brain, unlike machine brains, can be incomplete, that our brains do not need to understand everything with which they are presented because we also have minds that interpret the world for our brains. What we therefore need to recognize, and what work in AI unintentionally seems to show, is that humans are distinct, not because their brains store *more* information than machines, but because they can store *less* and get away with it. Our distinctiveness, in short, lies in the *unknowability* of the world around us to the brain within us.

Confirmation for the neurological notion that human brains process information, but need minds capable of supplying context, is provided by the ethnomethodological tradition in sociology. For Garfinkel (1967), for example, conversations are interesting not for what is said but for what is not said. Thus the words, "Dana succeeded in putting a penny in a parking meter today without being picked up," might be difficult for a computer to process because it would not know whether Dana was being lifted up to the parking meter or had not yet been met by his parents in their car. But even if a "natural" language program had anticipated this problem and could reject the incorrect meaning of "pick up" in favor of the correct one, it would be unlikely to interpret the sentence to mean what one of Garfinkel's students assumed it to mean: "This afternoon as I was bringing Dana, our four-year-old son, home from the nursery school, he succeeded in reaching high enough to put a penny in a parking meter when we parked in a meter parking zone, whereas before he has always had to be picked up to reach that height" (Garfinkel 1967, p. 38).

Research into human conversations, which was stimulated by Garfinkel, illustrates the difference between how human minds and machine brains talk. Although there are researchers who find in conversation analysis a method for understanding how machines and humans can better talk to one another (Good 1985; McTear 1985), the whole point of ethnomethodology is to analyze how people themselves develop the rules that structure what they do. Thus, as only one example, Schegloff and Sacks (1979) show how something as seemingly obvious as the closing of a conversation is in fact a socially negotiated process between the speakers. If it is true that "there are possibilities throughout a closing, including the moments after a 'final' good-bye, for reopening the conversation" (p. 262), then human agency is always a third party to talk between two human beings. It thereby follows, as Scheff (1986) has argued, that direct and exact translations between natural languages and computer languages will never be possible, since human language production, like all human activities, presupposes a "micro-world underlying all social interaction, [which] connects individuals in shared meanings and feelings, and also connects them to the social structure of their society" (Scheff 1986, p. 82).

In a similar manner, although Bateman (1985) has pointed out that nearly all of Alfred Schutz's (1967, pp. 80–81) concepts can be translated into AI concepts—that Schutz's phrase "stock of knowledge," for example, is the same as Minsky's (1981) frames or Schank's scripts (Schank and Abelson 1977)—the ethnomethodological impetus would seem to lead to an appreciation of how plastic our tacit understandings of the world tend to be. Schutzian phenomenology fills in the gaps that a formal analysis of grammatical rules can never fill: the everyday world provides the background or tacit knowledge that makes it possible to act in a contingent world, to act, as Dreyfus (1979) puts it, without a theory of how we act. Tacit knowledge, background assumptions, and practical reasoning are all features of mind that enable individuals to be rule-governed creatures even if they do not know what all the possible rules may be. The Wittgensteinian regress—the notion that the specification of any set of rules always contains a ceteris paribus condition that cannot be understood within the terms of the rules specified (Dreyfus 1979, pp. 56–57)—while always a logical problem, rarely becomes a practical human problem. We can define the situation because the situation is not defined. We can construct meaning because the meaning is not known. Having gone through a period in which it tried to escape from ambiguity, sociological theory is coming to appreciate it, in part, as Levine (1985, p. ix) states, because of "the recent ascendancy of computerized thoughtways." Ambiguity is essential to human communication—as well as to other types of human behavior—because it is precisely through the gaps in our ability to communicate through language exactly what we mean that the social order exercises its ability to bind us together in realms of shared meaning.

The differences between the knowing brain and the unknowing mind are illustrated by one of the activities that both machines and humans periodically undertake: playing chess. As Georg Simmel once pointed out, in a metaphor exceptionally appropriate to the age of AI, there are two conditions that would inhibit an individual from playing a game of chess. One is not knowing any moves. The other is knowing all the moves (cited in Heritage 1984, p. 61). Chess-playing programs developed by AI researchers cannot specify all moves; that is why heuristic rules were developed that eliminate nonsensical moves, making it possible for computer programs in the real world to play exceptionally expert chess. Yet let us grant one assumption of science, that, if something is theoretically possible, we can imagine it to be practically possible. When the perfect chess program is developed, the result is to stop "playing" chess: when all moves are known, it can no longer be a game. A minimal condition for gaming, as Erving Goffman once pointed out, is that "a prior knowledge of the players will not render the outcome a foregone conclusion." What makes a game a game is that interaction has taken place: "The developing line built up by the alternating, interlocking moves of the players can thus maintain sole claim upon the attention of the participants, thereby facilitating the game's power to constitute the current reality of its players and to engross them" (1961, p. 67). Winning games is something our brains do; playing them is something our minds do. (That people both play and play to win only means that they have both minds and brains.)

There would be no need for mind if, not only in the playing of chess but in all other human activities as well, human agents acted with complete knowledge of the consequences of their acts. If the self knows the consequences that will follow from any gesture, speech act, or form of behavior, it will no longer be a self.[2] Human forms of learning grow out of the uncertainty of what we do, leading us to rely on social practices, the cues of others, experience, definitions of the situation, encounters, norms, and other ways of dealing with uncertainty that enable mind to develop. One of the leading German philosophical anthropologists, Arnold Gehlen (1988, pp. 79–92), argued that because humans are more imperfect than

[2] As Turkle has written, "If mind is a program, where is the self? . . . In its challenge to the humanistic subject, AI is subversive in a way that takes it out of the company of rationalism and puts it into the company of psychoanalysis and radical philosophical schools such as deconstructionism. . . . Artificial intelligence is to be feared as are Freud and Derrida, not as are Skinner and Carnap" (1988, p. 245).

other animal species in their gestation periods—they remain infants far longer than other creatures do—their specific traits develop out of their need to compensate for the lack of what nature has given them. The same can be said for their brains. Imperfect, trial-and-error bound, hesitant—the human brain is incomplete in the absence of a social mind. It may therefore be the case that someday a computer will surpass any one particular human brain in its intelligence (although none have come even close so far), but what a computer is unlikely to surpass is the collective intelligence of assembled minds. We are not sure how we comprehend the world outside our brains, but we are fairly certain that we do not do it through a detailed set of algorithms specified in software or its functional equivalent.

IV. HARDWARE APPROACHES TO AI

The past few years have seen, within the AI community, a revival of hardware efforts, once associated with Frank Rosenblatt (1962) and his notions about perceptrons, now called parallel distributive processing (PDP), neural nets, or connectionism (see, e.g., Grossberg 1988; Hinton and Anderson 1981; Rumelhart et al. 1986; Mead 1988). Rather than model how a brain decides without entering into the way it decides, these efforts use certain understandings of neurological behavior to develop analogous data-processing systems (Mead 1988). Because the brain works so much faster than computers, these thinkers argue, it must be composed of many computational devices working in parallel fashion. And because the brain does not necessarily store its memory in specific locations, waiting to be activated by signals that enter the system, its architecture is better viewed as a series of nets activated by the connections that exist between them. In that sense, PDP approaches circumvent the most conspicuous flaw of earlier efforts, the use of a von Neumann machine instead of the brain itself as a model for human intelligence: "One important difference between our interpretation of schemata and the more conventional one is that in the conventional story, schemata are stored in memory. Indeed, they are the major *content of memory*. In our case, *nothing stored corresponds very closely to* a schema. What is stored is a set of connection strengths which, when activated, have implicitly in them the ability to generate states that correspond to instantiated schemata" (Rumelhart et al. 1986, p. 21; emphasis in original).

Two important considerations follow from this major shift in emphasis. One concerns rules and scripts. Researchers in the PDP tradition "do not assume that the goal of learning is the formation of explicit rules. Rather, we assume it is the acquisition of connection strengths which allow a network of simple units to act *as though* it knew the rules"

(McClelland et al. 1986, p. 32; emphasis in original). It follows that the machine—more accurately, in this kind of work, a set of parallel machines—can "learn," because it can react to ambiguous or incomplete instructions. While researchers in this tradition are cautious about making large claims for their work, they are convinced that machines can reproduce the human capacity to act in particular ways on the basis of past experience.

One example provided by researchers in this tradition helps illustrate what is new about this approach when compared to older forms of AI work. Suppose that a child is in the process of learning the past tense. The general rule is that we take the present tense and add "-ed." Following this rule, a naive subject would reason as follows:

jump	jumped
walk	walked
come	comed

To respond to such a difficulty, earlier research in AI would have begun a search for all exceptions to the general rule, specifying them as precisely as possible so that a machine would know, if asked to give the past tense of a verb, how to respond. But PDP works in the opposite way. It begins with what the naive subject would do, makes a mistake, corrects the mistake, and accumulates in the process enough associations that it eventually comes to learn when an "-ed" ought to be added and when some other form of the past tense is correct (McClelland et al. 1986, pp. 39–40). In short, the reasoning here is trial-and-error reasoning and is, in that sense, similar to the way humans think.

If software approaches to AI located intelligence in a set of instructions to a CPU, hardware approaches locate intelligence in a set of procedures that can activate connections. "Under this new view, processing is done by PDP networks that configure themselves to match the arriving data with minimum conflict or discrepancy. The systems are always taming themselves (adjusting their weights). Learning is continuous, natural, and fundamental to the operations" of a system (Norman 1986, p. 544). It is clear that, with the use of PDP methods, machines can adjust to unforeseen instructions, which they were not able to do under software conditions (although this ability to adjust is limited to certain rather artificial situations). Still, our concern in this article is not whether machines can "learn" but whether the way they respond to instructions has anything to tell us about the way human beings learn and respond to instructions.

At issue is the way that "learning" organisms relate parts to wholes. One of the root assumptions in AI research is that intelligence is manifested when enough very small bits of information are assembled together into something called knowledge. Working with the philosophical tradition inspired by Descartes and Hume, AI researchers believe that understanding the mechanics of neural networks enables us to solve what Hume called the problem of the homunculi, which asserts that it is impossible to understand what takes place in the brain by imagining that a little person exists inside of it giving it instructions, for then we would have to posit a little person inside the brain of the little person, and so on, in an infinite progression (Dennett 1978, p. 123; Pylyshyn 1981, p. 68; Edelman [1987, p. 45] believes his understanding of neuronal selection also solves the homunculi problem). Because workers in the PDP tradition look not at whole scripts but at *sub*systems, at the smallest units of communication possible, they respond to the Humean problem by fashioning a smart machine out of exceptionally dumb—indeed, the dumber the better—components.

> Homunculi are *bogeymen* only if they duplicate *entirely* the talents they are rung in to explain. . . . If one can get a team or committee of *relatively* ignorant, narrow-minded, blind homunculi to produce the intelligent behavior of the whole, this is progress. . . . Eventually this nesting of boxes within boxes lands you with homunculi so stupid (all they have to do is remember whether to say yes or no when asked) that they can be, as one says, "replaced by a machine." One *discharges* fancy homunculi from one's scheme by organizing armies of such idiots to do the work. [Dennett 1978, pp. 123–24; emphasis in original][3]

As Douglas Hofstader (1979) has put it, the paradox of AI is that "the most inflexible, desireless, rule-following of beasts" can produce intelligence (p. 26). They can, at least if one defines intelligence in ways compatible with AI's emphasis on the parts adding up to a whole. But what if human intelligence is otherwise, that is, what if our way of thinking and learning involves the continuous back and forth of wholes and parts?

It ought to be clear that the question of parts and wholes is a fundamental issue in sociological theory. Durkheim's notions about the division of labor—where each human agent, generally acting in ways unknown to other human agents, nonetheless contributes to the effective overall functional performance of the society—is but one formulation of an age-old problem of how parts and wholes interrelate. Durkheimianism, or any strong form of structuralist sociology, is an effort to understand how a smart organism—civilization or social structure—could emerge out of,

[3] Dennett is not writing here about connectionist approaches but about AI in general. His remarks, nonetheless, are especially relevant to hardware approaches.

if not necessarily dumb, then at least somewhat limited components, that is, people. In relying on a biological metaphor as the basis for his functionalism, Durkheim envisioned society as composed of hearts, muscles, heads, and other organs—all of which have tasks to perform, but none, save perhaps the head, has much consciousness or awareness of why it is doing what it is doing. It was precisely this sense of structures operating because of their homeostatic functions—inherited by Parsons from Durkheim—that led microsociologists, especially Garfinkel and Goffman, to pay more attention to individual human minds.

Just as sociological theory was led to a greater appreciation of how micro and macro interrelate, work in AI is coming up against the limits of the notion that a focus on the smallest possible parts will tell us something about the behavior of the whole. Minsky (1986), for example, although once associated with software approaches to AI, has, like his colleague Seymour Papert (1988), become sympathetic to the new approaches associated with PDP and connectionism. He proposes that we answer the question of how dumb components can make a smart machine by asking us to imagine that intelligence is a "society" composed of agents—such as the comparing agent, the adding agent, the seeing agent—each of which is ignorant of what the other agents are doing: "Each mental agent by itself can only do some simple thing that needs no mind or thought at all. Yet because we join these agents in societies—in certain very special ways—this leads to true intelligence" (1986, p. 17). The overlap with Durkheim here is striking, and, as with Durkheim, the question becomes whether PDP approaches can enable us to focus on aggregates—what Minsky calls "societies"—without attributing significant intellectual qualities to the parts that compose those societies.

It is worth noting in this context that Minsky not only reaches for the metaphor of "society" to talk about the whole, but also the term "agent" to describe the part. One reason that the interaction between parts and the whole seems to work for human societies is that human beings clearly possess agency: they can shift their attention back and forth from parts to wholes because they are autonomous agents capable of thinking for themselves. (It is the recognition of the power of human agency that led sociological theory away from an overdetermining structuralism.) Can a machine premised on parts that are as dumb as possible in any way replicate the way real human agents operate in the world? Just as software approaches could have manifested intelligence resembling human intelligence if they could have overcome one fatal flaw—the need to specify descriptions of the real world as thoroughly and unambiguously as possible—the ability of the hardware approach to approximate human learning hinges on one question as well: Does it make sense to apply the term agency to whatever is operating through a set of essentially dumb microprocedures that activate its states?

The role that agency plays in the case of human intelligence is underscored by the same neuroscientists who reject the CPU model of the human brain. True, they admit that the PDP approach is closer to what we know about how human brains work, yet they are by no means convinced that these new approaches will enable machines to model the brain (Reeke and Edelman 1988, p. 152). The reason has once again to do with the social nature of the human mind. Rosenfield (1988) has written that "the world around us is constantly changing, and we must be able to react to it . . . in a way that will take account of the new and unexpected, as well as our past individual experiences" (p. 8). The question, then, is how we come to "take account" of unexpected events. Rosenfield's view is in accord with the Meadian notion that because humans are meaning-producing creatures, their intelligence lies in their ability to interpret the meaning of the stimuli around them: "Fixed memory stores, we have already seen, cannot accommodate the factors of context and history. Computations—the use of procedures in a limited way—bring us closer to a better solution but still fail to explain a crucial aspect of our perceptual capabilities: how our past affects our present view of the world, and how our coordinated movements, our past and present explorations of the world, influence our perceptions" (p. 145).

The clear implication of the work described by Rosenfield is that human brains work the way they do because the signs they recognize are not merely representations of microparts but also interact with larger wholes in the culture outside of the brain. Human agency, in other words, is a central feature of human intelligence. The unit doing the thinking and learning must be capable of taking in the context of the whole if the parts are going to fit together in any coherent way. If this point of view is correct, then both PDP and connectionist approaches to AI must demonstrate that their machines can in some way model the human agent's capacity to understand the meaning of wholes before anyone can take the first step toward engineering a replica of the human brain. Yet, in the PDP view of things, it is precisely meaning that is sacrificed in order to specify microprocedures. As D. A. Norman (1986) puts it, "I believe the point is that PDP mechanisms can set up almost any arbitrary relationship. Hence, to the expert, once a skill has been acquired, meaningfulness of the relationships is irrelevant" (p. 544). Because "the interpretation of the process is *not* in terms of the messages being sent but rather by what states are active," it follows that "in general, *there is no single reason why any given cognitive state occurs*" (p. 546; emphasis in original). What we get in these approaches, even under the best of circumstances, is a machine that may resemble the human brain in an architectural sense but one still without the capacity of human brains to move backward and forward from microprocedures to macroawareness. These approaches come somewhat closer to what would be needed to reject the hypothesis of human distinctiveness but are still far from even the first stage of what would be needed to do so.

V. CONCLUSION

The founders of sociological theory were stimulated to think about the specifically human features of their societies because the intellectual air around them was filled with Darwinian thoughts. Contemporary sociological theorizing, in a very similar way, is inevitably going to be affected by the revolution in computing that marks our own age. It ought to come as no surprise that "artificial intelligence . . . is beginning to tread in waters very familiar to sociologists, while sociologists could soon find some of the methods and concepts of AI provide a novel, but reliable approach to their subject" (Gilbert and Heath 1985, p. 1). While there are some who question the relevance of AI to sociology (Jahoda 1986; Woolgar 1985; Oldman and Drucker 1985), there have also been attempts to apply the insights of AI to such diverse topics as Goffmanesque dramatological models (Brent 1986), ethnomethodology (Good 1985; McTear 1985; Pateman 1985), sociolinguistics and social cognition (Bateman 1985), and the sociology of medicine (Gilbert and Heath 1985; Weaver 1986).

One indication of the impact of AI on sociological theory is the availability of the processing power of machines to serve as a model for all the complexities of human interaction in societies. Beniger, for example, has argued that, because "every living system must maintain its organization by processing matter and energy," it follows that "information processing and programmed decisions are the means by which such material processing is controlled in living systems, from macromolecules of DNA to the global economy" (1986, p. 59). But by far the sociological theorist most influenced by the cognitive revolution stimulated by computers is Niklas Luhmann (1982, 1989).

Like all great theorists in the sociological tradition, Luhmann seeks to answer the question, What makes society possible? Modern societies, more complex in their economic, legal, and technological density than any that came before, particularly raise the question of how they can possibly reproduce themselves. Imagining the world as a flux of potentialities existing through time in the manner similar to Schutz (1967), Luhmann argues that we can obtain a meaningful grasp on the world at any particular moment only by thinking of societies as organized by the principle of "autopoiesis," or self-creation (Luhmann 1989; p. 17; see also Maturana and Varela 1987). "In the case of meaning-processing as well as living systems, autopoiesis has to be secured before all else. This means that the system exists if, and as long as, meaningful information processing is continued" (Luhmann 1989, p. 18).

An example of how information processing makes it possible to reduce complexity is provided by Luhmann's recent sociology of law (1989; see also the essays in Teubner [1987]). "Binary coding" is what makes self-referential systems possible. In the legal realm, for example, the binary code legal/illegal defines the entire realm of possibilities. "Since in any case only one of these two possibilities exists for the legal system—there is no third possibility—the schema contains a complete description of the world" (Luhmann 1989, p. 64). Since no binary code can change, however, all coding has to be supplemented by what Luhmann calls "programming." A program determines into which of the binary codes any particular act will fall. Thus *justice* is a programming category, while *legal* is a binary one. Legal regulation therefore occurs when the binary code generates a program that reflects back on itself, and so on, in the form of an eternally recurring braid. Society is possible because the whole system runs on its own:

> This form of legal regulation can be proclaimed as the protection of freedom, indeed as the promise of freedom. Viewed more prudentially it is a matter of a specific technique for dealing with highly structured complexity. In practice this technique requires an endless, circular re-editing of the law: the assumption is that something will happen, but how it will happen and what its consequences will be has to be awaited. When these consequences begin to reveal themselves they can be perceived as problems and provide an occasion for new regulations in law itself as well as in politics. Unforeseeable consequences will also occur and it will be impossible to determine if and to what extent they apply to that regulation. Again, this means an occasion for new regulation, waiting, new consequences, new problems, new regulation, and so on. [Luhmann 1989, p. 66]

In Luhmann's work, we have a picture of society organized to deal with complexity just as machines are programmed to process complex amounts of information. Yet the problems revealed by work in AI ought to give pause to making the analogy too exact. One can assume, for example, that a legal system will work algorithmically: rules will be generated by inputs that will reflect back and modify the rules accordingly. Such a model, however, faces two problems. One is that even machines, as I have argued above, do not necessarily work algorithmically and, it has been recently argued, neither do even apparently rule-driven human activities such as mathematics (Penrose 1989). But even if a legal system did work algorithmically, or even approximately so, it could only do so if the agents affected by the system were more or less automatic rule followers. Here the Meadian distinction between mind and brain becomes crucial. Brains, understood neurologically, can be imagined to be information-processing mechanisms that work by following programmed rules. But minds do not. What makes human intelligence different—and what therefore makes models of society based on analogies with the machine inappropriate—is that, in human societies, people alter the rules they are expected to follow by incorporating information from the contexts and situations in which they find themselves together with others.

The major differences between the way machines and humans relate to rules are summarized in table 1. The most extreme form of rule adherence is contained in the software approaches to AI, for there the rules are everything. The failure of such programs—or, more precisely, their success only on condition that heuristics, efforts to generalize about rules rather than to specify them, become the dominant way to realize them—indicates that concepts of agency based on the notion that an agent is an algorithmic rule follower and nothing else are impossible to conceive: here is where the Wittgensteinian regress will always assert itself.[4] We must, at the very least, posit the possibility of exceptions to rules, as the hardware approach to AI does, thereby introducing what its advocates call learning. It is clearly possible, as connectionist approaches have demonstrated, to make machines that will follow rules, even when the rules are ambiguous or unspecified. (In the words of Terry Sejnowski, "It's not like we're throwing rules away. This is a rule-*following* system, rather than a rule-*based* system. It's incorporating the regularities in the English language, but without our having to put in rules" [cited in Allman 1989, p. 186].) But human minds do not merely follow rules; they also make them. There has not yet been developed a machine capable of making the rules it will then follow.

An emphasis on rules, in turn, raises the question of how they are transmitted. If we distinguish between representations that mean only what they mean and representations into which other meanings can be read—my approach to this long-debated issue will, following Pagels (1988, pp. 192–94), call the former signs and the latter symbols—then machines can read signs, whereas minds can interpret symbols. We recognize symbols as whole configurations and can disassemble them to account for their parts, while signs are the individual elements that to-

[4] From an even more radically Wittgensteinian position, it is possible to challenge the notion that even the simplest machines following the simplest instructions are following rules, for the whole nature of what it means to follow remains problematic (see Shanker 1987).

TABLE 1

Artificial and Human (Social) Forms of Cognition

	Artificial		
	Software	Hardware	Social
Locus	CPU	Neural nets	Mind
Relation to rules	Rule following (algorithmic)	Rule excepting	Rule making
Language	Signs	Subsigns	Symbols
Meaning	Formal/notational	Procedural	Supplied

gether form a symbol. Given the complexity of their parts, symbols are open to interpretation; given the simplicity of their parts, signs are not. The meaning of a symbol does not exist within the symbol but has to be interpreted by the mind.

It is due to the difference between signs and symbols that machines and humans respond differently to questions of meaning. Meaning is formal and notational in some kinds of AI research, most especially those based on the software model. One searches for formal modes of expression that enable thoughts to be represented in terms of syntactical rules (or grammars) that can be rendered into computations. For this very reason, as Fodor writes, "the machine lives in an entirely notational world; all its beliefs are false" (1981*b*, p. 315). Fodor calls this "methodological solipsism"; machines process data *as if* there were referents in the real world to be interpreted, without, of course, ever interpreting them.[5] The development of PDP models reinforces the point, for these versions of AI do not, as the software versions sometimes did, make the claim that they are representing the real world. They use the real-world structure of the human brain merely as a model for a machine. For them, meaning lies in the strengths of connections between nets and nowhere else. Such an approach makes possible a greater mechanical facility with machines, but it cannot duplicate the particular form of intelligence we associate with use of the mind.

If these differences are accepted, it follows that the hypothesis of human uniqueness need not be rejected as a result of AI. Because we can make "a principled distinction between the study of mind (as properly

[5] This feature of AI research is elevated into a methodological principle by Dennett (1978, pp. 3–22), who argues that we can take an "intentional" stance toward machines, ascribing to them certain features without necessarily making an argument that they possess those features in reality.

conceived) and the study of brains and nervous systems" (Coulter 1983, p. 146), there still remain important differences between how people living in society think compared with how machines do. It follows as well that efforts to rely on AI understandings of cognition for models of how human societies work will miss the essential difference between human and other forms of intelligence. To be sure, systems approaches such as Luhmann's seem at first glance to receive support from the cognitive revolution associated with computers. Yet, far from equating the kinds of intelligence associated with all organic systems, AI, in an admittedly backhanded way, actually reinforces the hypothesis of human distinctiveness by calling attention to the ambiguity-resolving, incomplete, and meaning-dependent features of human minds. Work in AI stimulates sociological theorizing in many directions; those traditions with antisystematic inclinations—such as ethnomethodology and symbolic interactionism—receive just as much support from AI as those who would use what we have learned about machine behavior to model what we know about human behavior.

REFERENCES

Allman, William F. 1989. *Apprentices of Wonder: Inside the Neural Network Revolution.* New York: Bantam.

Bateman, John. 1985. "The Role of Language in the Maintenance of Intersubjectivity: A Computational Investigation." Pp. 40–81 in *Social Action and Artificial Intelligence,* edited by C. Nigel Gilbert and Christian Heath. Aldershot: Gower.

Beniger, Joseph. 1986. *The Control Revolution: Technological and Economic Origins of the Information Society.* Cambridge, Mass.: Harvard University Press.

Boden, Margaret. 1987. *Artificial Intelligence and Natural Man,* rev. and expanded ed. New York: Basic.

Brent, Edward E., Jr. 1986. "Knowledge-based Systems: A Qualitative Formalism." *Qualitative Sociology* 9 (Fall): 256–82.

Churchland, Patricia Smith. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain.* Cambridge, Mass.: MIT Press.

Churchland, Patricia S., and Terrence J. Sejnowski. 1988. "Perspective on Cognitive Neuroscience." *Science* 242 (November 4): 741–45.

Clancey, William J. 1984. "Use of MYCIN's Rules for Tutoring." Pp. 464–89 in *Rule-based Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project,* edited by Bruce G. Buchanan and Edward H. Shortliffe. Reading, Mass.: Addison-Wesley.

Coulter, Jeff. 1979. *The Social Construction of Mind.* London: Macmillan.

———. 1983. *Rethinking Cognitive Theory.* New York: St. Martin's.

———. 1985. "On Comprehension and 'Mental Representation.' " Pp. 8–23 in *Social Action and Artificial Intelligence,* edited by C. Nigel Gilbert and Christian Heath. Aldershot: Gower.

Dennett, Daniel C. 1978. *Brainstorms: Philosophical Essays on Mind and Psychology.* Montgomery, Vt.: Bradford.

Dreyfus, Hubert L. 1979. *What Computers Can't Do: The Limits of Artificial Intelligence,* rev. ed. New York: Harper Colophon.

Durkheim, Émile. 1973. "The Dualism of Human Nature and Its Social Conditions."

Pp. 149–63 in *Émile Durkheim on Morality and Society: Selected Writings,* edited by Robert N. Bellah. Chicago: University of Chicago Press.
Edelman, Gerald M. 1987. *Neural Darwinism.* New York: Basic.
———. 1988. *Topobiology: An Introduction to Molecular Embryology.* New York: Basic.
Edelman, Gerald M., and Vernon B. Mountcastle. 1978. *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function.* Cambridge, Mass.: MIT Press.
Elias, Norbert. 1978. *The Civilizing Process.* New York: Urizen.
Epstein, Cynthia. 1988. *Deceptive Distinctions: Sex, Gender, and the Social Order.* New Haven, Conn.: Yale University Press.
Feigenbaum, Edward A., and J. Feldman. 1963. *Computers and Thought.* New York: McGraw-Hill.
Feigenbaum, Edward A., and Pamela McCorduck. 1983. *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World.* Reading, Mass.: Addison-Wesley.
Fodor, Jerry. 1975. *The Language of Thought.* New York: Crowell.
———. 1981*a*. *Representations: Philosophical Essays on the Foundations of Cognitive Science.* Cambridge, Mass.: MIT Press.
———. 1981*b*. "Methodological Solipsism." Pp. 307–38 in *Mind Design,* edited by John Haugland. Montgomery, Vt.: Bradford.
Gardner, Howard. 1985. *The Mind's New Science: A History of the Cognitive Revolution.* New York: Basic.
Garfinkel, Harold. 1967. *Studies in Ethnomethodology.* Englewood Cliffs, N.J.: Prentice-Hall.
Gazzaniga, Michael S. 1985. *The Social Brain.* New York: Basic.
Gehlen, Arnold. 1988. *Man: His Nature and Place in the World,* translated by Clare McMillan and Karl Pillemer. New York: Columbia University Press.
Gerth, Hans, and C. Wright Mills. 1958. *From Max Weber: Essays in Sociology.* New York: Oxford University Press.
Gilbert, C. Nigel, and Christian Heath. 1985. "Introduction." Pp. 1–7 in *Social Action and Artificial Intelligence,* edited by C. Nigel Gilbert and Christian Heath. Aldershot: Gower.
———. 1986. "Text, Competence, and Logic: An Exercise." *Qualitative Sociology* 9 (Fall): 215–36.
Goffman, Erving. 1961. *Encounters: Two Studies in the Sociology of Interaction.* Indianapolis: Bobbs-Merrill.
———. 1974. *Frame Analysis.* Cambridge, Mass.: Harvard University Press.
Good, David. 1985. "Sociology and AI: The Lessons from Social Psychology." Pp. 82–103 in *Social Action and Artificial Intelligence,* edited by C. Nigel Gilbert and Christian Heath. Aldershot: Gower.
Greenberg, David. 1988. *The Construction of Homosexuality.* Chicago: University of Chicago Press.
Grossberg, Stephen. 1988. *Neural Networks and Natural Intelligence.* Cambridge, Mass.: MIT Press.
Heritage, John. 1984. *Garfinkel and Ethnomethodology.* Cambridge: Polity.
Hinton, Geoffrey E., and James A. Anderson, eds. 1981. *Parallel Models of Associative Memory.* Hillsdale, N.J.: Erlbaum.
Hobbs, Jerry R., and Robert C. Moore, eds. 1985. *Formal Theories of the Commonsense World.* Norwood, N.J.: Ablex.
Hofstader, Douglas R. 1979. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Basic.
Honneth, Axel, and Hans Joas. 1988. *Social Action and Human Nature.* Cambridge: Cambridge University Press.

Jahoda, Marie. 1986. "Artificial Intelligence: An Outsider's Perspective." *Science and Public Policy* 13 (December): 333–40.
Karelis, Charles. 1986. "Reflections on the Turing Test." *Journal for the Theory of Social Behaviour* 16 (July): 161–72.
Levine, Donald N. 1985. *The Flight from Ambiguity: Essays in Social and Cultural Theory*. Chicago: University of Chicago Press.
Luhmann, Niklas. 1982. *The Differentiation of Society*, translated by Stephen Holmes and Charles Larmore. New York: Columbia University Press.
———. 1989. *Ecological Communication*, translated by John Bednarz, Jr. Chicago: University of Chicago Press.
McClelland, J. L., D. E. Rumelhart, and G. F. Hinton. 1986. "A General Framework for Parallel Distributive Processing." Pp. 45–76 in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1: *Foundations*, edited by David D. Rumelhart, James L. McClelland, and the PDP Research Group. Cambridge, Mass.: MIT Press.
McCorduck, Pamela. 1981. *Machines Who Think*. San Francisco: W. H. Freeman.
McTear, Michael. 1985. "Breakdowns and Repair in Naturally Occurring Conversation and Human-Computer Dialogue." Pp. 104–23 in *Social Action and Artificial Intelligence*, edited by C. Nigel Gilbert and Christian Heath. Aldershot: Gower.
Maturana, Humberto R., and Francisco J. Varela. 1987. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: New Science Library.
Mead, Carver. 1988. *Analog VLSI and Neural Systems*. Reading, Mass.: Addison-Wesley.
Mead, George Herbert. 1934. *Mind, Self, and Society*. Chicago: University of Chicago Press.
Minsky, Marvin. 1981. "A Framework for Representing Knowledge." Pp. 95–128 in *Mind Design*, edited by John Haugland. Montgomery, Vt.: Bradford.
———. 1986. *The Society of Mind*. New York: Simon & Schuster.
Minsky, Marvin, and Seymour Papert. 1969. *Perceptrons*. Cambridge, Mass.: MIT Press.
Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, Mass.: Harvard University Press.
Murray, William R. 1988. *Automatic Program Debugging for Intelligent Tutoring Systems*. Los Altos, Calif.: Morgan Kaufmann.
Newell, A., and Herbert A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.
Norman, D. A. 1986. "Reflections on Cognition and Parallel Distributive Processing." Pp. 531–52 in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 2: *Psychological and Biological Models*, edited by David D. Rumelhart, James L. McClelland, and the PDP Research Group. Cambridge, Mass.: MIT Press.
Oldman, David, and Charles Drucker. 1985. "The Non-reducibility of Ethno-methods: Can People and Computers Form a Society?" Pp. 144–59 in *Social Action and Artificial Intelligence*, edited by C. Nigel Gilbert and Christian Heath. Aldershot: Gower.
Pagels, Heinz. 1988. *The Dreams of Reason*. New York: Simon & Schuster.
Papert, Seymour. 1988. "One AI or Many?" *Daedalus* 117:1–14.
Pateman, Trevor. 1985. "Using and Defending Cognitive Theory." Pp. 24–39 in *Social Action and Artificial Intelligence*, edited by C. Nigel Gilbert and Christian Heath. Aldershot: Gower.
Penrose, Roger. 1989. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. New York: Oxford University Press.
Pylyshyn, Zenon W. 1981. "Complexity and the Study of Artificial and Human Intelli-

gence." Pp. 67–94 in *Mind Design,* edited by John Haugland. Montgomery, Vt.: Bradford.
Reeke, George N., Jr., and Gerald M. Edelman. 1988. "Real Brains and Artificial Intelligence." *Daedalus* 117:143–73.
Rosenblatt, Frank. 1962. *Principles of Neurodynamics.* New York: Spartan.
Rosenfield, Israel. 1988. *The Invention of Memory: A New View of the Brain.* New York: Basic.
Rumelhart, D. E., P. Smolensky, J. L. McClelland, and G. E. Hinton. 1986. "Schematic and Sequential Thought Processes in PDP Models." Pp. 7–57 in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* vol. 2: *Psychological and Biological Models,* edited by David D. Rumelhart, James L. McClelland, and the PDP Research Group. Cambridge Mass.: MIT Press.
Schank, Roger C., and Robert P. Abelson. 1977. *Scripts, Plans, Goals, and Understanding.* Hillsdale, N.J.: Erlbaum.
Scheff, Thomas J. 1986. "Micro-linguistics and Social Structure: A Theory of Social Action." *Sociological Theory* 4 (Spring): 71–83.
Schegloff, Emmanuel, and Harvey Sacks. 1979. "Opening Up Closings." Pp. 233–64 in *Ethnomethodology: Selected Readings,* edited by Ray Turner. Baltimore: Penguin.
Schutz, Alfred. 1967. *The Phenomenology of the Social World.* Evanston, Ill.: Northwestern University Press.
Searle, John R. 1981. "Minds, Brains, and Programs." Pp. 282–306 in *Mind Designs,* edited by John Haugland. Montgomery, Vt.: Bradford.
Shanker, S. G. 1987. "Wittgenstein versus Turing on the Nature of Church's Thesis." *Notre Dame Journal of Formal Logic* 24, no. 4 (October): 615–49.
Shannon, Benny. 1989. "A Simple Comment regarding the Turing Test." *Journal for the Theory of Social Behaviour* 19 (July): 249–56.
Simon, Herbert A. 1969. *Sciences of the Artificial.* Cambridge, Mass.: MIT Press.
Sleeman, Derek. 1983. "Inferring Student Models for Intelligent Computer-aided Instruction." Pp. 483–506 in *Machine Learning: An Artificial Intelligence Approach,* edited by Ryszard S. Michaelski, Jaime G. Carbonell, and Tom M. Mitchell. Palo Alto, Calif.: Tioga.
Sleeman, Derek, and John Seely Brown. 1982. "Intelligent Tutoring Systems: An Overview." Pp. 1–11 in *Intelligent Tutoring Systems,* edited by Derek Sleeman and John Seely Brown. New York: Academic Press.
Teubner, Gunther. 1987. *Autopoietic Law: A New Approach to Law and Society.* New York: DeGruyter.
Turing, Alan. 1950. "Computing Machines and Intelligence." *Mind* 49: 433–60.
Turkle, Sherry. 1984. *The Second Self: Computers and the Human Spirit.* New York: Touchstone.
———. 1988. "Artificial Intelligence and Psychoanalysis: A New Alliance." *Daedalus* 117:241–68.
Waldrop, M. Mitchell. 1988*a*. "Toward a Unified Theory of Cognition." *Science* 241 (July 1): 27–29.
———. 1988*b*. "Soar: A Unified Theory of Cognition?" *Science* 241 (July 15): 296–98.
Weaver, Robert R. 1986. "Some Implications of the Emergence and Diffusion of Medical Expert Systems." *Qualitative Sociology* 9 (Fall): 237–57.
Weizenbaum, Joseph. 1976. *Computer Power and Human Reason.* San Francisco: W. H. Freeman.
Wenger, Etienne. 1987. *Artificial Intelligence and Tutoring Systems.* Los Altos, Calif.: Morgan Kaufmann.
Woolgar, Steve. 1985. "Why Not a Sociology of Machines? The Case of Sociology and Artificial Intelligence." *Sociology* 19 (November): 557–72.