The Journal of Mind and Behavior
Summer, 1981, Volume 2, Number 2
ISSN 0271-0137

Metatheoretical Issues in Cognitive Science

John A. Teske
The Pennsylvania State University
and
Roy D. Pea
Clark University

This paper is a "framing device" for cognitive science as a program, including critical discussions of issues at the heart of the field. The paper first lays out the basic assumptions and points to the roots of using a computer programming context for generating hypotheses about cognition and to the use of simulations of thinking in order to pursue theoretical work. Secondly, the paper addresses the parallel problems of level of analysis and the mind/body distinction, asserting the possibility of studying a symbol system apart from its embodiment. Third, the paper discusses the nescessity of internal models for dealing with intentional symbol systems of this sort. Fourth, the issue of knowledge representations is addressed, particularly with regard to the importance of knowledge structure-process integration. Finally, possible limits of the approach are examined, setting aside some misconceptions, and voicing some warnings.

The aim of this essay is to present a rough sketch of cognitive science, a multidisciplinary enterprise viewed primarily as an approach to cognition based on a "computational" metaphor. Since philosophical concerns with mind and cognition have in general been insulated from developments in cognitive science (with some notable exceptions; e.g., Goldman , 1978), our hope is to introduce cognitive science and provide some indications of the discipline's current directions. Our other aim is to address, from the perspective of cognitive science, some philosophical problems of longstanding significance which concern levels of explanation, the embodiment of mind, and the relationship between representation, intention, and knowledge use.

Cognitive science as a whole is a wide ranging and rapidly advancing field synthesizing aspects of artificial intelligence, cognitive psychology, philosophy, and linguistics. The unifying motif of the field is the attempt to study thinking scientifically. However, workers from these different fields have slightly different perspectives and concerns. The perspective

---

This research was supported in part by a National Science Foundation Fellowship to the first author. The first author also wishes to thank Rachel Joffe-Falmagne and Bernard Kaplan for advice and discussion contributing to this paper. Requests for reprints should be sent to John A. Teske, W-135, The Pennsylvania State University, The Capitol Campus, Middletown, PA. 17057, U.S.A.

taken here is rooted in cognitive psychology, particularly with respect to some of its metatheoretical concerns, although that perspective is not to be taken as a limitation. The boundaries of the various component fields of cognitive science are not only fuzzy, but often explicitly abrogated. As a result, this brief presentation may at times appear idiosyncratic, but the hope is that an important subset of the general philosophical and metatheoretical issues in the field can be raised.

In particular, we believe that the concept of an internal representation of events in the world is focal, and developments since Craik's (1943) seminal insights have demonstrated the central role this conception plays in explaining a wide range of cognitive phenomena. We will treat, in general, a set of problems in cognitive science which have also been of interest to philosophers. (For a treatment of a representative set of specific topics, see Johnson-Laird & Wason, 1977; for a brief introduction see Collins, 1977.)

This essay is comprised of five major sections. The first describes the guiding assumptions of the field. This section first points to the roots of using a programming context for generating hypotheses about cognition and to the use of simulation of thinking for theoretical work. The naiveté of ordinary notions of "machine" is indicated, and the ground is set for studying the human mind as a physical symbol system.

The second section is on the parallel problems of level of analysis and the mind/body distinction. Here we address the second basic assumption of cognitive science, the possibility of studying a system apart from its embodiment. Two issues are then raised and examined. The first of these has to do with the *possible* reducibility of symbol system to physical embodiment. The second concerns the related reducibility of psychological terms. Since both of these issues appear to involve similar irreducibilities, the section is concluded by addressing the appropriateness of using psychological terms to describe non-human systems.

Given that psychological terms are to some degree appropriate for describing the functioning of non-human systems, and that intention and subjectivity are important to that sort of functioning, the third section introduces the notion of *internal model*, a notion necessary for dealing with such systems. The constructive nature of these models or "knowledge representations" is described. The remainder of the section addresses issues related to the problem of selfhood, which is addressed via a special kind of internal model. The relationship between self and world models is examined, and a rough mechanical-purposive distinction introduced as part of these models. An attempt is made to clarify the problem of introspective knowledge via the models notion and some illustrations, and to provide a mechanism for addressing the difference between conscious/automatic, and the "implicit" processes of psychological functioning.

The fourth section confronts the issue of knowledge representation by asking how knowledge can be best represented for efficacious use. The distinction between the epistemological and heuristic sides of the problem is made; we then illustrate their interwoven character. This interweaving is so important for dealing with large bodies of knowledge that an intergration of knowledge structure and process is suggested. After examining the procedural-declarative controversy and pointing out the fundamental difference between these two views on the interaction/modularity issue, the notion of "frames" is introduced as a step toward an interactive-modular synthesis. We then discuss the symbol system as a whole and, in light of its self-modifiability, address the issues of generativity and development. Although the problem of large scale system changes is seen as central, and developmental considerations important, a "general principles" approach is down-played.

Having sketched some major overarching considerations of cognitive science, the final section turns to possible limits of the approach. First, some misconceptions are set aside: a distinction is made between simulation and reproduction, and a difference is indicated between knowlege by acquaintance and by description. Next, some current limitations are examined as to their likely permanence: the difficulties of dealing with tacit knowledge, and the tractability of some areas traditionally taboo to "mechanistic" treatment. Finally, some warnings are voiced: against a computer view of the "whole person," in terms of machines lacking human context, and against the "imperialism of instrumental reason." A last warning has to do with the potential problems of a *lack* of limits.

General Introduction to Cognitive Science

The "computational metaphor" has its roots at least as far back as Miller, Galanter, and Pribram (1960), who were among the first to draw explicit hypotheses about cognitive functioning from a computer programming context.[1] The approach grew in influence slowly at first, given the behavioristic leanings in much of psychology, but gathered momentum, christened as "cognitive science" in 1975 (Bobrow & Collins, 1975), a field which now has its own journal (*Cognitive Science* first appeared in 1977). The first Cognitive Science Conference was held in La Jolla, California, August 1979 (Norman, 1980), and the Sloan Foundation is

---

[1] Of course the history of thought about humans as computational machines extends far into intellectual history, with origins in the writings of La Mettrie (1747/1966), extended by Huxley (1874) and culminating in the science of cybernetics (Wiener, 1948) and Turing's (1950) influential essay on machine intelligence. But in its modern form, the computational metaphor is generally traced to the late 1950's (Newell & Simon, 1972).

supporting over a half-dozen programs of research in cognitive science and cognitive neuroscience (e.g., Posner, Pea & Volpe 1980) in America. (For a thorough account of the applications of this approach to cognitive psychology until recently, see Miller & Johnson-Laird, 1976).

What is meant by "computation" is not simply counting or any specifically mathematical operation, but rather any kind of effective procedures with which an intelligent achievement can be generated. Studying thinking by means of this approach generally involves, on a *theoretical* level, the attempt to simulate thought processes by programming a computer to perform them. Of course, theoretical formulations do not require actual implementation on a machine, but this is a powerful test of their coherence. Note, however, that the question of determining whether such formulations can account for how people actually think is separate and empirical. This is a methodological issue which will be set aside for our purposes here.

The idea of using technological artifacts as tools, theoretical and otherwise, to aid in studying ill-understood phenomena is by no means a new one.

> Just as the development of complex physical machines a few hundred years ago resulted in the elaboration of mechanistic models for every aspect of the human enviroment and humankind itself, our newfound ability to devise information-processing machines has inspired novel ways of thinking about what goes on when a person thinks, says, or does something. (Winograd, 1975a, p. 133)

However, our rather naive and stereotyped notions of "machines" hardly do justice of the idea and power of a "machine" as "an assembly of symbol-association and processing-controlling elements" controlled by the "networds of interlocking goal-formulating and means-ends evaluating processes" known as programs (Minksy, 1968, p. 10). The sophistication of computer programming has progressed at a rapid pace, so much so that programs are no longer adequately described as linear sequences of instructions; with the advent of heterarchical programming techniques, no longer structured as a hierarchical system of programs and subprograms, programs might be more appropriately viewed today as a set of "courts" which can call on each other for differenct areas of expertise. Given a machine with such a high degree of flexibility and sophistication, how might human thinking be studied?

Cognitive science is grounded in two basic assumptions. The first is that the human mind can be viewed as a *physical symbol system*; symbols being embodied as physical patterns which are components of larger structures consisting of tokens of symbols in physical relation. The system contains a collagation of these structures as well as processes operating on these structures to produce others, i.e., the system changes over time (Newell & Simon, 1976). The second assumption concerns the possibility that scientists can study the properties of such systems at a level of analysis abstracted from the details of their particular physical embodiment (Winograd, 1976). At this level we can no longer talk about physical structures; rather, the structure is more like what is meant by *mathematical* structure, physical only in the sense that it may be embodied in a set of equations written on paper. We can view psychological structures in the same sense as, for example, the flow-chart of a computer program.

The present paper is intended to provide a general metatheoretical outlook on cognitive science. However, the position being taken here is not uncontroversial, even within the ranks of cognitive science itself. Two of the fundamental presuppositions of this paper have been characterized by Newell (1980) as "obstacles to correct interpretation" of physical symbol systems: the "computer metaphor" and the "computer as tool kit." While Newell's opinion may be a little puritan, it deserves acknowledgement and reaction.

Newell's fear seems to be that considering the computer merely a metaphor or merely a tool kit for exploring the mind will weaken the theoretical claims of cognitive science. These fears are not unfounded. To the layperson, a metaphor is not something to be taken as seriously as a scientific theory. Cognitive science is indeed constructing "a scientific theory of mind, not different in its methodological characteristics from scientific theories in other sciences" (p. 179). However, Newells' fear is based on an over-literal conception of science which many laypeople share. Philosophers such as Black (1962) consider all theory and science to be fundamentally metaphorical. Other cognitive scientists (e.g. Lakoff, 1980) share this view, as do the present authors. Moreover, as psychological research is beginning to indicate (Werner & Kaplan, 1963; Verbrugge & McCarrell, 1977; Ortony, 1979) metaphorical comprehension may be more central than literal comprehension for understanding most phenomena. Therefore, "computational metaphor" seems to us a perfectly felicitous phrase for expressing our position; we do not deny that some of the similarities between human minds and physical symbol systems may be literal. We merely assert that the progress of research in cognitive science will consist in *discovering* such similarities, rather than *stipulating* their literality.

Concerning the computer as tool kit, Newell's fear is that the computer, as a tool for simulation, will be seen to have no different role and no greater significance for cognitive science than for any other science. Newell's argument is that it is the structure of the computer, as a general purpose tool, that (via its theoretical analysis) reveals the nature of symbolic systems. To this argument we would only add a proviso. The structure of the computer has revealed some important things about symbolic systems, but we would rather leave open the question of whether or not human beings are merely a subset of the class "physical symbol systems." However, we reserve a right to *also* consider the usefulness of the computer as a tool kit in the same sense that it is a tool kit for other sciences.

Levels of Analysis and the Mind/Body Problem

Within the cognitive science framework, intelligence is viewed as the ability to creatively manipulate symbols. Of particular interest within this framework is the attempt to understand the relationship between events in the world (including mental events) and the mental structures and processes involved in their production and/or comprehension. The setting and meaning of events are to be understood in terms of cognitive structures which are not peculiar to each particular kind of event. For example, procedural semantics (Johnson-Laird, 1976), that part of cognitive science concerned with language use, deals with the relation between linquistic objects or events and the symbol system, but not simply in terms of cognitive structures peculiar to language. The symbolic structures are *constructive*; that is, they are not prearranged arbitrary codes uniquely associated with each event in the world. "So long as we are required to map distinct stimulus types into elements of a prearranged set of arbitrary codes, we must assume that the organism already possesses an infinite set of such internal codes" (Pylysyn, 1973, p. 32). Rather, when events are experienced, these structures are built out of a finite set of elements and relations (the set of primitives) and/or by operations on other previously built structures. This construction is accomplished in interactive concert with an analysis of each event into its constituent parts and relations. The relationship of symbolic structure to event is then no longer arbitrary, but systematic, and a characterization of the symbolic structure in terms of the relations among its primitive elements and substructures reflects its formal relation to the event it represents. This relationship between symbolic structures and events is similar to the early Russellian notion of "structural realism" (cf. Halwes, 1974, Russell, 1926), in which the correspondence of knowledge to events in the world is seen in terms of structure only; in much the same way that vibrations in the air, the oscillations of eardrums, and musical notation may all embody the same melody. That melody *is* the structure. Nevertheless, we *can* also reference or indicate structures without embodying them: I can say for example, "that melody," thereby referring to the melody without embodying its structure. But that representation can only refer to the melody by virtue of the structured symbol system in which the phrase itself is embodied, and by which it is comprehened, i.e., that which allows access to an embodiment of the structure which is the melody (the language). These possibilities nonwithstanding, structural realism is probably more an ideal, a goal only; most of the time the structures are probably largely *imposed.*

It is extremely important to note that the "primitives" in which the symbolic structures may ultimately be grounded do not have to be simple physical features (cf. Charniak & Wilks, 1976, on the question of the adequacy of any single set of such primitives). Undeniably, theorists often work in a "top-down" manner, formally conceptualizing theoretical categories via primitives lacking context-free physical correlates. However, the key question is whether the ultimate goal is to relate physical patterns and context to those categories.

Two issues are involved here. The first is controversial even within cognitive science: Is the possibility and usefulness of studying a symbol system apart from its embodiment to be taken as a *fundamental assumption* of cognitive science, with reducibility either a logical or pragmatic impossibility, or as a *temporary directive,* with reducibility of the system to its embodiments the ultimate goal? The second issue is whether the structure of the physical symbol system itself is the same as the structure of physical events in the world, a question answered in the affirmative by the Wittgenstein of the *Tractatus* (1922). Both issues concern the reducibility of one level of analysis to another. The former addresses the reducibility of the symbol system to its embodiment. We will argue that such a reduction ignores the very structures of theoretical interest. The second issue is partially dependent on the first. If the symbol system was reducible to its embodiment, the symbol system would be identical in all essential aspects to physical events in the world, being little more than a convenient shorthand at best. However, since this possibility is dismissed, the second issue becomes a question of whether, and in what ways, further levels of analysis are necessary for understanding human thinking. We present the case that psychological terms are irreducible to physical terms. This being true and the structures of a physical symbol system being irreducible to any particular physical embodiment, we then explore the relationship between psychological terms and terms used to describe the structures and operations making up a physical symbol system.

Explanation and the Non-Reducibility of Symbol System to Embodiment

Although perhaps a bit presumptive, Johnson-Laird and Wason (1977) assert that: "An undisputed virtue of the computer is that it provides a metaphorical solution to the traditional dichotomy between the brain and the mind" (p. 7). The argument is that although the computer is an organized physical system, being a universal Turing Machine (capable of being programmed to realize any process that could naturally be called an effective procedure), what is crucial is not the *physical realization*, but the *logic* of operations. The brain is also an organized physical system; so "perhaps mental operations are merely its 'computations,' depending not so much on the physiology of nerve cells as on the logic of their operations. In order to understand human mentality, it may be more fruitful to discover the 'program' and 'plans' that underly it rather than their underlying physiological representation" (p. 7). Being physically embodied, a computer system cannot violate physical law. However, as Weizenbaum (1976) puts it, the game it plays is not determined by physical law, any more than the speed of movement or the tightness of a player's grip determines the outcome of a game of chess. Analogously, the physiological analysis of brain functioning is no more useful in understanding thinking than a detailed analysis of the electronic pulses in a computer would be for understanding the operation of a program: such analyses are simply on the wrong conceptual level. The model would be as hard to understand as the thing itself. For this reason, physical descriptions cannot be substituted for symbolic descriptions. Therefore, the fundamental theoretical goal of cognitive science becomes the understanding of intelligent processes independent of their particular physical realization, by understanding the structures and operations involved.

As Goldstein and Papert (1977) show (also see Boden, 1979), the cognitive science approach has a great deal of similarity with the genetic epistemology of Piaget—both in the interest in structures and operations as well as in the understanding that although intelligence would not be possible without some physical embodiment, the physical mechanisms involved are not the source of intelligence in the important structural sense. For Piaget the necessary biological hardware for intelligence is present at birth. Development is not construed as the emergence of new hardware, but the acquisition of knowledge by assimilation of and/or accommodation to events by general problem solving schemata. Some accommodations result in local improvements in particular schemata, others require much larger reorganization of the structures and operations involved; these latter reorganizations are what is referred to as transitions in developmental *stage*.

Winograd (1976) pushes the levels of analysis issue much further than the simple distinction between structure and embodiment. Using an automobile as his example of an object of explanation, he first makes the sort of distinction we have made:

> An automobile, like any physical device of similar size, operates according to the principles of Newtonian mechanics and classical thermodynamics. But this is clearly insufficient to explain how the automobile works. A physicist with a complete grasp of all the relevant theories may have no idea whatsoever about the behavior of an

> automobile. (p. 291)

The explanation a mechanic would provide is at a different level, not reducible to the physical but just as important, having to do with the form and function of the parts and subsystems, analyzable in a number of different ways (brake versus drive system, hydraulic versus mechanical versus electric—and their interactions). This is what we have been calling the structural level of explanation. Winograd also indicates two other distinct levels of analysis. The first is the "evolutionary," having to do with understanding the structure of a system by an examination of the sequence in which different parts are added and the system modified, e.g., understanding the role of pollution control devices in automobiles as a later addition to an otherwise structurally stable design. The second is the "social-economic," having to do with larger systems within which the particular system of interest evolved, e.g., understanding the evoluation of the current structure of the automobile via the development of systems of highways, the availability of fossil fuel, and human family structure (although some of these have functioned interactively with automobile evoluation).

Winograd takes a position against primary reliance on high level "general principles" in the kind of explanations we are interested in. With regard to the human language facility (and we presume this point to be extendable to other cognitive capacities), it is Winograd's assertion that:

> Much of the work in procedural semantics [cognitive science] is concerned with the design and structure of processes which take place within the human language user [symbol manipulator], assuming that most of the observable regularities are to be explained on that level rather than as consequences deducible from basic principles. (p. 292)

The point is that while simplicity and parsimony are always relevant considerations, if what we are studying is the result of complex interactions between multiple substructures, we ought to expect a little more complexity to our theories. For example, we would expect to be unsuccessful in analyzing the relationship of automobile speed to accelerator pressure and time if we attempted to explain the complex regularities directly on the basis of general principles without taking into account interactions between a number of different aspects of the system: engine response characteristics, transmission, engine vacuum, pollution control devices, and so forth. An averaging strategy simply begs the very questions of interest. There is no reason why it should not be just as important to the understanding of human mental functioning to map detailed cognitive mechanisms (i.e., the structure of the symbol system and its operations) as it is to the understanding of human biological functioning to map anatomy and physiology. Furthermore, human cognition may be as complex in relation to human biology as human biology is to the functioning of an automobile.

The immediate goal of cognitive science, then, is to attempt to divide human intellectual functioning into a set of structured systems and subsystems and operations upon those systems. How the divisions between systems will be articulated is ultimately an empirical question. Nevertheless, we can get a good idea of what sorts of divisions and interactions are possible by attempting to build computer simulations with similar structures to human cognition. We can only learn about these possibilities by keeping the entire system in mind. By constructing a theory of one component in isolation, we run the danger of forgetting that the form that theory takes has consequences upon the various other properties of the system. This is essentially the argument of Pylyshyn (1973) and Miller (1975) as to why Chomsky's generative-transformational grammar can have only dubious claims to "psychological reality"—because the theory was constructed without taking into account the larger system into which it must fit. Although a person's syntactic knowledge may have some independence from its use, it is unlikely that this is true for linguistic knowledge as a whole. The disciplines of sociolinguistics (Hymes, 1974) and developmental pragmatics (Ochs & Schieffelin 1979) certainly support this argument.

How does this explanation by interactive componential structure fit into larger contexts of explanation? Winogard (1967) provides some suggestions. First, the interactive structural components are probably what determine the psychologically interesting properties of the system. Turning to inherent limits or abstract capabilities is like trying to understand the operation of gasoline-driven, wheeled vehicles by pointing out that, being physical systems, they cannot exceed the speed of light. Second, while natural systems and conscious technological artifacts are often explained by the functions served by their particular structures (the approach primarily taken in cognitive science), there is also a complementary approach which concerns the range of alternative systems and the pressure operating on the system as a whole. Third, while a consideration of changes over time, evolution, development, and learning may be important to understanding the current state of a human symbol system, just as anatomy can be understood better by considering ontogeny and phylogeny, so far cognitive science has left this area relatively unexplored. Fourth is the issue of the comprehensability of a whole system. Traditional scientific methods of component isolation and experimental control are most useful where components of a larger system can be studied without much consideration to the larger systems of which they form a part—but it is likely that human cognition and language-use cannot be dealt with in such a fashion (Cole, Hood & McDermott, 1980). Given the complexity and interactive nature of the physical symbol systems which are probably necessary to model human intelligence, it is possible that we would need a theory of the simulation programs we build to get an overall understanding of them. While Pylyshyn (1974) believes the argument to be false he recognizes as cogent that "even if we could build powerful intelligent computers we might still be unable to satisfactorily understand the human mind because in a sense we might not understand the machines we had built" (p. 21).

Our answer to the first of the two issues raised above has hopefully been clarified. *The study of symbol systems apart from their embodiments is taken to be a rather fundamental assumption of cognitive science.* Clearly brains are not machines, but minds may be usefully viewed computationally. Furthermore, these "computational theories" are capable of a great deal of flexibility and seem to be quite adequate for dealing with much of the complexity in symbol systems. (According to Fodor, 1975, they may be all we have.) It is true that current theories are still less than adequate to account for the charm and variety of the rich pattern which is human mental life (for an extensive critical overview of computer simulation work, see Boden, 1977). Further, this fact is at least part of what drives the criticisms (Dreyfus, 1972; Weizenbaum, 1976) which we will address in a later section concerning the possible limits of the cognitive science approach. But one certainly does not want a theory which is more complicated or even *as* complicated as the phenomena to be explained, otherwise little if anything has been gained. The computational metaphor seems a powerful one.

> Why, then, should machine theories still evoke hostility in many psychologists? The answer seems to be that such theories raise the spectre of reductionism, the notion that mind (brain, life,...) can be explained (ultimately) by invoking no more than the laws of physics (whatever they may turn out to be). But this is a very strang objection indeed, for the whole drift of recent philosophy has been in the direction of arguing that machines (real machines!) are exactly the type of beast that *cannot* be explained in terms of the laws of physics, although they are, of course, not permitted to violate physical law. (Marshall, 1977, p. 484)

Hopefully, the argument here has served to fortify this claim and illustrate its place within cognitive science (for further references, see Polanyi, 1967; Putnam, 1967; and Fodor, 1975). Nevertheless, our second question remains: is the psychological functioning of a human being adequately characterized as such an information processing system?

System Analysis and the Non-Reducibility of Psychological Terms

What is the relationship between psychological terms and the terms used to describe the structures and operations making up a physical symbol system? There is no question but that workers in cognitive science use psychological terms in discussing what their systems do, describing them as interpreting, planning, setting up goals, constructing, selecting, and even commenting on their own activities. And writers such as Harré and Secord (1972) present a rather extensive argument as to why notions such as these are essential to a scientific study of human activity, an argument fortified by Harré (1974). But are such terms being used appropriately in describing the activities of physical symbol systems, or are they being used in some metaphorical or even misleading way?

Some discussion of psychological terms is necessary before this question can be answered. The claim is that many of our everyday psychological concepts are not illusory in the sense of being reducible to behavorial or physiological terms; psychological phenomena must be conceptualized as meaningful actions on the part of subjective agents rather than as causal processes in the natural world. The key is that we are dealing here with *intentionality*. The argument is that most of the sentences necessary to describe psychological phenomena have logical properties, called *intentional*, which are *not shared* by sentences sufficient to describe non-psychological phenomena (Chisholm, 1967).

Over a decade ago, Minsky asserted that the mentalistic terms used in describing much computer programming were *not* just superficial analogies and that "...at least some mentalistic descriptions of thought processes can be turned into specifications for the design of machines or programs" (1968, p. 1). General notions like interpretation, meaning, knowledge, representation, as well as more specific ones like purpose, plan, hypothesis, search, inference, and assumption have all been used in the design and implementation of physical symbol systems. Programs have been implemented which use purposive notions in debugging their own programs and writing better ones, and they can embody knowledge of the structure of the actions they, and others, "think" about. It is far from clear whether what the programs do could be described in other language; it appears that the more complexity and flexibility a program has, the more this is probably true (Dennett, 1978). It is also true that terms can be chosen to falsely represent a program as more intelligent (as when a program error message is described as an expression of anger), that psychological terms have a much richer significance when applied to humans (though the symbolic structures to which they refer may turn out to be identical), and that there are many psychological terms which cannot as yet be turned into specifications for the design of programs, although it is unclear whether this could ever be demonstrated in priciple. Nevertheless, the use of such terms points to the computational power many programs really have. Furthermore, while the psychological terms may not be used with the same depth or range as with human beings, and as such may be used only analogically, it is a scientific and open question as to how far that analogy might extend, or be extended and thereby increase our understanding of human psychology.[2]

Most importantly, the language we use in describing the operation of physical symbol systems, at even current levels of sophistication, *can* meet the criteria for intentionality (for a detailed review of these criteria, cf. Aune, 1967; Chisholm, 1967). Therefore, the language used to describe the operation of physical symbol systems can be sufficient to include psychological statements. We can describe a physical symbol system named "Vard," for example, and come up with statements fulfilling each of the criteria of intentionality. Vard can perform operations on structures which have no necessary correspondence to events or objects in the world, i.e., it can "think about" a "horse," a "mermaid," or a "future flood" (failure of existential generalization). Vard can contain in its data base a structure corresponding to one particular state of affairs in the world, without necessarily containing any other such structure, i.e., it can "believe" that Socrates was a philosopher and not that Freud lived in Vienna (nonextensional occurrence). Similarly, Vard can contain, as a substructure of some larger unit, a structure corresponding to some event or object in the world without necessarily also containing some other structure corresponding (or referring) to that same object or event, i.e., it can "believe" that "Dewey" is "Truman's successor" but is not "Eisenhower's" (referential opacity). Clearly, the correspondence of some structure with external events is irrelevant to the operations the system can perform on that structure, and the presence of a structure in the system requires no worldly correspondence (indeterminacy of embedded clause). Further, Vard is quite capable of attaching a substructure to any level of a hierarchical structure—e.g., to the hierarchy as a whole or, duplicated, to any number of subelements of that structure, i.e., it can "believe" that, in general, all x are y, or it can maintain separate "beliefs" about each x being a y (differential of insertion into quantified statements). Finally, while Vard may have the resources for transforming one structure into its logically equivalent form, this is not necessarily automatically done in any particular case, i.e., Vard can "think" (contain a structure formally equivalent to) "x or y" without "thinking" "not both not x and y" (specificity of logical form).

The computational approach suggests an alternative formulation to the Cartesian mind/body problem and the corresponding idealist-materialist impasse. Boden (1977) suggests that the question be formulated this way: "how is it possible for mental phenomena to be both irreducibly psychological and somehow wholly dependent on a mechanistic causal

---

[2] The use of analogies as explanatory vehicles is a complex question in its own right (e.g., Hempel, 1965; Hesse, 1963; Lakatos, 1970, 1972), but beyond our consideration here.

base (the brain and nervous system)" (p. 426). Minsky (1968) suggests that although the idealists were better prepared to handle the necessary abstract structures and interactions, they had no material ground for them, because of a too tightly limited stock of kinematic images. But with the growth of computer technology, a much more flexible set of images is available. It is now possible to distinguish, as Boden does, between different senses of "reductionist" or "mechanist," the usual antimaterialist invectives. In the traditional sense of "reductionist," psychological description and explanation were taken as shorthand for neurophysiological processes. But as we have seen, this is false: the notion of intentionality is essential to discussions of psychological reality, and is inexpressable in a vocabulary that lacks a distinction between subject and object. A more sophisticated sense of "reductionist" sees subjective phenomena as totally dependent on cerebral mechanisms, in much the same way as the symbol manipulating functions of a program are grounded in the detailed engineering of a particular computer. While the level of analysis may be limited to the program structure, it is important to remember that no program is actually *functional* without implementation on some particular physical system. Similarly, the usual "mechanist" approach abandons the subject-object distinction, viewing the explanation of behavior via meaning or purposive action to be a shorthand labelling at best, a complete mystification at worst. But a notion of "mechanist" based on the physical symbol system idea allows viewing subjective psychological phenomena as generated by bodily processes. (Such a view does not require subjective phenomena to be identical with bodily processes nor to be mere "effects" of bodily "causes"). The concept of an internal model or representation is central to understanding how subjective psychological notions can be embodied in an objective causal mechanism. To this issue we now turn.

Internal Models and Introspection

Intentionality or subjectivity can be attributed to human beings or to physical symbol systems because physical processes (bodily processes in the brain for human beings, electrical circuitry for computers) function as models, or representations, of actual and possible worlds for the entity concerned. These models or presentations are in no sense *copies* of actual events, but are based on a complex of interacting physical processes, which are describable at various levels of computation. However, description on an objective physiological or electronic level cannot express meaning or intentionality, because the role of the processes described in the functioning of the computer or the life of the individual is lost. Even in rather low level description, mechanisms are often described with regard to their function in an intentional world, e.g., the neurophysiological mechanism known as a "bug detector" in frogs. Without the use of intentional language it is not possible to speak of mental phenomena, which are essentially intentional. "Concentration merely on the physical mechanism of an intentional system ignores its intentionality, whether the system be natural or artificial in origin" (Boden, 1977, p. 430).

In effect no system, be it a human subject or a physical symbol system, can be said to "know" anything outside of itself without some kind of representation of this externality being present ("modeled") within the system. Cues, events, or objects in the world are in no sense simply "taken in" by the system. It is a truism that what is a cue for one system, program, person, or whatever, may only be "noise" for another. Without reference to the epistemological models used to interpret the imput, it would not be possible to identify the relationships between the cues and the structures built which link the representation to the domain being modeled. Cues, events, and objects are not simply detected but are constructed, and without an epistemological system actively imposing constructive procedures (schemes, frameworks) on outside input, there would be no cues, events, or objects; nor could there be persons, conceptions, or beliefs.

Feigl (1967) writes of three different sorts of problems related to the embodied-mind issue. The first is *sentience*: phenomenal experience is irreconcilable with a complete materialism. The inexplicability of sentience in physical terms is what leads to recognition of processes like perception as psychological rather than physical mysteries. This issue has been alluded to in our references to the constructive nature of arriving at knowledge of the world. The second problem is *sapience*: the existence of knowledge and meaning is inexplicable in physical terminology—and their existence has arisen with respect to the earlier levels of analysis. Such recognitions are also at the core of the second of the basic assumptions of cognitive science. The key here is that knowledge and meaning, being at different levels of discourse than that of physical elements, are not "locatable" in any physical sense at all; their embodiment may be physically locatable, but otherwise its "locatability" is only in terms of set membership and abstract structural relations. (Also see Osheron & Wasow, 1974/5). The third problem is *selfhood*. The usual scientific view of reality is that of a pluralistic physical universe containing only objects, but persons are singular entities (regardless of the degree to which their processes are unified)—logically the subjects of conceptual activity. It is this selfhood which is potentially most problematic for cognitive science, particularly with regard to its simulation efforts.

Are machines capable of truly self-directed activity? Even programs capable of modifying themselves are ultimately explicable in terms of their design—in accordance with the purposes of the human designers. So, whatever purpose, whatever subjectivity, a machine might be programmed to have can never really be intrinsic to the machine. In an equivalent sense, people pursuing ends other than their own are often compared to machines. Certainly the expression of intrinsic human needs are shaped and channeled by others: family, social world, culture. But those needs are intrinsic, genetically, to humans in a way that nothing ever can be for machines. However, this means little more than that computer intelligence will always be "artificial" in the sense that human beings will have final responsibility for it—regardless of the extent to which computers may ultimately be capable of evolving on their own. Nevertheless, computers can have all sorts of motives programmed into them, and as such the computer analogy can be a useful tool for increasing the understanding of the human mind. At this point no system is even near the complexity of interacting and competing (and often contradictory) motives and goals that characterize much of human thinking, and such complex phenomena as self-deception (Fingarette, 1967). That fact does not belie the usefulness of computer simulation, or the computer metaphor—the goal is to match aspects of human thinking and physical symbol system operations that *do* fit, and then find new similarities and differences as the analogy is developed.

In order to better convey how the models notion may help explicate the structure and processes involved in human knowledge and self-knowledge, we turn to Minsky (1968) who, in attempting to address the question of why introspection does not give clear answers with regard to the relationship between mind and body, clarifies a number of the issues involved. It will be seen that essentially subjective mental realities cannot only be acknowledged, but structurally elucidated.

When someone answers questions about an action without doing it, thinks about non-present objects or events, plans future activities, or comments on past events—knowledge of the world is being used. We can attribute the possibility of a person engaging in any of these processes to the possession by the person (P) of a model of the world (MW). Abstractly speaking, a distinction can be made between MW and the processes that operate, use, and build on it.

> However, one cannot really expect to find, in an intelligent machine, a clear separation between coding and knowledge structures, either anatomically or functionally, because (for example) some "knowledge" is likely to be used in the encoding and interpreting process. What is important for our purposes is the intuitive notion of a model, not the technical ability to delineate a model's boundaries. (Minsky, 1968, p. 426)

Broader questions about the nature of the world, activities requiring people to assess or operate on their *knowledge* of the world, and activities requiring self-knowledge can only be dealt with by a more complex structure. These sorts of activities cannot be dealt with via the simple MW. Instead, we must suppose that MW contains a model of the person (MP), that MP contains a model of MW (MMW), and that MMW in turn contains a model of MP (MMP). Although the MP would be sufficient to answer ordinary questions about oneself, and perform activities requiring a minimal sort of self-knowledge (e.g., reporting one's shoe size or monitoring one's position in space during a volleyball game), answering more general questions, making major self-evaluations, and perhaps even maintaining a self-concept or a view of oneself as a coherent entity, requires the MMP. Note that this idea of models, and models of models, does not necessarily lead to an infinite regress: higher level models can be very simplified and even vacuous. So if the system uses one of these models while engaging in some activity, given that the model can be simpler than the actual processes, or even distorted or incorrect in some cases (a system is subject to *at least* the same limits in modeling itself as in modeling the world)—then, just as with humans, mistakes and illusions are possible. Certainly it is impossible for any system, mechanical or human, to analyze everything it is doing at each step of an operation. To do so would preclude getting beyond one step. Therefore, abbreviated analyses are done with certain sorts of strategies, such that what is happening in the system may or may not be reflected veridically.

That it is often not a simple and straightforward matter to determine whether an activity requires MMW or MW is at least in part due to the question of how MMW and MW are related. The relations are not simply ones of physical substructure. Certainly the world "contains" the MW, so the MW must "contain" the MMW in the same sense. But if the MMW is only a *part* of the MW, then it clearly cannot be a *model* of the MW. On the other hand, MMW must include some means for referring to MW, just as the MW has its own correspondence with the world. The idea of a "part" is much more complicated for an operating computer than for a moving physical object. Through as little as a change in a single decision branch in the computer's program of instructions, the computer can behave like two very different machines in different circumstances.

> With interpretive operation ability, a program can use itself as its own model, and this can be repeated recursively to as many levels as desired, until the memory records of the state of the process get out of hand. With the possiblity of this sort of 'introspection', the boundaries between *parts, things,* and *models* become very hard to understand. (Minsky, 1968, p. 430)

We have run up against the levels of analysis issue again. Although we are trying to remain on a single level of *theory*, this time we find a number of levels in the system we are trying to understand. But this is as it should be; since we are human thinkers trying to better understand human thinking, the endeavor has a necessary reflexivity. Nevertheless it is clear that the notion "contained in" does not do justice to the relations between models, construed as programmable constructs.

In an attempt to elaborate the models notion to a developable level of sophistication, Minsky suggests that people generally have MWs divided into a couple of quasi-separate parts, concerned either with mechanical, geometrical, and physical aspects of the world or with goals, meanings, and social matters, and that the MP is also so divided. The claim is that we see and understand movements, events, and so forth in terms of mechanical force or human-like purpose, but rarely both. Phenomenologists (e.g., Schutz, 1967) have long urged this distinction as important to understanding human social activity. Cognitive scientists such as Wilks (1977) have made proposals for implementing this sort of dimorphism as an essential part of a language understanding system, as it represents a division present with a good deal of richness in both language and thought. The distinction between energetic/physical and informational/symbolic sorts of explanation can be characterized as part of the same dimorphism. The dimorphism is not a clear disjunction, but probably contains large areas of overlap and fuzziness: when psychological goals are blocked by mechanical obstacles, when we see emotional symbols in geometric arrangements or intentions in postural attitudes. The occasional inextricability of the two parts "reflects not so much any synthesis of the two kinds of explanation as it reflects the poverty of either model for description of complicated situations" (Minsky, 1968, p. 428). The differentiation of both this dimorphism and the different models which we spoke of above is one aspect of the growth of intelligence, a process which continues long beyond adolescence.

Given the above distinctions, it is easy to understand the common dualistic answer to questions about human nature. Such questions would be answered using MMP, which contains a representation of the roughly bipartite character of MP. A belief in a mind/body distinction is but a conventional way to express this distinction. However, since the separation is complicated and often indistinct, any further attempts to elaborate on a mind/body distinction are likely to be imperspicuous and unsatisfactory. Note that although some progress has been made in clarifying the complexity and multiplicity issues differentiating rational conduct and causal phenomena (by philosophers of psychology such as Harré, 1974, and Toulmin, 1974), there is probably no simple way of distinguishing them: Toulmin distinguishes seven stages of difference; Harré proposes an "enigmatic" class of episodes falling between what he calls "formal" episodes and biological episodes. Nevertheless, even given the likely necessity of a multichotomous view for a thorough and adequate understanding of the rational-causal continuum, a dichotomous

MP is probably sufficient to account for everyday thinking. To the extent that similar structures can be found in both parts of the dichotomy, a model of the system can eliminate the redundancy, but a reduction of one part to the other is far more inappropriate than a dichotomy i simplistic.

The conceptualization of human thinkers as symbol systems capable of self-modeling clarifies the role of introspection in the construction and evaluation of theories. Such a clarification is particularly important at the level of complexity involved in theorizing about human thinking: It is very difficult to rationalize away introspective data about goals, purposes, plans, meanings and social matters, largely because of our strong beliefs that thoughts very much like these introspections are what guide our activity (Miller, Galanter, & Pribram, 1960). Of course, this may well often be true. Effectively, however, it is only via *models* we have of our planning, intending, and so forth that we can produce introspective reports at all, and the match between these models and the processes themselves (or even our "scientific" models of these processes) is problematic. This is true even in cases of presumably rationally directed action where the model is involved in the process (to know *that* requires a higher order ability subject to the same pitfalls). There are no grounds for distinguishing, in introspective reports, between rationally directed action, and activity which only is *accounted for* as rationally directed in a certain way. Labov & Fanshel (1977) provide an example in discussing the inherent ambiguities of intonation patterns in speech acts; one can always deny the subversive intentions the listener imputes of them on the basis of such contours and claim their speech acts are *accounted for* in non-Machiavellian terms. People talking about themselves have exactly the same sorts of epistemological problems as do scientists in constructing theoretical models, except that in some cases a person's self-model may have an interactive and governing, regulatory, or monitoring role with regard to other processing structures of the system, much as if a scientist's theories influenced the object of study. Regardless of the role the self-model takes in the operation of a system, its validity has no guarantees. Even in cases where the role is to direct other parts of the system, nothing assures the quality of that direction. Clearly, then, introspective reports cannot straighforwardly validate, in any sense, theories we have about human thinking, but must be accounted for by a theoretical framework in much the same sense as we would account for any other publicly accessable activity, behavior, or report. The problem is that, with regard to our own thinking, it is fairly easy to believe that such thinking operates in a certain way, simply by assimilation to whatever theory or model we have at our disposal, be it common sense or scientific. We simply do not have a coherent body of theory capable of accounting for introspective reports. The models notion is a start, but we still know little about what aspects of which model at what level we can call conscious and which not—and under what circumstances different aspects of our knowledge are accessable.

Some illustrations may be useful to further clarify the "theory of introspection" problem. Let us draw on the computer as metaphor. Efficient storage of information in a data area is often messy, intertwined, and more and more convoluted the larger the data base; yet it is often important to be able to represent this information to a user of the computer in a fairly straightforward, orderly way. This concern is called *data base management* and may involve abbreviating, truncating, or simply ignoring substantial portions of the otherwise available information. Much the same reduction may occur in our representations of our thoughts and activities to ourselves. What is available to introspection (or consciousness) may be a similarly "managed" portion of our own knowledge base.

There is yet another useful distinction in the actual *running* of computer programs between "interpreting" and "compiling." Practically all programming is done in higher order languages that must be translated into the operating language of the particular machine on which the program is implemented. Both "interpreting" and "compiling" involve *translating* a program into machine language; the actual *execution* is a separate process just as understanding and obeying a command are separate for people (Johnson-Laird, 1976). However, "interpreting" involves individual translation for each step (or expression) of a program, effectively allowing a capacity for examining each step before executing it, whereas "compiling" involves translating the entire program such that, when executed, each step in the program follows automatically. In terms of our metaphor, we can see conscious activity as an interpretive process, and automatic or overlearned activity as the execution of a compiled program. Although the interpret/compile metaphor has been superceded by more precise accounts of system architecture, it is nonetheless useful for the general understanding being aimed for here. The metaphor is appropriate in terms of speed of operation as well as style of operation: a compiled program runs appreciably faster when executed than an interpreted program which, as we saw, is only executed one step at a time. Perhaps this metaphor provides a better conceptual handle on differences between activities like, say, tennis and chess. In tennis, because efficiency and speed of movement are more important, it is valuable to arrive at a state in which one's "tennis program" is compiled and running. The clumsiness of the novice is the interpretive use of a program. Interpretive use of a program, though, may be very important to debugging it, in this case, to improving one's game. In chess, on the other hand, examining each move individually and being aware of the possible developments and tactics in an overall strategy is what is important, although a debugging process is certainly important here, too; and certain kinds of substrategies or movements may be deployed semi-automatically. Furthermore, something like a compile-execute distinction might be a useful theoretical tool for approaching a problem like "internalization." Ultimately, the interpret-compile-execute sort of distinction may make even more intimate mind-body ties more approachable, like the psychological aspects of recovery from illness: if we see the usual operation of the body's healing mechanisms as a compiled-program, perhaps the positive infuence of a patient's active involvement is much like an interpretive debugging of small parts of the program. (For a theoretically and empirically elaborate account of a very similar distinction see Shiffrin & Schnieder, 1977, and Schneider & Shiffrin, 1977, on controlled versus automatic processing.)

Applying our metaphor more directly to language and social activity, it is probably the case that many of the processes by which we understand utterances or actions are largely unconscious or work via compiled programs. Many of these processes of understanding may also or simply be inadequately modeled within the human symbol system. Note that these are two separate issues.

The first issue has to do with the interpretation/compilation difference. Interpretation and compilation are two difference modes of operation within the system: interpretation is the slow, step by step execution so important for debugging (equatable with conscious thought); compilation is quick and automatic (or unconscious). This does not mean the latter lacks bugs and could not benefit from being run interpretively and debugged. (Debugging via compiler, as any introductory computer student knows, it a very tedious and difficult task—becoming virtually inhuman in difficulty with a program of any complexity.) This process of interpretation and debugging is what we are asked to do in psychotherapy, or anytime someone says "Watch what you are doing!" This issue of operational mode is distinct from the second issue.

The second issue has to do with the level and the particular adequacy of the required internal model *of* those very processes of understanding of which one may be attempting to speak. Those internal models represent what knowledge we have, and what knowledge we are capable of becoming aware of, but not what knowledge we will be aware of at any particular time. Ample evidence indicates that we often tell more than we can know about our own internal processes (Nisbett & Wilson, 1977). We now have the beginnings of a fuller understanding of both why and how introspection is often so misleading: We can only report about our internal processes to the extent that we have internal models of those processes. To expect anything else, to expect any kind of direct introspective access, would be epistemologically naive. We do often report about internal processes, just as we report our beliefs about external events. The question of the validity of such reports is equally open in either case. Verbal reports of internal processes, like any other verbalizations, are data which require explanation (cf. Ericsson & Simon, 1980, for an attempt to explicate some of the mechanisms by which such reports are generated).

As an example, consider reading a difficult passage. We only seem to have worthwhile introspections about our process of understanding when we hit trouble, when it takes awhile to figure something out (debug); when the model of our internal processes turns out to be useful in altering those processes. Even then there are things that we often cannot be aware of, given the character of our self-models. Most of the time we have no real awareness at all of how we went about understanding some stretch of discourse. It is during the times of trouble that change in usage, learning, adjustment take place. This observation is basic to both *structural* (Piaget, 1971) accounts of the disequilibria mechanisms of developmental change and *dialectical* accounts of developmental mechanisms (Riegel, 1976). Perhaps our understanding processes are different in the conscious mode, but if the same program can be interpreted or compiled might we not just be using the same processes in different modes? So, as a working hypothesis, we assume that some sort of inference process is involved in language comprehension.

> The distinction between conscious deductions and everyday inference is probably a reflection of a more general contrast that can be drawn between explicit and implicit inferences. The inferences that underly problem solving are often slow, voluntary, and at the forefront of awareness: they are explicit. The inferences that underlie the ordinary processes of perception and comprehension are rapid, involutary, and outside conscious awareness: they are implicit. (Johnson-Laird & Wason, 1977, p. 5)

In much the same sense that people use implicit inferences (the rules of which may differ from standard logic, cf. Braine, 1977; Osherson, 1977) in thinking and understanding language, implicit rules are probably involved in most realms of meaningful human activity. However, these implicit rules are certainly a lot more flexible than rigid rules of etiquette, and do not have the explicit institutional sanctions of laws. In facts, many of these implicit rules may be modified as they are used. The notion of a rule is a problematic one (cf. Collett, 1974; Harré, 1974; Pylyshyn, 1973; Toulmin, 1974). Part of the problem is that the term "rule" is often thought to imply some awareness of the rule on the part of the person following or conforming to it. Speaking of rules as "implicit" (Chomsky, 1965) seems to entail the converse. However, simply "acting in accord" with a rule reduces "rule" to something like "physiological law," or the level of a fictitious postulate for explaining behavior. On the other hand, people can become aware of rule-breaking, even though not constantly being aware of the rules broken, and in no sense being "caused" to necessarily obey them. Again, part of the answer seems to be that we are capable of having knowledge of which we are not always or even often aware. We also have first order models of the world which have aspects not fully represented in the second order models which would be used to report on them. Nonetheless these first order models might be capable, perhaps under some circumstances and not others, of governing or being used to direct our actions. Lacking the requisite second order knowledge, we are often incapable of reporting on this governance. A whole literature in a discipline designated as "ethnomethodology" is devoted primarily to uncovering the implicit rules an procedures which guide our daily lives but of which we are generally unaware (Garfinkel, 1967; Douglas, 1970; Filmer, et al, 1972; Cicourel, 1974; Turner, 1974; Mehan & Wood, 1975).

Cognitive scientists generally prefer the term "procedure" over "rule" because it does not have the strong connotation of requiring consciousness for it to be followed. We will return to the idea of "procedure" when we address the knowledge-representation issue in more depth. In any case, it is important to keep in mind that we are more often than not quite unaware of the subtleties and computational intricacies involved in even simple activities and achievements.

We now return to the question of whether or not machines are capable of self-direction, and to the larger issue of free will. Minsky (1968) points out the difficulty of determining how free will differs from caprice and suggests that the supposition of free will is a primitive defense mechanism resisting the emotionally unacceptable recognition of "compulsion" or "control." He proposes that we assume the MMW rules in some places, but in other places there is an element of randomness. Any time the unpleasantness of this proposal results in suggesting a third component, such as "will," we have only to recognize that component to be structureless and empty. Where any regularity or non-randomness is seen, procedures, rules of the MMW, are in operation. Free will simply means self-direction.

> When intelligent machines are constructed, we should not be surprised to find them as confused and as stubborn as men in their convictions about mind-matter, consciousness, free will, and the like. For all such questions are pointed at explaining the complicated interactions between parts of the self-model. A man's or a machine's strength of conviction about such things tells us nothing about the man or about his model of himself. (Minsky, 1968, p. 431)

The Representation and Use of Knowledge

For workers in cognitive science, as well as for researchers like Piaget, the processes of intelligence are determined by the knowledge held by individual subjects (Goldstein & Papert, 1977). Memory is the foundation of intellectual functioning. Here, intelligence is not seen as some separate crystalline element of the mind but as existing within a highly organized, intricate body of knowledge structures, in other words, the internal models we have been speaking of and the processes operating on and through them. Pylyshyn (1973) distinguishes two aspects of the problem of understanding intelligence in general, which correspond roughly to the distinction between models and processes: (1) what people know and how that is represented in the mind (the epistemological problem), and (2) how this knowledge is used despite various cognitive limitations (the heuristic problem). The epistemological aspect of the problem is the more difficult since, as we have seen, we have no direct conscious access to most of the knowledge which we actually use in intelligent activity; we cannot easily study the kind of knowledge which may underlie a large range of cognitive skills. Admittedly, studies in cognitive science to date have, by this characterization, focused almost entirely on the heuristic side of the problem, being concerned with the use of knowledge in rather particular and limited domains. In fact, given a clear-cut epistemological-heuristic distinction, computer simulation studies are quite far from and may never be adequate to address the epistemological side of the question. This is a possibility, not a necessity, and the possibility is grounded on a distinction that may never be clear—people's knowledge and how that knowledge is to be represented is very dependent on the *use* of that knowledge. Part of what people know is a repertoire of different ways to use and expand the knowledge they have, how to stretch it to make do in some situations, how to combine it and reorganize it, to recognize what its limitations are, and so forth. In part, this repertoire is captured by what Bateson (1972) describes as deutero-learning or learning *how* to learn. Such second order knowledge may be represented in ways that make its use more efficient and flexible. Given the massive amount of knowledge every human being has, if such knowledge were not represented in a readily usable form, humans would spend the majority of their time just figuring out what to do. Thus, epistemological and heuristic aspects to the problem of intelligence may be inextricably intertwined—in a way that makes an integrated procedural-heuistic approach to epistemology quite viable.

Certainly, for an individual to act in the world and continue to do so, knowledge must be retained over time. Otherwise, the individual would have little inkling of what to do when confronted with a situation, whether new or old. Some new situations seem less strange than others, and we face most situations as if they were variations on what is familiar. To the extent that we can consider and answer questions about nonpresent objects or events, we need some internal representation of them.

These kinds of abilities require memory. The question generally raised is how this memory is applied, changed, enriched, and integrated, either for shorter periods of understanding a single utterance or action, story or conversation, or for longer periods in which a person attempts to make sense of a period of his or her life or any large collection of events. Being able to utilize something learned in the past, or to make sense out of something occuring in the present may not be *explained* simply by reference to memory. Nevertheless, as long as the action accomplished or the situation understood can be called meaningful, memory is necessary. Memory is necessary because it *enables* us to relate present situations or events to situations or events occuring in our past. Memory requires *something* that remains over time with the individual. However, it is not the *thing* that is of interest; the thing, as we have argued before, is only the *embodiment* of memory. Memory is not a physical object and therefore is not locatable in any physical sense. Thoughts and memories are not "in the head" but "in" the mind, *in* denoting set inclusion rather than location.

Even so, how could it ever be anything but futile to get a grasp on the entirety of a person's knowledge. Presumably, such a grasp would be necessary for a *full* understanding of that person's behavior or action. Even to fully understand a person's comprehension of a single sentence could require research of vast proportion.

> It is indeed impossible to lay down all of the knowledge that a speaker of a language might bring to bear on producing or understanding a sentence, or to determine all of the factors which influence an individual piece of behavior. But it is equally impossible to specify the form and location of all the particles in any physical segment of the universe or to determine all of the forces acting on them. This does not preclude a science of physics. (Winograd, 1976, p. 265)

Similarly, cognitive science does not attempt to arrive at a complete map of human knowledge or cognitive structure, but rather is interested in exploring the operations *available* for deployment in tasks of reasoning, perceiving, understanding, knowing, and remembering. By viewing human intelligence as a symbol system, cognitive science hopes to construct a theory with some degree of generality, allowing representation of entities not all physical or real, using a wide range of symbol manipulating processes such that logic and mathematics might be seen as an abstraction from a subset, providing means for building onto the system via experience, and allowing most of knowledge to be "tacit" (cf. Polanyi, 1964) or unavailable to conscious self-examination.

The major question thus becomes how to represent large amounts of knowledge so that it can still be effectively used. This means that: (1) knowledge must be represented in a usable form, and (2) competence in different domains of knowledge must be able to interact (avoiding isolated components for greater efficiency) (cf. Goldstein & Papert, 1977). The problem is the virtual impossibility of clearly separating the representation of knowledge from the uses to which it is put (Norman, 1973). Even in answering a relatively straightforward question, we first investigate its sensibility, its referents, and the level of response expected (depending in turn on the questioner's motivation and knowledge). The commonality here with Austin and Wittgenstein's programs of ordinary language analysis is striking. In much the same sense "it is neither possible nor desirable to separate memory for action from the plans necessary to execute those actions"—the difference is more along the lines of the compile-execute distinction, i.e., whether an instruction is examined or activated. (Note that both examination and activation are procedures, and that even examination must involve some sort of representation of what is involved in activation and its possible results.) Newell (1972) notes that this and a number of other distinctions rely on a distinction between what *is*—the static, permanent, and object like (structures)—and what *happens*—the dynamic, transient, and transformation-like (processes). Among these are the distinctions between material and activity, data and program, language and interpreter, grammar and recognizer, and competence-performance. He suggests that the way out of most of the problems is to treat structure-process distinctions as relative to the level of analysis or operation of a system. A procedure on one level of a system may thus be analyzed as data on another. It is this idea of multiple levels of integrating knowledge-process and knowledge-structure that forms the basis for Bobrow and Winograd's KRL (1977), a formal computer language designed explicitly for representing knowledge.

The Procedural-Declarative Controversy

Pylyshyn (1973) points out that cognitive *processes* are often described by using sets of rules or procedures. However, turning to the epistemological side of the problem, we find that cognitive *representations* are not typically described this way. The "models of the world" accessed are usually likened to static structures. This is the usual distinction: cognitive processes as operations (characterized by rules or porcedures) on a static data base (characterized by declarative statements). Winograd (1975b) calls this the declarativist position, in opposition to the proceduralist position which contends that our knowledge is primarily a "knowing how" rather than a "knowing that" and is most appropriately seen in terms of programs for operating on the world (rather than a set of facts *about* the world). For the declarativist, knowledge processes are rather general principles capable of being applied to a wide range of particular domains. As we have seen, there is no way of deciding which position is correct because any piece of knowledge can be seen as a procedure or as data. On the one hand we can view the compiler or interpreter as the only "program" with everything else as data; on the other hand we can view even factual statements as programs which output truth values. It is of interest that philosophers of language such as Austin (1962) and Searle (1969), in choosing to view declarative statements as activities (speech acts), lean heavily toward the proceduralist approach. The crux of the issue is really which position stands to offer the greatest epistemological advantage. How is knowledge best represented?

Both declaratives and procedures have their special advantages. A declarative representation: (1) allows a "fact" to be used in a number of different ways without multiple representation, simply by applying general (e.g., deductive) principles; (2) allows easy addition to a knowledge base—facts are independent—in a procedural system, minor changes in a program may have many effects elsewhere in the system, so changing or adding is much more complicated; (3) allows efficient use of natural language, which is primarily declarative. A procedural representation: (1) allows much more straightforward representation of knowledge about actions—in the form of programs for accomplishing those actions as well as more efficient use of this knowledge (since it can indicate what knowledge is to be used rather than having to check through a whole list of statements for their truth values); (2) allows efficient expression of second order knowledge, for example, knowledge concerning the complexity and difficulty of some procedure, distinct from notions of truth and probability, which would be extremely cumbersome to express declaratively; and (3) allows recognition that much of our knowledge about a particular domain is heuristic, specifying particular strategies to be tried under particular conditions and allowing easy integration of such knowledge.

Clearly some kinds of knowledge are going to be best represented procedurally, some declaratively, but a straightforward synthesis is precluded by a fundamental difference between the two approaches on the problem of complexity and the issue of modularity/interaction (cf. Winograd, 1975b, pp. 191-192). The declarativists take symbolic mathematics as metaphor: axioms and rules of inference are entirely distinct, and axioms are logically independent—all changes are additive. The proceduralists take programming as their metaphor: interaction is primary, the programmer controls what is used and when subroutines have side effects on other pieces of knowledge and facts, the processes are interwoven—changes are not simple additions but *debuggings* or modifications of existing structures. The advantages offered by each are a result of these differing perspectives. Winograd (1975b) suggests one move toward a synthesis: building a declarative data structure hierarchically and then attaching, to various levels, procedures which could guide in selecting general procedures to use in a particular case. This approach attacks the modularity issue by interposing another layer of structure.

> Most of what the system knows is included in both a modular and an integrated form. The procedures for learning and debugging continually use general knowledge or programming to take the individual facts and combine them into the specific integrated procedures which do most of the system's deductions. Faced with a problem for which no specific methods are available, or the ones available do not seem to work, the system uses the specific facts with more general methods. There is no sharp division between specific and general methods, since there is an entire hierarchy of methods attached at all levels of the generalization hierarchy for the concepts in the problem domain. The most critical problem for the representation is to make it possible for this shifting between levels of knowledge to occur smoothly, without demanding that the programmer anticipate the particular interactions. (p. 209)

Frames and Frame Systems

To illustrate the sort of interactive modules spoken of here, it is necessary to flesh out some of the intuitions represented by what have been variously called "schemata" (Bobrow & Norman, 1975), "frames" (Minsky, 1975), "beta-structures" (Moore & Newell, 1974), or "networks"(Norman, Rumelhart, & LNR, 1975). Historically, these structures have been called "schemata" by psychologists. However, we will retain the current terminology of "frames." Nevertheless, since the notion of a schema is an epistemologically central concern to writers such as Barlett (1932) and Piaget (1971), among many others, moving toward a more detailed and formally explicit treatment of schema-like structures would seem to be an important endeavor given the often vague and intuitive usage of the notion by these and other writers. Note, however, that we will be merely reviewing some of the broader generalizations about frames rather than discussing the technical difficulties of really using such notions to construct crisp psychological theories.

Bobrow and Norman (1975) propose a system of active schemata put into operation either in the service of higher order purposes and expectations or to account for input data. These schemata are also capable of accessing each other via context-dependent descriptions. The goal of Bobrow and Norman was "to specify a memory structure that allows one schema retrieved from memory to suggest others that should also be retrieved, and that is so constituted that it yields analogical and metaphorical retrieval as a fundamental mode of its operation" (p. 113). Context-dependent descriptions only have to be precise enough to specify a referent relative to some context. The descriptions contain other descriptions as well as some constants or primitives which unambiguously retrieve a single schema. These are absolute reference points necessary for initial buildup of a memory structure. They would presumably consist, in the human, in various innate capacities and primitive sensory-motor schemata à la Piaget from which higher order capacities and descriptions could be built. Reduction of all descriptions to primitive conceptual elements (cf. Shank, 1972, 1975) would not be required: conflicts between levels of description would presumably be rare given the formation of new schema by specialization and alteration of older schema. Context-dependent reference provides six important features for the system: (1) *Efficiency*: only information necessary for specifying the referent in context is needed. (2) *Generalizability*: context can be changed without changing description ("the schema will refer to memory structures which have the same relative properties with respect to the new context as the originally intended memory structures had to the original context," Bobrow and Norman, 1975, p. 136); analogy and metaphor require no special operations. (3) *Approximation*: close matches can be retrieved and focus attention on context mismatch. (4) *Reliability*: the system always returns something, whether close match, analogy, or metaphor. (5) *Currency*: newly acquired information will be retrievable by an old description—allowing easy updating. (6) *Partial knowledge*: even if the description cannot specify a unique referent, the system continues searching, knows what it is searching for, and a referent can eventually (given more information) be found. Ambiguity is represented by the alternatives possible at one point.

In general, schemes or frames are *modular* and more or less self-contained, each containing its own procedures for operation in its specialty, the system as a whole being organized heterarchically. In any situation, a number of these frames may be operating concurrently, accessing each other and drawing on each other's results. This is essentially what is called *heterarachical organization* (termed "coalition" by Shaw and McIntyre, 1974), meaning that control of processing is passed back and forth between teams of "experts" each responsible for one particular aspect of a problem (see Boden, 1977, for a more thorough account). Of course, a totally independent sort of heterarachy runs the risk of becoming chaotic (though this may be closer to human thinking than we would like to believe), but a number of solutions are available, e.g., the proposal of a central "blackboard" for which every frame gets its information and on which it writes its results. We will return to this issue of central control shortly. Winograd's (1972) natural language understander, although limited to a very small domain, provides an early example of a computer system operating on a heterarchical basis. Goldstein and Papert's (1977) theoretical discussion of a frame system for understanding language provides some clarification. Basically, the set of frames for understanding a sentence consist of "framekeepers" for each important word; understanding a sentence involves the interaction of these as each tries to fulfill certain needs, e.g., trying to find the agent and instrument of some action. Deductions are made and expectations created on the basis of the frame's content. Larger context is provided by frames such as Shank and Abelson's (1977) "scripts" in which some particular activity is understood by mapping onto the actions specified by the frame (such that we can understand without being told, that eating in a restaurant involves ordering and paying for food). Thematic frames for even larger pieces of discourse might be provided by something like Rumelhart's (1975) story grammar, which represents the knowledge that links sentences into a coherent whole. (More recent empirical work on the role of story grammars in memory is decribed in Mandler and Johnson, 1977; Thorndyke, 1977.) Levin and Moore (1977) propose a similarly frame-like system for handling what they call "dialogue games." Each frame contains expected events and default conditions (i.e., they are assumed to have occurred unless otherwise specified) to fill in gaps and prepare the individual for what is to come. When expectations are violated, the system is capable of either trying another frame, altering or correcting the frame, or entering the particular conflict as an exception. With reference to written text, but applicable to other situations, events, and so forth,

> ...the process is not one of literally understanding the text, but instead is more one in which the text triggers rich, highly structured knowledge packets that supplement the literal content, provide expectations for the remainder of the story, and place the story in a context of related knowledge. Understanding is seen as essentially a process of evoking and then debugging existing knowledge packets. (Goldstein and Papert, 1977, p. 96)

In what sense do frames represent a synthesis of procedural and declarative knowledge? Frames are basically formalizable in terms of the nodes and labeled relations of semantic memory networks (as in Anderson and Bower's HAM, 1973 – further developed in Anderson, 1976, and in the formulations of Norman, Rumelhart, and the LNR group, 1975; see also some interesting suggestions of Scragg, 1975, on formulating frames as planes within a semantic space). These networks were originally designed as formal equivalents to a propositional or declarative base with higher level structure fixed, leaving slots for particular instantiations at lower levels. With the introduction of frame or schema-like notions, we begin to see the net not as uniform but as a highly structured set of context. These contexts can include default values of some parameters and ranges of expected alternatives, giving a "common-sense" knowledge of contexts. Finally procedural knowledge can be attached or integrated directly, using operations like that of a frame-keeper which contains procedures for filling slots in the frame. Additionally, this approach provides an alternative for coping with paradox and contradiction. The traditional formal approach is to try to find an axiomatization which does not lead to these problems. The alternative, certainly more realistic (at least intuitively with regard to human thinking), is to use principles which might lead to such problems but to refrain from dangerous processing directions, using "warning processes" much like sentries. Such an approach would be ridiculously ad hoc in a purely propositional formulation. "But if the knowledge system is represented in the first place as a set of interactive processes, the sentry is merely one more process" (Goldstein & Papert, 1977, p. 98). The epistemological advantage is in the introduction of multiple active agents: an item of knowledge, a concept, or a schema is no longer a passive object requiring manipulation but has active properties of its own.

The basic approach is thus to represent knowledge in highly structured but active packets having sets of operations to perform under various circumstances and capable of mutally accessing each other. One of the fundamental assumptions is that fewer situations than we think are really as novel as we think, and they are handled via these frames rather than by deduction from more general principles. Framekeepers monitor the possibility of various likely consequences of particular events or activities and are capable of inserting reasons, motives, purposes, and other aspects of events or activities which are not explicitly stated or obvious on the surface. These default insertions may sometimes be inappropriate, in which case alternatives must be suggested or corrections made (which the system is capable of doing). In general, however, our basic ability to understand, comprehend, and intelligently act in a variety of situations is dependent on a large bulk of background knowledge brought to those situations.

Kuipers (1975) provides a good summary of the important properties of frames. These are: (1) *Description*: A frame is a structure which provides a description which can be maintained over small changes in the observed. These frames are learned or elaborated from other frames—they organize observational information that would otherwise be incoherent and integrate it with other non-observational information. Elements of a larger frame can be expanded into frames in their own right. (2) *Instantiation*: A frame describes an object, situation, or event by gradually substituting actually observed values, for stereotype or default values or by selecting out of a limited set of alternative values. (3) *Prediction*: It is the frame's prediction that guides the instantiation, providing a sketch of what to look for. Such a sketch is left as a default for what cannot be or simply has not yet been observed. (Thus, a result can be provided at various stages of an analysis in progress.) Some default values are based on idiosyncratic experience and are easily replaceable; others are definitive or more essential to the frame. (4) *Justification*: Along the same lines, confidence may vary with other features, some being default, some only likely choices among alternatives, others being fairly clear observations. (5) *Variation*: Limits of variation and the possible dimensions of variation are specified by the frame. If a number of features too closely tax the limits, a correction or an alternative frame may be required. (6) *Correction*: If permissible limits are taxed or exceeded, the frame can access an appropriate replacement. In case of replacement, most of the information gathered (the description prior to finding anomaly) is still usable. Since the character or identity of an object, situation, or event is not known prior to framing and can only be known through some frame, this sort of correction is not rare, though the selection of any particular frame will be guided by accession from other frames or by a larger overarching frame. (7) *Perturbation*: Frames contain procedures for making small adjustments in descriptions when there are minor changes in the system or the external situation. (8) *Transformation*: With larger changes, other procedures propose new frames.

The Human Symbol System

At this point it is appropriate to make some comments about the human symbol system as a whole, within which scheme/frame structures might operate. Bobrow and Norman (1975) suggest that some kind of low-level sensory processing (perceptual-frame mediated) occurs, which passes results on to the ubiquitous high level conceptual activities of conscious humans. Because the central high-level mechanisms do not have a large capactity, it makes sense to suggest that most of the sensory input is simply assimilated into ongoing processes—things going more or less as expected do not require further processing. On the other hand, unexpected or particularly important new signals may require more attention. (Note that *no* event can be *entirely* unexpected—if this were the case it would not be processed at all. However, events can be expected with extremely low probability.) Sensory signals not currently requiring high level attention are ignored. In order to determine the importance of some low-level signal or event, some processing has to go on, for example, to determine semantic content.

Bobrow and Normal also propose some basic prinicples underlying the operation of the human symbol system. The first is that processing can be initiated (and driven) both by the high level conceptual structures (motives, goals, over-all plans) and by the low-level processing of events. The high-level structures operate on the lower level structures, fitting the latter into its expectations and opertions. The low-level processing mechanisms seek higher order structures into which event data can be fit. The second basic prinicple is that all incoming signals (i.e., anything the sensory system is capable of responding to) require some kind of processing, whether simply in a fit with expectation or in noting the failure of an expectation to be met. The last principle is simply that processing resources are limited: given the other two priciples, this accounts for the difficulty of simultaneous activites (Kahnemann, 1973) and the possibility of interruption of high level processing. On the basis of these priciples, Bobrow and Norman distinguish between resource-limited and data-limited processes. A task is resource-limited anytime an increase in processing resources (memory, effort, etc.) can improve performance. Once all processing is done or all possbile processing resources made available, only the quality of date (in terms of its "fit" with low-level schemes) can further affect performance. At this point the process is data-limited. One process can only interfere with another by infringing on the resources allocated for that process. Availability of resources can be increased for one task by reducing allocation for other tasks and/or eliminating one or more of those tasks.

Any kind of input event must be accounted for in some way—these events automatically initiate processing in terms of low-level descriptions to be fit into interpretive schemes, which are in turn fit into higher order schemes. Recent studies indicate that a good deal of such processing, such as in reading, takes place in *parallel* as well as serial order (reviewed in Posner, 1978). The procedural information associated with any scheme may require some other operations—usually involving low level decision processes and calling other expectations into play but sometimes requiring central-processing facilities (e.g., upon hearing your name). The more conceptual the operation, the more processing resources required. Consider driving a car while conversing. While the low-level processing required by the continual change in incoming sensory signals probably is enough for driving, the overall quality of driving will suffer: the higher cognitive aspects of driving (planning ahead, being prepared for decision points, anticipating possibilities) are not automatically invoked by sensory events and, because of resource limits, will deteriorate. The real flexibility of the system is that while all events must be dealt with, incorrect or very general accountings are usually quite adequate—because most sensory events are irrelevant to the course of our high level cognitive activity. If this sensory-level activity requires better accounting or sensory events are extremely anomalous, the schemes can be altered, clarified, or rejected. Kuipers (1975) suggests that cases of extreme anomaly not only call predicting frames into question but present an attack on continuity, thereby questioning the knowledge that led the processing so astray (e.g., one may conclude one is dreaming, has gone crazy, or has been led astray).

Given that all input must be accounted for—and resources are limited—the allocation of our limited resources provides a constraint on the processing of incoming information. In philosophy, the reorientation of epistemology to take account of such a psychological reality has been recently advocated by Goldman (1978). He shows how a view toward *human executability* constrains the plausible set of epistemological rules. As we mentioned before, attention is given to the unexpected or the anomalous: these are the priority events, and they will be processed to the point at which they can be accounted for. This means simply that some schemata must be found into which they fit; deeper processing will only be carried out if one of those schemata is relevant to whatever processing is of central importance to the system at the time. Events which are close to expectation are readily assimilated and, not requiring much in the way of processing resource allocation, do not require the attention of high level processing structures (which are quite likely concomitant with conscious awareness).

> We view the cognitive processing structure as one that consists of a multi-layered assemblage of experts. Each expert is a process that knows how to handle the data and suggestions provided it. When situations arise that an expert cannot handle, or when communication with the other experts that it knows about fail, then it passes on its information and messages to higher level processes. The entire system consists of a multiplicity of hierarchies of experts each expert working on its own aspect of processing, interpreting and predicting the data which are available to it, shipping requests to higher processes, and expectations of inputs to lower ones. (Bobrow & Norman, 1975, pp. 145-146)

Additionally, there must be some overarching considerations if the system as a whole is going to have any coherence at all. These overarching considerations would consist in some kind of procedures capable of selecting which activities to pursue, providing the longer term direction and goals so that a single line of activity could be pursued, providing some means for mediating the resource allocation conflicts among multiple activities, and providing access to alternative schema when one perspective is evaluated as unsatisfactory. Of course, there can still be more than one high level mechanism, each operating somewhat differently, but for maintaining even momentary unity, only one can be operating at any given time.

One of the most important features of the higher order control structures is without question the provision of a capacity for self-knowledge to the system. In order to be able to plan ahead, the system has to have some knowledge of its own functioning: what the system is capable of doing under various circumstances, what sort of problems the system has to watch out for, and so forth. Although, as we saw in the previous section, this knowledge may in various ways be distorted, over-general, or incomplete, the system can still use it to observe its own behavior, analyze its own actions, and thereby better direct its other activities. The importance of this higher level monitoring of cognitive resource allocation and operation (e.g., Neisser, 1967) cannot be overemphasized. It is the link between the vast amounts of knowledge, operational memory, and the specific requirements of the task at hand. Simply having the knowledge and ability required for performing some particular task is not enough. The system not only has to determine what is required for a task but must have the second order knowledge of system capabilities necessary for the application of those capabilities. Understanding the task, constructing the goals and subgoals necessary for task accomplishment, accessing the appropriate operational knowledge, regulating and modifying that knowledge in accordance with momentary changes and current task demand, are all essential to intelligent function. The system can then utilize self-observation to *learn* by debugging flaws in its own procedure, generalizing, or specializing techniques for particular problems. Norman (1973) even suggests an initial filtering of high level tasks on the basis of a system's knowledge of its own knowledge such that, e.g., the system (like a human) can respond fairly quickly to a question like "What is Charles Darwin's telephone number?" The system knows that it could not possibly know such a fact. Even an implementation of a system with as limited a world as Winogard's (1972) SHRDLU had a means of keeping track of the main goal it was working on and some of the subgoals: SHRDLU could not only answer questions about its own activities and place activities in contexts but also analyze and redesign its failure. The assembly, coordination, and integration of knowledge and operational ability in new ways is the central problem in both learning and development, as well as a prerequisitie capacity for handling new situations. The requirement is for active orchestration and subsequent regulation and monitoring of system operations toward goals defined by the system via its understanding of the task. This is what Flavell (1976), Brown (1978), and others refer to as "metacognition."

Some kind of metacognitive self-knowledge or "debugging" knowledge would certainly be essential for the acquisition as well as the operation of a program with the complexity of human thought and language. The knowledge of one's own strategies of action and thought, and the processes by which these are modified can be called "self-frames." These self-frames are what allow things like discussing the evolution of a plan through stages of debugging, providing the justifications for various aspects of such a plan, as well as simply summarizing a series of events. A system might know for example that linear solutions are tried before taking interactions between subgoals into account (cf. Goldstein & Papert, 1977). Sinclair, Jarvella, and Levelt (1978) review evidence that metalinguistic awareness provides assistance in the acquisition of the first language a child learns. In the real world no two solutions are exactly the same. To be able to process *any* situation, we need a sort of analogical or even metaphorical way of matching old situational strategies to present circumstances and the self-knowledge to debug old procedures to handle current considerations.

> Procedures, in order to allow debugging, are embedded in a web of declarative commentary specifying the purposes, requirements, bugs, and effect of programs. Procedural knowledge thus serves as a basis for action while propositional knowledge provides a basis for understanding the behavior of procedures and especially for knowing how to debug or change them. (Goldstein & Papert, 1977, p. 112)

We have come full circle. The discussion of frames has brought us to a tentative procedural/declarative synthesis and, further, has illustrated the difficulty of proposing a clear separation between a system's various levels of models. In the quote above, the system's operations are effectively embedded in its model of itself.

Generativity and Development in Symbol Systems

We now turn our attention to an issue which may be of central relevance to the whole enterprise: that of generativity and development. The centrality of this issue is particularly likely considering that intelligence may be defined as fundamentally little more than a capacity for development and self-modification—in the face of internal demands from the system as well as external demands from the environment.

The work in cognitive science on scheme and frame-like structures, as we have indicated, could be powerful in elucidating some of Piaget's ideas on the same subject by being a great deal more explicit and rigorous in detailing their operations and organization. However, little research in cognitive science has been directed towards what are much more fundamental problems for Piaget: How to simulate abstraction and the equilibration of structures? How to make the proper operations and concepts available at the right time? How to formalize the child's construction of its environment, and how to deal with the reciprocal assimilation and reflective abstraction by which old schemes are coordinated and new schemes formed? (Cellerier, 1972). Cellerier's criticism, of course, was directed against cognitive scientists making theoretical claims on the basis of low level simulations of experimental situations which actually do little more than elucidate what is occuring in a particular experiment. This criticism also predates most of the theoretical work on frames in cognitive science. Yet the framework does little more than hint at thorough solutions to these problems.

The central aspect of the generativity/development issue is whether a symbol system of the kind we have been characterizing is capable of generating new schemata and the heuristic procedures for accessing, instantiating, and operating on them. That is, can such a symbol system develop cognitively on its own, or will such development always be under the auspices of a human programmer? We have seen that, at least theoretically speaking, a computer-implemented symbol sytem should ultimately be capable of making adjustments in its frameworks for varying specialized usages and learning, via its self-model, from past applications and mistakes, even potentially altering some of its higher order control structures as a result of this sort of operation. Certainly, the interactions and cross-talk between a multiplicity of frames in a semi-layered organization can provide a good deal of flexibility. The problem is that so little is known about the nature of the cognitive representations which would enable the *generation* of relevant frames, appropriate problem-solving operations, and heuristic procedures when novel ones are called for.

Many cognitive scientists believe that the epistemological problem concerning the representation of knowledge disappears because representation in a symbol system consists simply in a set of frames and their associated heuristics of use. Pylyshyn (1973) contests this assumption, asking whether, for example, the list of heuristics enabling a judgment of algebraic well-formedness tells us what is known. He asserts that there is no reason to expect a limit to the number of heuristics that might be used. "If the student has learned the concept correctly he can surely keep coming up with more principles which describe the critical characteristics on which he based his decision in each instance" (p. 37). It follows from these premises that knowledge is independent of any specific set of heuristics. Pylyshyn suggests that it is this knowledge that is behind the ability to generate new heuristics. He does admit the possibility of a finite set of general heuristics being used in this generation but points out that such a set would be formally equivalent to a set of recursive definitions. Of course formal equivalence is not psychological equivalence, and cognitive scientists have found the heuristic approach infinitely more useful than a strictly formal approach.

Furthermore, there are a number of other reasons, pragmatic as well as theoretical, for devoting more energy to a frame-based heuistic approach. For instance, from whence does this higher order knowledge come which presumably guides the invention of new heuristics and schemes for dealing with situations, and how is it used? The frame-based approach provides the following possibility: Some new task or situation is confronted (not entirely novel, or the system would be unable to comprehend it at all) and a number of schemes are tried for processing the information. Some do not work, but others are suggested and combinations of schemes (some adjusted, some not), all mutually calling on each other at various times, are arrived at. During this process, the system's self-model has kept track of various errors and blind alleys and has provided debugging knowledge to help with the various adjustments and coordinations. We have now built a more complicated ("higher order") scheme which can be instantiated again at some future time (for performance of some similiar task), or which can be altered, debugged, added to, or coordinated with still other schemes in the buildup of larger structures. With more and more usage, a particular scheme can become generalized in application as the system becomes more versatile. In fact, a heuristic frame-based system would be capable of the sort of generativity of which Pylyshyn speaks, by virtue of the symbol manipulation operations on the currently available set of heuristics. The division between the heuristic and epistemological has become very fuzzy indeed. But this fuzziness should not be very surprising at all since the knowledge represented in the mind may not in fact be in usable form until the situation (internal or external) requiring the use of this knowledge is confronted, resulting in the "generation" of such knowledge via the combination, coordination, and debugging of the discrete knowledge packets known as frames. The assertion that knowledge is generated in this way does not mean the "knowledge" cannot be characterized by unified sets of formal rules and recursive definitions, simply that this "knowledge" is not appropriately seen as represented that way. Pylyshyn presents the same sort of argument in arguing against implementing Chomsky's model of generative grammar *as is* as part of a language understanding system: although Chomsky's model may formally characterize a certain aspect of our knowledge of language, performance considerations (i.e., *use* of knowledge) may require a different characterization of that knowledge.

Another aspect of the generativity issue also deserves emphasis: Human beings are probably not as creative and original as we would have ourselves believe. While some of what we learn we do generate on our own, we learn a lot simply by being told—either by being told or being given guidance, commentary, and suggestions about how old perspectives might be revised, different abilities combined and coordinated, and so forth. Our history and our culture (mediated naturally by other individuals and by our frameworks for understanding and using what they have to offer) provide much of what we know: much of the "generativity" or human intelligence may be on the cultural level, with the creativity of given indviduals rarely having more than minor impact. As we noted, the self-knowledge of the system is important for at least partially analyzing the system's own attempt to make sense out of the world. But even human beings are not extremely good at doing this—new ways of viewing the world, new problem-solving methods, and the development of new skills are often major cultural events (Minsky, 1968) or scientific revolutions (Kuhn, 1970). That most of our knowledge is not self-generated may have been what was behind the failure in the prehistory of cognitive science to develop minimal self-organizing systems (collections of components arranged in a weakly specified structure which would gradually adapt to their environments). These self-organizing systems missed the whole order of complexity introduced by a socio-cultural environment. Furthermore, there may be no such thing as a simple set of "basic principles" on which such a system could be based, or even general procedures for combining, coordinating and debugging frames.

Building a general intelligence does not necessarily require some small set of general principles. It is quite uncertain that there *can* be such "general principles" in dealing with human activity. Even Piaget (1972) goes so far as to speculate that adults may reach the formal operational level in areas of specialty (be it car maintenance or brain surgery) but not in less familiar areas. Goldstein and Papert (1977) propose the view that intelligence, *in general*, is not grounded in an individual's possession of a small set of general abilities, but rather in a well developed capacity for procedurally using large bodies of diverse forms of knowledge. Such a capacity may well *begin* with the application of uniform techniques to specific domains, as in the child who reacts to all manipulable objects with a "mouthing schema." However, this seems rather global and primitive. The development of intelligence involves the specialization within, interaction between, and organization of such general abilities. In its emphasis on complex interaction instead of reduction to simple rules, cognitive science is an organic science (cf. Winograd, 1977).

The above account does not discredit the theoretical importance of designing a system capable of aiding our understanding of developmental issues. The kinds of judgment, problem-solving abilities, linguistic and communicational capacities, and so forth which the human symbol system makes possible represent the operation of an immense complex of computational ability. Such a complex uses motoric, visual, spatial, and conceptual representations that took years of childhood to build, and this complex of human abilities no doubt transcends the complexities of any computer system yet implemented or yet implementable given the current state of the field. Nevertheless, the possibility of work on the interactions between frames and their gradual articulation and specialization (e.g., in dealing with the gradual differentiation of self and world frame systems), the central role of means-ends analyses in computer problem solving, the importance of assimilative and accommodative capacities to flexible functioning of a frame system and so forth, all contribute to a belief that a large number of developmental issues can be ultimately handled by the approaches of cognitive science. (For current work attesting to this belief, see Klahr and Wallace, 1976; Siegler, 1978.)

According to Cellerier (1972), both child and adult thought utilizes structured representation of task environments to compute possible courses of action. Our rules and structures for operation are constantly updated by new discoveries about the representation itself and how we go about exploring that representation. Structured representations are not permanently organized in a certain way but are reconstructed from stored information whenever a problem is faced: structures are in fact only a posteriori descriptions of the results of an evolving process. So far this could be an appropriate characterization of the human symbol system articulable within the context of cognitive science. However, Cellerier asserts that the central problem in both structural and processing theories is that of development. Cellerier sees development as a result of a change in the rules or procedures for reconstructing and/or presumably instantiating the schemes by which the system operated. The problem of development is extremely high level; cognitive science in its present state only hints at answers.

Limits of the Cognitive Science Approach

It is important to recognize the existence of some real and a priori limits to the usefulness of cognitive science as an approach to understanding human mental functioning. Therefore, it is important to present some provisos. However, before doing so, two other points must be discussed. First is the fact that a number of criticisms of the approach appear to be based on misconceptions. Second is the awareness that a large percentage of the current limitations of research in the field may only be temporary and do not present boundaries in principle.

Misconceptions

Most misconceptions in cognitive science concern whether it is possible to produce an entirely human or human-like intelligence via the computer implementation of a symbol sytem. Weizenbaum (1976), being one of the more articulate critics of this possibility, goes so far as to point out that modern computers (and their programs) are sufficiently complex and autonomous to warrant talk of them as organisms: computers are capable of sensing and affecting their environments, are modifiable by experience, and can have a sort of self-consciousness. With these kinds of capacities, there is really no boundary to the kind of intelligence a computer might exhibit. For Weizenbaum, the questions remains as to how much this intelligence could really duplicate the intelligence humans customarily exhibit. To most workers in cognitive science this question is misconceived: the important question being not whether a computer can think exactly as a human does or exhibit some sort of prototypically human intelligence but, more importantly, what we can learn about human intelligence by programming computers to do various tasks. The duplication of human intelligence is not a scientific goal. Reproducing a phenomena is neither necessary nor sufficient for understanding that phenomena. (Certainly the extraordinary fecundity of the human race and the resultant overpopulation cannot be taken as testimony to self-understanding.) The cognitive scientist's goal is not to reproduce human intelligence but to simulate it, using principles, operations, heuristics, strategies, organizational schemes, and so forth, that are intellectually accessible.

We have seen one possible limit to the humanness of computer intelligence in our discussion of purpose: the question of whether or not the "purpose" exhibited by a programmed machine (however much it is capable of self-programming) can ever be considered "intrinsic." The closer the performances of computers to human abilities the more squarely we must face the ideological decision concerning the *necessity* for purpose to be "intrinsic" in order to ascribe intelligence (Boden, 1977). Perhaps the result of such a decision may be foreshadowed by the idea that, aside from fairly basic survival needs, most of human "purpose" is no more intrinsic than that of a computer and its program. (The human programmers are parents, teachers, peers, and various other institutional and cultural agents.) In the case of both human and computer intelligence, there are features, goals, and purposes not directly intended by the programmer yet following from directly programmed elements. Once we drop the notion of a program as a rigidly defined sequence of events to be automatically executed, the metaphor of "human programming" loses most of its offensiveness.

Another argument that can be expeditiously set aside is one often presumed to follow from Gödel's incompleteness proof. This familiar proof shows that for any consistent system of logic there exists a statement, meaningful within the system, about which it is impossible to demonstrate truth or falsehood but which would be accepted as true by logicians. The proof has been interpreted to indicate the existence of statements which are possible for people but not possible for machines to know as true. But, Gödel's proof applies only to closed systems, in which all axioms and inference rules are fixed. Any system, human or computer, which is capable of learning new axioms and/or inference rules and can extend its own internal representations accordingly may be able to make truth-value decisions at one point in time that it could not make previously.

Weimer (1974) provides a reminder of the Russellian distinction between knowledge by description and knowledge by acquaintance:

> We are acquainted with phenomenal experience: the sights, sounds, smells, tastes, and touches of our sensory systems, our feelings and emotions, etc. We know these things personally, experientially; we are literally acquainted with them. But this is not the sort of knowledge that science discloses. (p. 435)

The knowledge of science, however, is knowledge by description of structural characteristics, not just first order properties of objects and other events.

> We don't *experience* the objects that science discloses, and yet, we *know* them as well as, if not better than, we know our own "raw feels." Knowledge, the discursive, propositional sort disclosed by both science and common sense, is not based or founded on experience, even though it ultimately refers back to the experiences of an observer. (p. 435)

This sort of knowlege includes all kinds of things which experience cannot directly address, from the activity of subatomic particles to epistemological questions.

Descriptive and acquaintance knowledge are closely intertwined. Although our acquaintance knowledge of experiences and feelings can develop nuances and sensitivities, experience is, without a doubt, modifiable by discursive knowledge. Moreover, we can indirectly "experience" objects of science; for example physicist *sees* a certain track in a cloud chamber *as* the behavior of a certain subatomic particle (Hanson, 1958). We experience *any* conceptual object in much the same way; for example, we *experience* a certain configuaration of wood and upholstery as a "chair." Then, even our discursive "knowing" represents a kind of experience dependent on a degree of familiarity ("acquaintance") with the symbols manipulated in that mode. Yet, though the "knowing" may be said to be "experiential," the knowledge (though it may be *of* experience) is not.

Interestingly, Dreyfus (172) has claimed that immediately meaningful aspects of experience are somehow quite distinct from events which gain their meaning from a theoretical or scientific context. As Pylyshyn (1974/5) points out, however, this is at odds with the prototypic notion of science:

> Science has always been concerned with showing that there is a fundamental uniformity in nature even though appearances are often to the contrary. For example, in Galileo's time it took an immense conceptual step to accept that there might be real physical obejcts which could not be seen directly with the naked eye. For Galileo to consider pointing his telescope toward the heavens required a great leap of faith in the uniformity of nature. It entailed the belief that there could exist objects which cannot be seen but which could be rendered visible by a process which is continous with seeing (i.e., it was not a magical image-creation but rather a rendering of potential image). In the subsequent scientific tradition people have come to believe that theoretical entitites such as molecules are *like* things which we can see even though in fact they can never be seen directly. In other words, *if* we could become aware of them in our perceptual or phenomenal field they would be fundamentally no dif-

> ferent from the objects which we see around us all the time. A very similar belief is held by cognitive theorists when they speak of such things as plans. The analogous claim here would be that there are certain strategies and thinking processes of which we are phenomenally aware and others of which we are not aware (some are so abstract that we can never be aware of them) but that the two are not fundamentally different entities. In other words, if we could become aware of them we would class them with the class of 'thoughts' and not with such things as biological processes or anatomical structures. (p. 25)

The motion of the phenomenological "given" is not at odds with a notion of unconscious processing. In the same way, though some of the inferences we make in our day-to-day lives are available to us, many are not. Further, many other "essentially intuitive" or "immediate" processes may merely be unexplicated.

Another argument which is often misdirected when applied to compute simulation work concerns the *digital* form which such simulation usually takes. The argument is that as a result of their digital form, such simulations are only really capable of dealing with discrete variables and are incapable of dealing with the multitude of continuous variables presumably necessary to characterize human activity. The misdirection of the usual argument has to do with the fact that discreteness of input *or* output is not an inherent limitation of the system: discrete steps can be made as small as desired. The error here seems to be, as Boden (1977) indicates, a confusion between a code and the information coded. Continuous information can be represented quite adequately by a discrete code. For example, a great deal of ambiguity and indeterminacy can be represented in the discrete 26 letter code of our alphabet; and handwriting, certainly "continuous," is only interpretable in terms of the discontinuous, discrete code of the alphabet. One cannot, simple because a code is discrete, infer that the information coded is also discrete. Anyone who has ever watched television can attest to the ease with which changes in discrete patterns of dots can be used to present continuous information. It may be true, as both Norman (1973) and Pylyshyn (1974/5) admit, that the customary digital representations (formal networks or logical repesentations) may be too complex and unwieldy to deal with some analog-like processes, an argument echoed by Boden with regard to physiological process. "A single simple analog of a situation can be equivalent to a large complex of logical statements about the same situation, as well as containing information not easily (or even not possibly) represented in formal language systems" (Norman, 1973, p. 147). Nevertheless, simulation work requires no thoroughgoing commitment to digital representations, and hybrid analog-digital systems are possibilities.

Yet, while analog computation is not *in principle* necessary, some digital computation probably is. As one major example, logical negation can *only* be conveyed in a digital code (Altman, 1967; Pea, 1978; Sebeok, 1962). A digital machine has a flexibility an analog set-up can never have because the latter cannot have contingent branches. Contingent branches allow completely different activities under slightly different circumstances, or the discontinuous capacity necessary for obeying commands such as *if X, do Y,* or the capacity to represent universals and concepts. In brief, an analog set-up can only represent particulars.

Possible Limits

So far, we have stressed what we consider to be the misdirected arguments. We next consider some possible limits to the cognitive science approach. Among these limits we consider how the approach may deal with embodied or tacit aspects of knowing, simulating social and emotional relations, and address the problem of altered states of consciousness, primitive thought, and the like. Most of these topics represent current limits in cognitive science, but the extent to which they *must* remain so is the open question.

One of the possible limits to a computer simulation approach concerns the extent to which the human body is tied to certain aspects of human intelligence. To the extent that a body of human form is necessary, the ability of a computer to simulate human intelligence will be limited (until bionic or android-like appendages and sensory-motor systems can be developed). However, as Pylyshyn (1974) points out, the primary role of the body is in the genesis of intelligence (à la Piaget), but once conceptual structures are built up, sensory pattens are grouped into perceptual integrals, intuitions formed of space, time, and causality, and motoric operations internalized—the body is no longer so intimately tied to intelligence.

> By the time he is an adult a person's intelligence depends on his possessing a body only in the obvious sense that his body contains the mechanisms in which intelligence is realized and provides means for perception, locomotion, etc. To claim otherwise is to suggest that a person who is paralyzed has lost his intelligence. (p. 17)

This returns us to a notion of "embodiment" which we have discussed before: the necessity for a symbol system to be embodied in some kind of physical form, human or mechanical.

Nevertheless, "intelligence" may be tied to a stronger sense of "embodiment." Although it is a possibility that the current state of an organism could be simulated without simulating the development of that state (as we have argued before), it is unclear how we ever have a steady state in an organism presumed to be continually developing. That is, given that any functioning of an organism occurs over a period of time, if that organism is said to be continuing to develop over that time, there is no steady state. To the extent that the organism's functioning (and development) over that period of time involves any interaction between the organism in its physical embodiment and the external world, may not the particulars of that interaction, and therefore that functioning and development, be in some sense more intimately dependent on the particulars of that organism's physical and biological properties than the notion of "embodiment" would imply? Furthermore, in humans, symbolic and formal operational intelligence is at the top of a "pyramid" containing concrete operational and sensorimotor intelligence, which never cease to exist and from which formal intelligence is rarely entirely independent (McNeill, 1978).

Additionally, although the present state of an organism may embody the entire effective history of the organism (such that simulation would presumably be possible without reproducing a developmental history), that "present physical state" most certainly also includes the development of musculature, physical coordination, and the peripheral aspects of the sensory-motor system. These "bodily aspects," then, would play a role in further development, including the short period of development which any enduring "state" would encompass. Central processing operations may be tied to (or at least be capable of interactively accessing) peripheral sensory and motor systems: for example, the production of visual imagery must, at some level, access the visual system. Even fairly deep and stable "knowledge, " such as a person's knowledge of how to ride a bicycle, or even how to hold a pen, or walk, is certainly tied to a motor-system.

It is possible that people "embody" some things that they in no sense "know" (to the extent that knowledge embodied in activity is different from that posited by an observer, even a self-observer); there are things which require a human body but which are not codable into information structures at all. It is this uncodable knowledge which is spoken of by Polanyi (1964) and others as "tacit knowledge," According to such writers, human reasoning (even in mathematics) uses tacit inference and introspectively unavailable global knowledge in a fashion crucial to determining the cognitive contents of focal awareness. Further, these tacit aspects of reasoning are indeterminate—they cannot be completely specified. It appears that large areas of human cognition are in principle not formalizable and as such not simulatable. In fact, Polanyi admits that most of this knowledge *can* be formalized but claims that the process of doing so can expand the powers of the mind, so that there will always be knowledge *as yet* unformalized, but that nothing in particular is unformaliz*able*. In this case, nothing denies the possibility of usefully articulating some aspects of tacit knowledge. The limit is simply the inevitability that part of our knowledge is tacit at any given time. Given the continual development of the human epistemological system this limitation is not particularly problematic; we can never be fully aware of what the systems are capable of doing because the systems are continually evolving new capabilities.

Social interaction is another domain often posing problems concerning the limitations and abilities of computer simulation work. While comparatively little work has been done involving the interaction of independently programmed systems, there is no reason why machines could not be programmed to interact with other machines, thus simulating interaction between people. Furthermore, even if a machine is to do something as simple as answering questions intelligently, it has to take into account the knowledge and expectations of the asker. For example, as Norman (1973) points out, a question like "Where is the Empire State Building?" would require a different answer by someone in a foreign country than by a tourist in Manhattan. And as Simon (1976) points out, it is implausible that the processes for handling social situations are very distinct from those for handling other sorts of situations: social choices *are* choices, attributions of agency *are* causal inferences (broadly speaking), perceptions of others *are* perceptions. In any complex situation information is built up and processed via thinking, judging, inferring, etc., while bringing to bear background knowledge and long term memory. Particular social situations may have particular task requirements; but so may any situation: situations may involve different scripts or frames but not an entirely different processor. The frames for social situations will only differ from others by their specific content and, perhaps, level of complexity.

In line with this, Dawes (1976) provides evidence for what he calls a "cognitive conceit." This results when people attribute errors in judgment to the intrusion of motivational ends rather than the inherent limits of judgment processes. Cognitive incapacity may be as much a mental as an emotional problem dependent on, for example, the utilization of different information by opposing sides in a dispute, inabilities to integrate information, systematic biases in estimating probability, inability to keep two analyzable dimensions in mind at one time, and so forth. Such limits lead us to place more faith in our intuitions than they deserve; for example, beliefs persist that clinician's interpretations are superior to statistical models despite the lack of supportive evidence, or that graduate admissions committees are more accurate than linear composites in making judgments despite evidence to the contrary. It appears that people are more confident with incorrect over-simplified memories of complex situations than with correct memories and more confident with judgments based on redundant information than with ones based on larger amounts of non-redundant information. We use all sorts of biasing and distorting heuristics when dealing with situations, social or otherwise, to make these situations easier to cope with; such errors simply occur more often in the more complex social situations:

> ...our own cognitive limitations may lead us to confuse the cumulative technological advances of our society with the power of a single human mind. The fact that a lot of us with the aid of a printing press, telephone and verbal communication can create an H-bomb does not mean that any of us singly can think very straight. (Dawes, 1976, p. 10)

The expertise of good judges, for example, may have more to do with simply knowing what things are important to attend to than with integrating information more efficiently; more complex judgments may be based on the evolution of memory structures such that inefficient conscious processing can be bypassed.

Abelson (1976) eleborates a theory of script processing applicable to attitude information and decision making. Scripts are a kind of frame representing a coherent sequence of expected events. According to Abelson, scripts are built up out of vignettes (much as a cartoon script is built out of pictures with captions) which may be established by being stored as single experiences (episodic), by collecting similarities in groupings of experience (categorical), or by abstracting features which may be more widely applicable (hypothetical). Scripts can be composed of any of these or in combination. An episodic script would be: "I stole a cookie and got spanked"; the generic script would be "doing forbidden things leads to punishment"; the hypothetical level is more flexible, including inferences and further abstractions which refer to different possible particular scripts. Abelson (1976) warns against the "academician's error":

> Because academicians devote (or aspire to devote) so much cognitive activity to formal operations on abstract materials, it is easy for them to fall into the view that such cognitive activity is generally characteristic. A contrasting view would be that concrete processing of episodic and situational material is often much more compelling. Because people are capable of formal operations does not mean that they prefer them. (p. 36)

Furthermore, use of more concrete scripts may be equivalent to a greater feeling of effectance—this is because being reminded of a specific concrete instance from the past is more evocative than assimilation to a general category or consideration of a complex set of "hypothetical" categories. In some settings, the concrete instances are more salient. Nisbett, et al., (1976) provide supportive evidence from the mental health realm: Apparently, depressives are more helped by concrete information about specific people than by information about "other people." Such abstract information may be more difficult to use without some concrete example. Abelson's script theory can not only cope with a lot of the nonformal vagaries of human reasoning but provides a handle on dealing with attitudes and cognitively mediated social behavior. Attitudes are simply the ensemble of scripts concerning some object, person or activity; cognitively mediated social behavior involves the selection of a script for a certain situation and taking a participant role in that script.

Apparently, computer simulation may be able to handle many cognitive aspects of social interaction, as well as the proceses of judgment, decision making, and attitude formation. A problem, however, according to Miller (1974), is that though the computer metaphor is relevant to cognitive theories, it has little place for the affective components of a person's acts. Certainly, emotion plays an important role in social interaction, and even in solitary activities like reading and writing we must take account of emotional factors. In activities which are primarily cognitive there are emotional components like interest or boredom. To what extent could a computer system have any understanding of emotion? To what extent can we learn about human emotions using computer simulation?

Seemingly, it makes little sense to attribute "feeling" to an inorganic system; however, we must also point out that there is a great deal more to emotions than affect. The seminal work of Shachter and Singer (1962) demonstrated that emotions are a function of situational cues plus the labelling of feelings, and subsequent work has carried these ideas much further. Emotion probably has a lot to do with interpreting situations via schemata with varying cognitive content (shame, pride, vanity, etc.). The bodily manifestation of an emotion may have more to do with strength of emotion than with the particular form an emotion takes. It is true that bodily manifestations might not be readily simulatable. Nevertheless, the theoretical goal is to represent the origins and effects of emotions rather than mimic their actual bodily form. The core structure of cognitive choices, which even as phenomenologically oriented a writer as deRivera (1977) finds *essential* to each emotion, are far more accessable to implementation on an artificial symbol sytem. To understand emotion it becomes important to articulate the background knowledge, the concepts, desires, abilities, plans, and so forth, in terms of which the emotion plays a role in a person's psychological life. It is via scriptal frames covering themes like comfort, rescue, violation, and so forth that a person's emotional life might be understood. Only through the rich interconnection between emotion, situation, activity, and other phenomena do we respond appropriately to friends, understand works of art, engage in productive work, and do any number of other essentially human actions and projects. Clearly, since virtually all emotion has some sort of cognitive content, and since we are able to speak of emotions in a shared language expressing their qualitative content, it is quite probable that a computer system could "comprehend" the "cognitive content," and "speak" of emotions, responses, and so forth with some facility and appropriateness. "...To say that a computer could have no real understanding of emotions—no matter how plausibly it used emotional *language*—on the ground that it supposedly cannot experience feelings, is to make a highly dubious claim" (Boden, 1977, p. 442).

Given a computer implementation that can make appropriate emotional interpretations of situations and other people's actions (whether directly or indirectly exposed to these), a computer could be designed to take into account the emotions of its human interactants. If a computer is going to simulate emotions, however, it has to be capable of placing itself as participant in emotion-laden scripts which have procedural implications for its own activities, symbolic and otherwise. Some emotions, and their corresponding (or identifying) procedural effects, may be crucial to the central controlling operations of the computer as a system. These sorts of effects may be quite important to *any* sort of intelligent system which must function in a world of such complexity. In a sense, many emotions can provide information about how to proceed: When confident we proceed with efficiency, things are working well, we do not have to be constantly checking out possible alternatives and dangers; at other times anxious checking of multiple possiblities may be in order. Sussman's (1975) HACKER simulation employs a distinction along these lines, working in CAREFUL mode when first attempting a new kind of problem, becoming more "confident" with greater experience. Emotional states like agitation or obsession might be represented by different sorts of search processes—breadth first versus depth first, for example. Other sorts of emotions, like humiliation and guilt, might be represented by quite culturally dependent sorts of themes and scripts—directing reoccurrence of certain mental operations, creating certain imperatives for action, and so forth. Many emotions provide warnings about present or future dangers to physical, psychological, social, or reflexively, emotional well being of the system. Such warnings may initiate searches for avoidance strategies, or provide suggestions of things to watch out for or keep in mind (particulars deserving attention or actions needing to be executed). These suggestions are admittedly speculative, but they do indicate directions for research and provide at least intuitive evidence for the usefulness of the computer metaphor in exploring various aspects of human emotional life. The essential cognitive or information processing component to emotion is also emphasized by some empirical work in psychology. For example, Isen, Shalker, Clark and Karp (1978) indicate that mood may be an effective retrieval cue, i.e., for a subject in a certain mood, certain sorts of cognitions are more accessable than others. The authors also suggest that mood may be appropriately conceptualized as a cognitive state. The claim is that the cognitive processes or changes which were originally thought to *mediate* the relationship between mood and behavior, actually produce, or are the same as, the mood itself.

A number of problems may be solvable by reference to alternative modes of processing, different means of accessing or constructing various kinds of information, and different procedures for dealing with or operating on information in various circumstances. One of these problems concerns the role of fringe consciousness in attentional focus: since the operation of an attention focusing device must depend on the whole field, it must include unattended parts (Posner, 1978). Some beginnings to solving this problem have been attempted making a "local context" more accessible to ongoing processing (e.g., Winograd, 1972). A capacity for taking different perspectives may be programmed into a system via the operation of different retrieval strategies operating on memory, allowing the kind of changes in recall following a shift in perspective found by Anderson and Pichert (1978). Similarly, different "states" of consciousness may not be *states* at all but rather different organizations of encoding, processing, and retrieval operations of a single "state of consciousness." As Kaplan (1971) points out, it is unlikely that a general theory of cognition can be established without considering all forms of mentation: that of infants, children, schizophrenics, archaic and preliterate people, the senile, and of oneiric, hypnogogic and hypnopompic "states." Work in this direction is really only now beginning.

Provisos

Having looked at some of the misconceptions about simulation work in cognitive science and at some of its possible limits, it is appropriate to close with some provisos and some warnings. First of all it is important to recognize that what we know about the mind at this time is little more than rudimentary, and any theoretical system currently formulatable must be only humbly proposed. Yet, in a number of areas, as we have seen, no real *in principle* statements about limits can really be made: we just do not know how far the approach can be taken. However, we must recognize the quantum leap in the claim that representations of conceptual (and other) structures *only* in the form of computer-manipulable data structures are what underly all of human thought. The safer road is to claim that symbolic structures of this form underly *some* of human thought. *Research to date does not support the claim that computer modeling can provide a view of the whole person.* To accept the extreme approach that humankind be viewed as no more than a species of the genus "information processing system" may be a grievous error. Human beings surely do a great deal more than simply process information. It would be more appropriate to reverse the genus-species analogy. One can sympathize with Weizenbaum's (1976) difficulty in seeing how a computer could understand "Will you come to dinner with me this evening?"

as an attempt by a lonely young man to overcome his shyness. Extant sytems are light years from this depth of understanding. And we can be equally touched, if not awed and humbled, as Weizenbaum was in this passage:

> Sometimes when my children were still little, my wife and I would stand over them as they lay sleeping in their beds. We spoke to each other in silence, rehearsing a scene as old as mankind itself. It is as Ionesco told his journal: "Not everything is unsayable in words, only the living truth." (p. 201)

We can hope that this awe and humility is as deeply felt by our fellow researchers in all fields directed toward a greater self-knowledge for the human species.

The major limitation on the simulation of human intelligence concerns the lack of a shared human context for an artificial system. Intelligence is meaningless without a domain, a frame of reference: an organism is, in a large sense, defined by the problems it faces and the goals that direct its behavior. Artificial systems are yet a long way from facing the problems people do, nor are they anywhere near a full and rich understanding of those problems. People are shaped to a large extent by biological pressures—for survival, for reproduction—and it is likely that only with an underlying grasp of these sorts of considerations can any full undersanding of human thought be achieved. To *simulate* human activity adequately, these concerns must be embodied in the underlying conceptual structures of implemented programs. "To what extent one can understand another person's ideology by theoretically representing it, as opposed to adopting and experientially entering into it, is of course a deep and difficult question" (Boden, 1977, p. 440). One can be overstrict about this, as we have seen Dreyfus to be, but it is probably true that failing to share a human form of life (including its particular embodiment, intrinsic interest, and the capacity to express emotion) may seriously limit the usefulness of artificial systems for understanding human intelligence. Some things we know as people are only knowable as a consequence of being treated as whole persons by other people. But even humans do not fully understand each other. This is not only because they can never share the totality of life experiences, but because they are "embodied" differently (for example, some feminists argue that even a sympathetic male cannot possible understand the issues of her concern). But within these limits a great deal of understanding can be achieved, and as long as we recognize that the computer is a tool for *our* understanding, a computer metaphor for human thought may cover significant ground.

In important ways, the real problem with the computer metaphor may be not its limitations—but its great power. Yes, human minds may literally *be* physical symbol systems, but what *kinds* of physical symbol systems? The goal is to specify the particular computational procedures and the forms and transformations of the internal representations used by human minds. Yet, these may be as changeable in humans as in programmable systems: the advantage of the computer as metaphor may be that it is a system approaching in flexibility the greatest all-purpose system we know—the human mind. And why shouldn't it be? The computer is a product of human intelligence, so why should not human beings make reflexive use of it?

References

Abelson, R.P. Script processing in attitude formation and decision making. In J.S. Carroll & J.W. Payne (Eds.), *Cogition and social behavior.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Altmann, S.A. The structure of primate social communication. In S.A. Altmann (Ed.), *Social communication among primates.* Chicago: University of Chicago Press, 1967, pp. 325-362.

Anderson, J.R. & Bower, G.H. *Human associative memory.* Washington, D.C.: Winston, 1973.

Anderson, R.C. & Pichert, J.W. Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior,* 1978, *17,* 1-12.

Aune, B. Thinking. In P. Edwards (Ed.), *The Encyclopedia of Philosophy* (Vol. VIII). New York: Macmillan, 1967.

Austin, J.L. *How to do things with words.* London: Oxford University Press, 1962.

Bartlett, F.C. *Remembering.* Cambridge: Cambridge University Press, 1931.

Bateson, G. A theory of play and fantasy: A report on theoretical aspects of the project for study of the role of paradoxes of abstraction in communication. In *Steps to an ecology of mind.* New York: Ballantine, 1972.

Black, M. *Metaphors and models.* Ithaca, NY: Cornell University, 1962.

Bobrow, D.G., & Collins, A.M. (Eds.) *Representation and understanding.* New York: Academic Press, 1975.

Bobrow, D.G., & Norman, D.A. *Some principles of memory schemata.* In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding.* New York: Academic Press, 1975, pp. 131-149.

Bobrow, D.G., & Winogard, T. An overview of KRL, knowledge representation language. *Cognitive Science,* 1977, *1,* 3-46.

Boden, M. *Artificial intelligence and natural man.* New York: Basic Books, 1977.

Braine, M.D.S. On the relation between the natural logic of reasoning and standard logic. *Psychological Review,* 1977, *85,* 1-21.

Brown, A.L. Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.). *Advances in instructional psychology.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Carroll, J.S., & Payne, J.W. (Eds.). *Cognition and social behavior.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Cellerier, G. Information processing experiments in recent research on cognitive learning—Theoretical implications. In S. Farnham-Diggory, (Ed.), *Information processing in children.* New York: Academic Press, 1972, pp. 115-123.

Charniak, E., & Wilks, Y.A. (Eds.). *Computational semantics.* New York: American Elsevier, 1976.

Chisholm, R.M. Intentionality. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. IV). New York: Macmillan, 1967, pp. 201-204.

Chomsky, N. *Aspects of the theory of syntax.* Cambridge: MIT Press, 1965.

Cicourel, A.V. *Cognitive sociology: Language and meaning in social interaction.* New York: The Free Press, 1974.

Cole M., Hood, L., & McDermott, R. *Ecological niche picking: Ecological invalidity as an axiom of experimental cognitive psychology.* Cambridge, MA: Harvard University Press, 1980.

Collett, P. (Ed.). *Social rules and social behavior.* Oxford: Oxford University, Department of Experimental Psychology, 1974.

Collins, A.M. Why cognitive science? *Cognitive Science,* 1977, *1,* 1-2.

Craik, K.J.W. *The nature of explanation.* Cambridge: Cambridge University Press, 1943.

Dawes, R.M. Shallow psychology. In J.S. Carroll & J.W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Dennett, D.C. *Brainstorms: Philosophical essays on mind and psychology.* Montgomery, VT: Bradford Books, 1978.

DeRivera, J.A. *A structural theory of emotions.* New York: International Universities Press, 1977.

Douglas, J.D. (Ed.). *Understanding everyday life.* Chicago: Aldine, 1970.

Dreyfus, H.L. *What computers can't do: A critique of artificial reason.* New York: Harper & Row, 1972.

Ericsson, K.A. & Simon, H.A. Verbal reports as data. *Psychological Review,* 1980, *87,* 215-251.

Feigel, H. *The mental and the physical.* Minneapolis: University of Minnesota Press, 1967.

Filmer, P., Phillipson, M., Silverman, D., & Walsh, D. *New directions in sociological theory.* Cambridge, MA: MIT Press, 1967.

Fingarette, H. *Self-deception.* New York: Humanities Press, 1967.

Flavell, J.H. Metacognitive aspects of problem solving. In L.B. Resnick (Ed.), *The nature of human intelligence.* Hillside, N.J.: Lawrence Erlbaum Associates, 1976.

Fodor, J.A. *The language of thought.* New York: Cromwell, 1975.

Garfinkel, H. *Studies in ethnomethodology.* Englewood Cliffs, N.J.: Prentice-Hall, 1967.

Goldman, A.I. Epistemics: The regulative theory of cognition. *Journal of Philosophy,* 1978, *75,* 509-523.

Goldstein, I. & Papert, S. Artificial intelligence, language, and the study of knowledge. *Cognitive Science,* 1977, *1,* 84-123.

Halwes, T. Structural realism, coalitions, and the relationship of Gibsonian, Constructivist, and Buddhist theories of perception. In W. B. Weimer & D.S. Palermo (Eds.), *Cognition and the symbolic processes.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974, pp. 367-383.

Hanson, N.R. *Patterns of discovery.* Cambridge: Cambridge University Press, 1957.

Harré, R. Some remarks on "rule" as a scientific concept. In T. Mischel (Ed.), *Understanding other persons.* Oxford: Blackwell, 1974, pp. 143-184.

Harré, R., & Secord, R.F. *The explanation of social behavior.* Oxford: Blackwell, 1972.

Hempel, C.G. *Aspects of scientific explanation.* New York: The Free Press, 1965.

Hesse, M.B. *Models and analogies in science.* New York: Sheed and Ward, 1963.

Huxley, T.H. On the hypothesis that animals are automata, and its history. In his collected essays, Vol. I. *Methods and results.* New York: D. Appleton, 1911. (Originally published in 1874)

Hymes, D. *Foundations of sociolinguistics.* Philadelphia: University of Pennsylvania Press, 1974.

Isen, A.M., Shalker, T.E., Clark, M., & Karp, L. Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology.* 1978, *36,* 1-12.

Johnson-Laird, P.M. Procedural semantics. *Cognition,* 1977, *5,* 189-214.

Johnson-Laird, P.M., & Wason, P.C. (Eds.). *Thinking: Readings in cognitive science.* Cambridge: Cambridge University Press, 1977.

Kahnemann, D. *Attention and effort.* Englewood Cliffs, NJ.: Prentice-Hall, 1973.

Kaplan, B. Genetic psychology, genetic epistempology, and theory of knowledge. In T. Mischel, (Ed.), *Cognitive development and epistemology.* New York: Academic Press, 1971, pp. 61-81.

Klahr, D. Steps toward the simulation of intellectual development. In L.B. Resnick (Ed.), *The nature of intelligence.* Hillsdale, NJ.: Lawrence Erlbaum Associates, 1976.

Klahr, D. & Wallace, J.G. *Cognitive development: An information processing view.* Hillside, N.J.: Lawrence Erlbaum Associates, 1976.

Kuhn, T.S. *The structure of scientific revolutions* (2nd Ed.). Chicago: University of Chicago Press, 1970.

Kuipers, B.J. A frame for frames: Representing knowledge for recognition. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding.* New York: Academic Press, 1975.

Lakatos, I. Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge.* Cambridge: Cambridge University Press, 1970.

Lakatos, I. History of science and its rational reconstructions. In R. Buck & R.S. Cohen (Eds.), *Boston studies in the philosophy of science* (Vol. 8). New York: Humanities Press, 1970.

Lakoff, G. Toward an experientialist philosophy: The case from literal metaphor. In D.A. Norman (Ed.), *LaJolla conference on cognitive science. Program in cognitive science,* UCSD, 1979.

La Mettrie. *L'Homme machine.* Paris: J.J. Pauvert, 1966. (Original publication, 1747).

Levin, J.A. & Moore, J.A. Dialogue games: Meta-communicational structures for natural language interaction. *Cognitive Science,* 1977, *2.*

Mandler, J.M., & Johnson, N.S. Remembrance of things parsed: Story structure and recall. Cognitive Psychology, 1977, *9,* 111-151.

Marshall, J.C. Minds, machines, and metaphors. *Social Studies of Science,* 1977, *7,* 478-488.

McNeill, D. *The conceptual basis of language.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Mehan, H., & Wood, H. *The reality of ethnomethodology.* New York: Wiley Intersciences, 1975.

Miller, G.A. Toward a third mataphor for psycholinguistics. In W.B. Weimer & D.S. Palermo (Eds.), *Cognition and the symbolic processes.* Hillside, NJ.: Lawrence Erlbaum Assoicates, 1974, pp. 397-414.

Miller, G.A. Some comments on competence and performance. In D. Aaronson & R.W. Rieber (Eds.), *Developmental psycholinguistics and communication disorders.* New York: New York Acadmey of Sciences, 1975.

Miller, G.A., Galanter, E., Pribram, K.H. *Plans and the structure of behavior.* New York: Holt, Rinehart, & Winston, 1960.

Miller, G.A., & Johnson-Laird, P.N. *Language and perception.* Cambridge, MA: Harvard University Press, 1976.

Minsky, M.L. (Ed.), *Semantic information processing.* Cambridge, MA: MIT Press, 1968.

Minsky, M.L. A framework for representing knowledge. In P.H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill, 1975, pp. 211-277.

Moore, J., & Newell, A. How can MERLIN understand? In L.W. Gregg (Ed.), *Knowledge and cognition.* Potomac, MD: Lawrence Erlbaum Associates, 1974, pp. 201-252.

Neisser, U. *Cognitive psychology.* Englewood Cliffs, N.J.: Prentice-Hall, 1967.

Newell, A. A note on process-structure distinctions in developmental psychology. In S. Farnham-Diggory (Ed.), *Information processing in children.* New York: Academic Press, 1972.

Newell, A. Physical symbol systems. *Cognitive Science,* 1980, *4,* 135-183.

Newell, A., & Simon, H. *Human-problem solving.* Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Newell, A., & Simon, H. Computer science as an empirical inquiry: Symbols and search. *Communications of the Association for Computer Machinery,* 1976, *19 (3),* 113-126.

Nisbett, R.E., & Wilson, T.D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review,* 1977, *84,* 231-259.

Nisbett, R.E., & Borgida, E., Crandall, R. & Reed, H., Popular induction: Information is not necessarily informative. In J.S. Carroll & J.W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Norman, D.A. Memory, knowledge, and the answering of questions. In R.L. Solso

(Ed.), *Contemporary issues in cognitive psychology: The Loyola Symposium.* Washington, DC: Winston, 1973.

Norman, D.A. Twelve issues for cognitive science. *Cognitive Science,* 1980, *4,* 1-32.

Norman, D.A., Rumelhart, D.E., & the LNR Research Group. *Explorations in cognition.* San Francisco: Freeman, 1975.

Ochs, E., & Schieffelin, B.B. (Eds.), *Developmental pragmatics.* New York: Academic Press, 1979.

Ortony, A. Beyond literal similarity. *Psychological Review,* 1979, *86,* 161-180.

Osherson, D.N. Natural connectives: A Chomskyan approach. *Journal of Mathematical Psychology,* 1977, *16,* 1-29.

Osherson, D.N., & Wasow, T. Task-specificity and species-specificity in the study of language: A methodological note. *Cognition,* 1974/5, *3,* 203-214.

Pea, R.D. *The development of negation in early child language.* Unpublished doctoral dissertation, Department of Experimental Psychology, University of Oxford, 1978.

Piaget, J. *The principles of genetic epistemology.* London: Routledge, 1972.

Polanyi, M. Life transcending physics and chemistry. *Chemical and Engineering News,* (21 Aug 1967), pp. 55-56.

Polanyi, M. *Personal knowledge.* New York: Harper, 1964.

Posner, M.I. *Chronometric explorations of mind.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Posner, M.I., Pea, R.D., & Volpe, B. *Mind and brain: Toward a scientific synthesis.* Report to the Sloan Foundation on the 1979 Sloan workship in cognitive neuroscience, New York. 1980.

Putnam, H. The mental life of some machines. In H.N. Castaneda (Ed.), *Intentionality, minds, and perceptions.* Detroit: Wayne State University Press, 1967.

Pylyshyn, Z.W. The role of competence theories in cognitive psychology. *Journal of Psycholinquistic Research,* 1973, *2,* 21-50.

Pylyshyn, Z.W. Minds, machines, and phenomenology: Some reflections on Dreyfus' "What computers can't do". *Cognition,* 1974/5, *3,* 20-42.

Riegel, K.F. *Psychology of development and history.* New York: Plenum, 1976.

Rumelhart, D.E. Notes on a schema for stories. In D.G. Bobrow & A.M. Collins, (Eds.), *Representation and understanding.* New York: Academic Press, 1975, pp. 211-236.

Russell, B. *Our knowledge of the external world.* London: Allen & Unwin, 1926.

Schank, R.C. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology,* 1972, *3,* 552-631.

Schank, R.C. (Ed.). *Conceptual information processing.* New York: American Elsevier, 1975.

Schank, R.C., & Abelson, R.P. *Scripts, plans, goals, and understanding.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Schneider, W., & Shiffrin, R.M. Controlled and automatic human information processing: I. Dectection, search, and attention. *Psychological Review,* 1977, *84,* 1-66.

Schutz, A. *The phenomenology of the social world.* Chicago: Northwestern University Press, 1967. (Originally published, 1932.)

Scragg, G.W. *Frames, planes, and nets: A synthesis.* (Working paper #20.) Castagnola, Switzerland: Institute for the Study of Semantics and Cognition, 1975.

Searle, J.S. *Speech acts.* London: Cambridge University Press, 1969.

Sebeok. T.A. Coding in the evolution of signaling behavior. *Behavioral Science,* 1962, *7,* 430-442.

Shachter, S., & Singer, J.S. Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, *69,* 379-399.

Shaw, R., & McIntyre, M. Algoristic foundations to cognitive psychology, In W.B. Weimer & D.S. Palermo (Eds.), *Cognition and the symbolic processes.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Shiffrin, R.M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review* 1977, *84,* 127-190.

Siegler, R. *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Simon, H.A. Discussion: Cognition and social behavior. In J.S. Carroll & J.W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Sinclair, A., Jarvella, R.J. & Levelt, W.J.M. *The child's conception of language.* New York: Springer-Verlag, 1978.

Sussman, G.J. *A computer model of skill acquisition.* New York: American Elsevier, 1975.

Thorndyke, P.W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology,* 1977, *9,* 77-110.

Toulmin, S. Rules and their relevance for understanding human behavior. In T. Mischel (Ed.), *Understanding other persons.* Oxford: Blackwell, 1974.

Turing. A.M. Computing machinery and intelligence. *Mind,* 1950, *59,* 433-460.

Turner, R. (Ed.). *Ethnomethodology.* Baltimore: Penguin, 1974.

Verbrugge, R.R., & McCarrell, N.S. Metaphoric comprehension. *Cognitive Psychology,* 1977, *9,* 494-553.

Weimer, W.B. Overview of a cognitive conspiracy: Reflections on the volume. In W.B. Weimer, & D.S. Palermo (Eds.), *Cognition and the symbolic processes.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Weizenbaum, J. *Computer power and human reason.* San Francisco,: Freedman, 1976.

Werner, H., & Kaplan, B. *Symbol formation.* New York: Wiley, 1963.

Wiener, N. *Cybernetics - or control and communication in the animal and the machine.* Cambridge, MA: MIT Press, 1948.

Wilks, Y.A. What sort of taxonomy of causation do we need for language understanding? *Cognitive Science,* 1977, *1,* 235-264.

Winograd, T. *Understanding natural language.* New York: Academic Press, 1972.

Winograd, T. Computer memories: A metaphor for memory organization. In C.N. Cofer (Ed.), *The structure of human memory.* San Francisco: Freeman, 1975a, pp. 133-161.

Winograd, T. Frame representations and the declarative/procedural controversy. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding.* New York: Academic Press, 1975b, pp. 185-210.

Winograd, T. Towards a procedural understanding of semantics. *Revue internationale de philosophe,* 1976, *3-4,* 260-303.

Winograd, T. On some contested suppositions of generative linguistics about the scientific study of language. *Cognition,* 1977, *5,* 151-179.

Wittgenstein, L. *Tractatus logico-philosophicus.* London: Routlege & Kegan Paul, 1922.