THE BEHAVIORAL AND BRAIN SCIENCES (1978), 1, 93–127
Printed in the United States of America

Computational models and empirical constraints

Zenon W. Pylyshyn
Departments of Psychology and Computer Science, The University of Western Ontario, London, Ontario, Canada N6A 5C2

Abstract: It is argued that the traditional distinction between artificial intelligence and cognitive simulation amounts to little more than a difference in style of research – a different ordering in goal priorities and different methodological allegiances. Both enterprises are constrained by empirical considerations and both are directed at understanding classes of tasks that are defined by essentially psychological criteria. Because of the different ordering of priorities, however, they occasionally take somewhat different stands on such issues as the power/generality trade-off and on the relevance of the sort of data collected in experimental psychology laboratories.

Computational systems are more than a tool for checking the consistency and completeness of theoretical ideas. They are ways of empirically exploring the adequacy of methods and of discovering task demands. For psychologists, computational systems should be viewed as functional models quite independent of (and likely not reducible to) neurophysiological systems, and cast at a level of abstraction appropriate for capturing cognitive generalizations. As model objects, however, they do present a serious problem of interpretation and communication since the task of extracting the relevant theoretical principles from a large complex program may be formidable.

Methodologies for validating computer programs as cognitive models are briefly described. These may be classified as intermediate state, relative complexity, and component analysis methods. Compared with the constraints imposed by criteria such as sufficiency, breadth, and extendability, these experimentally based methods are relatively weak and may be most useful after some top-down progress is made in the understanding of methods sufficient for relevant tasks – such as may be forthcoming from artificial intelligence research.

Keywords: artificial intelligence, cognition, cognitive science, methodology, empirical constraints, computer simulation

Introduction

The development of the field of artificial intelligence over the last two decades has been hailed by many as being of paramount theoretical importance to cognitive psychology (and therefore to the philosophy of mind which has been increasingly sensitive to the empirical sciences). Although the importance of this new technical tool is generally accepted, opinion regarding the nature and extent of its influence on the understanding of cognition varies considerably – even among those sympathetic members of the newly formed "Cognitive Science" fraternity. There are those who believe that computers are merely a convenience, a tool with which we can examine independently framed theories for their completeness and consistency. John Anderson and Gordon Bower, well-known computer users themselves, put it this way: "The computer is only a *computational tool* for explicitly checking the predictions of the theory, for determining whether all the specified mental processes are in fact fully specified and whether they can work together as claimed (Anderson & Bower, 1973, p. 143)."

Others believe that what we call intelligence or intelligent behavior is so constrained by the physical environment, the natural laws governing realizable physical mechanisms and the requirements of ontogenetic and phylogenetic development, that intelligence can be studied more or less directly and independently of how it is realized in detail in specific organisms or artifacts. Between these positions are a variety of views that have been examined by Reitman (1965) and Newell (1970).

Whatever view one takes regarding the particular role that computer systems play in the development of cognitive theory, there are many reasons for viewing the potential contribution of the computational approach with optimism. There are clear indications in the history of science (e.g., Butterfield, 1957) that periods of progress are coincident with major new technical and conceptual developments or sometimes, as in the case of Galileo's use of geometry, with taking an existing formalism seriously as a way of understanding the world. Similarly, philosophical understanding also rests on available conceptual tools. As Susanne Langer put it, "In every age, philosophical thinking exploits some dominant concepts and makes its greatest headway in solving problems conceived in terms of them (1962, p. 54)."

Several decades ago, the brilliant mathematician and computer pioneer John von Neumann pointed out that in the past science has dealt mainly with the concepts of energy, power, force, and motion and he predicted that " . . . in the future science would be more concerned with problems of control, programming, information processing, communication, organization, and systems (Burks, 1970, p. 3)." The precise nature of the *Weltanschauung* which ties together this syndrome of concepts is not yet clear (see, however, Simon, 1969; Newell and Simon, 1976). They seem to represent a move away from the study of material substance towards a more abstract study of form. The new conceptual tools leading up to the development of artificial intelligence and cognitive science are bound up with such notions as mechanism, information, and symbol.

Along with this general shift toward formalism (evident as well in analytic philosophy, linguistics, and much of biology) has been a more detailed working out of certain specific problem domains. One which has seen great progress in the last fifteen or twenty years is computer science. The formal analysis of algorithms and data structures in this science has given us new

 0140–525X/78/ZPYLY003$04 00/0

insight into the nature of *computation* and *process*. Artificial intelligence has, from the earliest days, formed part of the frontier of computer science. From the outset, people like von Neumann, Turing, Shannon, Newell, Simon, McCarthy, Minsky, and others recognized that the study of symbol processing in computer science and attempts to understand the nature of intelligent behavior were at some level inseparable.

But the question still remains in the minds of many psychologists: How close a connection can we expect between computational ideas and psychological theory? Is the relation between the two to remain at the level of exchange of concepts and a metaphorical focusing of attention or can it be more intimate? In particular, can a program be a psychological theory?

Artificial intelligence versus computer simulation

It has been generally assumed in discussions of such issues that there are two distinct brands of "intelligent" computational systems: Those designed primarily to do difficult tasks using whatever clever techniques are available and those designed primarily to simulate the human cognitive process. Now it is clear that one could sort different computational systems in a variety of ways. For example, one could use criteria such as the interests, training and methodological commitment, and goals of the designer. What is not so clear is what the difference between a clever artificial intelligence (A.I.) system and a computer simulation of a cognitive function would be if we controlled for: (a) generality (i.e., range of tasks carried out); (b) power (i.e., the level of performance of the system); and (c) the way in which we described the system (i.e., at what level of abstraction we cast our description of the process – from machine code to a flow chart or a general statement of the method in terms of the principles and rules followed). I see no compelling reason to believe that under these conditions (i.e., comparing systems equated on a, b, c) there need be any systematic difference between systems designed purely as A.I. artifacts and those designed as cognitive simulations. Of course in general, a system designed with one or the other of the goals as primary will differ on (a), (b), and (c) and these differences are worth some comment.

Generality and power. Consider first the issue of generality and power. Newell (1969) has argued that there is a fundamental inverse relation between the breadth or generality of a method (roughly the range of tasks to which it is applicable) and its power (roughly how well it does on problems to which it is applicable). This relationship arises in part because the narrower the range of tasks to which a method is applicable the more one can assume about the task in designing the method. Psychologists are familiar with this trade-off in the area of statistical methods for hypothesis testing where, in general, the more that can be assumed about the data (e.g., sampling distribution, true population mean, direction of sampling error, etc.) the more powerful the test can be made. This phenomenon is ubiquitous: The most powerful analytic tools (say in operations research) and the most powerful computational systems are ones in which advance knowledge about the task domain has to be built into the method. It is then typically up to the user to apply his general skill when he decides which method to use with which task.

It is the existence of these built-in assumptions that gives some methods their power. But this feature, plus the division of labor that leaves to the human agent the problem of deciding whether a method is applicable to a given risk, is precisely what makes some computational systems seem implausible as psychological models. Thus some of the criticism directed against Evans' (1968) geometrical analogies system, Winston's (1975) learning system, Waltz' (1975) scene analysis system, or Winograd's (1972) language comprehension system is based on such lack of generality. The high performance "expert" A.I. systems such as Dendral (Feigenbaum, Buchanan, & Lederberg, 1971), which infers chemical structures from mass spectroscopy data, or MYCIN (Shortliffe, 1976), which diagnoses infectious diseases, also exhibit the power-generality trade-off. Their power derives from a pragmatically circumscribed, relatively narrow range of problem-solving ability.

Such examples are taken by many as showing that systems designed primarily to accomplish a particular task (i.e., an A.I. system) need not do so in the same way that people would. While there may be some truth to this, our present point is that this is not primarily a claim about A.I. versus computer simulation but rather a comment about the more general phenomenon that performance may be purchased at the price of generality. This holds within theoretical psychology in general (where "performance" refers to experimental fit) just as much as within cognitive science. It is a basic fact of life that systems adequate for some task X may be qualitatively different from systems adequate for both task X and task Y.

While this qualitative discontinuity principle hangs over every scientific enterprise, it is of special concern in cognitive science where it has both a positive and a negative side. On the negative side the considerable premium which is sometimes placed on constructing systems with dazzling performance can blind one to the ultimate concern with generality. On the positive side, however, there is some reason to believe that in spite of this performance snare the computational approach is on a more secure footing than conventional theorizing in psychology. The reason for this lies in the recurring themes which continue to arise in very different areas of artificial intelligence application. Whenever sufficiently broad task domains are challenged, A.I. researchers find that they are preoccupied with problems of how to represent task relevant knowledge and how to organize control so that relevant portions of this knowledge are brought to bear when it is appropriate. Within these broad categories there are problems of efficient search, of exploiting problem constraints, of generating plans, of discovering and implementing useful heuristics, and of decomposing the system into perspicuous extendable modules. Very few of these problems are approached *ab initio* these days because they have been encountered and solved in various ways in many existing systems. The, perhaps surprising, finding has been that no matter what task domain one is concerned with the technical problems of organizing systemic complexity are very similar.

The recurrence of major problems of organization and representation of knowledge, and the organization and distribution of responsibility or control (c.f., Pylyshyn, 1978, in press) have produced the growing conviction among cognitive scientists that intelligence is not to be had by putting together language abilities, sensory abilities, visual abilities, memory, motivation, and reasoning (as the chapters of typical psychology textbooks suggest) but by bringing a large base of knowledge to bear in a disciplined way in all cognitive tasks. In view of the qualitative discontinuity principle mentioned earlier, the way a problem domain is decomposed for research purposes may be critical to the eventual generality of solutions. If the A.I. conviction turns out to be correct, it will mean that the conventional taxonomy of problem areas in psychology has been a misleading one, being based partly on a historical view of mental faculties and partly on the centrality of available research methods in shaping psychological research (c.f., Newell, 1973b). It may turn out, for example, that one cannot understand perception, reasoning, and memory independently of one another – that the general laws of cognition are at the level of abstract information handling principles (representation and control). Intelligence may be a phenomenon which appears when a large system of specific mechanisms and a large body of knowledge are organized along the lines of these abstract principles. Should this be the case, then work in artificial intelligence, even when apparently unmotivated by a desire to simulate human behavior, could nonetheless supply the foundations for a cognitive psychology.

Level of description. When a computational system is presented as a model of some aspect of cognition, there is always at least one major difficulty which has to be faced. This difficulty is related to the following question: How is one to compare processes that are operating in different media (i.e., processes with different physical instantiations)? How is one to judge whether the *same process* is running in two radically different material systems (even two different computers)?

One thing should be clear: One cannot answer this question by merely examining the two physical devices (or the device and the organism). One can no more compare a machine and a person than one can compare a plastic model of a molecule (say, DNA) with the chemical substance itself or the diagrams and formulae of Newton's Principia with pictures from an all-sky camera or any map with the territory it is supposed to depict. The model does not have its intended interpretation "written on its sleeve." The appropriate comparison in each case is not between two physical objects but between two carefully constructed *descriptions* of the objects. For example, one compares a description of the molecular model, which makes reference only to its three-dimensional structure and to the identity of its components (leaving out any mention of size, color, weight, taste, etc.), with a description of an object derived from x-ray diffraction photographs and other sources of evidence and mediated by a chain of theoretical deductions.

Similarly, in the case of a computational model the relevant object of comparison is not the computer program per se but a description of the computational process cast at some appropriate level of abstraction. The level of abstraction which is appropriate varies with the model and the goals of the theorist (i.e., the types of phenomena he wishes to explain). For example, some models have parameters representing such abstractions as aggressiveness (Colby, Weber, & Hilf, 1971) or the inferred motives of others (Schmidt, 1976). In these cases the details of the algorithm, say the general flowchart level, are clearly irrelevant. On the other hand there are models (e.g., Newell, 1973a) in which the operation of each individual rule (production) is taken to be an empirical hypothesis.

Whatever the model, there is clearly a level (or range of levels) of description appropriate for comparison. There are equally clearly levels at which it is inappropriate to make comparisons with the human cognitive process. One particularly acute difficulty with computational models is that the exact level that is appropriate is often unclear – even to the theorist. In physical models one is rarely tempted to, say, criticize a plastic model of a chemical molecule because unlike molecules of the substance being modelled it is inedible, or to decry Newton's mechanics because the planets are not the same color as his ink. But comparable errors *are* made routinely in criticizing computational models. For example, when people criticize computers as models because of the type of memory they have (i.e., location addressable), because they lack flexible motivation (c.f., Neisser, 1963), because of their speed or precision, or because of their serial operation, they are invariably victims of this fallacy. When a computational system is presented as a model, the implementation details are not part of the description under which the system is to be compared with the human cognitive process. Similarly, the appropriate description of the human process is typically one which does not include such properties as serial versus parallel processing. The latter *can* enter indirectly by virtue of different complexity profiles which may be exhibited by serial and parallel algorithms operating on a variety of different inputs (e.g., serial and parallel algorithms may exhibit different relations between time taken to perform a task and various task parameters such as size of input – see 3(b) below), but what is inappropriate is, say, to criticize a serial algorithm on the grounds that *in the brain* various events are taking place in different locations at the same time. (In fact, this is also true at the level of electronic activity in a serial computer, showing that what is parallel at one level of description may be best described as serial at another level.) Similarly, though errors in hardware are rare, systematic errors brought about by structural design features or limitations in the use of resources at the level of the algorithm are not uncommon (e.g., Feigenbaum, 1963; Newell & Simon, 1972; Anderson, 1976).

The issue of the appropriate level of description at which computational systems are to be evaluated remains a serious problem in all computational models. With large comprehensive systems the problem may become exceptionally acute. The SHRDLU natural language system (Winograd, 1972) is a program of over 200 pages of instructions in a variety of high-level languages. To present *this* as "a theory of language comprehension" is only slightly less absurd than to exhibit its author under that description. Clearly what is needed is something approaching a theory of the program; a description of the system which highlights the general principles underlying its operation.

Computational theories have a unique difficulty in this respect. Theories, such as those of physics, which explain phenomena by appeal to nomological laws, make a clear separation between fundamental laws and systems of calculation based on them. Such a distinction is not easily available in computational theories Complexity may be an endogenous part of this approach, not because intelligent behavior depends on very many factors (so does molecular motion) but because it may be that the most interesting aspects of intelligent behavior are more a function of the interaction of very many small components (e.g., a lot of specific knowledge) than the product of a few deep principles. However, this is speculative and only time will tell. This might, however, explain why the Newtonian style of theory (an axiomatic, quantitative, mathematical approach) has had almost no impact on cognitive psychology. Nonetheless, some abstraction of principles is essential if these systems are to mediate understanding. But it should be kept in mind that developing a system and describing its underlying principles are two distinct tasks. The literature contains many examples in which one but not the other of these tasks was well executed. (It *does* work both ways: One can have a good description – i.e., theory – of which the program is a poor exemplar).

Before going on to discuss empirical constraints on computational systems, it might be appropriate to comment briefly on the relation between the computational and the biological levels of analysis since this bears on the more general issue of the level of description discussed earlier. Although it is possible to simulate biological functions, artificial intelligence systems are invariably directed toward cognitive rather than biological states – i.e., computational states or expressions are given intentional or cognitive interpretations. Thus computational systems are *functional* models, in the sense understood in philosophy of mind (e.g., Fodor, 1975; Dennett, 1971; Cummins, 1975; Haugeland, next issue). Although the functional approach is the dominant one in psychology, it is viewed in a different perspective in computationally oriented cognitive science than in biologically oriented brain science. In both cases the more conventional type of functional explanation is viewed as an incomplete explanation, though for different reasons. For the former, a functional explanation is incomplete unless it is instantiated (or in a form in which it could be instantiated) in some physically describable artifact (the most convenient being a general purpose computer). We shall not comment on the reason for demanding this criterion beyond the remark that it forces a certain level of functional reduction which encourages, but does not ensure, mechanistic explanation, and it makes empirical exploration of a certain kind possible (see below; for more on this criterion see Pylyshyn, in preparation). For brain science, conventional functional explanation is incomplete unless the form of instantiation of functions in organic tissue is known. Thus functional analysis is seen by the latter discipline as a step towards a more complete, ultimately biological, explanation.

Although this is not the place to argue this difficult question in detail, it is perhaps appropriate to at least summarize some of the

beliefs held by cognitive scientists, though I clearly cannot speak for all of them. One position (the weak version) is that just as it is not only possible but highly enlightening to study algorithmic processes independently of how they are implemented (e.g., most of computer science is independent of which particular machines might be used), so there is much that can be gained in trying to understand cognitive processes independently of the biological mechanisms used to carry out those processes in various organisms. Another position (the strong version) is that cognition can *only* be understood at this level (i.e., in terms of an intentional cognitive vocabulary which includes beliefs, percepts, goals, etc.). This position claims that although cognitive functions are indeed carried out by biological mechanisms, there is an important sense in which questions about how we perceive, acquire knowledge, act in accordance with our goals and beliefs, and so forth cannot even be addressed in the vocabulary of neurophysiology. Even if we *could* say what process in the brain was responsible for some behavior, this would not explain the *act* because the relevant regularities and generalizations simply do not occur at that level of abstractness – i.e., they are not expressible in the neural vocabulary (just as we cannot express traffic laws in the vocabulary of mechanics). For example, there is no question that a causal biological chain of events intervenes between, say, being requested to "please open the door" and the ensuing behavior. But a completely different biological description is relevant for each distinct way of being requested (different linguistic forms, voices, intonations, modalities – whether spoken or written or presentated in sign language), each different type of door opening mechanism, and each different possible social or perceptual condition (e.g., whether the person who is making the request is seen as being disabled or as having his arms full or as holding a gun). It is only by going to a nonbiological vocabulary that such relevant generalizations as there may be in such situations can be captured.

Interestingly enough, the same is true of the relation between algorithms and computer hardware: Although the latter carries out the former, a description in terms of electronics would not capture the rules by which the computation proceeds. Even more importantly, however, the computational rules take the form that they do because their terms *represent* something (e.g., numbers, alphabetic characters, propositions, etc.) and *this* aspect cannot be captured by an electronic description. One reason for this is that the semantics of expressions (in the classical sense – i.e., how they are given an interpretation in some domain) depend on the syntactic form of the expression, and this is invariably lost in the translation to an electrical vocabulary. The same holds for the relation between the cognitive vocabulary and the brain vocabulary. Thus, the relevance of computation to cognition is not only that the former provides the relevant level of abstraction (or that it can be explored empirically by being run on a computer—however important that is) but that both cognition and computation are *intentional rule-governed phenomena.* (This slight digression has been rather sketchy but more detailed arguments for the claims are available elsewhere – see especially Fodor, 1975; Fodor, 1978; Dennett, 1971; Pylyshyn, in preparation; Haugeland, next issue).

Responsiveness to empirical constraints. Among the criticisms which have been leveled against the attention paid to artificial intelligence by psychologists is that A.I. is strictly a rational exercise in building formal systems, more closely related to pure mathematics than to psychology. The critics who hold such a view correctly point out that if computational systems are to be other than untrammeled works of the imagination, they must be responsive to nature. They must, in other words, be empirically constrained. Are A.I. models empirically constrained? Is the difference between A.I. and computer simulation simply that the latter is constrained by empirical observation while the former is not? I shall argue that both are subject to a broad spectrum of natural and logical constraints, the difference being primarily one of the relative priorities placed on different types of constraints and the methodological styles to which their designers subscribe.

There are several ways in which the "pure" A.I. style of research – not apparently motivated by psychological goals – can be seen as nevertheless empirically constrained. In the first place, empirical requirements enter with the intuitive acceptance of a problem as requiring intelligent action. The class of such problems is distinguished by criteria which are fundamentally cognitive. For instance, identifying certain classes of patterns by machine can be trivial. It is not an A.I. problem to recognize certain auditory frequencies, certain optical gradients, wave length combinations, and so forth. So long as the equivalence class of that pattern has a simple physical characterization, the problem of recognizing it is not one of A.I. What *is* a problem of A.I. is to devise ways of recognizing equivalence classes defined by psychological criteria: Classes such as the visual patterns corresponding to someone's face, to the presence of ordinary objects in a scene, to the sound of a spoken word, and so on. These are all equivalence classes of physical events for which there is a simple psychological (or phenomenological) description, but no simple physical description – i.e., they are a "natural kind" for humans in the sense that they taxonomize the world in a way that is relevant to capturing psychological and behavioral regularities. There would be little interest in developing an A.I. system to recognize a class of patterns which did not have a psychologically relevant (natural kind) description – no matter how technologically difficult the problem was.[1]

It should be noted here as well that if a person is capable of doing a certain task X which is judged to require intelligence, then any device that can be said to "do task X" must not only be doing the same task as the person but in some sense must also be doing X "in the same way" that the person does it. Here we are merely noting a systematic vagueness in the use of terms such as "in the same way" which is closely related to the problem noted earlier of finding the appropriate description under which to evaluate a model. This arises here because "doing task X" does not refer merely to the production of a record of behavior, since we would not say that a video recorder or any other recorder-reproducer of behavior "did task X." When one speaks of the ability to "do task X" one invariably has in mind some *class* of tasks and hence "doing task X" must refer to the contingencies under which one particular behavior is produced as opposed to another. Thus, if a person and a computer are both capable of "doing task X" there is some level of description of the two at which they are doing it "in the same way."

The distinction between what and how – between process and product – becomes relative to the description under which a system is examined and so it is inevitable that some empirical data is smuggled into an A.I. system for a class of tasks that is a natural kind for people (e.g., playing chess, drawing inferences, but perhaps not, say, the skill acquired by learning one rule from each of 50 different board games – which may be more like the ability to perceive "grue" and "bleen" – c.f., Goodman, 1955). The point is that, both for the notion of "pattern" in pattern recognition and for the notion of "problem" in problem solving, A.I. is interested in precisely those classes that form a "natural kind" for us humans—that is at least one of the defining characteristics of "intelligent" tasks and it does provide an empirical constraint on A.I.

The second sense in which empirical constraints enter into "pure A.I." is that programs can be thought of as actual experiments put to nature (see Newell and Simon, 1976 for an excellent discussion of this view). They allow for possible discoveries. As in modern physics the intervention of nature in providing empirical constraints on theory is subtle and infrequent (e.g., most of the recent discoveries in particle physics are more or less inevitable consequences of certain general mathematical constraints, such as the group theoretic structures of Guage theory). But the intervention of natural law does occur neverthe-

less and is crucial to the form the theory or the system takes in the long term.

As a "science of design" (see Simon, 1969), A.I. conducts experiments in its own characteristic way. The goal of designing a system to perform a certain function must be realized subject to the constraints imposed by nature. In a discipline which proceeds by attempting to synthesize new objects with specified functions the distinction between empirical and formal (systems) constraints becomes blurred. If a designer attempts to design a system according to certain general premises as to how such a system should operate, then there are many reasons (apart from errors or lack of effort) why such an attempt might fail. It might be that his analysis of the task requirements is flawed. It might be because there are crucial aspects of his understanding of the putative method which are incomplete. It might be that his view of how the system is to function is not in the appropriate form for implementation. The latter is an important consideration because there is always a way of describing a function which in some sense appears to address the question of *how* it is carried out but which is inappropriate for implementation because it is not at the required level of specificity or reduction. For example, one might propose that the method of playing chess consists of selecting the best move, or of selecting a move such that there exists a sequence of contingent choices beginning with the selected move and terminating in checkmate. Such a description glosses over the important aspects of the process and thus does not provide an explanation – and may indeed be found to be wrong (as in the chess example) when an attempt is made to fill in the missing details.

Regardless of the precise reason for a system failing to operate as expected (again barring actual programming errors) the scientist discovers *empirically* that the system fails by observing its behavior under a variety of conditions. Among the reasons for failure are some which *could* in principle have been predicted without experimentation with a program. But most failures are not of this kind because it is not the lack of formal completeness and consistency of the method which causes the failure but rather the inadequacy of the method for an empirically provided (and typically not completely specifiable) set of tasks. Even when the task *domain* is a formal one (such as symbolic logic), the task of *discovering solutions* in this domain is not understood in a formal sense and so lends itself to empirical discovery. In the latter case the completeness of a method can sometimes be demonstrated formally but the naturalness and efficiency of the method over interesting subsets of problems (again as provided by intuitive and hence cognitive criteria) is established by empirical exploration of the program.

The clearest examples of such discoveries occur in those functions of A.I. in which the system must interface directly to a physical environment – as in machine vision, speech recognition, and perceptual motor coordination (robotics). These have provided the clearest cases of discoveries about what kinds of relations must exist between various levels of the perceptual system (e.g., acoustical, phonetic, prosodic, syntactic, semantic) in order for the system to function. Similarly Waltz' (1975) success in designing a system for parsing a scene consisting of polyhedra with shadows hinged on his *empirical* discovery that, with a certain set of labels for elements of the image, the contraints on physically possible scenes were so great that in most cases a single correct analysis was mandatory.[2] Systems for language comprehension are similarly constrained by the empirical facts about the structure of language, the structure of the world, and the structure of cognitive systems which use language. Together, these constrain the possible form which a computational comprehension system can take. Furthermore, the attempt to build such a system is instrumental in discovering these constraints so that A.I. also provides a methodology for discovery.

Inasmuch as the empirical discoveries mentioned earlier concern the structure of what Simon (1969) calls the "task environment" (as well as the way this structure is cognized by the human observer), and inasmuch as such structures are a fundamental determiner of human cognitive processing, we see that even "pure" A.I. can hardly avoid making some contributions to cognitive psychology. However, there also remain some significant differences between how the A.I. researcher and the psychologist as cognitive simulator approach their largely overlapping goals. They owe their allegiance to rather different methodologies. When a psychologist claims that some people function the way his system does, he usually means that the computational system meets certain kinds of constraints that are considered central by the psychological community. In what follows we shall discuss some of these experimental constraints and examine where A.I. stands in relation to them.

Computer simulation: constraints from psychological laboratories

Apart from such obvious sources as known limitations of human information processing and the standard experimental techniques of hypothesis-testing, there are three major sources of specifically psychological empirical constraints on computer models. Elsewhere[3] I have referred to these three sources for convenience as *intermediate state evidence, relative complexity evidence, and component analysis evidence.*

In order to introduce the idea behind these sources of constraint we shall begin by posing the following question. Suppose someone produced what looked like a standard production model mechanical calculator and claimed that it constituted a model of human arithmetic skill. What grounds might you offer to counter this claim? Before examining several typical arguments one should note that a lot depends on exactly how the original claim was presented. As was pointed out earlier, the comparison of model and phenomenon is really a comparison between two descriptions. The description under which one should view the calculator would have to make clear, for example, that such physical properties as color, weight, size, and so forth are not meant to be part of the modeling function. On the other hand, let us suppose that the numbers entered into the machine and the numbers appearing in the display window are to be viewed as relevant. In addition to these there may be (as in any model) a substantial grey region where it is not obvious, *a priori*, whether certain aspects (e.g., the time taken to calculate an answer in our example) are relevant or not. Assuming then that we have our calculator and some reasonable description, what are some grounds for thinking that this is not an adequate model? Let us examine a few.

(a) If we give the device two numbers to add and examine it closely as it goes through its calculation (e.g., by slowing it down or stopping it periodically), we find that there are intermediate states in the computation in which all digits have undergone some change but none are yet at their final value. Subsequently, there are intermediate states in which the register contains some correct digits but these are scattered throughout the sum. And finally, a number of positions, again apparently scattered through the sum, arrive at their final value at the same time at the end of the calculation. Now we have at least three general characteristics of what might be called "states of knowledge" intermediate between the initial and final states which appear to be quite different from the intermediate states which people go through (although what the latter are is an empirical question). Methods for studying intermediate states in human tasks are very few and rather crude compared with more conventional experimental methods. They consist mostly of the analysis of thinking-out-loud protocols, supplemented by some inferences about missing states. However, such evidence does represent a unique source of constraint on models. The authoritative source on this methodology is the comprehensive book by Newell and Simon (1972).

(b) We could attempt to rank various arithmetic problems in order of their complexity *in the model*. Various complexity measures might be sought but two simple ones are the time taken to complete the task and the number of elementary operations (e.g., machine cycles) which they required. More complex measures might include some more abstract notion of elementary operation which applied not only to this one machine but to a class of such machines. Sticking to the simple measures, however, one finds that complexity in the model is independent of how many digits are involved in the addition and of which specific digits are used. Similarly, with multiplication one might note the relation between number of digits and complexity and observe that the latter does not depend strongly on which digits are involved.

We can now do the same for human subjects. The scale of complexity here can depend on such measurements as the time taken to complete the task or the frequency of errors. Again we can examine the empirical complexity as a function of variations in the task. We might observe that this complexity measure increases with the number of digits to be added, increases even more rapidly with the number of digits to be multiplied, and varies with the number of zeros in the problem as well as with the number of columns requiring "carries" in adding. Clearly, on such measures the two complexity orderings would not correspond very well.

(c) If we examine the model in finer detail, we can identify various subtasks which contribute to the total computational process. We might then be able to evaluate these subtasks independently. For example, it could be that processing individual columns conformed to data on human subjects even though the overall performance measures yielded an inconsistent pattern due to the way in which these are combined (e.g., from the fact that in the model they are done in parallel). If this were so, it should be posssible to show similar intermediate state or complexity effects on single column addition tasks. Data on human addition of pairs of digits show that the amount of time taken is not constant (as might seem subjectively to be the case) but depends on the smaller of the two digits (see Groen and Parkman, 1972). This is not the likely case in the model – even if we keep in mind that "time" may map into a more complex dimension in the model.

Component analysis of the task provides a powerful method of validating models. The nature of errors people make often helps to pinpoint loci of processing difficulty. Such bottlenecks can also be revealed frequently by deliberately stressing certain aspects of the system – for example, by examining what happens on progressively larger problems when external memory aids are forbidden compared with when they are permitted. Young (1973) has tuned his model of seriation in children by examining what specific differences occur when certain types of information are selectively provided or blocked during the task (e.g., by letting the child see all the blocks, or only those he has seriated or only the last one, the largest one, two at a time, etc.). Simon (1969) has argued that we might obtain the most discriminating evidence concerning human information processing by observing the human at the limits of his performance. If we could apply the stress to his performance in a selective manner, governed by the model we are constructing, we might be in an even better position to make direct use of the observations.

In addition to the types of empirical constraints discussed previously there are a number of other general categories worth mentioning. One is evidence regarding the malleability of behavior, which includes the ability of a system to assimilate new information, to accommodate to new environmental demands, and to learn by doing, as well as by being told or shown how. A *second* source of constraint is related to the *development* of intelligence – particularly ontogenetic, but also to some extent phylogenetic development. Development is not merely a process of incrementally adding new skills, but also involves radical reorganization of intellectual structures. The mechanism for such reorganization may never be discovered unless the evidence of developmental sequences is introduced as part of the empirical constraint on systems.

A *third* source of empirical constraint comes from the evidence of neuropsychology and allied disciplines. For example, a variety of evidence is available which bears on the problem of decomposing intelligence into partially independent functions. Evidence from pathologies such as aphasias, agnosias, and apraxias as well as evidence of localization of certain functions and of the organization of sensory and motor systems are all relevant to the question of how the total function might be decomposed.

Even if one accepts that the sorts of constraints on intelligent systems which we have been discussing are all potentially valid, there still remain very important questions of priority. If we apply all the constraints, there may be no place to stop, short of producing a human. A system through which one is to understand intelligence must necessarily be partial. The question of which sequence of approximations or of application of constraints one ought to adopt as a matter of principle, and consequently of what type of incompleteness one prefers to tolerate along the way, is one on which there is a wide spectrum of opinion. For example, Newell (1972) states his position on this question as follows:

"I will, on balance, prefer to start with a grossly imperfect but complete model, hoping to improve it eventually, rather than start with an abstract but experimentally verified characterization, hoping to specify it further eventually. These may be looked at simply as different approximating sequences toward the same scientific end. But they do dictate quite different approaches . . . (p. 375)."

This position, sometimes referred to as the "top-down" approach because it proceeds from a broad general system to well established specific details, is characteristic of the artificial intelligence approach as opposed to the experimental psychology approach. However, top-down is a relative term. The approach taken by Newell, Simon, and other cognitive scientists is much more top-downish than is typical in psychology. Nonetheless it is still much more data-oriented than is the approach followed in most artificial intelligence laboratories. The kind of "grossly imperfect" details necessary to achieve some degree of completeness over a broad domain is often more than most experimentally minded psychologists would be willing to tolerate. Part of the difficulty lies in what was earlier discussed as the problem of finding the appropriate description under which to view the system. But a large proportion of the contents of the more ambitious general systems, at this state in the history of A.I., represents genuine *ad hocery* stemming from ignorance of alternative computational methods. Hopefully at least these technical difficulties are temporary.

Conclusion

We have been discussing the relationship between artificial intelligence and psychology and in particular the general question of the nature of the empirical constraints to which models in cognitive science are subjected. It was proposed that even the purest cases of nonpsychologically motivated artificial intelligence are subject to a variety of empirical constraints imposed by the natural laws which determine both how the mechanisms operate and how physical environments behave. Other more directly cognitive constraints arise in the definition of tasks, since the notion of a task requiring intelligence is a psychologically motivated one.

In addition we have examined those empirical constraints which have traditionally held a special status in experimental psychology. These are relatively "low level" or detailed constraints for which there exist a variety of experimental methodologies. With the aid of metatheoretical assumptions, em-

pirical data gathered by such methods can be used by theorists to narrow down the range of admissible models. But what use are detailed constraints unless there are candidate models to discriminate among? While it is true that models accounting for a small set of such constraints are usually not difficult to produce (in fact there is a glut of such micro-models in experimental psychology), models which are sufficient for a broad range of tasks are nearly nonexistent (depending on your view of what constitutes a "broad range"). It can be argued that without at least two systems which meet sufficiency criteria for some psychologically interesting domain, data which constrain the fine structure of mechanisms must either remain waiting in archives or else be used to infer mechanisms which may have no role to play once the larger picture has been painted.[4]

The debate between those whose allegiance is to "top level" sufficiency conditions (the archetypical A.I. type) and those whose allegiance is to "low level" necessity conditions (the archetypical experimental psychologist) will continue so long as there are no general unifying principles (as there are in physics) to ensure that the two approaches will converge. If there is to be a unified science of cognition (and at least a few people believe that there will not – see Chomsky, 1975), then general unifying principles will have to be found at some level. There have been grand theoreticians in psychology in the past (e.g., Freud, James, Hull) who have sought such general principles with very limited success (as measured by the longevity of their theories). It is a distinct possibility that these attempts ran up against the same barrier that prevented the blossoming of physics during the two millenia that separated Aristotle and Galileo: The lack of a powerful technical tool to discipline and extend the power of the imagination. It remains to be seen whether the current optimism in the potential of computational systems to fill such a need in cognitive science is warranted.

NOTES

1. Of course it is not inconceivable that such goals may become relevant in other contexts in which it will be important to recognize a class of physical events which were equivalent with respect to some functional need of the device – say for the survival of a mechanism in some exotic nonhuman environment (e.g., in some long-term automated space mission). But then it becomes problematic to decide whether such a system exhibits intelligence. How does one decide, for example, whether a system for recognizing some X was doing it intelligently or not? One way to decide might be to ask whether the kinds of procedures demanded are similar to those needed, say, to recognize scenes or whether the task can be carried out largely noncomputationally as in a thermometer. The issue does get cloudy, however, when the conditions become sufficiently alien.

2. Waltz (1975) had eleven possible labels for edges. These included such labels as boundary, convex and concave interior, and crack and shadow edges. Of the hundreds of thousands of logically possible permutations for trihedral junctures some 500 are physically permissible for a Y-shaped juncture and only 70 for an arrow-shaped juncture. However if a pair of junctures share a common line, then that line must receive the same label at both junctures. This typically reduces the candidate labelings for that line to only a few which are further reduced by continuing to examine adjacent junctures and "propagating labels."

3. Most of the examples in this section are taken from Pylyshyn (1978, in press).

4. One must, in fairness, add a third alternative which at this time in the history of cognitive science may be more the rule that the exception. It may be that experimental psychology and artificial intelligence live in a loose but symbiotic relation in which each supplies a source of heuristic inspiration and ideas to the other. While this may not be the way either group would wish it, it may be a useful step in the courtship.

REFERENCES

Anderson, J. R. *Language, Memory, and Thought.* Hillsdale, N.J.: Lawrence Erlbaum, 1976.

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

Burks, A. W. Von Neumann's self-reproducing automatia: In A. W. Burks (Ed.), *Essays on cellular automata.* Urbana, Ill.: University of Illinois, 1970.

Butterfield, H. *The origins of modern science 1300–1800.* Toronto: Clark, Irwin and Co., 1957.

Chomsky, N. *Reflections on language.* New York: Pantheon, 1975.

Colby, K. M., Weber S., and Hilf, F. D. Artificial Paranoia. *Artificial Intelligence,* 1971, *2,* 1–25.

Cummins, R. Functional analysis. *Journal of Philosophy,* 1975, *72,* 741–765.

Dennett, D. Intentional systems. *Journal of Philosophy,* 1971, *68,* 87–106.

Evans, T. G. A heuristic program to solve geometric analogy problems. In: M. Minsky (Ed.), *Semantic Information processing.* Cambridge: M.I.T. Press, 1968.

Feigenbaum, E. A. The simulation of verbal learning behavior. In: E. A. Feigenbaum and J. Feldman (Eds.) *Computers and thought.* New York: McGraw-Hill, 1963.

Feigenbaum, E. A., Buchanan, B. G., and Lederberg, J. Generality and problem solving: A case study using the DENDRAL program. In: B. Meltzer and D. Michie (Eds.), *Machine Intelligence 6,* New York: American Elsevier, 1971.

Fodor, J. A. *The language of thought.* New York: Thomas Crowell, 1978.

Computation and reduction. In W. Savage (ed). *Minnesota Studies in Philosophy of Science,* Vol. IX.

Goodman, N. *Fact, fiction and forecast.* Cambridge: Harvard University Press, 1955.

Groen, G J. and Parkman, J. M. A chronometric analysis of simple addition. *Psychological Review,* 1972, *79,* 329–343.

Haugeland, J. The nature and plausibility of cognitivism. *The Behavioral and Brain Sciences* (next issue).

Langer, S. *Philosophical sketches.* Baltimore: Johns Hopkins Press, 1962.

Neisser, U. The imitation of man by machine. *Science,* 1963, *139,* 193–197

Newell, A. Heuristic programming: Ill-structured problems. In: J. S. Aranofsky (Ed.), *Progress in operations research, Vol. III.* New York: Wiley, 1969.

Remarks on the relationship between artificial intelligence and cognitive psychology. In: R. Banerji and M. D. Mesarovic (Eds.), *Theoretical approaches to non-numerical problem solving.* New York: Springer-Verlag, 1970.

A theoretical exploration of mechanisms for coding the stimulus. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory,* New York: Winston, 1972.

Production systems: Models of control structures. In W. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973a.

You can't play 20 questions with nature and win. In: W. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973b.

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

Newell, A and Simon, H. A. Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery,* 1976, *19,* 113–126.

Pylyshyn, Z. W. Complexity and the Study of Artificial and Human Intelligence. In M. Ringle (ed.), *Philosophical perspective in artificial intelligence.* New York: The Humanities Press, 1978), in press.

Towards a Foundation for Cognitive Science (in preparation).

Reitman, W. R. *Cognition and thought.* New York: Wiley, 1965.

Schmidt, C. F. Understanding human action: Recognizing the plans and motives of others. In: J Carroll and J. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Lawrence Erlbaum, in press.

Shortliffe, E. H. *Computer-based medical consultations: MYCIN.* New York: Elsevier, 1976.

Simon, H. A. *The sciences of the artificial.* Cambridge, Mass.: M.I.T. Press, 1969.

Waltz, D. Understanding live drawings of scenes with shadows. In: P. H. Winston (Ed.) *The psychology of computer vision.* New York: McGraw-Hill, 1975.

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

Winston, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill, 1975.

Young, R. M. Children's seriation behavior: A production system analysis. Unpublished doctoral dissertation, Carnegie-Mellon University, 1973.

Open Peer Commentary

The following are among those who will provide Continuing Commentary on this article in a later issue: J Bigelow, H R Otto, and G Pask Commentaries submitted by the qualified professional readership of this journal will also be considered for publication

Note: *Commentary reference lists omit works already cited in the target article (as indicated by op cit)*

by **John H. Andreae**

Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand

AI: another defense of the field. Whether Artificial Intelligence (A I) is "pure," "a rational exercise in building formal systems," or "genuine ad hocery," another defence of the field is unwarranted Equally unwelcome is another reminder of the current overoptimistic A I slogan that "intelligence is knowledge " This has taken the place of the earlier "intelligence is heuristics," which replaced the first slogan, "intelligence is learning " The present craze will soon pass, leaving behind a useful collection of knowledge structures to be added to the collection of heuristics and the meager collection of learning techniques Pylyshyn suggests that "intelligence is not to be had by putting together language abilities, sensory abilities, visual abilities, memory, motivation, and reasoning," but that it may appear when you put together mechanisms, knowledge, and abstract principles! The sooner we stop looking ahead to the second coming of intelligence (in machines) as a future event, the better Machine intelligence is here to stay and to develop, be it somewhat primitive My feeling is that the learning or acquisition aspect of intelligence is being neglected; you may feel that some other aspect deserves attention Let us not all jump onto the same bandwagon at the same time

Pylyshyn sees the "power-generality trade-off" as distinguishing A I and cognitive simulation efforts Now, both computers and human beings escape this trade-off Computers escape it by being programmable in an infinity of ways It is only the *programmed* computer that is constrained by the trade-off, not the computer itself Similarly, it is only the aging, educated human being who is so constrained All the famous A I systems are heavily preprogrammed so that they strongly exhibit the power-generality trade-off Only lip-service is paid to learning and knowledge acquisition, so each system lands firmly on the power-generality scale

As evidence of this lip service to learning, we can compare Pylyshyn's three main sources of empirical constraint (internal states, complexity, and task components) with his three "other general categories worth mentioning": learning, development, and neuropsychology Contrary to his earlier warning that computational models must not be criticised for details of memory, flexibility, speed, precision, or serial operation, we now find him recommending a comparison of internal states, complexity, and task components Surely, if the human being performs a task with learning and the machine happens to do it without learning, we can expect considerable differences in internal states, complexity, and task components To make matters worse, internal states are largely unobservable, complexity is unmeasurable, and the components of a task are likely to depend upon how it is learned

by **Michael A. Arbib**

Computer and Information Science Department, University of Massachusetts, Amherst, Mass 01003

The halting problem for computational cognitive psychology. Pylyshyn offers a number of convincing arguments for the claim that "Artificial Intelligence offers useful tools and concepts for the cognitive psychologist " Unfortunately, he insists upon the far stronger claim that empirical constraints actually force a convergence between AI and cognitive psychology The evidence does not support this

(1) The paper is almost totally devoid of any thoughtful analysis of the real workings of recent contributions to AI The only exception is an approving description of Waltz's discovery that constraint satisfaction can reduce the search space in parsing a scene consisting of polyhedra with shadows Most workers in machine vision hail this as an important contribution, but few would view it as a model of human polyhedral recognition, and most still seek alternative approaches which extend to non-polyhedral scenes The claim for AI/psychology convergence is further weakened when we look at the experience of the speech-understanding group at Carnegie-Mellon University They constructed two systems – HARPY (the Markov-chain model) met performance constraints far better than HEARSAY (which had subsystems more closely tied to our ideas of human functioning)

(2) But Pylyshyn has an answer to this last example – an answer blatant in its implicit rejection of testability He labels (without further analysis) most of the best work in AI (including Winograd's language understanding system, DENDRAL and MYCIN – it's not clear that anything is left in AI when the exclusion principle is followed) as having power that derives from "a pragmatically circumscribed, relatively narrow range of problem solving ability " Pylyshyn admits that such systems need not accomplish a task in the same way people would; but then claims that if a system is general enough, it will be human-like "It is a basic fact of life that systems adequate for some task X may be qualitatively different from systems adequate for both task X and task Y " A marvellous escape clause, setting up an infinite regress! Each time we show Pylyshyn an AI program that performs well, but in a non-human way, he just has to say "Ah, yes, but though it does tasks X, Y, and Z, it doesn't do W, so *of course* it's different " The halting problem is unsolvable!

(3) I do agree with Pylyshyn that AI can and does provide valuable input to cognitive psychology But to view generality, as Pylyshyn does, as "the ultimate concern" of AI, and to refer to attempts to get efficient programs to work on well-defined problem-domains as "the performance snare" is to miss the point of much of the best work being done in AI today The aim of many workers is to build high-performance systems – and such workers seek generality not as part of a program, but as a set of concepts to aid the sharing of methodology between diverse projects The generality, then, is not in the program, but in the science But, alas for Pylyshyn's case, differences of approach far outweigh the commonalities Is it really helpful to cognitive psychology to receive from AI the (bland) general principle that "organization of knowledge is important" when there is so little agreement on the style of its representation? Semantic nets, production systems and schemas (in their diverse forms) can enrich the psychologist's study of a system, but there is no "convergent evolution" to a single approach which AI and cognitive psychology models *must* use

(4) Pylyshyn offers an excellent discussion of the different levels at which a model must be evaluated. Yet he is grossly inconsistent In arguing for the convergence between AI and cognitive psychology he notes (causing great mental anguish to those, such as myself, who believe that the brain is relevant to psychology) that "the appropriate description of the human process is typically one which does not include such properties as serial versus parallel processing" and "[it] is inappropriate to criticize a serial algorithm on the grounds that *in the brain* various events are taking place in different locations at the same time " Yet when talking about experimental verification of a model, Pylyshyn offers measures of task complexity and the lesion data of neuropsychology – the very data ruled out by the above quotations! Perhaps my feeling that strategies of parallel computation (cooperative computation in my jargon) are at the heart of "computational cognitive science" is well-founded after all

In summary, a good psychological model *must* confront data In doing so, it will probably have peculiarities not contained in a high-performance AI system (In building a high-performance AI system, one would use the computer's arithmetic capabilities In a cognitive model, human fallibility in addition may be a key criterion) The design of the cognitive model may benefit greatly from the tools and concepts of AI – but the model may still have special features to fit it for the constraint of testing against human performance In the list of references, Pylyshyn refers to his forthcoming book *Towards a Foundation for Cognitive Science* One hopes that the unsupported generalizations of the present paper will be replaced by a detailed accounting of the diverse, even inconsistent, approaches within AI, with a guide to their wise deployment to specific problems of cognitive psychology

by **Margaret Atherton**

Department of Philosophy, City University of New York, Brooklyn College, Brooklyn, New York

The artificiality of computer models. Quite often people express doubts about the relevance of work on computers to psychological explanation by claiming that computer simulation is really nothing more than artificial in-

telligence Pylyshyn seems to be saying that arguments like these are insufficient to reject work with computers as a way of gaining insight into psychological processes He thinks all work done on A I can provide information about how the mind works, even when the original intention was not to capture some psychological process, but merely to get a computer that could perform some task or other, by whatever means appropriate Thus Pylyshyn seems to think A I and C S will not in the end be readily distinguishable because A I will turn out to be a kind of C S More strictly, he thinks the difference is one of research styles, and that the more useful method for psychology is really that of A I

An important reason why it is commonly thought that A I can tell us little that is interesting for psychology is that, unlike C.S , work in A I proceeds unhampered by any constraints about how people do things Thus A I is claimed to be little more than a formal system Without any empirical requirements imposed on the solution to an A I task, there is no reason to believe the system says anything about the world, that it describes something that has psychological reality Pylyshyn wants to argue, however, that there are ways in which A I tasks are subject to empirical constraints, ways which make it possible for the worker in A I to make discoveries about how people do things If Pylyshyn were right, then he would be presenting a way of overturning an important objection to C S as a method of doing psychology For people often complain that since the empirical constraints on C S consist in our knowledge of how people do things, the resulting programs are nothing more than restatements in computer terminology of what we know already But Pylyshyn suggests that an A I approach provides sufficient constraints and thus perhaps will allow us to make genuine discoveries about cognitive processes

Pylyshyn is anxious to show that A I work does not constitute a formal abstraction because he thinks that from looking at computers it has been possible to develop a general model for intelligent functioning He wants to preserve for us the possibility that a computer can be used to produce a big overall picture of how the mind works, from which it might be possible to derive a "unified science of cognition " He suggests that work on A I gives rise to a way of looking at the mind in which "one cannot understand perception, reasoning and memory independently of one another – that the general laws of cognition are at the level of abstract information handling principles " I am not sure precisely how to understand Pylyshyn at this point If he means that it is futile to try to discuss each mental ability in isolation from each other, then this is undoubtedly true, but no one needed a computer to consider such a possibility Plato and Aristotle, after all, each had models of mind designed to exploit the interdependence of different mental functions But Pylyshyn may also mean that perception, reasoning and memory are each of them severally reducible to "abstract information handling principles " If this is the case, however, there is a grave danger that this result is an artifact of using computers and of assuming that people operate as computers do

What Pylyshyn seems to think should discourage us from drawing this conclusion is his conviction that there are empirical constraints on all work done on computers Since there are grounds for saying these systems are not constrained by purely formal requirements alone, there will be reasons to believe the resulting systems are also theories that might describe something psychologically real Pylyshyn presents two examples of such empirical constraints The first is that in choosing tasks to set the computer, our choice is constrained by ideas about what kinds of tasks require intelligence in the human to be solved What we take to be, for example, a pattern or what we take to be a problem is constrained by our notions of what forms a "natural kind" for humans Thus we ask the computer to recognize faces or play chess, rather than to recognize patterns or solve problems that lack a "psychologically relevant (natural kind) description " But for this kind of constraint to be genuinely empirical, it has to be true that our notions of natural kinds constitute restrictions on our psychological capacities Pylyshyn gives examples of non-natural kinds of tasks: he cites the skill defined by learning one rule from each of fifty board games or the ability to perceive Goodman's grue/bleen But while these are skills that humans on the whole don't learn, that is not the same as saying they are skills humans *can't* learn And if, for example, Goodman's entrenchment theory were correct, then that blue/green is a natural kind rather than grue/bleen could depend in part upon a vast series of social facts, a previous history of which "kinds" have been recognized, and not solely upon psychological constraints Thus focussing on what people *do* does represent an abstraction from what they *can* do, an abstraction which need not be motivated by any particular psychological facts There would be no special reason for assuming a system that had been designed to recognize blue/green was subject to empirical restraints such that we could argue the process it embodied constituted a psychological theory I suspect that identifying a description of a natural kind with one that is psychologically relevant depends on identifying what we do naturally as what we've been programmed to do But this is to assume the validity of the computer model

Pylyshyn's second kind of case hinges on the fact that we design computers to operate in nature, therefore requiring them to be subject to natural constraints This is certainly true But the natural constraints can only serve as reasons for taking computer systems to be psychological theories if nature imposes the same constraints on humans that it does on computers One of Pylyshyn's examples concerns the constraints provided by physically possible restrictions on descriptions of edges But this environmental restriction is psychologically constraining only if humans process in terms of edges Without reasons for believing this is so, such a constraint wouldn't hold for *any* description of the "task environment," unless there were an assumption that humans deal with the environment in the same way a computer does

I am not convinced therefore that work in A I has provided a sufficient reason for taking the computer as a model of mind Its suggestiveness as a metaphor has to be set against the possibility that enquiry will become deadlocked through sticking to a particular conception of how the mind works I fear that Pylyshyn's examples of allegedly empirical constraints might be evidence of just such a deadlock For Pylyshyn's constraints could be empirical rather than merely formal, I have suggested, only if we assume we know more about how the mind works than we are really entitled to assume In particular, they encourage us to imagine the mind works like a computer But the possibility that minds don't work the way computers do must be allowed to be just as live a possibility if the computer model is to be subjected to genuinely empirical tests

***by* L. Jonathan Cohen**

The Queen's College, Oxford University, Oxford, England

Rational reconstruction of inferential processes – a task straddling the AI – CS boundaries. Pylyshyn argues against overstatement of the difference between artificial intelligence and computer simulation as types of research-project in cognitive science His argument is reinforced if we ask the question: Within which of the two types of research-project should we include the construction of programmes to represent methods of reasoning that are attributed to human beings on the strength of logical analysis or philosophical reconstruction rather than psychological experiment? Examples of such methods of reasoning might be natural deduction (e g , in accordance with the rules of Gentzen, 1934) or inductive learning (e g , in accordance with some elementary strategy for evaluating probabilities) Cognitive competences of this kind seem to straddle whatever frontier there may be between the traditional domains of artificial intelligence and computer simulation On the one hand, they are not necessarily a proper target for computer simulation, since psychological experiment may be able to show that relevant human performance tends to be incompetent in certain characteristic respects For example, some psychologists (Wason & Johnson-Laird, 1972, see also Johnson-Laird, this *Commentary*) have claimed to show this in relation to deductive reasoning, and others (Kahneman & Tversky, 1973, 1974) have claimed to show it in relation to probabilistic reasoning On the other hand, the methods of reasoning in question are not put forward as artefacts of philosophical ingenuity Rather, they are held out as natural diamonds of human reason: the only contribution that the philosophers will admit to is some systematic cutting and polishing Thus some logicians (Kneale & Kneale, 1962, p 539) claim that Gentzen 'has in fact presented logic in a fashion more natural than that of Frege, Whitehead and Russell;' and others (Reichenbach, 1949, p 337ff) concern themselves with which admissible interpretations of the mathematical calculus of probability 'are used ' Nor need there be any shortage of what Pylyshyn calls 'intermediate state evidence' in relation to these methods of reasoning It is of the essence of a natural deduction system to reveal intermediate links in a chain of reasoning (though admittedly strategies for seeking or selecting those links will not be part of the system)

Indeed, in some areas, experimentally revealed performance data, which

might afford a basis for computer simulation projects, cannot be properly interpreted for this purpose unless an adequate logical or philosophical theory of the underlying competence is available Yet in constructing a theory of, say, probabilistic reasoning from the fragmentary data afforded by such sources as rational intuition or the history of science, philosophers can hardly avoid supposing that the competence in question adopts the most efficient and economical means to achieve its characteristic tasks A philosopher has no alternative but to proceed on that supposition in trimming and polishing the superficially incoherent and heterogeneous data so as to organise them into a unitary theory But the supposition seems more germane to research in artificial intelligence than to computer simulation

An example will help to clarify the point A computer simulation project for elementary probabilistic reasoning must obviously take account of Kahneman and Tversky's experimental data But what is the correct interpretation of these data? According to Kahneman and Tversky, one of the tendencies to which their subjects were particularly prone was to judge an outcome by the extent to which it was representative of (i e , similar to) the evidence rather than by its correctly calculated mathematical probability Since the subjects were asked to judge a probability, Kahneman and Tversky interpret this response of their subjects as being fallacious But the validity of their interpretation rests on the assumption that all reasoning about probabilities has to conform to the principles of the mathematical calculus of probability, and this assumption turns out to be a rather questionable one Let us call probabilities that so conform 'Pascalian' ones, in tribute to the mathematician who first began to theorise about them Then there are certainly just as good grounds for saying that various kinds of non-Pascalian probability are possible as for saying that non-Euclidean geometries and non-Zermelian set theories are possible Indeed it can be shown (Cohen, 1977) that, when the central modern tradition in inductive logic (the tradition of Bacon, 1620, and Mill, 1843) is adequately refined and systematised, it generates a theory of non-Pascalian probability according to which the reasonings of Kahneman and Tversky's subjects about representativeness are not at all fallacious Though there are many kinds of problems about which it is natural and proper to think in terms of Pascalian probabilities, there are others (in particular some of the questions put to Kahneman and Tversky's subjects) about which it is equally natural and proper to think in terms of Baconian probabilities The arguments for this conclusion (in Cohen, 1977) are characteristically philosophical and have as their main premises a logical reconstruction of the proof procedures that are conventionally acceptable to lay juries in Anglo-American law courts, and an analysis of reputable experimentalists' reasonings about controls, hidden variables, and so forth But the conclusion is scarcely in doubt Indeed, when Locke (1690, Bk IV, ch XVI, para 12) spoke of analogy as the ground of probability, or Hume (1739, Bk I, Pt III, sec XIII) remarked that 'in proportion as the resemblance decays, the probability diminishes,' they already anticipated this conclusion, albeit in a loose, impressionistic way The Baconian probability of a given situation's having a certain characteristic (e g , of a particular individual seed's germinating) can be graded by the extent to which the causally relevant circumstances of the given situation resemble the causally relevant circumstances of another situation that does have the characteristic It is just unfortunate that, whereas Pascalian probability managed to attract the attention of mathematicians as early as the seventeenth century, Baconian probability has had to wait until the present decade for the formalisation of its logical structure At any rate, there is now available a powerful general theory of inductive competence (i e , of the ability to learn from experience) that is open to computerised representation The theory lacks all the artificial constraints that infect range-theoretical inductive logics (like that of Carnap, 1950) and inhibit the use of such logics as models of actual human reasoning But it still seems hardly fruitful to ask whether construction of the relevant programme would belong to computer simulation or to artificial intelligence The pointlessness of that question is a further argument in support of Pylyshyn's thesis

REFERENCES

Bacon, F. *Novum Organum.* London, 1620.

Carnap, R. *Logical Foundations of Probability.* Chicago, Ill., University of Chicago Press, 1950.

Cohen, L. J. *The Probable and the Provable.* Oxford, Clarendon Press, 1977.

Gentzen, G. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift.* 39:176–210, 405–31. 1934.

Hume, D. *A Treatise of Human Nature.* London, 1739

Kahneman, D., and Tversky, A. On the Psychology of Prediction. *Psychological Review.* 80:237–51. 1973.

Subjective Probability: A Judgment of Representativeness. In: C.-A.S. Staël von Holstein (ed.), *The Concept of Probability in Psychological Experiments.* Dordrecht, Holland, Reidel, 1974.

Kneale, W., and Kneale, M. *The Development of Logic.* Oxford, Clarendon Press, 1962.

Locke, J. *An Essay Concerning Human Understanding.* London, 1690.

Mill, J.S. *A System of Logic.* London, 1843.

Reichenbach, H. *The Theory of Probability* (trans. E. H. Hutten and M. Reichenbach). Berkeley and Los Angeles, Calif., University of California Press, 1949.

Wason, P.C., and Johnson-Laird, P. N. *Psychology of Reasoning: Structure and Content.* Cambridge, Mass., Harvard University Press, 1972.

by **Steven Cushing** *and* **Norbert Hornstein**

Research Staff Higher Order Software, Inc , 806 Massachusetts Ave , Cambridge, Mass 02139 and Department of Philosophy, Harvard University, Cambridge, Mass 02138

Software systems, language, and empirical constraints. Pylyshyn's paper seems to present the following point: (1) It is possible that work in AI might contribute theoretically to the development of cognitive theory

(1) can be interpreted in at least three different ways: (2) The results heretofore attained in AI research have *in fact* contributed to cognitive theory in some domain and hence it is reasonable to believe that this sort of research will continue to make fruitful contributions to the articulation of cognitive theory (3) Though no concrete results have yet been produced, nonetheless, the constraints on AI research are of a piece with cognitive psychological constraints and hence could be expected to lead to fruitful insights for cognitive theory (4) There is nothing in AI research which *would prevent* it from making a contribution to cognitive theory

It is clear that interpretations (2)–(4) make progressively weaker claims and that the interest of (1) will be directly proportional to how strongly it is interpreted (2) would be the most interesting from the perspective of cognitive theory, as it would make the strongest possible claims concerning the relevance of AI research for psychology, while (4) would seem to be of substantially less interest, as it says little more than that it is not logically impossible that AI research will lead to psychologically interesting findings

Dresher and Hornstein (1976, 1977, 1978) argue that (2) is incorrect because AI research into language has *in fact* made no contribution of any kind to a scientific theory of language, due to its not having enunciated any general principles and hence none of any psychological interest Weizenbaum (1976) and Dresher and Hornstein (1976) challenge (3) by pointing out that the aims of AI research into language are of dubious psychological interest, as these aims, despite all protestations to the contrary, are essentially technological, not scientific, and hence tend to lead away from a consideration of the issues which would be of such interest Contrary to what Pylyshyn seems to say, the nature of the constraints in AI research into language results in more "than a difference in style of research"; it makes such work of dubious value for any psychological theory of language, as it focuses attention away from the psychologically and scientifically interesting questions

There is a stronger and weaker version of (4) The weaker version, which claims that it is not a logical contradiction to believe that work in AI might lead to illuminating results for a psychologically interesting theory of language, is correct, but so minimal as to be virtually devoid of interest A second, stronger claim, is that in some broad sense AI and cognitive psychology deal with the same entities – computable functions and computational processes – and hence work in AI cannot help but say something of interest for cognitive theory The problem with this, however, is that, although it is correct, it is hard to see why dealing with the same sorts of things *at this level of generality* would lead to interesting theoretical overlap Both plumbing and fluid dynamics deal with water in motion, but there is little reason to believe that breakthroughs in the former will contribute significantly to theories of the latter

This aside, however, there is some truth to this second claim, and it is worth sketching the conditions which would make work on computational processes of psychological interest AI work on language and mind presupposes a parallelism between minds and information-processing systems, a

parallelism that is also assumed or argued for by some non-AI researchers (Miller and Johnson-Laird, 1976 [See also Johnson-Laird, this Commentary]; Fodor, 1975 *op cit*) We agree that such a metaphor can be useful in gaining insight into the nature of cognitive processes, but we think that this is possible only if at least two important conditions are satisfied First, we think that any adequate computer model of the mind will have to be formulated not in terms of *programs*, as in current AI work, but in terms of *software systems*, as discussed by some authors in the newly emerging field of software engineering We find it highly implausible that complex mental processes can be modeled adequately in terms of sequential lists of instructions, (that is, programs, by definition) The mind is a highly complex system of related and interacting, but essentially autonomous components, and it seems likely that some of the more interesting generalizations concerning its structure and operation will involve the interfaces between these components at least as much as the individual programs that may make them up Not surprisingly, it is precisely in regard to their interfaces that some of the more interesting properties of software systems have emerged (Hamilton and Zeldin, 1976)

Second, we think that such an approach to the study of mind would require a genuine *theory* of software systems, rather than the sort of ad hoc programming that is endemic to current AI work Dresher and Hornstein (1976, 1977, 1978) argue that natural language-related work in AI has been generally devoid of explanatory value because of the ad hoc character of the programs involved Most of this work, as we noted above, seems to be not scientifically, but technologically oriented, i e , geared toward developing machines that can process sentences of natural language, rather than seeking general principles that can serve as genuine scientific explanations of linguistic phenomena The ad hoc character of computer programming has become a serious problem much more generally, however, especially in connection with the specification of large and very large software systems, and it is precisely this problem that motivated the development of software engineering in the first place While this motivation has also been primarily technological, e g , minimizing cost in the development of large systems, one of its aims is to develop a general theory of software systems that accounts for their essential properties in a principled way Such a theory might very well be of genuine scientific interest, precisely because of its concern for general explanatory principles

Such a theory would characterize the notion "possible software system" and, in accordance with the parallelism mentioned above, could thus be taken equally as characterizing the notion "possible mind," just as the notion "possible grammar" as a device that generates structured strings of objects is viewed in linguistics as providing an abstract characterization of the notion "possible language " Given such a formal characterization of "possible mind," we might then be able to constrain it in accordance with known empirical facts to get a notion of "possible X mind," where X = "human" or any other species, just as linguists try to constrain grammars to get a characterization of "possible human language " The notion "possible human mind" might then be further constrainable in accordance with the idiosyncratic facts of an individual's culture and life experience, giving us an explanatory account of an individual human mind and its associated behavior The difference between this approach and the one current in AI would be essentially the same as that between generative and taxonomic linguistics Rather than starting from scratch, as it were, and building up programming systems ad hoc, we would be beginning with a *principled* account of the *sort* of entity we assume the human mind to be and then narrowing that account down in accordance with empirical facts to determine precisely *which* entities of that sort the mind really is

Hamilton and Zeldin (1976) argue, in effect, that the notion "possible software system" can be formalized in terms of three theoretical constructs – *data types*, the kinds of entities that systems operate on or produce; *functions* (or *operations*), the entities that operate on or produce the members of data types; and *control structures*, the relationships in accordance with which functions can be decomposed or combined – and that each of these constructs can exist on various *layers*, which are strikingly reminiscent, in concept, to the "levels of description" of generative grammar They also provide a formal methodology for representing these constructs abstractly, in terms that are entirely independent of a system's implementation in particular configurations of hardware or resident software (operating systems, etc) To the extent that their theory does, in fact, capture the notion "possible software system" (see Peters and Tripp, 1977, for some relevant remarks), we can presumably take it as equally a theory of "possible mind," and proceed to constrain it accordingly

Cushing (1977a, b, c), in fact, argues that the algebraic characterization of data types that is incorporated in Hamilton and Zeldin's theory (Cushing, 1978) provides a revealing model for the semantic lexicon of a natural language, as one component of the human mind The model incorporates an empirical claim as to where in the lexicon we would most naturally expect to find constraints, as part of a general characterization of the kind of subcomponents that make it up Cushing argues that the semantic lexicon is a heterogeneous algebra (Birkhoff and Lipson, 1970) and that such issues as the dispute over lexical decomposition vs meaning postulates receive a natural and revealing reformulation, when viewed in this light

We will not speculate here on how fruitful this kind of research might ultimately turn out to be, because that can be determined only by time and further work We do think, however, that something along these lines is a necessary prerequisite to a computer-based model of cognition It may, in fact, turn out that the mind is not a computational device at all and that entirely new concepts will have to be developed to account adequately for its operation Our point is simply that any adequate theory of mind will have to base itself firmly on the search for general explanatory principles, and that this applies to computationally-based theories as much as to any other

REFERENCES

Birkhoff, G. and J. D. Lipson (1970) "Heterogeneous Algebras," *Journal of Combinatorial Theory*, 8:115–133.

Cushing, Steven (1977a) "Lexical Decomposition and Lexical Algebra," presented to the MIT Workshop on Language and Cognition, February 1977.

Cushing, Steven (1977b) "Lexical Functions and Lexical Algebra: A Software Model for the Semantic Lexicon," presented to the MIT Workshop on Lexical Representation, December 1977.

Cushing, Steven (1977c) "An Algebraic Approach to Lexical Meaning," presented to the Annual Meeting, Linguistic Society of America, December 1977.

Cushing, Steven (1978) "Algebraic Specification of Data Types in Higher Order Software (HOS)," to appear in the *Proceedings, Eleventh Hawaii International Conference on Systems Sciences*, Honolulu, Hawaii, January 1978

Dresher, B. Elan and Norbert Hornstein (1976) "On Some Supposed Contributions of Artificial Intelligence to the Scientific Study of Language," *Cognition*, 4:321–398.

Dresher, B. Elan and Norbert Hornstein (1977) "Reply to Schank and Wilensky," *Cognition*, 5:147–149

Dresher, B. Elan and Norbert Hornstein (1978) "Reply to Winograd," to appear in *Cognition*.

Hamilton, Margaret and Saydean Zeldin (1976) "Higher Order Software – A Methodology for Defining Software," *IEEE Transactions on Software Engineering*, SE-2, 9-32.

Miller, George and P. N. Johnson-Laird (1976) *Language and Perception*, Harvard University Press, Cambridge, Massachusetts.

Peters, Lawrence J and Leonard L. Tripp (1977) "Comparing Software Design Methodologies," *Datamation*, November 1977, pp. 89–94.

Weizenbaum, Joseph (1976) *Computer Power and Human Reason*, Freeman, San Francisco.

by **Daniel C. Dennett**

Department of Philosophy, Tufts University, Medford, Mass 02155

Why not the whole iguana? I have no disagreements worth mentioning with Pylyshyn's paper, but would like to explore two comments of his

"There have been grand theoreticians in psychology in the past (e g , Freud, James, Hull) who have sought general principles with very limited success," Pylyshyn suggests, because they lacked "a powerful technical tool to discipline and extend the power of the imagination " And now for the first time we have the tool that might permit us to express and test at least *sketches* of unified cognitive theories of whole creatures, the sort of theories to which Freud et al aspired Moreover, as Pylyshyn observes, the users of that tool have come to a consensus of sorts that theories of the whole creature are what is needed:

"The recurrence of major problems of organization and representation of knowledge, and the organization and distribution of responsibility or control have produced the growing conviction among cognitive scientists that intelligence is not to be had by putting together language abilities,

sensory abilities, visual abilities, memory, motivation, and reasoning (as the chapters of typical psychology textbooks suggest) but by bringing a large base of knowledge to bear in a disciplined way in all cognitive tasks "

Very true, but then why have cognitive scientists persisted in attempting to model sub-subsystems with artificially walled-off boundaries (not just language understanders, but nursery-story-only understanders, for instance)? Why are they not trying to model whole cognitive creatures? Because a model of a whole human being would be too big to handle; people know too much about too many topics, have too many interests, capacities, modalities of perception and action One has to restrict oneself to a "toy" problem in a particular domain in order to keep the model "small" enough to be designed and tested at a reasonable cost in time and money But faced with the conclusions quoted above, why not obtain one's simplicity and scaling down by attempting to model a whole cognitive creature of much less sophistication than a human being? Why not try to do a *whole* starfish, for instance? It has no eyes or ears, only rudimentary pattern-discrimination capacities, few modes of action, few needs or intellectual accomplishments That could be a warm-up exercise for something a bit more challenging: a turtle, perhaps, or a mole A turtle must organize its world knowledge, such as it is, so that it can keep life and limb together by making real time decisions based on that knowledge, so while a turtle-simulation would not need a natural language parser, for instance, it would need just the sorts of efficient organization and flexibility of control distribution you have to provide in the representation of world knowledge behind a natural language parsing system of a simulated human agent such as SHRDLU

Perhaps there are good reasons for not pursuing such projects I suspect that one of the *real* reasons such projects are not pursued is that in order to design a computer simulation of a turtle you would have to learn all about turtles, and who wants to go to all that trouble, when you already know enough about yourself and your friends (you think) to have all the performance data you need for the human mini-task of your choice? Moreover, only people who also knew a great deal about turtles would be knowledgeable enough to be impressed by your results

Considering the abstractness of the problems properly addressed in A I (Dennett, 1978), one can put this attitude in a better light: one does not want to get bogged down with technical problems in modeling the cognitive eccentricities of turtles if the point of the exercise is to uncover *very* general, *very* abstract principles that will apply as well to the cognitive organization of the most sophisticated human beings So why not then *make up* a whole cognitive creature, a Martian three-wheeled iguana, say, and an environmental niche for it to cope with? I think such a project *could* teach us a great deal about the deep principles of human cognitive psychology, but if it could not, I am quite sure that most of the current A I modeling of familiar human mini-tasks could not either

REFERENCES

Dennett, D. C. Artificial Intelligence as Philosophy and as Psychology. In: M. Ringle (ed.), *Philosophical Perspectives in Artificial Intelligence.* New York, The Humanities Press, 1978.

by **Zoltan Domotor**

Department of Philosophy, University of Pennsylvania, Philadelphia, Pa 19174

AI: model-theoretic aspects. Many of the controversies affecting the foundations of artificial intelligence (A I) and cognitive science appear to be a result of attempts to answer too sophisticated questions too quickly Rather curiously, in A I it has become a custom to adapt and adopt insidious concepts (such as inductive inference, inference by analogy, self-knowledge, self-awareness, association, causality, and many more) to tell fancy stories about what "intelligent" programs do Marvin Minsky and Seymour Papert claim in their M I T memo No 252 [*op cit* Leibovic, this *Commentary*] that "[The A I] ideas pass a fundamental test that rejects many traditional notions in psychology and philosophy; if a theory of Vision is to be taken seriously, one should be able to use it to make a Seeing Machine!" Poor philosophers! Since they typically cannot read/write programs, they may never find out that many of their perennial problems have actually been solved by others

It is reassuring to read Pylyshyn's interesting paper suggesting that this sort of thing could not happen so fast to psychologists because they are still the supreme power in providing the necessary *empirical constraints* for *computational systems* Biologists should stay out of this enterprise, at least for the time being

What are computational systems and what are empirical constraints? This is just it! These concepts are quite vague and unstabilized, and so are the inferences and arguments based on them!

Let me try first to deal with the aforementioned concepts along the lines of traditional philosophy of science and then comment on some of Pylyshyn's claims and arguments Due to limitations on space and time, all I can do is give a sketch

By a scientific system (frame) one usually means (among philosophers) a (formal) theory together with a bundle of intended applications (intended models) Correspondingly, a computational system (hopefully) may be conceived as a (large complex) program accompanied with a family of intended interpretations Thus, for example, physicists endow mathematicians' differential equations with physical meaning by providing adequate intended physical models and, *mutatis mutandis*, A I specialists do the same with programs Put conversely, a physicist associates mathematical structures with physical systems or objects of his concern in order to tease out interesting theorems with physical relevance Again, cognitive specialists associate computational structures with their psychological entities *Empirical truth* is then understood as a composite of *mathematical/computational truth* within the intended models and the *empirical adequacy* of the models While one hardly ever questions the validity of theorems of mathematical physics, one often wonders about the appropriateness of the intended models Similarly, foundationally the possible bugs in programs are irrelevant; what counts is the adequacy of the intended models the A I programs are supposed to describe

The foregoing conceptual scenario remains incomplete until we bring in *models of data* and intersystem relationships, that is, *representations* of one scientific/computational system within another Models of data come with error and preprocessing structures whose discussion requires extra care (often neglected by A I experts) Representations are best understood as maps with two components The first (syntactic) component assigns to sentences/program units of the represented scientific/computational system unique sentences/programs of the representing system in such a way that logic, inference, and laws are preserved, perhaps at the cost of destroying the similarity types of the basic nonlogical components The second (semantic) component relates the represented intended models to the representing models By means of this sort of intersystem gadgetry one can understand many of the crucial relations between levels of description, such as reduction, equivalence, and passage from micro- to macro-theorizing

It seems to me that Pylyshyn's empirical constraints in cognitive theorizing are the physicist's or philosopher's intended models Indeed, if a cognitive scientist wants to write a program for simulating face recognition and whatnot, he has to pack his intended models with psychological structure How else could he claim that he is studying cognition? Thus in this case the claim that one needs empirical constraints is trivially valid Also, in this context the business of levels of description is no different from that in physics There is no need to create a new philosophy of science here

However, in A I theorizing, I will argue, Pylyshyn's claim regarding the necessity of empirical constraints supplied by psychology is doubtful To set the initial intuitions and inspiration on the right track, in A I one may start with psychological constraints, but along the way, as research progresses, these constraints become typically not only unnecessary but even irrelevant In a bootstrap fashion, A I structures start to take care of themselves with their own internal foundations, laws, and life, with no direct recourse to psychological limitations The most recent results in theory-formation programs and program-formation programs rely on artificial intelligence and programming experience, and not on philosophy or psychology Just as the first design of airplanes was strongly dependent on the dynamics of wings and flight of birds, and later the development of propeller and jet aircrafts, helicopters, and rockets went its own way with no connections to ornithology, A I takes a similar course of development

Pylyshyn seems to use the term "intelligence" too anthropocentrically I do not see why every form of intelligence should be measured by a psychological yardstick Specifically, I disagree with his highly qualified claim that "if a person and computer are both capable of 'doing task *x*,' there is some level of description of the two at which they are doing it 'in the same way' " How do you know, for example, that I am executing a multiplication 19×18 in the same way you do? It so happens that I simply remember

the result My A I colleague's younger son seems to use a safe guess You may use one of the problem decompositions $19 \times (18 + 1) - 19$ or $(19 - 1) \times 18 + 18$ or Computers could use something still different, in a binary system I do not see how we do it "in the same way" even after stretching the meaning of the "same " Mind you, this is just a jejune example!

I submit that there is an urgent need for reliable foundations of artificial intelligence, powerful enough to render secondary many of the philosopher's speculations on what computers cannot do, the psychologist's anxieties about A I 's intellectual imperialism, and the current high-flown theorizing of A I connoisseurs Naturally, such foundations should be established within A I itself It is encouraging to observe that several of the overused concepts such as knowledge and intelligence are currently undergoing rapid evolution and redefinition, with increasingly less relevance to philosophy and psychology Aristotelian physics was stuck with psychologically constrained concepts of the class of warm and heavy, and it took many centuries of complex abstraction to pass to physical structures that today have a life of their own The single most powerful example of this century is the Fregean separation of logic from psychology

Until the bootstrap effect takes over fully, we will be attracted to and tolerated in advising the A I fraternity regarding what to do and how to do it

by **Hubert L. Dreyfus**

Department of Philosophy, University of California, Berkeley, Calif 94720

Empirical evidence for a pessimistic prognosis for cognitive science. Pylyshyn's paper provides a plausible list of empirical constraints on computer models of human cognitive processes Among these constraints, evidence concerning the generality of a model's performance plays a crucial role when we come to assess the claim that a given computer model contributes to the understanding of human behavior I have no quarrel with the criteria Pylyshyn sets forth; what puzzles me is his claim that "there are many reasons for viewing the potential contribution of the computational approach with optimism " Pessimism seems to me more in order, since the particular programs Pylyshyn mentions all fail the generality test

The obvious obstacle to Pylyshyn's position is that the only unqualified successes in A I : Samuel's checkers program, DENDRAL, MYCIN, and MATHLAB are all examples of domain-specific knowledge engineering Pylyshyn, however, dismisses such an objection as merely "a comment about the more general phenomena that performance may be purchased at the price of generality," and adds that "some of the criticism directed against Evans's (1968) geometrical analogies system, Winston's (1975) learning system, Waltz's (1975) scene anaylsis system, or Winograd's (1972) language comprehension system is based on such a lack of generality " Using these programs as examples, let us see if the criticism is justified

To begin with, putting the objection in terms of *generality* misses the point It is not the specificity of these programs that disqualifies them as explanatory models of human psychological processes; it is the *non-generalizability* of their clever exploitation of domain-specific properties Waltz's program depends, as Pylyshyn notes, on the working out of the possible real world permutations of 11 labels for edges If we add one cylinder or cone to the scene, the whole program breaks down For this very reason the success of the program provides no evidence that edge intersection analysis plays a role in human scene perception What makes such a suggestion implausible is not, as Pylyshyn would have us believe that Waltz's program has purchased power at the price of specificity and so lacks *generality,* but rather that the program has clearly exploited specific features of rectilinear objects, and so cannot be *generalized* to other sorts of scenes

Waltz wisely never claims psychological relevance for his work, but Winston does In fact, he claims that "learning requires in someone the same skills illuminated in [my] theory" (Winston, 1975 *op cit*) But Winston's program suffers from an even more damaging sort of non-generalizable specificity than does Waltz's Whereas in respect to Waltz's program Pylyshyn can consistently, if implausibly, hold that, after all, scene perception may be the result of a combination of "a large system of specific mechanisms," the Winston program works only because it has excluded from its task domain the very ability it is supposed to explain The program can "learn" a simplified geometrical concept like arch only if the programmer makes explicit, pre-selects, and pre-weights a small set of relevant features such as "left-of," "standing," "supported by," from which the program can then build its description But there is no clue as to how the program could be extended to cover this essential work of discriminating, selecting, and weighting Again the problem is not lack of *generality,* but rather the fact that the original success depends on tricks which preclude generalization to essential aspects of the task domain

It is a fact worth pondering in this connection that when they originally presented their programs Evans, Winston, and Winograd all suggested that the methods they proposed could be generalized (and Winston has made the same claim for Waltz's program), yet in no case has the generalization been forthcoming Among these A I researchers, only Winograd subsequently addressed himself to the issue of generalizability, and his conclusion agrees completely with my critique rather than with Pylyshyn's defense

> "Current systems, even the best ones, often resemble a house of cards The researchers are interested in the higher levels, and try to build up the minimum of supporting props at the lower levels The result is an extremely fragile structure, which may reach impressive heights, but collapses immediately if swayed in the slightest from the specific domain (often even the specific examples) for which it was built " (Bobrow & Winograd, 1977)

Pylyshyn's argument that from even the most domain-dependent program *something* can be carried over to others, namely that "researchers find that they are preoccupied with problems of how to represent task relevant knowledge and how to organize control so that relevant portions of this knowledge are brought to bear when it is appropriate" only shows that all A I programs sooner or later run up against the same wall This is about as encouraging for the unity of Cognitive Science as the fact that alchemists, trying to distill gold from dirt, all became preoccupied with making hotter and hotter furnaces That this persistence in thinking all that was needed was more heat control led to steady progress in the development of heat resistant retorts may well be true and gratifying, but hardly shows that those seeking to use heat to transmute the baser metals were on the right track Likewise, the assumption that a large knowledge base and better control of the resulting mass of facts holds the key to success in A I may lead to better data structures, but may hide the fact that many human capacities are involved in intelligent behavior, and that the attempt to represent what human beings perceive, imagine, feel, desire, and skillfully do as a body of rule-governed facts may be totally misguided

Pylyshyn's point that physics flourished when Galileo discovered the appropriate formalism for describing physical motion provides no grounds for disregarding A I 's difficulties, since the question is whether intentional behavior, involving as it does self-interpreting entities acting in concrete situations, can be captured in any abstract set of rules And it only begs the question at issue to assert that the pioneers in A I *"recognized* that the study of symbol processing in computer science and attempts to understand the nature of intelligent behavior were at some level inseparable" (my italics) Precisely what is to be tested is the *hypothesis* that intelligent behavior can be understood as "symbol processing "[1] Yet Pylyshyn holds that "the relevance of computation to cognition is that both cognition and computation are intentional *rule-governed phenomena*" (my italics) Only such a question begging assumption could account for Pylyshyn's optimism in the face of A I 's persistent failure to meet his first, and most important, empirical criterion

NOTE

1. It is to Newell and Simons' credit that in their article, "Computer Science as Empirical Inquiry," cited by Pylyshyn, they formulate this precise point, not as a fact to be recognized, but as an hypothesis to be tested.

REFERENCE

Bobrow, D. and Winograd, T. An overview of KRL, a knowledge representation language, *Cognitive Science,* 1977, 1:4.

by **Helen Goodluck**

Department of Linguistics, The University of Massachusetts, Amherst, Mass

Levels of evolution and psycholinguistic evidence. In commenting on Pylyshyn's paper, I will direct my remarks toward some of his observations on AI research mainly as pertains to the study of natural language, since this is the area with which I am familiar

Pylyshyn's review of some of the relevant parameters for evaluation of AI

systems is particularly valuable in that it contains a clear statement of a problem that I think is the basis for some of the reservations linguists and psychologists may have about the contribution made by AI research to their field(s) (For a critique of some AI systems for natural language processing, see Dresher and Hornstein, 1976)

Pylyshyn observes that: "The issue of the appropriate level of description at which computational systems are to be evaluated remains a serious problem in all computational models " As Pylyshyn notes, certain levels (such as the mechanical operation of computers or the "program per se" are clearly inappropriate to evaluate an AI system as a representation of a cognitive function However, just what the appropriate level for evaluation is remains undetermined The reason that this indeterminacy may give pause to a linguist or psychologist is that it makes it difficult to evaluate exactly what part of the contribution made by AI research to the elucidation of a problem in a given area results from the application of principles particular to the discipline of AI rather than from general research strategies common to AI and other disciplines

For example, Pylyshyn makes the following comments on AI language comprehension systems: "Systems for language comprehension are constrained by the empirical facts about the structure of the language, the structure of the world, and the structure of cognitive systems which use language Together, these constrain the possible form which a computational comprehension system can take Furthermore, the attempt to build such a system is instrumental in discovering these constraints so that A I also provides a methodology for discovery " The first two sentences of this quotation are fairly uncontroversial; not just an AI system, but any attempt to model language comprehension will be constrained in this way However, it is not clear that the claim that AI provides a "methodology for discovery" amounts to more than the statement that study of the data of natural language by the researcher may lead him to discover previously unnoticed phenomena, i e , to more than the observation that "nothing quite concentrates the mind as having to build such a model [as one in the form of a computer program]" (Johnson-Laird, 1977, p 212, and this Commentary) Although progress has been made in modeling natural language phenomena familiar through the work of linguists and psychologists, little has been discovered about natural language in the course of AI research on systems modeling language comprehension However, suppose that such discoveries are made Without a well-defined level of evaluation for AI systems, it is difficult to know whether a discovery is the direct result of the system of analysis If this level turns out to be one that defines concepts that play a crucial role in discoveries made in the course of AI research, and these concepts are distinct from those available within other disciplines that treat the same data, then AI will indeed have some special claim to furthering the study of cognition

The lack of a well-defined level of evaluation for AI systems does not in principle make the task of evaluating the potential of a system as a source of predictions and discovery an impossible one One approach might be to look at the extent to which the success of a computational system in simulating cognitive phenomena results from properties of that system not shared by other computational systems For example, experimental evidence from psycholinguistics suggests that while there appears to be some support for parallel processing as a factor in the comprehension of semantic ambiguity involving the interpretation of grammatical relations such as 'subject' and 'object,' there is considerably less evidence for parallel processing of syntactic ambiguity (For a review of some of the literature, see Fodor, Bever, and Garrett, 1974, Ch 6, pp 361–367 and fn 4) A model for sentence comprehension must reflect this distinction in some way Suppose that of two AI models for language comprehension one reflects the distinction between the processing of semantic and syntactic ambiguity, and the other does not, or can do so only in a complex way, and that this follows from differences in the computational design (rather than differences that could equally well be features of non-computational models) The model that reflects the natural language phenomenon would be a serious candidate as a source of discovery in language processing insofar as it proved a superior model for some known phenomenon A fruitful approach for AI research, and hence for the areas of cognition that it deals with, may be the comparison of properties of AI systems that accurately simulate known characteristics of cognitive systems with those that fail to do so, or can do so only by resort to complex ad hoc devices From such a comparison, the correct level of evaluation of AI systems that Pylyshyn observes to be lacking may be approached

REFERENCES

Dresher, B. E. and Hornstein, N. On some supposed contributions of artificial intelligence to the scientific study of natural language. *Cognition*, 1976, 4:321–398.

Fodor, J. A., Bever, T. G. and Garrett, M. F. *The Psychology of Language*. New York: McGraw-Hill, 1974.

Johnson-Laird, P. N. Procedural semantics. *Cognition*, 1977, 5:189–214.

by **Leon D. Harmon**

Department of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio 44106

Introspection, black boxes, and machine equivalence. John von Neumann noted that once *any* process can be clearly and unambiguously described, a computer program can be written to represent that process to any desired degree of accuracy and completeness But the rub lies in his further opinion that the primary language and logics of the nervous system must be structurally different from the symbolic manipulations we formally use in our conventional analysis (von Neumann, 1958) If so, one should not be surprised that in many cases our external languages permit only crude and artificial approximations, which result in but partial success in modeling the deeper processes

Because of von Neumann's first point, Pylyshyn's question, "In particular, can a program be a psychological theory?" is answered affirmatively However, that is a relatively uninteresting and trivial question compared to three other matters that Pylyshyn discusses and upon which I wish to comment: one relates to serial *vs* parallel processing; a second is concerned with modeling levels and black-box equivalence; the third addresses modeling constraints imposed by introspection

With respect to the first matter, Pylyshyn states that it is inappropriate "to criticize a serial algorithm on the grounds that *in the brain* various events are taking place in different locations at the same time " I disagree In a sense Pylyshyn's point is defensible; for computation as it is usually understood, the serial-parallel conflict ceased to exist theoretically with Turing's proof (Turing, 1937) that a single tape manipulating one character at a time could compute any computable number However, for our present purposes it is not so important to consider computable numbers as to consider whether or not some theoretical or practical constraints might apply differently in the two (serial, parallel) cases

Two considerations follow: 1) In principle the brain may do more than or be different from a simple computer; thus the question is open as to whether a serial digital machine is adequate to imitate the brain (if that is our aim) 2) Parallel processes of sufficient complexity may require serial representation that is so extensive as to be absurd It may be that the lifetime of the universe is not a long enough period to compute serially all the operations of a brain's lifetime (assuming serial digital equivalence were possible)

Consider a simple estimate of the serial digital computation requirement for simulating a small portion of the brain The cerebellum constitutes about 10 percent of the central nervous system in man According to present estimates, this is 10 percent of roughly 10^{11}, or about 10^{10} cells. A small fraction of those cells (about one ten-thousandth) constitutes the Purkinje cells These 10^6 Purkinje cells are estimated to have as many as 300,000 individual synaptic inputs each

Suppose now we form a rough upper-limit estimate for the number of pulses (idealized, simplified spikes) that could occur and should be handled in a century of brain function in our serial representation For the Purkinje-cell inputs alone we have:

3×10^{11} connections $\times$ 10^2 pulses/sec per connection
$\times$ 3×10^7 sec/yr
$\times$ 10^2 yr/century $\simeq 10^{23}$ pulses/century

Now if we take a conservative computer simulation rate of 10^7/sec, then we will require about 10^{16} sec or 3×10^8 years just to represent these pulses Notice that we have considered *only* the input signal traffic We have not mentioned continuous variable, non-linear interactions within each cell, or output considerations, to say nothing of specifying the interconnection matrix Note, too, that all of this, in addition to the nearly 10^9 years needed for simplified input signal representation, would be only for 10^6 cells out of 10^{11} – one part in 100,000 It thus follows that the entire lifetime of the universe may not be sufficient to compute the lifetime action of one brain by serial digital techniques

It seems reasonable to suppose that for many modeling representations of the brain, serial algorithms themselves are wholly inappropriate Of course for representation of *some* processes (notably the conscious introspective step-by-step plodding thinking we do in overt problem solving), serial techniques can be fine But I suspect that the really interesting and potent CNS operations (like pattern recognition, language synthesis and analysis, creativity, emotions, insight, etc) which are manifestly unavailable to conscious introspection, are beyond the reach of our serial analysis and/or possibility of representation Finally, we appear to be in the unfortunate bind that we almost totally lack parallel computing concepts or theory So far we have been able only to visualize relatively simple parallel multiples of serial processors That is probably far removed, indeed, from the few-step, great-parallelism processes that appear to work in nervous tissue

Pylyshyn's second and third topics that I wish to comment upon, black-box equivalence and introspective constraints, are sufficiently related that discussion together seems appropriate

Pylyshyn correctly observes that systems designed to do a particular task in, say, artificial intelligence, need not do so in the same way that people would Obviously if the automaton successfully replicates a particular human behavior, there is black-box equivalence *at that level* Now what is in that black box at some much lower level could be electronics, mathematical expressions, wheels and levers, or copper pipes and oil, but most certainly not living neurons The issue of interest, of course, is whether at some intermediate level – say, gross algorithmic or functional subsystem – there may be similarity or even equivalence

I agree with Pylyshyn's view that there are many levels of modeling comparison which are inappropriate, but outside of the grossest (behavioral) level, I feel (unlike Pylyshyn) that most, if not all, comparisons probably are inappropriate This owes principally to the fact that we really do not know much, if anything, about our internal processes True, we can make statements about our beliefs, percepts, goals, and so forth, but what can we elucidate about our esthetics, speech and scene analysis, or language synthesis, for example?

Thus I take issue with Pylyshyn's subsequent statement that any device doing a certain task *x* must be doing it in the same way that the person does it This disagreement takes into account Pylyshyn's proviso that " if both person and computer are both capable of doing task *x* there is some level of description of the two at which they are doing it in the same way " I do not see how, aside from tautology, such an assertion can be defended A single counter-example will suffice: a computer can be programmed to recognize multi-font typewritten alphanumeric characters in many different ways The literature abounds with examples Clearly at *most* one of these systems does it "in the same way " All others obviously do it differently And I doubt that even one comes close to modeling human procedures

Pylyshyn's observations on the utilities (and dangers) of empirical constraints in model making are well taken Without such constraints infinite numbers of models are possible Unfortunately, if the numbers of degrees of freedom or states of a system are large compared to the numbers of constraints on those states obtained through finite observation, an infinite number of models is still possible One thus is well advised to view models with restraint and nervous systems with humility

REFERENCES

von Neumann, J. *The Computer and the Brain*. New Haven: Yale University Press, 1958.

Turing, A. M. On computable numbers with an application to the Entscheidungs-problem, *Proceedings London Mathematical Society, 2(42)*, 230–265, 1937.

by **John Haugeland**

Department of Philosophy, University of Pittsburgh, Pittsburgh, Pa 15260

The problem of generality. One can "model" just about anything, from economic cycles to weather patterns, on a suitably programmed computer; and such models can be a powerful aid to scientific investigation But psychological modelling seems special Whereas there is no danger of any computer containing *actual* market crashes or cyclones, getting one to display *actual* intelligence is the whole idea So psychology is unique, in that studying appropriate computing systems can be thought of as studying "the real thing," though modified and perhaps simplified in various ways Pylyshyn, I think, takes this view, summing it up with the remark: " both cognition and computation are *intentional rule-governed phenomena* " Accordingly, he also says that artificial intelligence and cognitive simulation differ in "little more than style of research," and that " even 'pure' A I can hardly avoid making some contributions to cognitive psychology " A I and C S are both studying cognition and intelligence, but they differ in which modifications and simplifications they will tolerate at the outset

This is an elegant position, presented with a wealth of examples and insightful observations, but I think problems emerge when we look closely at the supporting arguments First, since not all of computer science counts as "cognitive science," one wants to know why artificial intelligence research does Pylyshyn anticipates this question in his section "Responsiveness to empirical constraints " The gist of his long reply is that "intelligence" (or "task requiring intelligence") is an anthropomorphic notion, and that an integral part of that notion concerns the relevant "task environment " So one could hardly build any intelligent artifact without *ipso facto* discovering a lot about *human* intelligence, including, at least, the task environments that determine human cognition

Incredibly, however, the only specific citation in this section is to "Waltz's (1975, *op cit*) success in designing a system for parsing a scene consisting of polyhedra with shadows " But Waltz's system "succeeds" only by exploiting tricks that are utterly idiosyncratic to polyhedra. It is as if I built a machine that "visually" (that is, optically) "parsed" scenes of mixed fruit by analyzing the distinctive fine structure of their absorption spectra Would my successful gimmick have to count as a discovery *on any level* about fruit identification by people (or its "task environment")? If not, then why do Waltz's gimmicks fare any better? The mere fact that people can recognize polyhedra and fruit on sight does not prove that either of these machines has anything whatsoever to do with psychology; nor does it make any difference that one was put together in a room labeled "A I Lab " A theoretical discussion of what would make a difference would just bring back all the issues about performing "in the same way as people do," which Pylyshyn is trying to downplay

Second, most A I programs employ "gimmicks" of one sort or another, that is, techniques that are utterly ungeneralizable because they depend entirely on the peculiar quirks of a special class of cases Pylyshyn has this in mind when he introduces his "qualitative discontinuity principle" in the "power-generality tradeoff"; and he acknowledges that it "makes some computational systems seem implausible as psychological models " But he goes on to say that the computational approach is on a comparatively secure footing anyway because no matter what A I workers try to do, they find themselves preoccupied with the same problems: "how to represent knowledge and how to organize control " Presumably this is supposed to reassure us that, after all, certain fundamental principles do transcend those awkward qualitative discontinuities But it sounds to me like defending witchcraft on the grounds that, no matter how diverse the projects, the same problems of incantation and spell-casting recur *Of course* the computational approach leads to a preoccupation with representation and control – that is practically a definition of it I will not be reassured until there is some *independent* reason to believe that gimmickry can be transcended

Finally, there is a problem about how exactly A I and C S are supposed to differ (in research style or whatever) In his last four paragraphs, Pylyshyn suggests that A I is a "top-down" approach compared to C S , which is more "bottom-up" (though still not as bottom-up as the approach of traditional experimental psychology) A top-down approach is one that seeks "some degree of completeness over a broad domain," that is, an account that roughly captures the general features in a wide spectrum of data, even at the expense of blurring local details Or, the relative emphasis can shift, until at the opposite (bottom-up) extreme, you have many locally impressive but narrowly isolated "micro-models "

The trouble is that no A I system has ever displayed even a hint of broad generality; *at best,* they are impressive in narrowly isolated domains (and Pylyshyn admits as much when he says they purchase performance at the price of generality) Why this contradiction? Well, if A I is to be on a common scale with empirical psychology at all, it has to be at the opposite end, because it self-consciously spurns traditional psychological data Thus it can offer *at most* broad general principles If the computational approach to psychology eventually pans out (and we all agree that this remains to be seen), *then* the top-down/bottom-up taxonomy will be vindicated By the same token, however, the failure so far of A I to have any broad general suc-

cesses, or even to come up with any interesting generalizations beyond its own premises, must be counted as *prima facie* evidence against the approach

by **P. J. Hayes**

Department of Computer Sciences, University of Essex, Colchester, Essex C04 35Q, England

Doing AI but saying CS. This is a most stimulating paper, full of insightful remarks I wish I had written most of it But I think Pylyshyn doesn't quite carry his main thesis, which I take to be that cognitive simulation (CS) and artificial intelligence (AI) are really only different styles of doing the same sort of empirical research

I will urge a slightly different view, one which I think accords better with the way in which many workers in AI and CS view their own activities While agreeing with Pylyshyn that AI problems are defined by cognitive or psychological criteria, by and large (although it is difficult to say quite what 'problem' Lenat's [1977; see also this *Commentary*] program is solving, for just one significant example), I suggest that there is a crucial difference in the kinds of hypotheses they test Pure AI hypotheses have the form 'this behaviour *can be* realised by the following computation ' Pure CS hypotheses have the form 'this behaviour, of this organism, *is* realised by the following computation' Both are hypotheses relating behaviour to computation (in a sufficiently broad sense) and both are empirical But the kinds of empirical test they are subject to are different AI hypotheses are tested by implementing the algorithm and seeing how well it works – Pylyshyn correctly notes that running a program is often a real experiment CS hypotheses are more difficult to test, as Pylyshyn's insightful discussion of some of the rather weak and inconclusive techniques illustrates It is important to realise that his three 'sources of empirical constraint' on computational theories apply only to CS hypotheses, not to AI hypotheses Thus, I suggest, the differences between AI and CS are not mere side-effects of divergent methodological allegiances, but reflect a real difference between the kinds of hypotheses being tested, between what AI-ers and CS-ers are trying to do AI gets its problem statement from psychology, but its criteria of adequacy and success from engineering and computer science; CS, on the other hand, is wholly a branch of psychology The 'levels of description' problem, noted by Pylyshyn, crops up in both areas: it concerns exactly what is meant by 'the following computation:

I have sketched the 'pure' AI and CS formulations, and examples of both can be found in the literature But many workers explore various sorts of compromise or mixed position: AI types often use introspection or psycho-experimental results as at least a source of inspiration; CS types sometimes use, explicitly or implicitly, an argument of the form: 'this algorithm works, and I can't think of any other that would; so there aren't any others that work; so this must be the way the organism does it', to justify their AI activity by CS criteria More interestingly, there is a conscious agnosticism embraced by many AI workers, of the general form that since our knowledge of, say, human cognitive processing is so poor, it is not worth distinguishing the two kinds of hypothesis: to discover how people work, we might as well just ask how they *possibly could* work Something like this view is the most popular among the AI community It is reflected in Pylyshyn's more sophisticated, and I think correct, idea, that CS progress may have to wait until more top-down constraints on what computations could *possibly* mediate cognitive activity are made available from progress in AI CS is, after all, much harder than AI To discover how people work, maybe we *have* to first discover how they could possibly work

There is a stronger thesis, that to do AI is to do CS already, since there could be only one way to implement cognitive behaviour as an algorithm One who embraces this view can regard himself as investigating the abstract principles of intelligence I think I see something of this view echoed in Pylyshyn's early remark: 'I see no compelling reason to believe there need be any systematic difference between systems designed purely as AI artifacts and those designed as cognitive simulations' I tend to agree, as do many other AI workers, but it is important to realise that this *is* a nonobvious thesis and *a priori* even rather unlikely After all, it is a commonplace of computer science that a given behaviour can often be realised by a variety of algorithms, each of which can certainly be implemented in a variety of ways I wonder why it seems so plausible? Is it – horrid thought – a rationalisation of the inner conflict that we want to do AI (it's easier and more fun) but we want to be seen to be doing CS (it's more respectable)?

REFERENCE

Lenat, D. (1977) AM: an artificial intelligence approach to discovery in mathematics as heuristic search. Memo SAIL AIM-286, Stanford University.

by **P. N. Johnson-Laird**

Department of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England

The correspondence and coherence theories of cognitive truth. Philosophers often distinguish two theories of truth: the correspondence theory and the coherence theory An assertion is true according to the correspondence theory if it corresponds to some state of affairs in the world It is true according to the coherence theory if it coheres with some set of assertions constituting a general body of knowledge The same sort of distinction appears to underlie the methodological difference between experimental psychology and artificial intelligence Psychologists want their assertions to correspond to the facts; intuition is a notoriously fallible guide to the facts; hence, the truth is best revealed by forsaking the armchair for the laboratory and undertaking controlled and detailed observations Artificial intelligencers want their assertions to fit together in a coherent and comprehensive way; intuition is a notoriously fallible guide to consistency; hence, the truth is best secured by forsaking the armchair for the computer and the construction of large-scale computer programs A viable cognitive science, however, needs theories that both cohere and correspond to the facts Clearly, some sort of rapprochement between experimental psychology and artificial intelligence is required, but how is it to be effected?

The answer according to Zenon Pylyshyn is that psychology should move in the direction of artificial inteligence He argues that its theories are already empirically constrained and that the detailed constraints discovered in the laboratory will not lead to a science of cognition until some general unifying theoretical principles have been established at the computer console It is true, of course, that computers have had a scant effect on the course of psychological theorizing (with a handful of striking exceptions) Unfortunately, it is also true that reasoned argument is seldom responsible for an abrupt change in behaviour, not even in the methodological habits of scientists Precept is less powerful than example Hence, much as I should welcome a growing adherence to Pylyshyn's principles on the part of psychologists, I am not sanguine about its likelihood

Psychologists are suspicious of the idea of developing their theories in the form of computer programs Given the current state of knowledge, such programs inevitably involve a large number of *ad hoc* and simplifying assumptions They may embody empirical constraints, but they often lack empirical consequences They do not yield the sort of predictions that can be tested in the laboratory, and tests are usually otiose since the program can generally be seen to be false in psychological terms Consider the fate of old AI programs They are not refuted At best, their cleverest ideas are ripped out of them and embodied in more advanced systems; at worst, they persist as hulks drifting in an intellectual vacuum Why should psychologists devote a massive number of man-hours to constructing large, complicated and obviously erroneous models when there are plenty of other workers in the AI fraternity who are prepared to do so? This seems to be an unanswerable objection

Yet, I do not believe that the moral is that psychology should continue in its present ways There is a desperate need, as Pylyshyn emphasises, for unifying theoretical principles After twenty years, the experimental study of human information processing has not yielded them, and seems unlikely to do so How should we proceed?

Pylyshyn draws our attention to the matter of description: one does not criticize a model of a chemical molecule because, unlike the molecule itself, it is not edible He argues that what is needed is "something approaching a theory of the program; a description of the system which highlights the general principles underlying its operation " Of course, there is a perfectly good medium for describing such principles, the ordinary everyday language of psychology After all, much of our knowledge of AI programs comes from reading such descriptions of them Cynics sometimes say that these descriptions often go beyond what is actually embodied in code In fact, I

consider this alleged shortcoming a virtue – or, rather, I take it as the central commandment of the following set of methodological principles for psychology

Psychology needs general theories, and they should be developed and couched in the vernacular of the discipline There is an awkward problem of scale here, but one can only hope that psychologists will recover some of the comprehensiveness of the illustrious founders of the subject Such theories naturally tend to be vague No matter Explicit models of parts of them should be developed in the form of computer programs The primary aim of such a program should be neither to simulate human behaviour nor to carry out a difficult task by the ingenious exercise of artificial intelligence On the contrary, the point of the program should be to develop the general theory Hence, only a small part of the theory should be tackled at any one time, and the program should be small-scale and easy to modify It should embody principles and eschew *ad hoc* patches, or at least allow the theory to be easily discerned In my experience, the development of such programs is a truly dialectical process, which leads to revisions in the general theory, and which can even suggest experimental tests of general theoretical principles (e g , see Miller and Johnson-Laird, 1976; Johnson-Laird, 1977; Steedman and Johnson-Laird, in press; Johnson-Laird and Steedman, in press)

Such an approach will not supplant artificial intelligence: there are some discoveries that can probably be made only by developing large-scale programs However, experiment and computer programs offer at best limited methodologies, and it is unlikely that either on its own will elucidate the nature of human mentality The experimenter's concept of truth contains the latent danger of his becoming a Gradgrind, whose only concern is to establish the facts The programmer's concept of truth contains the latent danger of his becoming a Flat-earther, whose only concern is to maintain the internal consistency of his ideas Like Pylyshyn, I believe that our best hope is to bring the two methodologies together Unlike Pylyshyn, I believe that this goal is best achieved by cutting programs down to the psychologist's size

REFERENCES

Johnson-Laird, P. N. Psycholinguistics without linguistics. In N. S. Sutherland (Ed.) *Tutorial Essays in Psychology. Vol. 1.* New Jersey: Erlbaum, 1977.

Johnson-Laird, P. N. and Steedman, M. J. The psychology of syllogisms. *Cognitive Psychology,* 1978, 10:64–99.

Miller, G. A. and Johnson-Laird, P. N. *Language and Perception.* Cambridge, Mass.: Harvard University Press; Cambridge: Cambridge University Press, 1976.

Steedman, M. J. and Johnson-Laird, P. N. A programmatic theory of linguistic performance. In P. T. Smith and R. N. Campbell (Eds.) *Proceedings of the Stirling Conference on The Psychology of Language.* London: Plenum, in press.

by **K. N. Leibovic**

Department of Biophysical Sciences, State University of New York at Buffalo, Buffalo, N Y 14214

The problem of validation. Computers not only compute but also simulate In Artificial Intelligence (A I) computers can be made to simulate "learning," "choice," and other "intelligent" operations as defined by the programmer We must define in advance (and possibly interactively) the programs that the computer is to execute, including random moves, trials, optimizations with respect to criteria, and so forth

The computer per se is not endowed with intelligence, unlike living things (at least a few) Per se computer "intelligence" can serve no function, while biological intelligence has a paramount survival function It is surely of interest to study the behavior of apes and men in a comparative way It is of no interest whatever to investigate how a computer could "run a maze for food reward" except as a means for evaluating preconceived hypotheses or possible predictions of a model generated with or without the aid of a computer

Thus, the computer is basically a different "animal" from the one whose intelligence we wish to study. Moreover, the computer is constrained by its language and its rules of operation as expressed in the programming rules These are different from the language and operation of animals and men, although a correspondence may be set up with the aid of a model Therefore, if we can discover general principles of information, representation, and control in A I we may ask whether analogous principles operate in the brain, although there is no a priori reason why such a correspondence should exist

The preceding arguments imply that when a computer and a person "do a task *x* in the same way" this can only be meaningful with respect to the equivalence of computer and person in a specific model It is stated that the "clearest examples of discovery" are those where A I interfaces directly with the environment, for example, in speech or pattern recognition or sensorimotor coordination As Pylyshyn says, one can discover through simulation (e.g , Waltz, 1975) "that with a certain set of labels the constraints (are) so great that a single correct analysis is mandatory " But in what sense is this different in principle from discovering by simulation that within certain parameter ranges a differential equation has oscillatory solutions? In both cases our discovery can tell us something about psychology only if the correspondence between the model and biological intelligence is valid It is precisely this problem of validation that qualifies any conclusions about psychology when a computer does "a task X in the same way" as a person

Psychology is in crisis: Freud and his contemporaries produced for us great new insights that were, however, qualitative and not universal, unlike physical "laws " Psychology has been reaching for quantification through statistics and computation Unfortunately, the tools as such demand, but do not of themselves yield, meaningful quantification A quantitative psychological theory cannot exist without the appropriate quantitative data of the biological substrate Once these are known, computer simulation and A I can make a significant impact Without these data, the value of A I can be to suggest possible viewpoints and, perhaps, negative existence theorems (Minsky & Papert, 1969; Leibovic, 1976)

In conclusion, therefore, the distinction between A I and biologically valid cognitive simulation is more than a matter of style.

REFERENCES

Leibovic, K. N. Brain Models. *Encyclopedia of Computer Science and Technology,* vol. 4. New York, Marcel Dekker, 1976.

Minsky, M., and Papert, S. *Perceptrons.* Cambridge, Mass., M.I.T. Press, 1969.

by **D. B. Lenat**

Department of Computer Science, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pa 15213

On astrophysics and superhuman performance. A close friend of mine, an astrophysicist, came to stay for a few days recently As I observed him work, I found to my astonishment that he *sat at a desk* and – what is even more significant – he *used pencil and paper* to carry out his research, even as I do Clearly there must be some deep commonality between AI and astrophysics, if we employ such similar experimental tools Under subtle questioning, he revealed that his research was, just as mine, continually evaluated against a large set of observed data; i e , both fields are subject to *empirical constraints*

This parody could be continued, but the lesson is clear by now: surface similarities are often merely that The author sees "no compelling reason" for a difference between pure AI artifacts and cognitive simulation systems Such a difference is fundamental: each system is constrained to fit a (somewhat idealized) model of a particular kind of information processing system In the cognitive simulations, this is of course a model of the human IPS (involving STM, LTM, their peculiarities as represented by psychological data, etc) In AI artifacts, the analogous constraint is much weaker and indicates tailoring the system to the underlying "machine" (computer + language) architecture – and *its* peculiarities As the architectures are radically different, so are the classes of algorithms that may be taken as primitive, as natural, as *feasible* I am not quibbling about "implementation details," but rather fundamental distinctions that can and must be drawn at all but the highest (and most vacuous) levels of description At any nontrivial level, man and machine are *not* doing X in "the same way "

This difference manifests itself again when we ask which tasks are suitable for investigation, and what are the criteria for success In cognitive simulation, the goal is to match human performance – including human error and imperfection In AI, there is no corresponding ideal: distinctly superhuman performance is much more desirable than precisely human performance For example, my research centers about automating the discovery of powerful new heuristics useful to mathematicians This is an activity which can be

done only by the very best mathematicians, and even then only rarely (on the order of once in several decades) The negligible number of humans capable of performing it makes it no less attractive to attack by AI Of course it would not be fit for cognitive simulation, since it is not what the author calls "a natural kind" of task

The motivation for actually writing computer programs is different as well The psychologist attempts to validate a theory, and he builds his program solely to *run* it The AI researcher, however, builds his program to *experiment* upon it; it is his species of laboratory animal He probes, mutates, exacerbates – and observes the resultant effects This experimentation is oriented toward discovering the sources of the apparent intelligence exhibited, pointing the way toward ever more powerful (not necessarily more human-like) mechanisms and system designs In this regard, AI is much more an empirical science than cognitive simulation

There may be much that the two fields can benefit from sharing, even if their commonality is not quite so deep as some of us desire After all, I still get along splendidly with the astrophysicist

by **Christopher Longuet-Higgins**

Center for Research on Perception and Cognition, Laboratory of Experimental Psychology, University of Sussex, Falmer, Brighton BN1 9QG, England

On describing cognitive processes. There is practically nothing in Pylyshyn's article with which any reasonable man could disagree, but this will not stop unreasonable men from doing so In particular, he is obviously right to stress the distinction between different levels of description of intellectual processes, whether these are taking place in someone's mind or inside an electronic computer Even in the computational case it is imperative to distinguish among (1) what David Marr [*oper cit*, Ullman, this *Commentary*] describes as the "method" by which a given problem might be solved, (2) the computer program in which the method is embodied, (3) the machine code instructions into which this program is translated by the compiler, and (4) the (virtually indescribable) physical events inside the hardware of the computer that accompany the actual running of the program Each of these has its analogy in the description of a human intellectual process: (1) at the level of "method," a very direct comparison may be possible, for example in the case of arithmetical multiplication, which either a human being or a computer can perform either straightforwardly or by looking up logarithms (2) is rather more problematical: computer programs are directly available for inspection in a sense in which mental routines clearly are not Pylyshyn is therefore quite right to stress that a computer program, which is a putative simulation of a human intellectual process, cannot itself be regarded as a theory of that process (I will come back to this point later) When we come to (3) and (4), the cognitive psychologist must at the moment admit to being out of his depth Plainly there is a sense in which the patterns of pulses travelling to and fro in the computer have their analogue in the nerve impulses travelling through the brain, but any resemblance between the detailed patterns must be purely coincidental As for (3), it is a moot point whether one should attempt to find any biological analogy to the concept of a machine code; possibly the routines stored in the cortex (or wherever) might qualify for such a comparison, but we have no idea of the representation in which these routines are actually stored

Perhaps the most immediate challenge to the artificial intelligence worker who is attempting to describe the effective procedure by which the human being solves a given intellectual problem (having chosen, intentionally or not, his method of doing so) is to discover an appropriate language in which to specify the detailed processes – be they serial, parallel, or both – that mediate cognitive tasks in general No existing computer language could qualify for this purpose, but some high-level languages seem to capture, better than others, certain essential features of human cognition Thus LISP is plainly a more interesting language from this point of view than FORTRAN, say; there is nothing in FORTRAN that corresponds to the dual nature of a LISP expression, which can be regarded either as a piece of text to be manipulated or as a piece of program to be run In any convincing account of human thought a similar distinction will probably be crucial John McCarthy's [cf this *Commentary*] insight into this matter is still, in my opinion, the most important contribution that has yet been made by computer scientists to our understanding of information processing in general But clearly one needs to go very much further, not least in the direction of clarifying our ideas about parallel processes, and the extent to which they must be kept in separate compartments if deadlocks and similar disasters are to be avoided It may well be that when we have a much better understanding of the semantics of natural languages we shall find that something quite similar to natural language is the best vehicle of expression for precise ideas about cognitive processes in general; but this is a wild speculation

One thing that Pylyshyn does not say very loudly, perhaps for diplomatic reasons, is whether or not he thinks there is any real distinction between artificial intelligence and cognitive simulation on the one hand, and theoretical psychology on the other It could be argued that psychology is or is not a science according as one can or cannot claim the existence of an underlying theory Botany graduated from a taxonomy to a science when evolutionary theory came of age, and in psychology a similar transition may be in progress Most readers of this journal will probably take it for granted that psychology desperately needs a coherent body of ideas and nontrivial generalisations about cognitive processes in general and those of human beings in particular; the very idea of a complex process cries out for clarification in computational terms But it is not enough merely to make such a claim: what we all have to do is to try to spell out in detail what we believe may happen when someone utters a sentence or interprets a visual input; when we have a few more interesting candidate accounts of such skills we will be able (perhaps) to catch ourselves employing new and promising ideas about cognition – ideas that are less likely to emerge from reflections of a more general and philosophical nature

by **John McDermott**

Department of Computer Science, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pa 15213

On A.I. as psychology: now and then. Pylyshyn begins his article with the claim that the only significant difference between Artificial Intelligence (A I) systems and psychological models is that their designers have different methodological allegiances As Pylyshyn develops this claim, it becomes clear that he recognizes that many (and perhaps most) current A I systems do not qualify as psychological models; thus, assigning some reasonable interpretation to the claim becomes problematic I think what he is saying is that although the methodology of A I is different from that of cognitive psychologists who build computer models, and although A I systems do not (for the most part) currently qualify as psychological models, nevertheless in time, these systems will almost inevitably become psychological models If this is Pylyshyn's view, then I think he misunderstands the A I enterprise In order to focus attention on what I think are his most serious misconceptions, I will consider some of the support that he offers for his position within the context of the following two questions: (1) To what extent are current A I systems psychologically interesting? (2) To what extent will future A I systems be psychologically interesting? Pylyshyn lists three kinds of evidence (intermediate state, relative complexity, and component analysis) that can be used to determine the adequacy of a psychological model Since few current A I systems satisfy any of these criteria even minimally, I assume that Pylyshyn's answer to the first question would be that current A I systems are psychologically interesting only in the weak sense of being experiments for the discovery of mechanisms of intelligence (some of which will, it is hoped, turn out to be the mechanisms that humans use) Pylyshyn's answer to the second question, however, would be that most A I systems of the future will be psychologically interesting in the strong sense of being psychological models

I cannot find in Pylyshyn's article convincing support for what I am claiming would be his answer to this second question (which may, of course, be because it is not the answer he would give) He observes that much of the manifest dissimilarity between A I systems and psychological models might disappear if we controlled for generality and power and for the level of abstraction at which the descriptions are given But this observation allows him to conclude only that there are not necessarily any differences between the two Later he elaborates on the methodological allegiances of the A I system builder and the psychological model builder According to Pylyshyn, one crucial difference between the two methodologies is that the A I system builder takes a top-down approach (i e , is concerned with sufficiency conditions), whereas the psychological model builder takes a bottom-up approach (i e , is concerned with necessity conditions). Pylyshyn couples this observation with the other one and reaches the much stronger conclusion that it is likely that the differences between A I systems and psychological models

will someday disappear His premises appear to be: (1) A I system builders are attempting to build intelligent systems that are both general and powerful; (2) There is only a limited number of ways in which an intelligent system that is both general and powerful can be built Thus his conclusion: the sufficient conditions of the A I system builder will turn out to be the necessary conditions of the psychological model builder

The most serious flaw in this argument is Pylyshyn's assumption that A I system builders are concerned (at least ultimately) with systems that are general as well as powerful It is not clear why he makes this assumption He argues that A I is interested in those classes of tasks that form a "natural kind" for human beings; perhaps he infers from this that A I must be interested simultaneously in all classes of tasks that form a "natural kind" for human beings But surely such an inference is not valid There are many people who claim to be in A I who have no (higher) goal than to build a system that can do a single class of tasks intelligently; undoubtedly these people will be unmoved by Pylyshyn's assertion that they should have an "ultimate concern with generality " Pylyshyn could, of course, claim that the reason that A I system builders must have an ultimate concern with generality is that a general system is always more adequate (and thus more desirable) even for a specific class of tasks than is a specialized system But he offers no support for such a claim

Let me modify my second question and then point out a second flaw in Pylyshyn's argument: To what extent will future A I systems that are general as well as powerful be psychologically interesting? For Pylyshyn to claim that this subset of A I systems will be psychological models, he must provide some support for his premise that the number of such systems is necessarily small Pylyshyn offers one piece of evidence He observes (perhaps somewhat prematurely) that the space of solutions to A I 's central problems – system organization and knowledge representation – is fairly well understood Within this space of solutions, only a few appear appropriate for general systems since such systems impose a variety of constraints on their designers that more specialized A I systems do not Thus Pylyshyn's claim that future A I systems that are general as well as powerful will qualify as psychological models has some plausibility However, at the present time, it is not clear that the constraints are all that severe; it may be that many future general systems will be as unpsychological as the specialized A I systems of today

by **Allen Newell**

Department of Computer Science, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pa 15213

State-of-the-art constraints. As to the underlying theses of Pylyshyn's paper, how could I have any quarrel? The computer is for psychology more than a computer; it is an embodiment of a symbol system, thereby having immense theoretical significance Work in pure AI, that is to say, work motivated by non-psychological considerations (an odd sense of purity, come to look at it), is subject to constraints that increase its chances of contributing to understanding intelligent action in humans Mostly this happens because the task structures are so-called natural kinds for humans (I would have been pleased if Pylyshyn had clarified the notion of "natural kind"; it is one of those pregnant quasi-technical terms, like "competence," which for me obscures almost as much as it helps) Much is explained in the practice of science – here, cognitive science – by distinguishing alternative approximating sequences for approaching empirical truth And so on All of this makes eminent good sense to me I find it easy to hang loose on the exact adequacy of his description of the enterprise Such meta remarks are useful in a rough and ready way, nothing more should be asked of them However, I dare say others may not see it that way Some no doubt will be considerably more uptight (intellectually), wishing, in the words of the *Arabian Nights,* to spend their days in the durbur, bidding and forbidding

Let me pursue the theme of hanging loose It is valuable to realize that much that appears to be of general methodological and philosophical concern is just a reflection of the current state of the art of a science (often, its state of obscurity) Consider the fretting that has occurred over what claims a program-qua-theory makes Programs often contain many mechanisms just to make the simulation operational, mechanisms that seem in no wise distinguishable from other mechanisms that are theoretically relevant In other types of models, the irrelevant seems more easily distinguished, such as the color of the ink used to write equations Now, it is a reasonable presumption that psychology will progress to models of the information processing architecture of human cognition That is, positing a basic invariant structure within which symbolic processing occurs will become a matter of course Given this, the claims about what is psychologically relevant will gradually become clearer and the whole question will become moot, except at a few special junctures In short, there is no general issue of moment, only getting the substantive state of the science sufficiently advanced

Consider, to continue this example, the plausible outcome that the architectures putatively describing man have some quite idiosyncratic features that impress themselves in various ways on most of the processing that man does – giving it characteristic shape, even if not determining exactly what can and cannot be computed Such features might be the (characteristic) ratios of read to write times in long-term memory, or the underlying representation out of which more general knowledge structures must be encoded (e g , memory might be event oriented) Then to work on human cognition, as opposed to "pure" AI, would be to focus attention within this class of architectures As our knowledge of this special class increases, the separation between it and other classes of computational mechanisms can be expected to increase, until there would be little doubt about which enterprise a particular scientist was engaged in This does not really blunt the general point that Pylyshyn makes about the relevance of all of AI to cognition The situation simply reflects the structure of Computer Science generally, which fractures into subdomains both by task (partial differential equations vs graphic displays) and architectures (array processors vs networks of microcomputers) These separations imply that knowledge has locality, but do not gainsay that the whole science gradually gets itself together by understanding what is true in its subdomains and how to generalize them Nothing fundamental is at stake and there seems to me nothing special about the pure-AI vs human-cognition issue to distinguish it from, say, the uniprocessor vs multi-microcomputer issue Sometimes, as with optical computers, the underlying medium and the specialized processing is so strong that for long periods of time it appears as a distinct field But watch what will happen if optical computing ever achieves the ability to support general symbolic processing Zip! it will be assimilated into the main stream of computer science as just another technology, which both teaches us some special lessons and yields one more instantiation of what we have come to understand about information processing The point is that nothing in this seems more than the garden variety evolution of scientific and technical knowledge

The mutterings about languages of intention and their essential separation from the underlying language of neurophysiology can provide a last example As Pylyshyn notes, this same separation exists between the symbolic level of computer architecture (the programming level) and the lower levels (the logic or circuit levels, the latter probably being most analogous to the neural level) I am not sure what content lies behind Pylyshyn's statement that " the computational rules take the forms that they do because their terms represent something ( ) and this aspect cannot be captured by an electronic description " Within computer structure there is no issue that I know of corresponding to the one raised here (and more generally in philosophical psychology) It is rather clear in what sense the same exact system can be completely described at each structure level Electronics engineers never find it appropriate to get into the hassles with programmers that are evidenced by the discussions of the language of intentionality Elements of confusion are to be solved just by the individual computer scientist's thinking a little bit harder about the specific case at hand This all happens, of course, not by any virtue of computer scientists and engineers They are mortal and of limited rationality like the rest of us It happens because of the state of art – the matter is all laid bare to be seen and understood by one fool as well as another

In a great article, whose exact reference I have mislaid, but which was entitled "There is plenty of room at the bottom," Feynman observed how easy biology would be if one were simply small enough to go and look Much of what passes for the profound arises from the accidents of the veil nature happens to have passed between us and the phenomena we wish to understand When finally seen in a clear light, the substance of the matter makes all larger issues moot It suggests that such considerations as Pylyshyn lays before us, should be taken freely but without much concern about their details They are commentary, helpful to focus the attention and alert the mind The issues that are ultimately engaged will come clear and obvious in the morning light of some substantive advance

by **Andrew Ortony**

Center for the Study of Reading, University of Illinois at Urbana-Champaign, Urbana, Ill 61820

Cognitive psychology, artificial intelligence, and cognitive simulation. Pylyshyn's interesting and provocative paper contains two main claims One of them he simply asserts For the other, he offers several arguments, none of which, I think, is very convincing

The first claim is that the field of Artificial Intelligence (AI) is generally recognized as being of "paramount theoretical importance to cognitive psychology" As stated, I think this is false Practitioners of AI sometimes overestimate the relevance of their efforts to psychology, perhaps because they have conceptions of psychology that are different from those of cognitive psychologists But it is the psychologists, not the computer scientists, who must make such evaluations Now, while many psychologists have adopted the computer metaphor for human information processing, they nevertheless seem to be largely ignorant of AI in Pylyshyn's rather general sense of it So, *contra* Pylyshyn, it can be argued that AI is not (at least, not yet) generally accepted as being of paramount importance to cognitive psychology at all This conclusion can be reached either by denying that the importance is *generally accepted,* since many cognitive psychologists know very little about AI, or by denying that it is of *paramount importance,* since if it were, it would be more influential than it is This may seem a small point, but the underlying issue is crucial to Pylyshyn's paper

The gist of what Pylyshyn has to say revolves around his second claim Briefly, it is that there is no essential difference between "pure" AI and cognitive simulation, a claim that has an important connection with Pylyshyn's first There is a version of the first claim that is true, namely that *cognitive simulation* is of paramount importance to cognitive psychology In fact, I think that Pylyshyn's most serious mistake is to lump together pure AI and cognitive simulation under the general heading of AI This is because what is true of cognitive simulation in particular is often not true of pure AI, and thus not of AI in general Consequently, the viability of the distinction between cognitive simulation and pure AI becomes of central importance

Pure AI comprises systems designed to attain a particular behavior, or range of behaviors, regardless of the mechanisms employed to do so Typically they need only be concerned with what Minsky (1975) refers to as "sufficiency" Cognitive simulation, on the other hand, comprises systems intended to simulate, as far as possible, the processes that humans employ in the generation of such (intelligent) behavior Pylyshyn believes that there are no important differences between the two, provided that one controls for three particular variables when making comparisons About each of these variables – generality, power, and level of description – he has some very interesting, and often insightful things to say But he believes that people (like me) who maintain that there are important differences between pure AI and cognitive simulation, are actually deluded as a result of their neglect of those variables He claims that if we equate a pure AI system and a cognitive system on each of these variables, then "there need be (no) systematic difference" I suspect that this amounts to saying that if we equate any two theories of the same phenomenon on these or similar variables, the theories will turn out to be variants of one another, and will be logically equivalent The *reductio ad absurdum* if one takes any pair of things and then equates them in the respects in which they differ, their differences will always go away! But, if the things really are different then they cannot be so equated and such efforts are not only unwarranted, but also doomed to failure

The crucial difference between pure AI and cognitive simulation always remains In AI there is no necessary intention that the program be an embodiment of a psychological theory *at any level of description,* while in cognitive simulation there is It is true that pure AI may by chance embody a psychological theory, but only by chance, and, for the most part, it is absurd to try to construe the mechanisms employed as instantiations of theories about human performance Symptomatic of Pylyshyn's misconception is his belief that: " if a person is capable of doing a certain task *x* which is judged to require intelligence, then any device that can be said to 'do task x' must not only be doing the same task as the person but in some sense must also be doing *x* 'in the same way' that the person does it" I think that some simple examples show that this is not so Consider first the machine translation efforts of the '50s Machine translation could be called AI on the grounds that language translation requires intelligence Yet nobody would believe that a program that attempted word-for-word translation by dictionary look-up could be regarded as an implementation of a *theory of human translation,* even if it worked Or consider the mechanism of finding lexical items in the "internal lexicon" Longuet-Higgins and I (Longuet-Higgins and Ortony, 1968; see also Longuet-Higgins, this *Commentary*), devised an algorithm for a theoretically optimally efficient search strategy that involved representing the lexicon as a tree of letter sequences It would be ludicrous to maintain that the mechanism employed was supposed to model the one used by people Finally, consider the tens (if not hundreds) of programs that have been written to play chess or engage in automatic theorem proving Many of them utilize so-called "brute-force" techniques They generate as many continuation board states or inferences as are physically possible within the constraints imposed by machine size and speed But, this bears no significant relationship to the mechanism employed by human problem-solvers Such programs can only be said to be doing the task in the same way as a human in an utterly trivial and useless sense

The upshot of all this seems to me to indicate the falsity of Pylyshyn's claim that AI is subject to empirical constraints It is false because he equates pure AI and cognitive simulation while, in fact, only the latter is subject to such constraints Pure AI typically is not and need not be

So, Pylyshyn, while rightly distinguishing between the use of computers to simulate processes, and the use of computers to simulate products, then blurs the distinction and attributes to the latter, characteristics that belong only to the former Furthermore, he fails to note one of the most important benefits to be derived from cognitive simulation, namely, that it provides an opportunity to experiment with, and to model a large number of complex, interacting processes The dynamic characteristics of computer modelling provide a much better way of representing the complex interactions that are so characteristic of human cognition than do the static representations afforded by more traditional modelling techniques It is precisely in its incapacity to deal with real-time thought processes in a general way that cognitive psychology has its greatest weakness, while it is precisely in its capacity to model such processes that cognitive simulation has its greatest strength Thus, unlike pure AI, cognitive simulation makes a fine bedfellow for cognitive psychology It is an encouraging sign that their offspring, Cognitive Science, is coming of age

REFERENCES

Longuet-Higgins, H. C. and Ortony, A. The adaptive memorization of sequences, in D. Michie (Ed.), *Machine intelligence 3,* Edinburgh, Edinburgh University Press, 1968 (pp. 311–322).

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

by **Juan Pascual-Leone**

Department of Psychology, York University, Downsview, Ont , M3J 1P3 Canada

Computational models for metasubjective processes. Although A I *and* Computer Simulation (henceforth jointly abbreviated AIS) lead to specific programs and programming systems which can be considered to embody theories of cognitive information processing, the sense in which this embodiment exists is not clear As Pylyshyn indicates, AIS systems are so complex that it is often difficult to know which aspects or components of the program are theoretically relevant and which ones are not In addition AIS programs are usually local, rather task-specific in nature, embodying much more of the structure of the task at hand than of general unifying principles applicable across types of tasks (for any given type of subject) What is needed, and usually does not exist, is a theory of the program – in Pylyshyn's words, "a description of the system which highlights the general principles underlying its operation"

This criticism of AIS programmes *qua theories* has been raised by Weizenbaum (1976) with great clarity (see also Pascual-Leone, 1976a, 1976b and 1977a, 1977b) In spite of these difficulties Pylyshyn believes that the AIS approach "is on a more secure footing than conventional theorizing in psychology" He reaches this conclusion despite the fact that programs must be task specific and, therefore, the general principles of human processing can hardly be discovered by way of writing programs, although extant programs could be used as data (together with psychological data) to infer these principles

I disagree with Pylyshyn's quoted statement and believe that his faith may be the result of a historical artifact, i e , the long and excessive domination of empiricism in modern technologically-based and technologically-financed

science The immense heuristic and epistemological/historical importance of AIS work for cognitive psychology could be largely due to the not-yet-quite-past dogmatic domination of empiricism in psychology This, often naive, empiricism has led to psychological theories of the local (situation or task specific) variety Organismic functional laws or principles which could only be inferred empirically, by the power of the imagination, from characteristic patterns of performance (empirical invariants) found across types of situations, have thus been ignored by the scientific establishment As a result, despite the pioneering work of psychoanalysts, Gestalt psychologists, Piagetians, etc it became common in the forties and fifties to confuse the process (the organismic generating operations) with the performance The resulting lack of representations for the process and the concurrent emphasis on the description of performances, led to two unfortunate trends The first epistemological trend was the theoretical reduction of all performances to two types: *learned performance* (i e , transfer of learning) and *innate performance* (i e , maturation) In this manner the concept of problem solving via a mental computation which generates (constructs) a *novel performance* became lost, until it was re-invented and made empirically explicit by the AIS revolution The second erroneous epistemological trend brought about by excessive empiricism, and which AIS work has eliminated, is the naive equation of theoretical prediction with empiricist generalization AIS work has re-invented and empirically demonstrated the possibility and power of rationalist prediction, i e , prediction of novel performance via process-theoretical computational methods

These achievements of the AIS revolution and its future contributions to psychology could be jeopardized, however, by a new brand of dogmatic empiricism (the computational or A I empiricism) which Pylyshyn implicitly criticizes but also reflects in his paper Pylyshyn's statement that the AIS approach "is on a more secure footing than conventional theorizing in Psychology" is only warranted if this approach is contrasted with the old empiricist psychological methods – the methods which the AIS revolution has helped to correct When the AIS approach is compared with the possibility of a modern constructive rationalist psychology (a new psychology informed by the AIS computational methods but free from the constraints imposed by the computer, which constructs process models for the sole need of experimental or structural prediction), Pylyshyn's statement is no longer so certain

A constructive rationalist approach to cognitive psychology would be one in which explicit process models are constructed of the mental operations produced by the subject in connection with tasks Unlike AIS models, these purely psychological process models would be used exclusively to make rationalist predictions of quantitative, qualitative or quantitative-structural characteristics of subjects' performances New process-descriptive languages may be convenient for this purpose; molar ("macro") languages able to reflect, at a suitable level of generality, subjects' true psychological constraints (the *metasubjective* constraints) inferred from their performance across types of situations Actual execution of these *metasubjective models* on a computer ("however important that is" – to use Pylyshyn's fitting remark) should necessitate the translation (transcription?) of the model to a less compact, but probably more explicit, computer-compatible language Process models formulated with metasubjective languages could serve as descriptions of the functionally-molar, intentional processes (i e , the metasubjective processes) which govern the subjects' psychological intercourse with their environment

The metasubjective simulation models could help to provide the inductive basis needed to infer the general principles of organismic computational power, as well as the functional characteristics of the organism's intentional structures A recurrent theme in Pylyshyn's criticism of AIS methods is the lack of clarity regarding general principles and the lack of clear separation between relevant (intentional) and irrelevant (calculational) aspects of the AIS programs These metasubjective models could also serve to connect more closely the AIS theoretical enterprise with the empirical constraints and methods of conventional psychology Failure to develop these metasubjective simulation procedures could condemn the promising AIS revolution to the isolationist/empiricist fate of other past "revolutions " The best safeguard of cognitive science may be in the open-minded constructive rationalism of its scientists For technology is no more free of error than the mind of its user

REFERENCES

Pascual-Leone, J. A view of cognition from a formalist's perspective. In K. F Riegel and J. Meacham (Eds.) *The developing individual in a changing world.* The Hague: Mouton, 1976. (a)

Metasubjective problems of constructivé cognition: Forms of knowing and their psychological mechanism. *Canadian Psychological Review,* 1976, 17(2):110–125. (b)

Review of Klahr and Wallace's "Cognitive development, an information processing view." *Child Development Abstracts and Bibliography,* 1977, 51(5,6):251–252. (a)

Constructive problems for constructive theories: A checklist for developmental psychologists. Contributed to the invitational conference of the Institut für die Pädagogik, Kiel, 1977. To appear in H. Spada and R. Kluwe (Eds.) *Developmental models of thinking.* New York: Academic Press, in preparation. (b)

Weizenbaum, J. *Computer power and human reason.* San Francisco: W. H. Freeman, 1976.

by **Karl H. Pribram**

Department of Psychology, Stanford University, Stanford, Calif 94305

On behalf of the neurosciences. Pylyshyn's analysis of the differences in approach between the extremes of artificial intelligence (A I) and cognitive psychologists (C P) is a most interesting one In a sense, A I scientists take the interaction between their own intelligence and that of the computer as the experimental datum, while C P scientists use the computer as an "in vitro" tool to check out "independently framed theories for completeness and consistency" much as the biochemist uses his test tubes Pylyshyn argues well that the experiments undertaken by the A I community are not altogether that different from those carried out by C P s, since the latter are also deeply influenced by the computer as a model of cognitive processing

I have no argument with either Pylyshyn's analysis or with his conclusions But I do want to raise a caveat Both A I and C P scientists are apt to construct "unrealistic" models of human cognitive processing unless they give serious heed to developments in the neurosciences In fact, heed is paid, but often there is a lack of seriousness whenever the neurological data become the least bit complex

Do I mean by this that I believe that the neurosciences are going to provide *the* models for cognitive processes? Certainly *not* Pylyshyn clearly states his position, "for psychologists, computational systems should be viewed as functional models quite independent of (and likely not reducible to) neurophysiological systems," and in part I agree with him Nonetheless, I feel that someone in the "Cognitive Science Fraternity" must attend to the neurosciences sufficiently to know whether the cognitive machine-language is appropriate to the machinery of the brain For high-level languages, this concern may not matter, but when such rudimentary processes as pattern recognition, decision and memory storage are involved, attention to wetware becomes nontrivial As a case in point, current computers are essentially serial processors while much of the brain operates on parallel processing principles The memories of current computers are location addressable while the brain's memory is more probably content addressable True, one can overcome these differences by appropriately programming, but to the extent that computers as presently constituted are to be taken seriously in modelling cognitive processes because they provide previously unrecognized insights, we do not yet know the power of alternative configurations that might embody principles of brain function in cognition additional to those now available

In short, my plea is that cognitive scientists continue to be problem-oriented rather than succumbing to the lure of technique The computer is powerful both as a model and as a tool, and a cognitive discipline centered on this instrument is certainly commendable Pylyshyn has made the case that the two subdisciplines involved with the cognitive functioning of computers might fruitfully court each other My view would be that such courtship might be incestuous and breed a monstrous cognitive science Perhaps better that each branch of the family court one of the neurosciences – the resulting offspring of such a union might prove healthier and therefore in the long run more viable

by **R. S. Rodger**

Department of Psychology, Dalhousie University, Halifax, N S , B3H 4J1, Canada

Computer-specific methods Pylyshyn writes that research on Artificial Intelligence and Cognitive Simulation is, "directed at understanding classes of tasks which are defined by essentially psychological criteria " It could be argued that it is not "classes of tasks" that research helps us to understand

but methods of operation It seems quite possible that "methods" not previously in use by humans may be discovered through research on A I It is even conceivable that after discovery they might remain useful for computers and impractical for humans The history of mathematics may be said to be replete with examples of newly discovered methods for solving problems; so I see no reason why A I research may not yield results of a similar kind The danger is that, once discovered, the new methods may be dismissed as trivial Remember the story of the mathematician proving a new theorem to his colleagues and, half-way through, claiming that the next step was trivial After a heated discussion for half an hour, everyone agreed that indeed the step was trivial!

Cognitive Simulation, as a psychological endeavour, would not aim at developing methods for their own sake, but insist that they represent, at an *appropriate* abstract level, processes used by humans If we adopt the author's terms of reference: compare systems with the same generality, the same power and having the same abstract description, it is difficult to see how one can deny his conclusion that the distinction between A I and Cognitive Simulation is little more than one of style and the ordering of goal priorities Imagine a human system and a machine system with the same generality and power but differing in certain details such as their intermediate states, relative complexity and the components they use For example, they step through problems differently and distribute their time and their errors differently Pylyshyn might argue that this description is not sufficiently abstract because, "there is some level of description of the two at which they are doing it 'in the same way ' " An experimental psychologist, interested in simulating cognitive processes, would not accept so high a level of abstraction as *appropriate*

A good A I system will, among other things, make rather few errors, but a good simulation makes just the correct number of errors of just the human kind at just the right places!

Equating research on A I with that on Cognitive Simulation appears to be a return to the doctrineof Protagoras, that "Man is the measure of all things " In spite of the obvious attractions, for a psychologist, in bringing the A I branch of Computing Science under the psychologist's umbrella, I believe this topic should be allowed to develop in its own direction, untrammelled by the limitations of human thinking, at least insofar as computers can avoid them and still communicate with their programmers

by **Roger C. Schank**

Department of Computer Science, Yale University, New Haven, Conn 06520

AI vs. CS: a methodological distinction. One of the more significant dividing lines in AI has always been the one separating those whose primary interests are to find out how people work from those who are more interested in getting machines to be smarter The problem with this dividing line is that it is a methodological one more than one that rigidly defines goals All researchers are happy to shed light on a problem in which their interest is only secondary after all

One issue that needs discussion with respect to Pylyshyn's paper, then, is whether this methodological difference produces systems that are in some way essentially different The answer has to be that it can and often does, although sometimes it does not Pylyshyn has argued the "does not" side of the issue, so here it is necessary to show that the "does" side also exists

The questions are, then, does the working methodology one chooses affect one's results; and can there be a difference in results? In other words, are there AI programs that are clearly not, and could never be, psychological theories?

There are general lines that must be followed in an argument of this sort First, we must discuss those AI programs that could not in principle be psychological theories Second, we must deal with those that could be but can be shown not to be

The class of AI programs that can in no way be psychological theories are those that rely on hardware devices for their front ends that are radically different, in terms of the output they produce, than the human devices for which they substitute Such devices exist in speech understanding systems to process the incoming speech signal, and in automatic vision systems to process scenes In both of these cases, there would be no problem if the hardware necessary for the task in any sense approximated the accuracy of the human device that does the same job But, in fact, they in no way come close As a result, the AI researchers that do speech and vision have had to develop systems that must recover from errorful data as soon as they start out This affects the entire system in such a way as to cause a major part of the program to be worrying about issues that simply do not come up when people are met with the same situation Thus, people almost never mistake the words "king" and "queen," both because of the expectations and knowledge they have attached to those words and because their phonological shape is quite different However, these words are similar enough phonologically for an imperfect hardware device to confuse them, and a great deal of effort must be put into methods of both detecting mistakes and recovering errors Similar problems occur in vision systems as well

Vision and speech constitute two major areas of AI, then, that do not use, nor really care a great deal about, psychological theories of human processing They are in no sense simulations and probably should not be

In areas such as the processing of written textual material, which constitute a large part of AI activity these days, we find a great many researchers concerned with doing cognitive simulations (Pylyshyn's term) as opposed to artificial intelligence (in the sense that Pylyshyn uses that term) Our research at Yale has always had the goal of being both a cognitive simulation and artificial intelligence, so of course that puts us right in the middle of the argument that Pylyshyn tries to make I am thus forced to disagree with the distinction that Pylyshyn sets up (his point exactly of course) and at the same time argue that the distinction sensibly exists for others

In the processing of written text there are two methodologies that correspond to Pylyshyn's distinction One tries to model human capabilities and the other just tries to do the job I do not see how both of these approaches can be psychological theories (in any serious sense of that term) at the same time (Pylyshyn could argue here that they are just competing psychological theories, but that argument would allow everything anyone ever thought about a process to constitute a theory of a human process Clearly there are many ways computers do things that no one in his right mind would equate with psychological theory)

To take one example of a place where such theories compete, we can look at the use (or each of the uses) of syntax in processing sentences There are some natural language programs that use totally separate syntactic parsers and worry about meaning later In opposition to them are the programs that do very little syntactic processing, all of which is fired off by expectations about meaning Now it seems to me that while the latter approach is a possible (albeit not validated) psychological theory, the former simply cannot be so regarded Is it a serious suggestion about humans that they analyze a sentence solely by grammatical rules applied to syntactic categories of words, never considering the meaning of those words until they have made a complete pass through the sentence looking at the syntactic structure only? Such a parsing logic, used in various forms by researchers in natural language over the years, treats a sentence as something that can be scanned through before the meaning of a word is even considered Certainly the researchers who do such things believe they are doing AI, but there is no psychological theory then

My point is simply this: There are those of us for whom AI and cognitive simulation are clearly the same thing But this is purely a methodological assumption about how one should go about creating and testing such a theory I believe strongly in that assumption in doing my own AI research, but there are those who do not subscribe to it and are thus not creating psychological theories; and there are those who ought not subscribe to it, considering the tasks they have at hand

by **Thomas W. Simon**

Department of Philosophy, University of Florida, Gainesville, Fla 32611

The AI/CS distinction and theory evaluation. Because the distinction between artificial intelligence (AI) and computer simulation (CS) is generally uncritically accepted, Pylyshyn's challenge to that distinction is welcomed, if for no other reason than that it jars our intuitions out of their complacency AI and CS have yet to be completely accepted into the mainstream of scientific thought On the one hand, like most of the "pure" studies today, the "purer" aspects of AI research are being challenged for their lack of relevancy What could a chess-playing machine possibly tell us about human thinking if there is no concern to have the machine play chess like a human? CS, on the other hand, still is not accepted by many as a legitimate means of psychological theorizing Perhaps, some of Pylyshyn's arguments will help to overcome some of these obstacles

Nevertheless, Pylyshyn may be in danger of blurring the distinction too much In fact, it is not clear what would constitute a "systematic distinction" between AI and CS for Pylyshyn, since all those *differentia* he relegates to style are just those that clearly and systematically distinguish the two If AI research is only constrained by very general empirical laws in focusing on problems that constitute "natural kinds," but CS is constrained by much more refined sources stemming from psychological methodology, then that seems to be as much of a distinction of kind as one could hope to find Admittedly, AI research *may* uncover descriptions of processes which are useful in explaining human thinking (particularly if in order to do a task as well as a human an AI device needs to do it in the same way) Yet, that does not distinguish AI research from analyses of poetry, which may yield similar fruits It is just as likely that AI research will hinder our understanding of the human mind There may be spinoffs from AI into psychology, but there is not any stronger systematic link between AI spinoffs and any others to psychological explanations

Important differences do exist between a discipline that is concerned to simulate the way humans think as well as the end-product and one only concerned to replicate or improve upon this end-product Although it is true that: " any device that can be said to 'do task x' must not only be doing the same [intelligent] task as the person but in some sense must also be doing x 'in the same way' that the person does it," in many cases its truth is only attained by trivializing the claim A chess master and a duffer both do the same task, but not at all in the same way, except in the vaguest sense of "in the same way " More bizarrely, the wind, a bulldozer, and a human can all demolish a house, but only in the most trivial sense do they accomplish this task in the same way

Moreover, the AI/CS distinction is not adequately blurred in terms of the generality/power trade-off Criticism of the admitted power but lack of generality of many computer systems is not simply "a comment about the more general phenomenon that performance may be purchased at the price of generality " Rather, it is a very serious charge against both AI and CS, particularly in light of the fact that researchers in both seem to place more emphasis on developing systems rather than on specifying the underlying principles The goal of cognitivism in psychology ought to be undermining a previous unfortunate inverse relationship between power and generality Furthermore, if "the most interesting aspects of intelligent behavior are more a function of the interaction of very small components (e g , a lot of specific knowledge) than the product of a few deep principles," then the continuity between AI and CS becomes even more unlikely For, intuitively, the more that two radically different means are divided into component parts the less the likelihood of finding similarities between them At some abstract, deep level there might well be some similarities between automatic and human door-openers, but when the two means are divided into sub-processes, the differences become more apparent

Turning to the problem of evaluating CS methodology, Pylyshyn's discussion does go beyond the rather meager fruits of the Turing test and the like Yet, sources of evidence such as intermediate state, relative complexity, and component analysis are fairly weak Methods for studying intermediate states in human tasks, consisting of introspective reports plus inferences to intermediate states, are not only "very few and rather crude," but may also be inherently so Complexity evidence merely indicates being on the right track (but nowhere near the station) toward uncovering the principles of mentation because it is too easy to envisage two different devices accomplishing the same task in the same time frame but by radically different means Component analysis, although the strongest source of evidence, is inferential, which may well be a polite way of talking about guesswork A likely check on this guesswork is at the neural level, which cognitivists sometimes all too readily dismiss

Even if we know how to evaluate a computer simulation, it is still not altogether clear what aspect of the simulation should be the object of evaluation While Pylyshyn is correct to note that " the relevant object of comparison is not the computer program per se but a description of the computational process cast at some level of abstraction," the problem of developing a means for specifying the appropriate level of abstraction remains An immediate difficulty is that any number of theories can be extracted from the same program Also, the extracted features are often recast into verbal form so that one of the main advantages of simulation, precision, is undermined Yet, despite these and other limitations CS does provide at least the first steps towards constructing a scientific theory of the human mind

***by* Aaron Sloman**

Cognitive Studies Programme, School of Social Sciences, University of Sussex, Brighton, England

Artificial intelligence and empirical psychology. If I conjecture that the sum of the first n odd numbers is always a perfect square, I can test this with $n = 1$, $n = 2$, $n = 3$, etc Is this an empirical investigation? If I use a computer instead, is it being used for an "empirical exploration"? These would not normally be called empirical investigations, unlike running the same programs to test a computer

Consider two definitions: An investigation is empirical$_1$ if it is based on examination of individual cases, but not if it uses a general proof It is empirical$_2$ if (like physics and geology) it is concerned with objects in the world of experience, and not merely (like number theory and theory of computation) with formal abstract structures

What Pylyshyn is really saying is that some work in Artificial Intelligence is empirical$_1$, like some mathematical explorations In other words, AI often uses "formal," but not "substantive," empirical investigations

Experiments with an AI program might be empirical in both senses They could reveal a failure of the program to understand something it was intended to be able to cope with (a formal empirical discovery) or they could show that people sometimes use language in a fashion not previously noticed (a substantive empirical discovery) Similarly, a vision program may fail where it was intended to cope, or it may fail in tasks the programmer had not realised most people could cope with

Pylyshyn suggests that empirical investigations can show "what kinds of relations must (sic) exist " Substantive empirical investigations might show what *can* exist, but "must" in this context presupposes a formal demonstration The example mentioned, namely Waltz's program, proves nothing about what *must* exist Moreover the power of his label-set is a formal, not a substantive, empirical discovery, whose status as an explanation of human abilities depends on the unavailability of anything better

AI versus computer simulation A divergence between AI and simulation systems is predictable Contrast (a) behaviour based on considerable expertise, built up over many years, like linguistic or perceptual skills, with (b) the floundering, exploratory, non-expert behaviour of beginners struggling with puzzles, like the novice logician, chess-player, or child seriator Only the latter incompetent behaviour is easily amenable to observation Deeply-compiled expert skills involve rapid and complex processes not available to introspection or laboratory observation So the "simulators" will tend to concentrate on (b), unlike the AI fraternity

But AI programs are still relevant to psychology, since they are testable by their generality, extendability and ability to account for the fine structure of phenomena When adequate theories of human learning emerge, it may be possible to test some AI models by asking if they could be built up by processes typical of human learning (Studying *how* infants learn is distinct from studying *what they learn when*, as in pre-computational psychology) AI work tends to produce deeper insights into human processes than simulation studies, since expert behaviour, not fumbling problem-solving protocols, is most characteristically human

Is parallelism relevant? Admittedly, a serial *computer* may be no bar to studying brain processes since parallelism can be simulated as closely as required But it is clear that many human abilities involve parallel processing at a cognitive level, e g , a child producing number names, pointing at different objects, and monitoring the two processes to keep them in phase But even if theories without such parallelism aren't adequate explanations of *how* we do things, they are steps towards formalisations of the tasks we perform and the information required for this

What is a "natural kind"? There are simple algebraic tests for straightness of a line yet they probably have little to do with how people perceive straightness and other shape properties – an important unsolved AI problem So the existence of a non-intelligent solution to a problem does not preclude the possibility of solutions using (human) intelligence

Pylyshyn uses the notion of a pattern or problem being a "natural kind" for humans Has he forgotten the variability of human beings, and the extent to which a "natural kind" may depend on a cultural context, like the symbols people can recognise easily? A particular class of patterns or problems which now does not form a "natural kind" for humans may one day form part of a widely practised skill Consider the sight-reading of piano music

Can we observe computational processes? Pylyshyn assumes that we can observe intermediate states But when people or programs produce pro-

tocols or answer questions about their strategies this may give misleading information about their normal functioning Further, much information about what is going on, (e g , about indexing strategies, matching procedures, rules for parsing and interpreting) may be quite inaccessible to processes concerned with external communication and global decision-making Procedures may have been compiled into "unreadable" lower level languages, and sub-processes may use "private" work-spaces Opening the machine to look at its innards would be like trying to understand a very high-level program by examining its machine-code compiled form

Empirical constraints in common sense As part of our ability to communicate and our self-knowledge about skills, beliefs, habits etc , we share encyclopaedic knowledge about what people can do We use it when we gossip, read novels, judge others, or make plans But psychologists often think that unless they do experiments they are not scientists, so they rarely attempt to analyse and codify this knowledge, as linguists and philosophers do

Since people doing AI have fewer hang-ups about being scientists, they are more willing to start from common knowledge about what people can do (e g , understand English, interpret drawings, plan actions, etc) We can often test explanatory models by noting how their performance falls short of what we know people can do, without relying on new experimental results Thus there are empirical constraints on AI theories embedded in common sense

Until AI can account for most of what ordinary people know about people, there may be no urgent need for new psychological data, except in the rare cases where two different models appear to be equivalent in explanatory power, generality, extendability, etc (Crucial psychological experiments may not always be feasible)

Of course, common-sense is often mistaken But, although it is not a good source of *laws* (indeed, it is doubtful whether there can be laws of psychology) it is a good source of information about *possibilities* – that is, things people can do The discovery and explanation of possibilities is a major feature of the progress of science And it is possibilities (abilities, capacities, skills) rather than laws that AI is mostly concerned with explaining

This will be lost on most psychologists until their training problems are revised to "how is this possible?" instead of "why does this occur?" The latter encourages a search for correlated conditions instead of explanatory mechanisms

How top-down is A I ? Pylyshyn suggests that AI workers try to devise complete systems He should have said they try to devise *working subsystems* A *complete* intelligent system would be a teachable robot, with moods emotions, etc In relation to the task of designing a person, AI work is mainly bottom-up, not top-down, since most computer models deal with a small subcomponent of some human ability (e g part of the ability to interpret pictures, understand stories, etc) Pylyshyn meant that AI work first explains general features of an ability, and later adds refinements to explain details But this is misleading, since relative to long-term goals, AI work is bottom-up, not top-down

This has serious risks By reflecting on processes typical of complete human beings, and on interactions between subsystems, we can formulate constraints which current computer models violate These "interface" constraints may be more significant than constraints generated by the underlying computer – the brain

For example, what you *see* can remind you of something, generate changes of mood, help you solve a problem, teach you a concept, help you understand a conversation, etc What features of a vision system are necessary for this, and what features are required by the other systems? Further study may show that existing program structures are grossly inadequate, even if their factual content (e g , about image- and scene-features) is correct But until bottom-up explorations have generated much more technical know-how, it may be premature to switch to the top-down mode and try designing complete systems, except in occasional philosophical moments

by **N. S. Sutherland**

Centre for Research on Perception and Cognition, University of Sussex, Brighton, Sussex, BN1 9QG, England

Task constraints and process models. If a neuroanatomist were to present us with a complete wiring diagram of an individual nervous system specifying every synaptic connection and the probability of a nerve impulse being generated for every possible combination of inputs at each synapse, we would have a complete model of cognition and the control of behavior The model would be virtually useless: so too would be the tape of an AI program punched in machine code Because of the limitations of our own cognitive systems, we can only understand any complex system in terms of higher level concepts that describe the outcome of blocks of lower level operations (Some examples of such concepts are "lateral inhibition," "receptive field," "depth first search," "list structure," "recursive procedure," and "relaxation method ") I cite the hypothetical example of the neuroanatomical model to illustrate Pylyshyn's claim that the goal of cognitive psychology, as of all science, is understanding: models of mental functioning, including computer models, are only of value insofar as they further our understanding It follows that the principles behind AI programs are of importance for the psychologist, not the detailed implementation Pylyshyn goes on to argue that "if a person and a computer are both capable of doing the same task there is some level of description of the two at which they are 'doing it in the same way' " This claim may be correct, but it glosses over the difficulty of how to decide what is the appropriate level of description

Pylyshyn argues that computer programs must at some level of description correspond to mental functioning since the task constraints are the same for both systems If a computer program interacts directly with the real world through sensory channels that correspond to our own, it will be subject to the same task constraints as ourselves In visual perception the constraints are imposed by the relationships between the optic array and the 3-D environment that gives rise to this array according to the laws of physics and optics By attempting to write programs that infer a 3-D description of polyhedral bodies from line drawings, a number of workers in AI (e g , Clowes, 1971; Mackworth, 1973; Waltz, 1975) have considerably added to our knowledge of how aspects of the optical array are related to bodies in the scene Huffman (1971) has also provided important insights simply by reflecting on projective geometry, and without writing any programs: nevertheless, some of the constraints specified by AI workers would probably not have been discovered without the discipline of writing programs, and to this extent AI can be thought of as a useful way of specifying task constraints Even in scene analysis programs, however, it does not automatically follow that the constraints used by programs are the same as those used by people For example, Waltz's vision program appears to have a richer knowledge of the depth information that can be provided by shadows than does the human perceptual system (compare Fig 2 32 in Waltz, 1975, p 65, *op cit*)

Moreover, the discovery of task constraints is not the only task facing the psychologist: he should also seek to specify the processes that take advantage of these constraints in order to carry out the task Not all psychologists would agree that this is a sensible goal for psychology J J Gibson (1966), for example, has articulated very clearly the need to specify task constraints, but appears to believe that once that is done nothing remains to be discovered In this tradition, Neisser (1976) talks of information being "picked up" from the optical array and used to guide behavior: it is true that Neisser also acknowledges the existence of vaguely defined "schemata" but he appears to believe that they determine what information is "picked up" rather than how it is used Unlike Gibson and his followers, most psychologists are interested in process models Computer programs are at the moment the only tool we have for giving rigorous expression to such models, for proving their formal adequacy and consistency, and for investigating their formal limitations But it is here that the difficulty glossed over by Pylyshyn arises: how do we decide under what description the processes embodied in the computer model are the same as those embodied in the brain?

This question can be illustrated by a further example from vision Whereas Waltz's program works by listing for each kind of 2-D vertex the possible combination of kinds of 3-D edges to which it could correspond, Mackworth has written a program that uses the same information about 3-D edges in a much more general way: the direction of a line in the picture constrains the plane in which the corresponding edge in the scene lies Armed with this knowledge, Mackworth's program arrives at a more highly constrained description of the 3-D scene than do those of Clowes or Waltz Moreover, the processes embodied in the program are, at least superficially, very different from those used by Clowes or Waltz The constraints used in all three programs are based on projective geometry and are to that extent similar, but

they are used in different ways and it is hard to design experimental tests of which method is used by the human brain

In their experimental investigations, experimental psychologists rely mainly on error and reaction time data Most AI programs either do not make errors or make very different errors from people; since the relative lengths of time taken by a computer to carry out different processes are unlikely to correspond to those taken by the brain, it is difficult or impossible to extract predictions about human RTs from such programs Pylyshyn is aware of this difficulty, but tends to dismiss it through his example of a hand calculator as a poor model of how people do arithmetic The example is misleading since a hand calculator is computationally very simple, avoiding the difficulties that arise when one tries to decide how far a really complex program models some complex psychological process simply

It could be argued that this difficulty is the most serious current obstacle to progress in cognitive psychology It should, moreover, be noted that it is not unique to theories couched in terms of programs It applies with equal force to any complex model of mental functioning and to many that are not so complex: consider, for example, the failure of experimental psychologists to specify with any degree of certainty the intervening processes occurring in such a simple paradigm as the Sperling memory search task, or the chaos that currently reigns in the field of semantic categorization It may indeed be that we will have to be content for some time to come with very general descriptions of the processes mediating cognition: it could even be that many of the details will only be supplied as a result of advances in neurophysiology The most successful visual preprocessing program (Marr, 1976) directly simulates in its first stage the known physiology of receptive fields and its subsequent stages exhibit why this is a useful way in which to commence the analysis of the retinal image

In summary, then, a successful AI program constitutes an existence proof that the information supplied to the program is sufficient for executing the task It is, therefore, a useful way of investigating task constraints, but the constraints used by the program are not necessarily identical to those used by people Although AI remains the only way of specifying and investigating formal models of complex processes, it is in practice extremely difficult to decide at what level of description a computer model applies to the processes underlying human cognition This difficulty should, however, not turn psychologists away from computer programs since in practice exactly the same problems arise when they try to test their own more loosely formulated models of cognition

REFERENCES

Clowes, M. B. On seeing things. *Artificial Intelligence*, 1971, 2, 79–116.

Gibson, J. J. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.

Huffman, D. A. Impossible objects as nonsense sentences. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 6*. Edinburgh: Edinburgh University Press, 1971.

Mackworth, A. K. Interpreting pictures of polyhedral scenes. *Artificial Intelligence*, 1973, *4*, 121–137.

Marr, D. Early processing of visual information. *Philosophical Transactions of the Royal Society of London B. Biological Sciences*, 1976, *275*, 483–524.

Neisser, U. *Cognition and Reality*. San Francisco: Freeman, 1976.

by **Michel Treisman**

Joint Institute of Aeronautics and Acoustics, Department of Aeronautics, Stanford University, Stanford, California 94305

On the relation between AI and CS: the heart of the problem. Pylyshyn has analysed the differences between the disciplines of artificial intelligence and cognitive simulation He argues that these amount to no more than a "difference in style," i e , they are unimportant The basic argument is that in both areas the tasks tackled are defined by "essentially psychological criteria " This is a phrase that should always arouse suspicion, but in the present case it is offered a reasonable definition: it is taken to define tasks that involve taxonomies that normally require humans for their application, such as identifying the different visual patterns corresponding to a given person's face

To this an artificial intelligence theorist might object that his discipline is essentially different from cognitive simulation First, the aim is to reduce "psychological criteria" to non-psychological criteria, i e , taking the example above, to make it possible to define equivalence classes in terms of the classificatory operations of an automaton rather than those of a human Second, in method: he is willing to employ processes or devices that are known not to parallel those underlying human cognition An example might be the creation of large libraries of templates to solve complex recognition problems

An answer to this, based on Pylyshyn's arguments, might be that what appears incompatible on one level of description may be no problem on another The incorporation of templates in a program may be no more of a difficulty than the use of transistors in the computer on which the program is run, if one chooses to take a sufficiently global view of the overall process as a model for some corresponding cognitive activity Indeed, a procedure that may seem an inappropriate analogy for one task may find an application to another Templates are out of fashion to explain visual pattern perception But suppose I ask, "What is 8 times 7?" It is not unlikely that you will rearrange this as "7 times 8" in your head before finding the answer Access to the internal multiplication table appears to be governed by a lock that recognizes a limited number of keys

The question might be rephrased: "Do programs produced by artificial intelligence practitioners provide models for cognitive processes?" To this the AI theorist above will answer No: because in designing his programs he does not attempt to make them correct representations of the processes and processing that may underlie cognitive attainments He is willing to include features, such as unrestricted and error-free memory, which do not pertain to human cognition but which work, or help his programs to work And, at the level at which his interest lies, these features are important elements in characterising his programs But his opponent, standing on Pylyshyn's ground, will reply Yes: because AI programs are relevant to the problems of understanding cognitive performance They address the same inputs, or tasks, and provide, or attempt to provide appropriate answers Even a model that proves eventually to be wrong may be useful and revealing Further, as Pylyshyn says, if a person and computer both "do task x," there is some level of description, if only we make it sufficiently general and sufficiently vague, at which they do it "in the same way "

To this conflict there is no resolution, because whether we classify the contrast between an emphasis on correctness, or the assignment of priority to relevance, as no more than a "difference in style," or as much more than that, is itself a matter of personal style, and so unresolvable

But this unsatisfactory conclusion is not unique to the AI-cognitive simulation debate It is a feature of all sciences that may be applied to human problems (or some other specialised field) and also more generally For example, consider an imaginary debate between the "mechanical physiologists – men such as Harvey, who recognized the heart to be a pump, and the anatomists who defined bones and muscles as levers and cables – and the common run of engineers The latter might argue: "We are in different fields, pursuing different aims For we desire only that our devices work and solve everyday problems We use wood and metal, not bone And we have no compunction in employing wheels and gears, which do not exist in the human body " To this the mechanical physiologists might rely: "Nay, we are brethren For we employ our knowledge of the principles of statics and dynamics, as do you Your devices may serve as models for us: Would Harvey have understood the heart to be a pump if he had not previously comprehended the working of the common water-pump? At our level of description the difference between bone and metal is no matter Nor need we be overly concerned at the difference between wheels and flail joints For there is a level of description, sufficiently general, at which a man and a car are one: both consume energy to produce forward motion and in this sense both function 'in the same way ' "

by **Shimon Ullman**

Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 545 Technology Square, Cambridge, Mass 02139

A.I. systems and human cognition: the missing link. Are A I systems likely to make significant contributions to the theory of human cognition? It is Pylyshyn's optimistic view that even "pure" A I systems, completely unmotivated by experimental evidence, "can hardly avoid" making such a contribution He argues that although such systems do not use experimental evidence and natural constraints explicitly, their success depends nevertheless

on "a variety of experimental constraints imposed by the natural laws which determine both how the mechanisms operate and how physical environments behave " Various constraints are thereby "smuggled into" pure A I systems, rendering them pertinent to the study of human cognition I accept the argument but not the conclusion An important link is missing which, if ignored, will probably hinder the contribution of "pure A I " to the theory of human cognition If experience gained from A I systems is to help in constructing theories of human cognition, the constraints discussed by Pylyshyn have to be addressed *explicitly* They have to be isolated, their individual influence assessed and incorporated in a systematic theory There are two main reasons why such an approach is needed One stems from the need for what I shall call "mapping-over the relevant conclusions," the second from the need to distinguish between what I shall call type 1 and type 2 theories

1 Mapping-over the relevant conclusions A typical program faced by the cognitive scientist is to understand how the human cognitive system carries out a certain task X The "pure A I " researcher, on the other hand, might try to construct an artificial system that is also capable of achieving X As Pylyshyn points out, constraints from a number of sources affect the possible ways in which the task might be accomplished The constraints discussed by Pylyshyn can be divided into four main categories: (C1) Constraints inherent in the task domain (C2) Constraints that stem from general computational considerations These include computability, aspects of complexity, and the abstract rules of representation and control (C3) Constraints imposed by the particular computation method ("algorithm," "software") in use (C4) Constraints imposed by the mechanism, i e , the physical characteristics ("hardware") of the system

For example, the fact that a standard pocket calculator presents only the first 8 decimal digits of the square root of 2 0 is the result of constraints of type C4 The fact that this number cannot be represented by any finite decimal belongs to the realm of C1 Whether the number is rounded-off or truncated is a problem that falls under C3 Of the above four groups, C1 and C2 are common to *any* system that achieves X, while C3 and C4 might be system-dependent It follows that *to be of value to cognitive theory, the properties of the artificial system determined by C1 and C2 have to be separated from the effects of C3 and C4* Unfortunately, this separation becomes impracticable unless the system has been developed with such a goal in mind Pylyshyn's analysis offers a two-stage solution to this problem In the first stage, large "pure A I " systems will be constructed In the second stage, theories will be developed to shed light on the incomprehensible programs of stage 1

An alternative approach that seems to me more plausible is to develop A I -type models from their conception in an "exploratory" rather than "pure performance" mode, aiming explicitly at isolating and revealing fundamental principles Furthermore, the identification of causal links between certain constraints and assumptions on the one hand and specific properties of the system on the other is not only *useful* in developing a theory, it also constitutes a *part* of it For example, Marcus' work [Marcus 1977] on the syntactic recognition of natural language suggests a link between specific properties of the language parser and Chomsky's "specified subject" and "subjacency" constraints [Chomsky 1975] If his analysis is correct, this link by itself seems to me an interesting part of the theory of language

To sum up the first point: Identifying and exploring the effects of various constraints, assumptions, and engineering decisions (that have to be made whenever a system is implemented) are essential for the development of cognitive theories and for the mapping-over of relevant conclusions To accomplish this task, exploratory programs, whose behavior is well-understood, and which address the above problems explicitly, are needed

2 Overlooking type 1 theories An important question raised by Pylyshyn is: "Can a program be a psychological theory?" The answer is probably positive in a trivial sense: Once a theory has been formulated, it can be cast in the form of a computer program Less trivial, however, are the following aspects of the question: (1) Does the writing of programs provide a *useful* means for the development of theories? and (2) When is a computer program an advantageous way of *expressing* a theory?

I do not wish to examine these questions at length, only to point out some implications to the problem at hand Marr [1977b] offered a distinction between what he called "type 1" and "type 2" theories of information-processing problems In a type 1 theory the underlying principles ("competence") are clearly distinguishable from implementation details ("performance"), giving rise to a relatively "clean" and concise formulation The type 2 theories are less elegant and more complex They might be required, for example, for the description of systems whose behavior is determined by the simultaneous action of a large number of processes The theory that accounts for such a system and predicts its behavior might be as complex as the system itself Newton's Gravitation theory is an example of a type 1 theory that explains the complex motion of the planets The problem of protein folding is cited by Marr [1977b] as a likely candidate for a type 2 theory In the study of cognitive capacities, it is often unclear whether or not the system under investigation has a type 1 theory However, it is important not to overlook type 1 theories when they do exist *The danger of using "pure A.I " systems in the development and the expression of cognitive theories is their tendency to concentrate on "type 2" and overlook "type 1" theories*

Pylyshyn offers the view that a unified type 2 theory might underly such faculties as language, perception, reasoning, and memory I do not share this view, at least as far as it concerns perception Various attempts have been made at devising uniform schemes of representation and control that will assist visual perception as well as other intelligent tasks [Freuder 1976] The lesson from this and other work [e g , Hanson and Riseman 1976] is, as far as I can tell, that such general studies have limited applicability to the theory of visual perception (at least in its present stage) Problems of representation do indeed play an important role in the study of vision [Marr and Nishihara in press], but visual representations seem to be shaped primarily by the specific visual tasks rather than by "abstract information handling principles " In contrast with Pylyshyn's view, it seems to me that in the realm of perception type 1, task-specific, principles can be formulated A careful examination of the problem domain and the study of the individual effects of various constraints will be more effective in unraveling these principles than "pure A I " systems An early example of the "explicit exploration mode" was the analysis by Waltz [1975, *op cit*] of scenes containing polyhedral objects Although there are no reasons to believe that his method of interpretation is applicable to human vision, it goes far beyond preceding work (e g , Guzman 1968), by making explicit the use of constraints imposed by the physical world Some additional examples of evolving type 1 theories in the A I study of perception are Marr's theory of early visual processing [1976] and of occluding contours [1977a], Ullman's [1976] method for detecting light sources and the theory of visual motion interpretation [1977], the theory of human stereo vision by Marr and Poggio [in preparation] and Stevens' [1977] study of local parallelism

Finally, it should be pointed out that identifying the contributions of the various constraints is also important for the success of A I systems *per se* One reason for this is the need to *evaluate partial results* A I systems often achieve only a partial success This is understandable since they often aim at comprehensive "top-down" goals (e.g , the study of children's story comprehension, [Charniak 1972]) But this style of research requires that a certain discipline be adopted to enable the accumulation of partial knowledge [McDermott 1976] A partially successful system usually has both desirable and undesirable properties If the system has been developed in pure-performance rather than in exploratory mode, successive improvements might prove impracticable, leading to the situation described by McDermott [1976]: "After five theses have been written, each promising with fuzzy grandeur a different solution to a problem, people will begin to doubt that the problem has any solution at all Five theses, each building on the previous one, might have been enough to solve it completely [p 8] "

It should be emphasized that A I is potentially a powerful tool for isolating constraints and for exploring the effects of individual mechanisms Compared with the psychologist, the A I researcher has a better control over what is included in his system, and he can trace with greater ease the internal stages that lead to a certain behavior Recent studies that have successfully exploited these advantages are Marcus's [1977] study of natural language parsing and Fahlman's [1977] investigation of the utility of a fast set-intersection mechanism

In conclusion: I share Pylyshyn's view that A I can provide a powerful tool in the exploration of cognitive theories I do not believe, however, that such contributions from A I will come inadvertently, as a spin-off of pursuing "pure A I " goals

REFERENCES

Charniak, E. Toward a model of children's story comprehension. ***M.I.T. A.I. Technical Report AI–TR-266,*** **1972.**

Fahlman, S A system for representing and using real-world knowledge *M.I.T. PhD. Thesis,* Department of Electrical Engineering and Computer Science, 1977.

Freuder, E. Computer system for visual recognition using active knowledge. *M.I.T.A.I. Technical Report AI-TR-345,* 1976.

Guzman-Arenas, A Computer recognition of 3-D objects in a visual scene. *M.I.T. A.I Technical Report AI-TR-228,* 1968.

Hanson, A. and Riseman, E. A progress report on VISIONS: representation and control in the construction of visual models. *COINS Technical Report 76-9,* The Univ. of Mass. at Amherst, 1976.

Marcus, M. P Theory of syntactic recognition for natural language. *M.I.T. Ph.D. Thesis,* Department of Electrical Engineering and Computer Science, 1977.

Marr, D. Early processing of visual information. *Phil. Trans. Roy. Soc. B, 275,* 483–534, 1976.

Analysis of occluding contours. *Proc Roy. Soc. Lond B, 197,* 441–475, 1977a.

Artificial intelligence – A personal view. *Artificial Intelligence, 9,* 37–48, 1977b.

Marr, D. and Nishihara, K. Representation and recognition of spatial organization of 3-D shapes. *Proc. Roy. Soc. Lond. B* (in the press).

Marr, D. and Poggio, T. A theory of human stereo vision. In preparation.

McDermott, D. Artificial intelligence meets natural stupidity. *Sigart Newsletter, 57,* 4–9, 1976.

Stevens, K. Computation of locally parallel structure. *M.I.T. Artificial Intelligence Memo 392,* 1977. (Also to appear in *Biological Cybernetic*).

Ullman, S. On visual detection of light sources. *Biol. Cyber 21,* 205–212, 1976.

The interpretation of visual motion *M.I.T.Ph.D. Thesis,* Department of Electrical Engineering and Computer Science, 1977.

by **Walter B. Weimer**

437 Bruce V Moore Building, University Park, Pa 16802

A.I and the methodology of scientific research: some cautions and limitations. While Pylyshyn's concern to increase the credibility and utility of A I research to the study of cognition is laudable, he makes or implies several methodological claims which cannot go unchallenged I believe that considerable "evidence," in the form of cogent arguments long available from the philosophy of science and the methodology of scientific research, militates against the position he presents, and restricts the manner in which A I can contribute to a hoped-for science of cognition Consider four classes of constraints that greatly change the relationship between A I and cognition from what Pylyshyn proposes

The role of technology in science Although it need not be in principle, A I is limited in practice to the use of available computation systems (computers and their programs) Thus Pylyshyn asks "Can a program be a psychological theory?" and later asserts "What is needed is something approaching a theory of the program " So a program alone cannot be a theory: It can, at best, be a model which instantiates a theory, and thus its role is to provide data which must be assessed in the light of methodological criteria Thus computation, like all technology, is never of decisive *theoretical* import in science: It can constrain and change the data base, but it is never sufficient to guarantee theory change (see Weimer, 1976) No matter what data it provides, A I need not change any psychological theory of cognition, no matter how "relevant" those data Thus we must reject Pylyshyn's claim that A I could perhaps "supply the foundations for a cognitive psychology," because no matter what transpires in A I , it will not be a "foundation" for any science Science has no foundations that are other than conventional agreement (See Popper, 1959, 1963; Kuhn, 1970; Lakatos, 1970; Weimer, 1975)

Data and the inference to theory Even when correctly interpreted as models to provide data for the domain of cognition, there is absolutely no reason to suppose that computation data (despite their quantitative precision or "axiomatic" appearance) are superior to any other form of data (including mystical intuition) in leading to theoretical principles It is not, as Pylyshyn's abstract asserts, that "the task of extracting the relevant theoretical principles from a large complex program may be formidable," but rather that it is simply impossible The leap from data to theory must always be made by theoretical imagination (tacit knowledge, call it what one will) rather than by some explicit or formalized inductive inference procedure This is so because there are always an indefinitely large number of theories that entail any amount of data *and* are in conflict with one another, and there is no way to determine, from evidence or data, which of those theories is more defensible (see Maxwell, 1975; Weimer, 1977) The computer can't think for us, and no matter how much data it provides it can't make our theoretical work any less difficult

Interjected caveat: the danger of symbolic precision If one appreciates the methodological constraints that rule out any algorithm or "logic" of scientific discovery, the notion of data as foundations for science, etc , it is easy to see that sheer quantification or symbolic precision is of no merit in science Early opponents of the computer as a panacea were often shouted into silence by the scientistic taunt that precision and exactness were the "essence" of science, and opposition thereto must obviously be "unscientific" and obscurantist It is commendable that Pylyshyn carefully avoids any such scientism (as Hayek, 1952, called it), although his remarks on the lack of success of "grand theoreticians" in psychology and the trend toward "formalism" in specific domains leaves one in doubt as to why precise theorizing is nonexistent in psychology The answer may lie in some problems of complex phenomena which Pylyshyn's account does not address

Complex phenomena and the limits of explanation Sciences which deal with phenomena of low complexity (such as physics) explain those phenomena by subsuming particulars to covering laws, and those laws in turn to theories which "deduce" them This account was proposed by positivists and logical empiricists as *the* nature of scientific explanation (and codified in the well known Hempel-Oppenheim "covering law" model) But as Hayek (1967) pointed out (initially from studying the complex phenomenon of the market place in economics in the 1920s) the phenomena of the social, psychological, and biological sciences require another type of explanation, explanation of the principle, rather than explanation of the particular They do so because it is impossible to achieve explanation of the particular in highly complex systems Later, and independently, von Neumann (1966) argued the same thing in his pioneering research in automata theory, which showed that in a system of high complexity the simplest model of a phenomenon is at least as complex as the phenomenon itself Thus our understanding of complex systems such as cognition will be limited, in principle, to accounts which provide what Hayek called explanations of the principle (see Shaw, 1971; Weimer, 1978) The "new" kind of theory which Pylyshyn correctly sees to be necessary for cognition will not be of the axiomatic form associated with Newtonian physics because that latter theory is one of explanation of the particular for simple (low complexity) phenomena The trend toward formalism will be irrelevant to the study of complex phenomena insofar as it is limited to explanation of the particular What we require for "cognitive science" to be other than a promissory note is not just more use of the computer, or functional and intentional specification of variables, but an understanding of complex phenomena Pylyshyn's "new theory to explain the program" will not explain particulars, but must instead specify the principles according to which the system operates Thus it would appear that A I will be of value if it can help us to achieve explanations of the principle for cognition, not because it is based on top-down analysis (all theoretical science must be), nor because it is intentional (even behaviorism is), etc [see Haugeland et al , next issue]

REFERENCES

Hayek, F. A. *The counter revolution of science: Studies of the abuse of reason.* Chicago: University of Chicago Press, 1952.

Studies in philosophy, politics and economics. Chicago: University of Chicago Press, 1967.

Kuhn, T. S. *The structure of scientific revolutions.* Chicago: University of Chicago Press, 1970. (Revised Edition).

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, 91–195.

Maxwell, G. Induction and empiricism: A Bayesian-frequentist alternative. In G. Maxwell & R. M. Anderson (Eds.), *Induction, probability, and confirmation.* Minneapolis: University of Minnesota Press, 1975, 106–165.

Popper, K. R. *The logic of scientific discovery.* New York: Harper & Row, 1959.

Conjectures and refutations. New York: Harper & Row, 1963.

Shaw, R. E. Cognition, simulation, and the problem of complexity. *Journal of structural learning,* 1971, 2:31–44.

Von Neumann, J. *Theory of self-reproducing automata.* Ed. by A. Burks, Urbana: University of Illinois Press, 1966.

Weimer, W. B. The psychology of inference and expectation: Some preliminary remarks. In G. Maxwell & R. M. Anderson (Eds.), *Induction, probability, and confirmation.* Minneapolis: University of Minnesota Press, 1975, 430–486.

Manifestations of mind: Some conceptual and empirical issues. In G. G. Globus, G. Maxwell, & I. Savodnik (Eds.), *Consciousness and the Brain* New York: Plenum Press, 1976, 5–31.

Scientific inquiry, assessment, and logic: Comments on Bowers and Mahoney-DeMonbreun. *Cognitive Therapy and Research,* 1977, 1:247–255.

Hayek's approach to the problems of complex phenomena: An introduction to the theoretical psychology of *The sensory order.* In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes,* Volume II. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978. In Press.

by **Yorick Wilks**

Department of Language and Linguistics, University of Essex, Colchester C04 3SQ England

Artificial intelligence and real constraints. Pylyshyn's paper addresses a most difficult and important issue: Do Artificial Intelligence (A I) systems, whether programmed on computers or not, have anything of value to offer to cognitive psychologists? I take the point of departure of his argument to be the passage where he sets out the view, in order to attack it, of those who answer no to the above question on the ground that "A I is strictly a rational exercise in building formal systems, more closely related to pure mathematics than psychology " Pylyshyn argues in the paper that A I systems are in fact empirically constrained in many ways, and that investigation of these ways lends no support to the supposed distinction between A I and cognitive simulation, on the ground that the latter is constrained by empirical psychological observation, while the former is not

Since I find myself in agreement with Pylyshyn's substantive claims, while at the same time admiring the breadth of perspectives he has been able to draw on these issues, my commentary will be restricted to expository and peripheral matters

On the expository level, I think Pylyshyn relies too much on distinctions he has drawn elsewhere, but has no space to explain in this paper One example I found particularly acute: a central claim for Pylyshyn is that both cognition and computation are "intentional rule-governed phenomena," but instead of some exploration and exposition of this interesting claim the text whisks us off to texts by Dennett and Fodor et al Doing that might be alright if the meaning of the key phrase were clear, but it is not, and most particularly the word "intentional" in this context is a clique usage that is liable to mislead It may even cause unwary readers to think it is no more than a typo for "intensional," and who could blame them, since such contexts as "an intentional vocabulary which includes beliefs " show that the word *does* have much of the meaning of "intensional" as well as the standard philosophical sense of "having to do with intentions " I am not accusing Pylyshyn of confusion here, only arguing that the sense he assigns to this key term must be provided if it is to bear the weight he seems to want to put on it

Let me now make brief comments on four closely related issues in the paper, which go beyond mere expositional difficulties Each issue demands far more extended and careful discussion than I can give it here

1 *Same task performance implies "doing things the same way"* This issue is near the heart of the relation of A I and psychology, and Pylyshyn's claim is essential to his argument that "even 'pure' A I can hardly avoid making some contribution to cognitive psychology": " if a person and a computer are both capable of 'doing task X' there is some level of description of the two at which they are doing it 'in the same way' "

The most obvious trouble with this claim is the existence of what are prima facie counter examples – the successful sentence analyzer that processes English right to left – unless Pylyshyn's phrase "some level of description of the two" is so interpreted as to *make* the claim true; that is, in the standard philosophical jargon, unless Pylyshyn's claim is *analytic* But that would be dubious procedure in a paper about the power of empirical constraints!

Moveover, in an interesting section later on, Pylyshyn discusses three "level of description," arguments that might cause us to say that a standard pocket calculator was not a proper model of the human arithmetical process I cannot find any precise point in that discussion where he avoids the trap he has set himself with the above quoted claim, according to which there must be some level of description at which the calculator *is* doing things "the same way" as we do

2 *A model's non-independence of biological implementation (in brain science) and of mechanical implementation (in cognitive science)* Pylyshyn distinguishes these two notions of non-independence and notes that, in the latter case, machine implementation "makes empirical investigation of a certain kind possible " I agree about the importance of this, apart from the reservations of (3) and (4) below

However, Pylyshyn goes on immediately to introduce a weak and strong version of the thesis of the independence of cognitive models and brain mechanisms, one of which he seems by implication to hold But he also appears to support the non-independency thesis on machine implementation of cognitive models

There is no inconsistency so far, but it does seem to arise when he, at one and the same time, uses an *in*dependency thesis on machine implementation as support for an independency thesis on cognitive models and the brain: " just as it is highly enlightening to study algorithmic processes independently of how they are implemented so there is much that can be gained in trying to understand cognitive processes independently of biological mechanisms "

3 *Difficulties about the status of the implementation of A I models* Pylyshyn correctly believes that the machine implementation of A I models, or theories as they should be called, is important He also seems to believe, in a quite straightforward way, that A I *experiments* have been done, though I think this requires considerable documentation and demonstration. This claim is important to Pylyshyn because he wishes to present A I as a highly empirical subject, although I believe that all his major points still carry without this claim that A I experiments have actually been done

There is already a clear sense in Computer Science in which a theory is distinguished from a program embodying it, contrary to Pylyshyn's worry that " a clear separation between fundamental laws and systems of calculation based on them is not easily available in computational theories " And in a clear philosophical sense it can be shown that programs can be models of theories of a cognitive science type, as I have tried to set out (in Wilks, 1974) So Pylyshyn, in my view, has no need to stretch the word "experiment" to include current A I work so as to argue the empirical status of A I , for that can be independently established along other lines in his paper

A very complex issue relevant to this point, one that Pylyshyn touches on but which I would very much like to see him develop further, is the significance for his argument of proofs of program correctness: that is, strong implementation independence for A I theories These do not exist for any really interesting A I programs, but would it affect Pylyshyn's claims about the empirical status of A I (and so confirm the psychologists' fears about A I that I quoted at the beginning of this commentary) if they did?

4 *A I and discovery* Pylyshyn refers to the work of Waltz as constituting "empirical discovery" in A I I believe that it in no way belittles Waltz's considerable achievement to deny that the discovery came, as Pylyshyn claims, "from empirical exploration of the program," at least in the sense of running it in anything like an experiment Even if I am right, this does not tell against the relevance of A I for cognitive psychology, nor does it deny that Waltz's work is properly called A I

From the little I know of vision work, it seems to me that Waltz's observations, to which Pylyshyn refers, could have been made without writing a program at all, and so the program functioned for him only in the older sense (i e , NOT the A I sense) of "heuristic": as "an aid to discovery "

It should be clear by now that I agree with Pylyshyn's claims, but believe that at several points he has misdescribed A I work, and in a way that he need not have done; for his arguments would have stood, even with A I in the slightly less empirical and more confused state that I believe it to be in It will, I hope, be clear that I found Pylyshyn's paper enormously stimulating, and a very encouraging pointer to the future relations of A I and psychology

REFERENCE

Wilks, Y. One small head: remarks on models and theories in linguistics. *Foundations of Language,* 1974, *11,* 1–21.

Author's Response

by **Zenon Pylyshyn**

The A.I. debate: generality, goals, and methodological parochialism

1. Prelude

It is gratifying to have my essay receive such wide interdisciplinary attention. Though the commentaries vary widely in their emphasis and approach, collectively they raise most of the contentious issues about the foundations of cognitive science that have been debated inside and outside the discipline. In responding, I have in most cases chosen to address a number of recurring themes rather than to reply directly to specific authors, thereby decreasing the temptation to indulge in nit-picking arguments.

As a preface to my response I must point out that many of the commentaries were directed at points that I had simply not undertaken to defend in my brief essay. Thus I had not gone into the question of what Ullman refers to as "constraints imposed by the mechanism" or what I call elsewhere (1978) the "architecture of the cognitive virtual machine." Nor had I assumed the task of defending the technical achievements of A.I. against the field's perennial detractors (e.g., Dreyfus, Haugeland, Cushing and Hornstein). I shall nonetheless make some brief remarks on both these issues later in this response.

As a final prefatory remark I should attempt to clear up one general misunderstanding concerning the goal of my essay. It was not intended to threaten anyone's job security by suggesting that A.I. and psychology were identical! The differences beween the two disciplines are obvious enough not to require comment. My remarks were addressed to that subset of the manifold goals of the two enterprises that is specifically directed toward the intellectual task of casting some light on the nature of intelligence or cognitive functions. Here I am making precisely the distinction Ullman refers to as "explanatory" versus "pure performance" mode. As the examples Ullman cites show, this distinction cuts across the A.I.-psychology discipline boundary.

But clearly a large proportion of the efforts of both disciplines is directed in various directions other than explanation and understanding – such as developing tools, solving specific technical puzzles, and applying the resulting insights and techniques to various engineering enterprises (e.g., behavior modification, automated postal sorting systems, etc.). Thus the fact (on which McDermott focusses in his commentary) that many A.I. workers "have no (higher) goal than to build a system that can do a single class of tasks intelligently," and indeed get along very well with little interest in generality, is really beside the point. I wish them every success, as I do the innumerable psychologists who likewise have different goals from those with which I was concerned in my essay. Nonetheless, there *is* a deeply rooted and widely accepted goal in both camps directed at understanding the underlying processes of intelligence, and this goal does have to take generality seriously, if for no other reason than that it may well be the defining characteristic of the phenomenon of intelligence and may be the sole criterion able to exclude "ad hocery" (see also section 4).[1]

As a further footnote to this distinction between the body of work being pursued under the aegis of some particular disciplinary affiliation and the specific task of providing an understanding of intelligent action I should acknowledge a possibility raised astutely by Domotor. He observes that while A.I. may have started out with a rather deliberate interest in human intelligence, it could develop an autonomous independent view of its subject matter which would take it in a quite different direction in the future. He cites as a parallel historical case the development of logic as a subject independent of psychology. This is an intriguing possibility, closely related to the competence-performance distinction in linguistics. There is also some evidence for such a possibility in A.I. from such work as John McCarthy'sttempt to study the epistemological as distinct from the heuristic problems of A.I. In general, however, I do not believe that such a pursuit will form the basis of an autonomous A.I. discipline. One reason is to be found in the very quotation from Minsky and Papert that Domotor cites at the beginning of his commentary: viz, A.I. is committed to the criterion of realizability. This in turn puts it firmly in contact with the heuristic problems – with pragmatics, resource-limited computation, and with behavior that relates to human performance. Unless the discipline changes radically it will be staunchly unwilling to rest on the sorts of idealizations accepted in logic and linguistics. As a consequence, its commitment to both explanation and the creation of working systems will locate it somewhere between a design discipline and a pure science.

Nonetheless, Domotor's point is well taken, it is clearly logically possible for some future A.I. program to pursue a less anthropocentric notion of intelligence and hence to be of less direct relevance to psychology than present projections suggest. But then there would still remain the computational task (perhaps no longer called A.I.) whose goal it would be to develop natural intelligent (N.I.) systems and that, much to the dismay of some members of the psychological community, might still place little emphasis on the data of psychological laboratories as opposed to more general sufficiency constraints. This would presumably not prevent people like Ortony from protesting that N.I. was unimportant to psychology on the parochial grounds that "it is the psychologists, not the computer scientists who must make such evaluations."

2. Methodological parochialism and the equivalence of processes

The last point, concerning the parochialism inherent in the views expressed by many of the commentators, deserves some special attention because it is both widespread and apparently quite unconscious. The widespread belief that A.I. is concerned merely with designing *a* method of doing a task while psychology is concerned with how people *actually* do it is reminiscent of an old debate that some psychologists conducted with generative linguists. In the latter case the cry was, "You linguists are merely concerned to provide some method of characterizing the structure of a language, whereas we psychologists want to find out how people *actually* represent and process language." The grain of truth in these claims should not obscure the fact that the main distinction among linguists, psychologists, and A.I. researchers (at least as pertains to these particular arguments) is the class of evidence to which they assign a special or privileged status in their practice. However, there is no a priori reason why reaction time, frequency of errors, extendability of the theory to broader domains, intuitions of ambiguity or well-formedness, or the ability actually to perform a theoretically interesting class of tasks should be taken as exclusive if one's goal is to understand cognition. Clearly, all these criteria can be relevant to understanding cognitive phenomena, and none of them can claim to represent a more direct path to psychological truth. Thus Arbib is quite correct in asserting that "a good psychological model *must* confront data," as is Johnson-Laird in insisting that such systems must "both cohere and correspond to facts." But to claim further that certain specific kinds of data must be given priority is simply to reveal one's affiliation, or bias, or deliberate research strategy. It is not an argument that A.I. systems are thereby less eligible candidates for psy-

chological models than are psychologists' systems: Both are partial explanations, though they may be partial in different ways.

As another example of how this parochialism manifests itself, consider what seems to me to be a rather straightforward remark I make in my essay to the effect that doing the same task versus doing a task "in the same way" is not as obvious a distinction as some people might assume. This remark was based on the observation that to specify what constitutes "doing a task" is really to specify certain constraints on a class of methods. Nearly one-third of the commentators singled out this remark as "obviously" false. Counterexamples cited ranged from different ways of demolishing a house, playing chess, and doing multiplications to recognizing typewritten characters, translating languages, and looking up words in a dictionary.

A number of lessons can be learned from examining several of these examples. First, some computer methods (e.g., language translation by word-for-word transliteration) are simply bad methods in that they do not work. They are bad as psychology and bad as A.I. They are not responsive to the demands of the task and nothing more can be said about them than that one has no more guarantee of winning by playing on a computer than by playing in a laboratory. Second, methods such as those for recognizing characters may be useful for narrow practical applications, but may be totally inadequate as general accounts of vision. This conclusion, however, is one that would be just as obvious by A.I. criteria as by psychological ones, and would not provide any basis for arguing divergent goals for A.I. and psychology. Finally, the chess example provides an excellent illustration of the main point I was driving at in my remarks regarding how specifying what constitutes "doing a task" constrains the class of possible methods.

Chess shares with many A.I. tasks the property of not being a very well-specified task. What a chess master knows, how this knowledge is organized, and how it must be accessed and operated on in order to meet the general criterion of "playing well" or of "making the correct move" are open research questions. *Any* program that plays very good chess is a contribution to specifying what the task of playing chess *is*. It constrains the set of methods that correspond to "playing chess well." Doing experiments on good chess players (Chase & Simon, 1973) also does this, but uses a quite different point of entry. Both are incomplete specifications and both provide constraints. I see no principled difference between these approaches, though clearly they represent very different research strategies. The A.I. worker *may* be satisfied in stopping at an incomplete understanding of the chess-playing task, provided his program performs well. He may stop short of inquiring what the general principles are in virtue of which high performance was achievable and may still be applauded by some of the A.I. community. On the other hand, the psychologist *may* in turn be satisfied with an incomplete understanding of chess masters' competence, provided his model predicts latencies and error data. He may stop short of asking how his various hypothetical mechanisms could function together to to produce expert play and yet he may still publish his results in a respected psychological journal. But one should not conclude from these differences in tolerance for incompleteness that the two enterprises are divergent. On the contrary, a desire to understand how general principles, particular knowledge, and task demands fit together in the context of chess lies at the intersection of both approaches.

The basic point in this discussion can be put another way. The reason that too much is made of the distinction between doing the same task and doing it in the same way is that there is a presumption that "the same way" is a clear and well understood notion. It is not. What is at issue here is nothing less than the very difficult notion of the semantic equivalence of processes. Such an equivalence relation is highly abstract, as Scott and Strachey (1971) have shown. To suppose that knowing that two systems (or an organism and a computer) are doing the same task will further allow us to say whether they are doing it in the same way is to suppose that we have general agreement on what constitutes semantic equivalence (strong equivalence) for cognitive processes. Though the goal of establishing strong equivalence is undeniably important, to claim that one class of constraining methods (e.g., reaction time, error rate) is superior or more direct than another (e.g., generality of mechanisms, intuitions, logical analyses of the task, extendability or learnability of the method, and any of the class of "intrinsic constraints" discussed by Moore & Newell, 1974) is simply to reveal one's methodological allegiance.

3. Underlying architecture

In my essay I placed considerable emphasis on the notion of level of description or level of aggregation, though the topic was not described as fully as elsewhere (1978). Longuet-Higgins focusses attention on this issue when he speaks of the need for "an appropriate language in which to specify the detailed processes." For a language defines a *virtual machine architecture* – a set of commands to be taken as primitive, and certain specifications concerning the way in which these commands interact, the resources available to the system, and various other restrictions and facilities governing the processes that can be expressed in the language and are hence executable on the virtual machine. The architecture determines both what algorithms can be carried out and their relative complexity functions over different inputs. For example, a classical Turing machine cannot carry out a tree sort algorithm, which requires a register architecture, while a register machine in which arithmetic operations are not primitive cannot execute the usual binary search algorithm, and so on. A virtual machine of special interest to cognitive psychology is one that contains exactly those operations that are cognitively primitive and that conforms to appropriate resource limitations characteristic of human processing. Elsewhere (1978, in preparation) I have suggested a number of criteria for determining whether a proposed operation is a suitable candidate for a cognitive primitive. Briefly, this relies on the notion that if an operation is too macroscopic, it will fail to account for certain observed differences among cognitive phenomena, while if it is too microscopic, it will fail to capture significant cognitive generalizations or rules.

Clearly, a great deal more needs to be said concerning this idea of a cognitive virtual machine architecture. Recent work by Allen Newell (1972, 1973a) is directed explicitly at its design. An adequate architecture would not only address the issue raised by Ullman of isolating and explicitly addressing the separate sources of constraint on computational systems, but it would also resolve the level of description problem, as Newell observed in his commentary. Put in terms that linguists frequently prefer, one might say that such a virtual machine would represent the minimally powerful formalism for expressing humanly realizable processes, and so it would, like universal grammar, capture universal properties of mind. It would also address what Pascual-Leone refers to as "metasubjective constraints."

Now, although the design of such an architecture is clearly a goal of special concern to psychologists, programming such a virtual machine to carry out intelligent tasks is exactly the same kind of problem as is faced in writing A.I. programs subject to the constraints of any other virtual machine (e.g., LISP). Thus, the enterprise does not resolve the debate raised in section 2 concerning the question of which additional process constraints should receive priority.

In addition to inferring the primitive operations that make up the mental architecture, one may also be interested in abstracting the more general properties of the class of processes adequate for a particular task at hand. For example, one might wish to specify some intrinsic constraints that such processes should meet (along the lines of Moore and Newell, 1973) or to attempt to

give an abstract characterization of the *method* that a particular function demands (what Marr, 1977, calls a type I theory and what I have referred to in another context as "requisite computations" – Pylyshyn, 1972). One might even go further and attempt to develop an abstract mathematical theory of certain classes of procedures or data structures (as some of the work in theoretical computer science attempts to do). All of these enterprises are undertaken in A.I., although they may perhaps not attain the same prominence outside the field as, say, specialized high-performance systems.

In view of the diverse levels of abstraction at which computation is studied within A.I., it is puzzling to find Cushing & Hornstein presenting what they seem to feel is an original proposal that "any adequate computer model will have to be formulated not in terms of programs, as in current A.I. work, but in terms of software systems." This will no doubt come as a surprise to researchers in A.I., who from the beginning of the discipline have taken the notion of software systems and their attending problems of control structures, run-time structures, modularity, extensibility, interfacing, data structure design, and so on, as the core technical problems of A.I. (see for example, the early paper by Newell, 1962, or the review of the goals of A.I. language design by Bobrow & Raphael, 1974). If Cushing & Hornstein think that they have stumbled onto a new idea with their observation that "the mind is a highly complex system of related and interacting but essentially autonomous components," then they simply have not been reading the A.I. literature (a relevant sample of which might include Hewett, 1977; Bobrow & Winograd, 1977, all the work on extensible languages, on production systems, on distributed computation, and in fact just about anything on the problems of organizing computation systems for modularity and extensibility). The same applies to the authors' claim that we require higher-level languages and theories of software systems. Such pursuits are the bread and butter of theoretical A.I. The sort of mathematical characterization of software systems the authors cite is just one of a large number of approaches to the theoretical understanding of large interacting systems, and only time will tell whether it has any special merits in contrast with, say, the work on distributed computation implicit in the studies of speech recognition, or the analyses of computation on the HYDRA system (both being carried out at Carnegie-Mellon University). In any case, it is clear that the sort of venture Cushing & Hornstein have in mind is not being neglected within A.I., although it is not obvious (nor have they proposed an argument to support the position) that progress at such an abstract level will contribute much in the absence of experience gained from attempts to develop specific systems designed to perform specific tasks.

4. Generality and general principles

Arbib argues that my generality criterion leads to an infinite regress since it is always possible, when confronted with a system that seems unreasonable as a cognitive model, to say that it was not sufficiently general. But this is only circular the way all inductive hypotheses are. To provide evidence for the hypothesis that generality leads to convergence, one does not examine a single case but one attempts to show that more general systems tend to be more plausible as cognitive models (e.g., contain more terms and processes that can be given psychological interpretations). This seems to be clearly the case at the lower extreme of the generality dimension. However, Arbib is correct in claiming that generality may not be in the program but in the "... set of concepts to aid the sharing of methodology between diverse projects." But the concepts that aid in sharing methodology must be ones that reveal the principles whereby high performance can be achieved. Insofar as we design systems with such principles in mind, attempt to isolate or make transparent their sources of performance, and use subsystems and methods also used by other processes, we obtain what I would consider to be a more general system, for generality does not refer only to the performance of one isolated program. To the extent that a program is an instance in a more general framework, it can be related to or interfaced with other programs and, more importantly, it becomes, in principle, more extendible. That is the crucial sense of generality. And incidentally, in this sense I consider it quite likely that the HEARSAY system (which Arbib refers to as an example of a more cognitively oriented system) will also prove to be a more *general* system than the Markov-chain model HARPY (at least if closely parallel early developments in machine translation can be taken as a guide).

5. Serial versus parallel computation

Arbib, Harmon, Pribram, and Sloman raise the issue of serial versus parallel computation. There is no question but that this distinction is an important one. Whether or not a process involves parallel computation is really a question about how resource limits apply to it and is thus a material distinction. My point was that this is only relevant insofar as it constitutes a functional distinction. This is why I emphasized that structural evidence (e.g., "in the brain various events are taking place in different places at the same time") does not bear directly on this distinction. I offered as a counterexample the observation that we would find the same to be true in a *serial* digital computer if we poked electrodes into it. The reason that structural evidence cannot decide the issue is that below the level of computational primitives (i.e., below the level of virtual machine architecture) one is concerned with questions of implementation, admittedly interesting in their own right, but not questions of algorithms or methods, and hence not questions of whether the *cognitive* process is serial or parallel.

As a footnote to this point I might remark that Harmon's numerical argument for parallel processing is not sound, for related reasons. The only relevant events are computational event-types and not physical event-tokens. This can be highlighted by noting that the numbers game with respect to brains can also be played to show that "the entire lifetime of the universe" would be insufficient to simulate the activities of one computer by another. One has but to enumerate all the physically discriminable states of the target computer and all the physical connections through which causal effects could propagate. But of course only a minute fraction of the physical parameter fluctuations and causal chains are relevant to the device functioning as a computer. Since we have little idea of what the computationally (i.e., cognitively) relevant physical events are in the brain, we cannot simply count anatomical units indiscriminantly and hope to draw meaningful conclusions about cognition's being serial or parallel.

6. What is "empirical"?

Sloman, Leibovic, and Otto* question whether A.I. can be considered empirical, as opposed to merely formal. Sloman is most specific in distinguishing formal from substantive explorations and discoveries. I do not wish to make too much of the term empirical, and perhaps Sloman's distinction is a useful one. But I think it is worth pointing out that the distinction between these two types of discovery is precisely the difference between what philosophers have called analytic and synthetic truth and is subject to the same arguments as those used against that distinction. For instance, Quine (1951) pointed out that the fact that some propositions seem to be analytically (or formally) true, and thus independent of how the world actually is, may simply reflect our lack of creativity in imagining how it might otherwise be. I think Waltz's discoveries are a good example of this. Sloman takes the Waltz discovery to be "merely a formal empirical discovery" –

implying that it has nothing to do with how the world actually is but merely reveals something about a formal system. But one could (with difficulty perhaps) imagine a world in which objects, reflectances, sensors, and so forth, were such that Waltz's system of labels did not converge on an interpretation of the scene. One might then be inclined to say that Waltz had made a substantive empirical discovery, just as it is an empirical discovery that the identity of a phonetic stimulus cannot be established from physical properties of a segment of the speech signal temporally localized at the point where that phonetic stimulus is perceived to occur. In fact, the types of knowledge (e.g., phonologic, syntactic, semantic) that must be brought to bear in identifying a phonetic stimulus were empirically discovered in the course of designing speech recognition systems. (The word "must" here has the force of a claim that the theoretical generalization is empirically true – not that it has been formally demonstrated. Indeed, further research could conceivably show it to be false.)

7. The goals of psychology and A.I.

Lenat's rather whimsical parody suggesting a parallel relation between A.I. and astrophysics on the one hand and A.I. and psychology on the other misses two crucial points that render it quite irrelevant. The first is that whatever might be their similarities, A.I. and astrophysics investigate completely different phenomena, while quite the opposite is true of A.I. and cognitive psychology, as I have argued. The second is that although A.I. and astrophysics may employ the same tools (even computers) it would be a mistake to consider the computer as a tool in the same sense when used in A.I. and cognitive psychology. But Lenat appears to assume exactly this in his claim that the psychologist "builds his program solely to run it" in order "to validate a theory." But this is a gross oversimplification of the cognitive science enterprise. It assumes, for instance, that a theory is first constructed, making use of experimental data, and then a computational model is built to instantiate it. Apart from some very simple and uninteresting cases, this is not at all what happens. The attempt to understand some phenomena in terms of methods sufficient to exhibit them, to design such methods subject to general constraints as well as some known facts (not necessarily stemming from psychological experiments), to examine the (partial) system for sources of inadequacy, and to redesign the system in the light of discoveries made along the way all contribute to the basic *content* of the model and typically precede the development of a psychological theory. The situation is in fact exactly parallel to the story Lenat gives of what the A.I. researcher does.

Lenat has a clear understanding of the A.I. enterprise, but his commentary betrays a stereotyped and seriously flawed image of what the cognitive psychology enterprise is about – an image that, I should add, is also shared by a large number of psychologists. These prejudices are sufficiently serious and widespread to merit some extended response. First, Lenat claims that cognitive simulation's concern is "to *match* human performance – including human error and imperfection." The status of evidence regarding human errors was raised by many of the commentators. Some, like Sloman and Pascual-Leone, feel that too much emphasis is being placed on error-infested performance by "computer simulation" psychologists. Others, like Sutherland and Treisman, appear to emphasize its importance. Arbib seems to take the same position as Lenat does when he speaks of the need to account for "human fallibility." But the view that a central task of cognitive psychology is to duplicate errorful performance of humans is, to say the least, a very misleading way to look at the goals of that science.

Essentially, cognitive science seeks to understand, not to *match* anything (not withstanding the ubiquity of the "variance accounted for" criterion, a methodological throwback from the positivist era). It does this by searching for general principles and showing how these, in combination with particular knowledge, particular goals and tasks, and particular mechanisms, are able to account separately for different aspects of a phenomenon. Errors and imperfections are not the primary phenomena to be accounted for; rather, it is the competence to deal with the task. The importance of error data is that they provide clues as to how this competence may be realized within certain kinds of resource-limited mechanisms. In other words, errors tell us something about the way the algorithms and the architecture fit together – just as errors in real computers can do the same (though in that case they are called "bugs" and we try to eliminate them). The reason errors are helpful in this respect is that the constraints of the underlying architecture are most visible when the system is forced to operate at its limit, that is, when errors are induced.

Second, Lenat claims that A.I. is interested in superhuman performance or performance that may rarely be exhibited and then only by the very best experts; the performance of such tasks "would not be fit for cognitive simulation." This is the myth that psychologists attempt to account for the typical or the modal cases. But that is simply false in general. If one wants to understand language phenomena, one studies competent speakers. If one wants to understand chess skill, one studies chess masters (when they are available). If one wants to study mathematical skills, one studies expert mathematicians, as did Hadamard and Wertheimer. Again, as in the case of the use of error data, evidence gathered from beginners or ontogenetic evidence gathered from children is useful either to infer subskills directly, to conjecture as to the set of possible methods, or because it is of intrinsic interest. But to understand cognitive processes is not to account for the typical or most frequent or the bungling cases.

Another interesting observation that casts doubt on the view that psychology needs to proceed by working up from data to theory to implementation is provided by Cohen's commentary. He gives a nice example of the central role played by the logical analysis of a task domain. It raises an issue I did not discuss in my essay but that provides another argument in favor of the A.I. approach to cognitive modelling. The way in which experimental data are interpreted is extremely sensitive to one's implicit normative system. Whether behavior is to be construed as appropriate to a task or aberrant depends on one's understanding of the task and its goals as well as on the normative system one adopts. Thus one might argue that both A.I.'s contribution to task analysis and philosophy's contribution to the analysis of normative systems (e.g., the Pascalian versus non-Pascalian probability case cited by Cohen) ought to be pursued prior to the attempt to devise a model from the experimental data.

8. The achievements of cognitive science

In spite of the fact that my essay was clearly not intended as a defense of the accomplishments of either A.I. or cognitive psychology, a number of commentators argued that my main points were invalidated by what they considered to be the lack of achievements of A.I. Although I tend to agree with Andreae that a defense of the field is not necessary, a number of points raised by several commentators in their attack on A.I.'s achievements bear some comment.

Dreyfus correctly takes generalizability, rather than the mere generality of one particular program, as the appropriate criterion for discussion. But then both he and Haugeland go on to argue that the sorts of systems I alluded to are incurably nongeneralizable. I do not wish to enter into arguments concerning the merits of particular systems. I suspect that many of those cited by Dreyfus are, indeed, not generalizable without some major redesign. The point I wish to make is that this is not an issue on which one can pronounce without a deeper understanding of the systems in question and the (often implicit) principles determining their performance – and possibly even calls for some effort at

attempting to *build* the more general systems. What carries over from one system to another as a technical "saving" will rarely be obvious from a casual examination.

To take a specific example, consider the Waltz system, which was cited by both Dreyfus and Haugeland as a paradigm case of nonextensibility. Haugeland draws a parallel between a system like Waltz's, which succeeds only "by exploiting tricks that are utterly idiosyncratic to polyhedra," and a system that identifies fruit by analyzing their absorption spectra (we shall put aside the fact that it is almost certainly not possible to do this in general, since being visually similar to an apple – as opposed to applesauce or apple pie – is not the sort of property to which spectrograms respond). In neither case, Haugeland claims, should this count as a discovery about identification by *people*.

Two issues are relevant to this example. First, to show that fruit recognition by spectral analysis is not an adequate model of visual recognition one need not, as Haugeland claims, reintroduce the question of "how people do it." It is a straightforward question of the extendibility of the method at least to some independently plausible cognitive subdomain or cognitive faculty (e.g., recognition of outdoor scenes or recognition of faces or English sentences.)

Second, it is not obvious from a superficial description of the process how far a particular method will generalize. Part of the problem relates to our earlier remarks concerning the difficulty in specifying the equivalence of methods. Thus while it may be that a system using the particular set of labels developed by Waltz will only work for polyhedra, it is not obvious a priori that, for example, a superset of very similar labels could not be an essential component of a vision system, or, even more relevantly, that the method of constraint analysis or label propagation developed by Waltz might not be a very general method of wide applicability, one embodying important cognitive principles (Winston, 1977).

Note that even the crude but seminal system of Waltz gives an interesting account of certain human visual phenomena, such as the perception of certain ambiguous and anomalous figures. It also shows how certain dispersed sources of local evidence can be exploited to resolve potential global ambiguities, and suggests ways in which locally parallel processing might operate in the visual system; all in all, it is a nontrivial contribution to understanding aspects of perception from a system built from very limited considerations and using what appear, on the surface, to be mere "gimmicks." The obvious moral is that whatever the ultimate verdict on the generality of the principles embodied in a system such as Waltz's, one is not entitled to dismiss the discovery that a certain method works well in a restricted domain merely on the basis of superficial observations as, for example, that it seems to depend on "peculiar quirks." The issue is a scientific one not to be resolved by an a priori approach.

But there is an even more general theme that recurs in the commentaries of Dreyfus, Cushing & Hornstein, Goodluck, and Haugeland. It is that A.I. is a poor approach to understanding cognition because it has failed heretofore to discover any general principles. For example, Goodluck claims that in A.I. "little has been discovered about natural language." This is a frequently made and much debated allegation. Its chief value lies in pointing out that the measure of a discovery depends a lot on what one considers a "natural language phenomenon" and on what one is willing to count as a discovery or a relevant principle.

For example, Goodluck cites the distinction between processing syntactic ambiguity and processing semantic ambiguity, and suggests that a system reflecting this difference by virtue of some properties of the computation itself would be a useful model. Unlike some other arguments directed against computer models, this proposal at least has the virtue of being quite specific in proposing a phenomenon of processing that would count as relevant. The trouble is that properties such as the one suggested abound in most computational models: In fact, exact analogues of the phenomenon in question can be seen in many computational systems. Without taking a stand on particular proposals I would simply note that countless examples of just this sort have been frequently cited in support of, say, the ATN parsers, without arousing much enthusiasm among A.I.'s detractors. These people always have the option of considering the cited properties to be adventitious to these systems, and hence not affording a principled account of the phenomena. And they may be right. But it does highlight the problem discussed earlier, of separating principles from the details of particular implementations.

Haugeland and Dreyfus go even further in their indictment of A.I. Haugeland claims that A.I. has not even "come up with any interesting generalizations beyond its own premises." It is not clear what kind of response these authors would accept as answering their challenge. They might find it instructive to try to persuade a nonbeliever of the intellectual achievements ("beyond its own premises") of, say, philosophy, linguistics, economics, or any other controversial discipline. The outcome is invariably the same: What counts as an achievement in one field can be dismissed with no difficulty by an unsympathetic outsider as merely meeting incestuous internal criteria. Thus, the philosophers' distinctions are viewed as word games by many scientists, who deny that any cumulative progress has been made, say, by philosophy of mind; the discovery of syntactic rules and even putative syntactic universals are very often considered irrelevant by psychologists and some A.I. researchers on the grounds that in using language one does not appeal directly to such principles but makes inferences based on semantics and pragmatics, which readily override syntactic regularities, and so on. Progress is usually judged by the degree to which findings are relevant to the problems as formulated within a particular approach.

Thus it is not surprising that many of the fundamental discoveries of A.I., such as those roughly categorized under the heading of principles of representation (e.g., data structures, procedural representations) and principles of control (e.g., pattern-evoked procedures, forward versus backward chaining, context mechanisms) are dismissed by Haugeland and Dreyfus as merely symptoms of A.I.'s preoccupation with programs. And of course this is true in the same sense that the discovery of syntactic regularities follows from a preoccupation with formalism and with conditions of well-formedness. What this way of putting it misses is that the source of motivation for the discoveries in no way invalidates the insights so gained. It is reminiscent of the story of the sailor in the crow's nest of a ship who cried out that there were icebergs ahead. He was ignored by those on deck on the grounds that he was moved to say *that* because he was up *there* rather than down on the deck with the others. And the deckhands were of course quite right – had he been on the deck he would have said something quite different!

The problem of justifying a pursuit is made easier if there are dramatic practical achievements (such as airplanes and atomic energy) to point to. That is in part why some A.I. researchers have focused their energies on "expert systems" (e.g., MYCIN, DENDRAL) whose performance is indeed impressive. But just as important is the discovery that *search* has certain fundamental properties and tradeoffs, that associating specific information with processes that transform representations rather than with the representations themselves greatly influences access efficiency (a fundamental point missing in neobehaviorist mediational theories), and that the systematic withholding and releasing of processes (such as compilation, interpretation, drawing inferences, initiating searches) is one of the cornerstones of intelligent problem solving as well as of language comprehension. There are also a number of very general principles which are only gradually being articulated within the A.I. community. Many of these have been part of the implicit art of the A.I. field but have not been distilled and articulated as general principles. Newell & Simon (1976 *op. cit.*) mention some of these, which they refer to as "laws of qualitative structure," and argue that

such general principles have historically had profound effects on the development of science.

After observing at rather close range the squabble among transformational linguists, psychologists, philosophers, and A.I. researchers – all of whom share immeasureably more of a common world view than they care to admit – I can only conclude that to ask a discipline to defend its existence by enumerating its successes in a manner acceptable to people outside that discipline is supreme folly. No a priori argument for anything as general as an entire discipline can have much validity. Even the often cited case of alchemy does not seem relevant to me since whether alchemy failed as a science (as opposed to one of its rather specific subgoals having to be revised) is not clear, and even less clear is whether the arguments advanced against it at the time were sound (even assuming that the conclusions were valid).

9. Transduction and nonhuman intelligence

(a) Transduction. The form in which one assumes that a system makes contact with an environment is critical in determining the shape of any resulting theory of perception and cognition. Thus, if one assumes that there are primitive transducers for texture or for parallax motion or for objects located "in the coordinates of the environment" (as some would put it), one gets Gibsonian perception theory; if one assumes that there are primitive transducers for operands and reinforcers, one gets behaviorism; and if one assumes that there are primitive transducers for lines and vertices, one can get successful block-world perception theories. On the other hand, simply accepting the point-intensity transduction of video cameras as primitive may place the unnatural burden of perception on low-level problems and on the wrong starting elements. The problem is exactly the same as that discussed earlier in connection with choosing the right primitive cognitive operations. Thus Atherton is correct in pointing out that constraints imposed by the environment are relativized to the transduction primitives through which the environment's effects are presumed to be felt. Artificial intelligence criteria will prevent the excesses of Gibsonianism and behaviorism, but could lead to unreasonable processing if inappropriate transduction primitives are assumed (as in Haugeland's example of absorption spectra). My contention is that in order to carve up the perceptual world in the right way (i.e., in a way that captures its cognitively relevant features such that sensory patterns are partitioned into phonemes, objects, movements, colors, and so on), we will be driven to design the correct transducer functions (an example of which may be Marr's "primal sketch"). Even though we may be forced to accept point-intensity transducing hardware and Fourier analyzers as input devices, we should still treat the low-level *functions* as the virtual transducers. The actual front-end hardware should bear the same relation to perception as machine instructions bear to cognitive processes – that is, they should be thought of as simulating the underlying cognitive architecture. Thus, in reply to Schank's point concerning the special status of A.I. systems that use certain specialized front-end hardware, I must reiterate the earlier point that the electronics are irrelevant to the modeling – only the cognitive, virtual machine architecture and algorithms written for it are significant.

(b) Primitive creatures, evolution, and learning. There was much interest in the early days of A.I. in the prospect of designing systems that in some sense were self-structuring (either through explicit training, as in the Perceptron work, or indirectly through some sort of simulated evolution). The basic deficiencies of this approach were recognized rather early in the game when it was realized that an environment cannot induce the rich required structure unless at least two conditions could be fulfilled. The first is that the environment must communicate back to the system more than merely its success or failure in meeting overall criteria of performance. The second is that the system has to possess an initial structure that gives it the capacity to make use of that information appropriately (as McCarthy has put it, in order for a system to learn it must be capable of being told, i.e., it must already possess an epistemologically adequate representation scheme). This in part answers Andreae's complaint that learning is being neglected (Newell & Simon, 1972 *op. cit.*, discuss other strategic reasons for this neglect). Attractive as is the prospect of cracking the "complexity barrier" (as Winograd has termed it) by letting the environment control the growth of complexity, it seems that this will not happen until much more is known about the representation of knowledge.

Another approach to dealing with complexity is suggested by Dennett. He offers the proposal that instead of treating carefully delimited microworlds (such as the blocks world), or treating complex open-ended domains in a superficial manner, we might set ourselves the task of modeling the entire cognitive system and the entire environment of less sophisticated creatures – and specifically artificially created ones. This is an extremely intruiging possibility which, if carried out successfully, would be in keeping with Domotor's idea that it might be possible to investigate intelligence in a less anthropocentric manner. I am aware of very few proposals along this line (apart perhaps from the class of "self-adapting systems" alluded to above). One intruiging early study was Toda's (1962) proposal to "begin with an environment, and attempt to design a subject with minimal qualities to function effectively in this environment." Although Toda's work was carried out in a game-theoretic framework, rather than in A.I., the basic idea may be well worth exploring again today. The danger, as in any attempt to cut complexity down to manageable proportions, is that one is constantly pursued by the "qualitative discontinuity principle," which I mentioned in my essay, and one thus always runs the danger of simplifying away the interesting problems, as has happened in empiricist psychology. But it may well be worth a try – there is no sure and easy way in this business.

10. Conclusion and neuroscientific coda

I am left with the strange feeling that only a few of the disagreements expressed by the commentators are substantive. In a number of commentaries (e.g., those of Ullman, Johnson-Laird, Hayes, Longuet-Higgins, Treisman, Newell, and others) the only debatable point is the impression some of these writers have that what they say might be in conflict with my views. In a number of other cases I do not see that much is left after the misinterpretations are cleared up. In fact, some commentators even discovered indirect ways of making my point. For example, Ortony says that my claim that there might be no substantive difference among systems for some domain if they were equated for power, generality, and level of analysis is equivalent to claiming that if we equate two things on the only way they differ then they will be indistinguishable. But of course that is simply to say that he agrees with my proposal that those are indeed the main dimensions of difference!

A few commentators (Sutherland, Arbib, Harmon, Leibovic, Pribram) were uneasy about the status I assigned to neurophysiological evidence in the cognitive science enterprise. Although I am skeptical as to whether such evidence can ever be relevant to specifying the *algorithms* that humans use, I can see that it might be useful in constraining the architecture of the virtual machine.[2] But the history of the subject does not encourage one to expect strong constraints from neuroscience, except perhaps at the level of transducers, or perhaps in suggesting a useful taxonomy of skills, methods, or even "mental faculties." Even the work of Marr, which pays close attention to neurophysiology, does not seem to me constrained in any material way by that line of evidence (and here my impressions are not in agreement with

Sutherland's). Of course I do not wish to prejudge the eventual possibility of more powerful constraints – especially if neurophysiological theory upgrades its theoretical approach to include consideration of more computationally relevant mechanisms. Thus I would endorse Pribram's advice that *some* cognitive scientists should "court one of the neurosciences."

NOTES

Asterisks indicate commentary is to appear in a forthcoming issue.

1. It might be noted that the charge of being *ad hoc* could be made equally well against proposals in any of the cognitive disciplines. Thus an *ad hoc* A.I. system is one that does not contain general mechanisms applying outside the arbitrary and narrow class of problems for which it was explicity designed. But this is also true of most micromodels that fill the psychological literature. To the extent that such models are developed in direct response to an empirical result arising from a very particular experimental paradigm, they have no more claim to offer theoretical insight regarding a cognitive process than do narrow A.I. systems such as, say, those designed for recognizing printed characters – notwithstanding the fact that the former derive from experiments. In both cases generality is crucial.

2. This does not, by the way, apply to data from various pathologies such as aphasias and agnosias. Such data represent rather direct evidence for the functional (not structural or anatomical) taxonomy of certain cognitive skills, and hence have implications at the level of algorithms. Furthermore, such evidence is quite independent of any neurophysiological interpretations that might be placed on its etiology.

REFERENCES

Bobrow, D., and Raphael, B. New programming languages for artificial intelligence research. *ACM Computing Surveys*, 6:155–174, 1974.

Chase, W. G., and Simon, H. A. Perception in chess. *Cognitive Psychology*, 4:55–81, 1973.

Hewett, C. Viewing control structures as patterns of passing messages. *Artificial Intelligence*, 8:323–64, 1977.

Moore, J., and Newell, A. How can Merlin understand? In: L. Gregg (ed.), *Knowledge and Cognition*. Hillsdale, N.J.: Lawrence Erlbaum, 1973.

Pylyshyn, Z. W. Competence and psychological reality. *American Psychologist*, 27: 546–52, 1972.

On the Explanatory Adequacy of Cognitive Process Models. Contribution to M.I.T. workshop on representation. Jan. 20–21, 1978 (Mimeo).

Quine, W. V. O. Two dogmas of empiricism. *Philosophical Review*. 1951.

Scott, D. S., and Strachey, C. *Toward a Mathematical Semantics for Computer Languages*, pp. 19–46 of Proceeding of the Symposium on Computers and Automata (ed. J. Fox), Polytechnic Institute of Brooklyn Press, New York, 1971.

Toda, M. The design of a fungus-eater: a model of human behavior in an unsophisticated environment. *Behavioral Science*, 7:164–83, 1962.

Winston, P. H. *Artificial Intelligence*. Reading, Mass.: Addison-Wesley, 1977.