AI Magazine Volume 4 Number 1 (1983) (© AAAI)

Psychological Studies and Artificial Intelligence

Martin Ringle

Department of Computer Science
Vassar College
Poughkeepsie, N Y 12601

Editors' Note: Should Artificial Intelligence strive to model and understand human cognitive and perceptual systems? Should it operate at a more abstract mathematical level of characterizing possible intelligent action, independent of human performance? Or, should it focus on building working programs that exhibit increasingly expert behavior, irrespective of theoretical or psychological concerns? These questions lie at the heart of most current debate on whether AI is a science, an art, or a new branch of engineering In fact, some researchers believe it is all three and consequently build systems that perform some interesting task, arguing for the "theoretical significance" and "psychological validity" of the approach. Although AI in general draws strongly from all three approaches, most individual research projects consider one to be central and thereby determine their objectives and research methodologies

One of the reasons for establishing this column is to discuss and clarify the various research paradigms that characterize AI The present contribution by Marty Ringle argues in favor of the cognitive modelling approach. In fact, it assumes the cognitive psychology paradigm as central and suggests that AI research would benefit from closer adherence to the data and methods of psychological research We welcome contributions in support of other research methodologies in AI, as well as discussions comparing or evaluating the various approaches to research in our field —*Jaime Carbonell and Derek Sleeman*

Abstract

This paper argues for the position that experimental human studies are relevant to most facets of AI research and that closer ties between AI and experimental psychology will enhance the development of booth the principles of artificial intelligence and their implementation in computers Raising psychological assumptions from the level of ad hoc intuitions to the level of systematic empirical observation, in the long run, will improve the quality of AI research and help to integrate it with related studies in other disciplines

Experimental Data Versus Intuition

TO WHAT EXTENT should research in artificial intelligence utilize results from experimental studies in psychology? Some AI workers (e.g , McDermott, 1980)deny the relevance of human studies altogether; others advocate the use of such studies and some (e.g., Schank and Riesbeck, 1981) insist that improvements in certain areas of AI, such as natural language processing, depend crucially on knowledge of human performance These differing attitudes reflect fundamental differences in the ways that AI researchers view their discipline.

A distinction is sometimes drawn between AI's engineering concerns and its theoretical concerns. As an engineering discipline, the goal of AI is to develop and implement

Research for this paper was conducted at the University of Chicago Center for Cognitive Science under a grant from the Alfred P Sloan Foundation I am grateful to Jaime Carbonell, Lance Rips, Janet Krueger, Nancy Stein and Thomas Simon foot their helpful comments on an earlier draft of the paper

machines which are capable of performing operations by *any* available means. Accuracy, efficiency, flexibility, and reliability are the principal criteria of success for such systems and information about human performance seems to be neither necessary nor desirable.

As a theoretical discipline, AI attempts to define the principles of intelligent behavior, with specific emphasis on the structural and processing constraints imposed by physical realization. From this perspective, human performance clearly seems to be relevant to AI research and one would expect that close cooperation between experimental psychologists and AI model builders would be the rule. In fact, it is the exception

A look at the history of AI shows that while program design and description has always relied on elements of human psychology, assumptions about cognitive processes have usually been drawn from the intuitions or introspective analyses of AI workers, rather than from empirical studies. Celebrated programs, such as Samuel's checker-player (Samuel, 1959) were devised, implemented, and tested without reference to any correlative studies of human performance. Perhaps part of the reason for AI's early disinterest in human studies was a lack of faith in the power of psychology to illuminate issues in the structure of intelligence. We find evidence of this in Gelernter's remarks concerning his geometry-theorem proving program (Gelernter, Hansen and Loveland, 1963):

> We shall not labor the question as to whether our machine is indeed behaving intelligently in performing a task for which humans are credited with intelligence. The psychologists offer us neither aid not comfort here; they have yet to satisfactorily characterize such behavior in humans, and have rarely considered the abstract concept of intelligence independent of its agents (p. 154)

Little has changed over the years. Although many AI workers have taken human studies into account, it is still quite common to overlook the value of empirical psychology and to rely instead on intuition or introspection. Ironically, even researchers who explicitly acknowledge the importance of human data omit psychological studies from their work. Winston (1979), for example, suggests that the first of five steps in building an AI learning program is "to observe or define some learning competence to be understood." (p. 351) Yet Winston's account of his program for replicating the use of simile in teacher-student interactions, makes no reference to any related studies in educational psychology (e g., Klausmeier's work on concept acquisition, Cf. Klausmeier, Ghatala Frayer, 1974). Instead, Winston relies on his own intuition to define the competence involved in such behavior. (Cf. Winston, 1979, p. 353) As ingenious as Winston's program is, it would be vastly more compelling if it rested on an empirical foundation derived from psychology.

This paper argues for the position that experimental studies are relevant to most facets of AI research and that the closer ties between AI and experimental psychology will enhance the development of both the principles of intelligent behavior and their implementation in computers. Raising psychological assumptions from the level of ad hoc intuitions to the level of systematic empirical observations, in the long run, will improve the quality of AI research and help to integrate it with related studies in other disciplines

Arguments for the Relevance of Psychological Studies

Precision of task definition and description. From nearly three decades, AI workers have been building (and reporting on) systems which employ processes which they call reasoning, understanding, problem-solving, decision-making, planning, concept-formation and so forth The meanings of these terms often differ from project to project, creating a patchwork of terminological customs. For example, the notions of "concept," as used in the description of semantic networks, has no fewer than five major interpretations. (Cf. Brachman, 1979)

Such informal use of psychological terms gives rise to three problems. First, it makes comparisons between different AI projects extremely difficult. (This is well-illustrated in the case of organization and inference procedures for semantic networks) Lacking explicitly agreed upon meanings for such terms, AI workers in the past have been free to describe their systems using any terms which seemed appropriate.

The second problem is the nature of task specification. AI workers are fond of using psychological terms to explicitly link their programs to aspects of human cognition and thereby to provide them with a pre-established theoretical framework. For example:

> We subscribe to the hypothesis that as people read and understand text, they construct a multi-level mental representation of its content, with the most concrete level at the bottom of the conceptual structure Thus, at the lowest level of this mental structure are the sentences and phrases of the text, while the representation becomes more concise and abstract at higher levels. This concept of *understanding* text is embodied in our Prolog text grammar (Silva and Dwiggins, 1980, p. 20; author's italics)

Although Silva and Dwiggins are not central figures in AI, the error they make in attempting to provide a sanction for the use of the term "understanding" is a common one in AI research. The problem with their approach is that although it may be true that human beings generate multi-level representations during text comprehension, such understanding involves far more than this. Merely saying that lexical items are represented at a "low" level of a conceptual structure and that "higher" levels contain representations which are "more concise and abstract" casts little light on the nature of understanding. Yet the reference to a psychological process, and the mention that the task which the program performs, produces the illusion that the task which the program performs is well-specified. McDermott pinpoints the methodological weakness of this tactic:

> If a researcher tries to write an understanding program, it isn't because he has thought of a better way of implementing this well unerstood task, but because he hopes he can come closer to writing the first implementation. If he call the main loop of his program *understanding,* he is (until proven innocent) merely begging the question He may mislead a lot of people, most prominently himself, and enrage a lot of others. (McDermott, 1976, p. 4: author's italics)

Defining a program (or task environment) in this way is misleading because it pretends to appeal to a psychological model when in fact it does not. This not only obscures the inherent lack of precision in the task specification, butbut it generates the third problem: implicit reference to associated cognitive properties. When a psychological term such as "belief" or "understanding" is casually used to define or describe an AI program, it is easy to succumb to the temptation to ascribe a cluster of related cognitive properties to it. This is almost unavoidable, since the pre-theoretic or lay of psychological terms ordinarily occurs only with reference to human (or animal) cognition. It is difficult to make sense of many of these terms without presupposing an integrated background of cognitive phenomena but, when we use them in an AI context, such a background is absent. One consequence is that we expect more from AI programs than they are able to give. More importantly, however, the blurred use of (lay) psychological terms tends to hide the differences between machine behavior and human behavior and encourages AI workers to make questionable claims concerning the theoretical significance of their programs. Thus Silva and Dwiggins, for example, describe the role of the knowledge representation structures in the Prolog system by saying that they are "used by [the] text grammar to derive content representations approximating a human's understanding of a text." (1980, p. 20)

As numerous argument have demonstrated, however, no current natural language program comes close to "approximating a human's understanding of text." (See Dreyfus, 1978; Odell, 1981; Ringle, in preparation, for discussions of this issue.) Claims of this sort can be taken seriously only if we are willing to accept an unanalyzed and rather weak interpretation of "understanding." The attempt to draw specific theoretical connections between program performance and psychological concepts, without providing a rigorous account of the relevant similarities and differences, adds nothing to significance of the program and instead makes the work appear to be more naive than it actually is.

These problems can be alleviated to some extent by making use of the language of cognitive psychology. The are several reasons why such terminology is helpful:

1. Unlike psychological words which are used in ordinary discourse, the vocabulary of cognitive psychology is a technical one which has evolved in the context of experimentation and analysis While some terms possess multiple meanings, their degree of vagueness is far less than that of ordinary language terms The use of such terms AI workers would permit easier comparisons and greater continuity among related programs. At the very least, it would discourage AI workers from using the psychological terms of las discourse, thus reducing the degree of vagueness of program specifications and descriptions.
2. Terms in cognitive psychology are often given operational definitions which minimize theoretical presuppositions. For example. when dealing with linguistic tasks, psychologists are more likely to describe specific behaviors such as relativization, wh-movement, passivization, pronominalization, et., rather than broad faculties such as "language understanding" If intentional properties are implied in such descriptions—and very likely they are—there is at least a well-defined performance which can be used as a focal point for analysis For example, if a computer can successfully associate noun phrases and sentences with related pronominals then its behavior can be accurately described as "anaphora resolution," regardless of whether or not it actually "understands" any of what it reads. The use of such a term limits the nature of the claim being made about the significance of the computer's behavior and encourages us to focus our attention on the actual structures and processes rather than on constellations of (related) human activities.
3. By using the more theoretically neutral language of cognitive psychology, and thereby emphasizing detailed information-processing descriptions of program behavior, we reduce the temptation ot extrapolate from program behavior to claims about implied cognitive models. In addition, the differences between the properties of the AI program and the related human activity are more visible from this perspective and thus more susceptible to further analysis (There are, however, problems with this suggestion which we discuss below.)

The Use of Human Data as a Methodological Heuristic

The distinction between AI an an engineering discipline and AI as a theoretical inquiry is usually drawn by reference to the methods used, rather than to the tasks considered. For example, one could take an engineering approach to visual pattern recognition, game-playing, or natural language understanding, just as one could take a theoretical approach to tree-searching, automatic-programming, or robotics. The difference between the two approaches to AI is a matter of the method and intent of the researcher, not the task specifications or programming techniques.

There are several objections which AI workers of the engineering persuasion might raise against the use of human studies:

1. It is often easier and more efficient to devise an ad hoc strategy for solving a particular problem than it is to obtain appropriate and reliable results from human subjects.

2. It is sometimes difficult or impossible to devise experiments which will produce human behavior for an isolated task; the data which can be obtained, therefore, may cover more aspects of a phenomenon than the programmer wishes to deal with and thus are of limited value
3. There is no guarantee that the process involved in cognition which are supported by protoplasmic hardware and an as-yet-undetermined architecture will be at ll relevant to the algorithmic process of electronic digital computers operating under the relatively primitive architectures currently is use.
4. In some cases (e g., theorem-proving) we already know that human performance is inferior to machine performance and that existing algorithms cannot be improved by analyzing data from human studies.

These are each strong objections. Methodologically it *is* often easier to devise an ad hoc strategy for a particular problem than it is to obtain relevant human data. But the efficiency of such a solution may be superior only in the short run. To solve each successive problem of a similar nature, new ad hoc techniques must be developed and thus the on-going enterprise begins to lose its efficiency. (Early attempts to build natural language front-end for database systems provide numerous examples of this. More recently, an emphasis on generalizability (e.g., Hendrix and Lewis, 1981; Carbonell and Hayes, 1981) yielded highly portable front-ends.) One can expect to find in the analysis of human cognition powerful mechanism which are flexible enough to accommodate a wide variety of different tasks. From the little we know of human cognitive processing, it is clear that the mind operates in an extremely efficient way, taking advantage of shortcuts, such as inductive generalization, analogic inference, error approximation, etc., wherever possible. Analyzing and emulating human methods of information processing may thus be viewed as an investment in long-term efficiency.

The classic example of this is the development of the General Problem Solver (GPS) by Newell and Simon (1963). By examining human techniques of problem-solving, Newell and Simon identified powerful heuristic methods of means-ends analysis, recursive sub-goal generation, and difference-reduction, which provided the basis for much of the work in planning and robotics of the past two decades. The significance of GPS—and one of the reasons it received so much attention in AI—is that it demonstrated how input from human psychology could provide new directions for AI research

The second objection, that it is sometimes difficult to isolate specific components of human cognition, is also valid. Here too, however, there is an interesting tradeoff between short-term economy and long-term success Consider, for example, the development of research in speech recognition. Early attempts in this area concentrated on bottom-u[acoustic/phonetic template matching for isolated words. An assumption of this research was that the transition from isolated word recognition to continuous speech recognition could be achieved by adding level of complexity to exist ing systems. Human signal processing seemed to be abl to contribute little to this area of investigation because o the difficulty of segregating acoustic/phonetic processes fron other (cognitive) processes which humans employ in wor((and speech) recognition.

It was precisely the multiplicity of human processes however, which prompted researcher to design speech understanding systems with parallel levels of analysis. The ARP/ speech understanding systems of the nineteen seventies such as HEARSAY-I (Reddy, Erman, Fennell Neely, 1976 and HWIM (Bruce, 1982) utilized conjoint constraints o phonetics, phonemics, morphemics, syntax, semantics, pragmatics, and discourse to achieve relatively high recognitioı accuracy for connected speech with a vocabulary size of : thousand words or more. In other areas as well, the inabilit to isolate particular cognitive process for psychological investigation may force researchers to adopt a more global viev of a phenomenon and thus produce more design strategies.

The third objection, that human data may not b relevant to AI program design because of essential difference in the underlying hardware and architecture, ahs been raise by critics of AI, such as Dreyfus (1978) and Searle (1980) There are two replies we can make to this objection:

1. It is true that there are profound differences in th hardware of the brain and the hardware of the digital computer; it is also true (or extremely likely) that there ar equally significant differences in their respective architectures. Yet, unless one is able to demonstrate the precis relationships which exist between hardware, architecture and intelligent behavior, there is no prima facie reason to assume that the particular hardware or architecture of humar beings is essential for intelligence. Merely noting that ther are differences between brains and computers is not enough the critic must show how the physical or logical feature of a system exercise necessary constraints on the functiona properties of that system and, moreover, he must also show that hose functional properties cannot be realized in a systen with alternative physical or logical features. Otherwise, th possibility of a functional replication remains, thus making the objection vacuous.

2. Suppose the critics turn out to be overwhelmingly correct about the importance of the differences in architectur and hardware between brains and computers. What then? Rather than being a reason for abandoning human studies it provides even stronger grounds for examining them. If th relationship between a particular physical or logical property and specific aspect of cognition were demonstrated, it migh well prompt AI researcher to redesign the tools of thei studies. (For example, the likelihood that the brain is a parallel, rather than a serial, processor has stimulated a grea deal of interest in parallel architectures (see Fahlman, 1979) and may eventually spawn an new breed of AI systems.

Artificial Intelligence, like any branch of science, in some sense is defined by the tools it uses. Historically, AI has focused its efforts on the instatiation of intelligence in serial digital computers. Yet there doesn't seem to be a special *theoretical* commitment to this type of hardware and, if data from psychological (or psychophysiology) were to occasion a revolution in hardware or architecture, AI would very likely make the transition without changing its goal or its self-image. The use of human studies, therefore, had advantages for AI regardless of whether or not the differences between brains and solid-state circuitry turn out to be crucial.

Human Performance as Validation Tool

The development of a program in artificial intelligence is always an evolutionary process which involves many iterations of the design-test-redesign cycle. Testing is frequently task oriented, with the performance of a program being evaluated against a well-defined set of parameters for speed, cost and reliability. But some tasks, particularly those involving elements of context identification, can be tested more perspicuously against human performance. In the area of problem-solving, for example, the behavior of the program—that is, the machine trace of the steps taken to reach a solution–can be compared to a record of the actions and reports of human subjects confronted with the same problem. This technique of protocol analysis was first used in the development of GPS (Ernst, Newell, 1969) but it continues to be a powerful method for evaluating AI programs. Luger, for example, describes the use of protocol analysis in the development of the MECHO system, a program designed to solve problems in mechanics (Luger, 1981). Differences between traces of MECHO's performance and subject's protocols are used to modify the program and to improve its ability to focus on important feature-relations within problems.

The use of experimental data to validate AI research goes far beyond the fine-tuning of particular programs, however. In areas such as natural language processing, where AI has begun to develop detailed models of memory organization, parsing, story understanding, and so on, data from psychology can be used to guide program development and to evaluate the plausibility of theoretical constructs. When such constructs are adequately specified at the functional level, they have predictive consequences which can be used in the design of cognitive experiments. A good example of cooperation between AI and cognitive psychology in this respect is the development of scripts.

The script concept in natural language processing was first introduced by Schank (1975) and later developed by Schank and Abelson (1977). In broad terms, a script is a stereotypical sequence of events, which includes a standard group of characters, props, entry and exit conditions, and a typical setting. It can be used both as a memory structure for storing event knowledge as well as a processing structure for story understanding. A presupposition of the original script concept is that scripts reside in long-term memory as fixed structures and that there is a separate script for each class of stereotypical events.

Scripts were implemented in a variety of programs, such as SAM, QUALM, FRUMP, and POLITICS, (Cf. Schank and Riesbeck, 1981) with impressive results. They quickly became a focal point for research in natural language processing, attracting a good deal of attention both inside and outside of AI. The concept of a script was so well defined that psychologists such as Smith, Adams and Schorr (1978), Owens, Bower and Black (1979), and others were able to devise experiments to test its psychological reality. In one experiment, conducted by Bower, Black and Turner (1979), subjects were presented with a series of stories which instantiated the same "underlying script" (For example, subjects would read stories about a visit to a doctor, a dentist, or a chiropractor, as instances of the "visit to a health professional" (script). Though similar in many respects, each story would contain some events not included in the others During a recall test, it was observed that subjects regularly reported events to be in one story when in fact they occurred in another. This sort of confusion, though quite common, could not be explained in terms of the original script concept.

Prompted by this empirical work, Schank reassessed the notion of scripts and formulated a new approach to long-term event memory. The similarity of features of different scripts, such as the script for DENTIST VISIT and the one for DOCTOR VISIT, suggested that there might be more basic memory structures common to both of them. Such structures could be accessed as needed, during story processing or event perception, to form a temporary, synthetic "superscript," corresponding in many ways to the original event script. Unlike original scripts, however, these superscripts would decay after processing, leaving a trace of the salient features of the event. Schank (1982) termed these basic memory structures Memory Organization Packets (MOPs). MOPs constitute a significant advance over scripts since they can be individually triggered by the content of a narrative, rather than by a set of uniform entry conditions As a result, stories which do not fall into the typical script-governed paradigm can now be processed and events from such stories can be perspicuously stored.

Schank's efforts to accommodate the script recall confusions reported by Bower, Black and Turner produced a new constellation of memory and processing models and sparked the development of new projects in AI (see, for example, Lebowitz, in press). This is precisely the sort of beneficial contribution that empirical model validation can make to AI.

Conclusion

The aim of this paper has been to argue for the relevance of psychology to AI. It should be noted, however, that psychology also stands to benefit from a closer association with AI. The development of simulations of cognitive processes provides the cognitive psychologist with new territory for empirical explorations, as well as powerful tool for framing new theories. In particular, AI is in a good position to shed new light on dynamic processing constraints and other implementation defined aspects of cognition. There are, however, limitations to the interactions between the disciplines and it is worth mentioning two of them.

The domain problem. While most projects in artificial intelligence focus on activities which have correlates in human cognitive behavior, some do not. For example, *automatic programming*, the process whereby a routine can extend or modify itself, has no clear counterpart in human cognition (except at a very abstract level). The design and optimization of programming languages for artificial intelligence is another area where psychological studies are not immediately relevant. The same hold true for some areas of automatic deduction, sensory processing, and robotics

It would be wrong to assume that psychological studies ought to be applied to every domain of AI research. Yet it is difficult to draw a precise line between domains which are appropriate and those which are not. Problem-solving, for example, could fall on either side of the line, depending upon the nature of the problem set being considered.

The grain problem. Even in those areas where human studies *are* applicable, there is still a difficulty with what Pylyshyn (1979) refers to as the "grain problem." When we attempt to devise or validate computer programs be referring to human performance, it is necessary to specify the precise level of behavior, structure, or internal process of the intended correlation. Pylyshyn illustrates the grain problem in his discussion of the comparison between the "mental arithmetic" of a person and the calculations of a computer:

> Should 'state knowledge' in the mental arithmetic example correspond to individual integers, as they would appear if the problem were done by writing down intermediate steps, or to finer states—perhaps even corresponding to the steps involved in accessing a stored addition tables? In the latter case, how do we obtain evidence of these states? . What are to count as primitive operations in determining complexity of computation in the model? What reasons do we have for expecting complexity scaled in this manner to correspond with latencies? Such an expectation only follows if one makes additional simplifying assumptions about what contributes to human response latencies (1979, pp. 51-52)

Pylyshyn reminds us that evidence from cognitive psychology is theory-laden and that even the most transparent behavioral results are still embedded in a highly interpreted framework. When we begin to peel away those aspects of human performance which are irrelevant to a correlated AI program, we may be left with elements which are still too complex to be unequivocably compared to machine processes (or structures). In the past, this complexity has often been ignored. As a consequence, AI programs whose behavior resembled human behavior only in the most superficial ways have been used as grounds for extravagant claims about artificial intelligence and cognitive simulation. The difficultly of correlating the grain of an AI program to the grain of a particular set of data in psychology is a clear invitation to sloppy interaction between the two disciplines.

Unfortunately, there is no simple or generalizable solution to the grain problem. If human studies are to be used in AI research, a special effort must be made in each case to insure that correlations are not merely gratuitous. Perhaps the best way to achieve this goal is to promote research projects which include both AI workers and psychologists. An example of this sort of teamwork is the recent analysis of dialogue by Robertson, Black and Johnson (1982) which "involves a coordinated synthesis of results from an AI program, and a naturalistic empirical study." This group developed a model for the determination of speaker intentions and topic in dialogue by testing features of an implemented program on human subjects. In the long-run, perhaps cooperative projects of this erhaps cooperative projects of this sort will become the norm in AI research for, as Schank clearly points out:

> It is absolutely crucial that AI researchers and psychologists, as well as cognitively oriented linguists, begin to work together on issues facing us. To pretend that we are interested in different things is folly--we are all working on the nature of the mind The fact that we bring different tools to bear on this subject is terrific. So much the better for getting potentially different results and thus learning from each other (1980, p. 282)

References

Bower, G , Black, J B , and Turner, T. J. 1979. Scripts in Text Comprehension and Memory, *Cognitive Psychology,* 1:177-220.

Brachman, R J 1979. What's in a Concept: Structural Foundations for Semantic Networks. In N. Findler (ed) *Associative Networks,* New York, Academic Press.

Bruce, B. C. 1982 Natural Communication Between Peron and Computer. In W Lehnert and M. Ringle (eds) *Strategies for Natural Language Processing,* Hillsdale, NJ, LEA, Inc , 55-87.

Carbonell, J and Hayes, P 1981. Dynamic Strategy Selection in Flexible Parsing, Proceedings of the Association for Computational Linguistics.

Dreyfus, H. L. 1978 *What Computers Can't Do,* New York, Harper and Row.

Ernst, G. W. and Newell, A. 1969. GPS: *A Case Study in Generality and Problem Solving,* New York, Academic Press.

Fahlman, S. E.k 1979. *NETL: A System for Representing and Using Real-World Knowledge,* Cambridge, Massachusetts, MIT Press

Gelernter, H , Hansen, J. R and Loveland, D. W 1963. Empirical Explorations of the Geometry-Proving Machine Reprinted in E Feigenbaum and J Feldman (eds) *Computers and Thought* New York, McGraw-Hill, 153-167.

Hendrix, G. and Lewis, W. 1981. Transportable Natural Language Interfaces to Databases, Proceedings of the Association for Computational Linguistics, 159-165

Klausmeier, H. J., Ghatala, E S. and Frayer, D A 1974. *Conceptual Learning and Development,* New York, Academic Press.

Lebowitz, M., Correcting Erroneous Generalizations. (In press.) *Cognition and Brain Theory*

Luger, G. F 1982 Mathematical Model Building in the Solution of Mechanics Problems: Human Protocols and the MECHO Trace, *Cognitive Science* 5:55-77

McDermott, D. 1976 Artificial Intelligence Meets Natural Stupidity, *SIGART,* 57:4–8

McDermott, D 1980. Methodological Polemic, presentation to the Artificial Intelligence Society of New England, Yale University.

Newell, A and Simon, H. 1963 GPS: A Program that Simulates Human Tthought. In E Feigenbaum and J. Feldman (eds), *Computers and Thought,* New York, McGraw-Hill, 279–296.

Odell, J. 1981 Are Natural Language Interfaces Possible? IBM Technical Report, Systems Research Institute, TR 73-024.

Owens, J., Bower, J. and Black, J., The 'Soap Opera Effect' in Story Recall, *Memory and Cognition,* 7:185–191.

Pylyshyn, Z. 1979. Complexity and the Study of Human Artificial Intelligence. In M Ringle (ed) *Philosophical Perspecctives in Artificial Intelligence,* New Jersey, Humanities Press, 23–56.

Reddy, R., Erman, L., Fennell, R. and Neely, R. 1976. The HEARSAY Speecch Understanding System: An Example of the REcognition Process, *IEEE Transactions on Computers* 25:427–231.

Ringle, M., The Case for Natural Language Processing: A Response to Philosophical Objections, in preparation.

Robertson, S. P., Black, J. B and Johnson, P N 1982. Intention and Topic in Conversation, *Cognition and Brain Theory,* 4:302–326.

Samuel, A. L 1959 Some Studies in Machine Learning Using the Game of Checkers, *IBM Journal of Research and Development* 3:211–229.

Schank, R. 1980. Language and Memory, *Cognitive Science* 4:243–284.

Schank, R. and the Yale A. I. Group. 1975. SAM-A Story Understander, Research Report No. 55, Yale Computer Science Dept., New Haven.

Schank, R. and Abelson, R. 1981. *Scripts, Plans, Goals and Understanding,* LEA, Inc.

Searle, J.R. 1980. Minds, Brains and Programs, *Behavioral and Brain Scciences* 3:417–457.

Silva, G., and Dwiggins, D. 1980. Toward a Prolog Grammar, *SIGART,* 73:20–25.

Smith, E , Adams, N. and Schorr, D., Fact Retrieval and the Paradox of Inference, *Cognitive Psychology,* 10:438–464

Winston, P.H. 1979. Learning by Creating and Justifying Transfer Frames. In P. H. Winston and R. H Brown (eds) *Artificial Intelligence: An MIT Perspective* Cambridge, Massachusetts, MIT Press, 347–376.