Information and Computer Science

Toward a Taxonomy of Methodological
Perspectives in Artificial Intelligence Research

Rogers P. Hall
Dennis F. Kibler

Technical Report 208

September 1983

TECHNICAL REPORT

UNIVERSITY OF CALIFORNIA
IRVINE

Toward a Taxonomy of Methodological Perspectives in Artificial Intelligence Research

Rogers P. Hall
Dennis F. Kibler

Technical Report 208

September 1983

Department of Information and Computer Science
University of California, Irvine

Toward a Taxonomy of Methodological Perspectives in Artificial Intelligence Research

Rogers P. Hall
Dennis F. Kibler

Department of Information and Computer Science
University of California, Irvine

Abstract

This paper is an attempt to explain the apparent confusion of efforts in the field of artificial intelligence (AI) research in terms of differences between underlying methodological perspectives held by practicing researchers. A review of such perspectives discussed in the existing literature will be presented, followed by consideration of what a relatively specific and usable taxonomy of differing research perspectives in AI might include. An argument will be developed that researchers should make their methodological orientations explicit when communicating research results, both as an aid to comprehensibility for other practicing researchers and as a step towards providing a coherent intellectual structure which can be more easily assimilated by newcomers to the field.

Introduction

Over a quarter century since its inception, the field of artificial intelligence (AI) has yet to propose a commonly accepted statement of purpose or description of conventional research practices. Studies are reported in a wide range of publications, some particularily oriented towards the field (eg. *Artificial Intelligence*) but others directed towards different research areas (eg. *Behavioral and Brain Sciences*), resulting in a profusion of literature which is difficult to encompass for students and practitioners alike. If the study of AI is to be considered (and conducted as) a scientific endeavor rather than an amorphous enterprise whose subject matter is constantly shifting or even disappearing as results are incorporated into other fields, one might profitably ask if distinct methodological perspectives can be identified which might organize some of the current confusion of efforts. Perhaps, as others have pointed out, "there are undoubtedly some views of AI that are more fruitful than others ... We ought to be guided by the most productive paradigms" (Nilsson, 1982, p. 2).

This paper will present a variety of perspectives for viewing AI research which have been previously reported in the literature and attempt to condense those perspectives into a useful epistemological framework. No attempt will be made to evaluate the relative efficacy of differing perspectives, however, we will argue that researchers in AI should make their methodological perspectives explicit when publishing research results as an aid to comprehensibility both for practitioners holding alternative perspectives and for newcomers to the field.

Are research perspectives important in AI?

In a relatively new field of intellectual exploration, adherence to (or even identification of) methodological perspectives might be considered somewhat counterproductive. The flexibility with which investigators could approach a tremendous variety of potential research questions might be seriously constrained if the means of investigation were rigorously defined. There are indications, however, that the field of AI is suffering from a lack of direction which a clear explication of such methodological orientations might provide.

Although this paper will not address the developmental maturity of AI as a scientific endeavor, some description of what organizing research perspectives provide for scientific activity seems in order. "Perspective" in this paper refers to a disciplinary matrix or context within which researchers practice their trade. This concept of a matrix is similar to the most general notion of a "paradigm" as advanced by Kuhn (1970) and includes sets of shared values within a common conceptual vocabulary (Weimer, 1979). It is assumed that multiple, competing paradigmatic orientations can exist simultaneously in a single field (Masterman, 1970). Hence proponents of differing perspectives often "talk past each other" when discussing aspects of their work in terms of those different perspectives.

First, the perspective of a particular community of researchers can be seen to guide their selection of appropriate research problems or phenomena to be examined in much the same way that a map defines boundaries in a geographic territory being explored. Having selected a problem for study, the shared perspective specifies how acceptable research should be conducted and indirectly provides a medium of communication (ie. journals and conferences) through which research findings can be examined by the community as a whole. This critical evaluation of findings proceeds in accordance with a shared set of criteria for gauging the quality of research efforts. Finally, the research perspective provides a framework within which new participants in the research community can assimilate the skills necessary for conducting and evaluating research.

Many of the above-mentioned functions of an organizing research perspective might seem lacking in the field of AI. There is considerable disagreement over what constitutes important research problems as, for example, evident in the plethora of divergent views collected in the SIGART 1980 survey on knowledge representation. Results of the survey revealed no clear consensus on what "knowledge" was to be represented or what "representation" entailed (Brachman, 1980). This sort of confusion extends into the selection of appropriate research methodologies, leading to arguments like the persistent methodological squabble between "scruffy" and "neat" views of AI research (Bundy, 1982). Perhaps as a result of the above problems, published studies turn up in a dizzying variety of journals, conference proceedings and books which are hard to follow even for the experienced AI enthusiast, much less neophytes. Assuming that readers can find published studies which relate to their interests, there seems to be no consensus as to what sorts of qualities constitute "good work." In the opinion of one leading AI researcher, this confusion over appropriate problems and methods has reached a point where most recent work submitted to conferences has been rejected as "junk" (Schank, 1983). Finally, the effects of such confusion must have consequences for the training of new members of the AI research community, forcing them into what amounts to isolated apprenticeships with curricula which have been described as "scandalous" (Nilsson, 1980a) in their lack of formal methodological training.

A description of AI research as being in such a serious state of disarray might suggest that organizing research perspectives currently do not exist in the field. It will be the contention of this paper that several such perspectives do exist, but they are seldom explicitly stated and in some respects represent opposing views of AI research. As a result, the process of critical evaluation with respect to problem choice, method, and quality of results suffers, and the study of AI is difficult to present as a coherent academic discipline.

Perspectives in print

In this section, a chronological series of research perspectives previously appearing in the AI literature will be reviewed. No claim will be made for an exhaustive enumeration of published methodological speculations, but major voices will be presented both in terms of characteristic approach to AI research issues and exemplary studies.

One of the first detailed considerations of research strategies in AI comes from Allen Newell (1973). Research orientations in AI are broken into three classes and exemplary studies are presented for each. First, Newell describes a class in which the exploration of intelligent functions provides the major research focus. In this approach, a problem task is chosen which is assumed to require intelligent (typically human) behavior, and a computational mechanism is proposed which is sufficient to support the accomplishment of this task. As an exemplar, Newell cites Green's

(1969) work in automatic theorem proving. The second class characterizes AI as a science of weak methods where "weak" is intended to designate general usefulness of a particular technique across a variety of problem domains despite low information content with respect to any particular domain. The methodological approach here is to describe a general technique or method, demonstrating its effectiveness across problem domains in the hopes of eventually establishing a collection of general methods which are useful in the construction of "intelligent" systems in much the same way that numerical methods are applicable across a wide range of problems. An exemplar of this strategy would be Newell's (1969) examination of ill-structured problems. Newell's final class is a view of AI as theoretical psychology in the sense of viewing human cognition as the performance of an information processing system. In this class, a computer model of some cognitive process is proposed and then validated by comparison with features of the human behavior being modelled. An exemplar of this third class would be Newell and Simon's (1972) analysis of cryptarithmetic problem-solving in which program control structure is compared with human strategies evident in problem-solving protocols.

An incompletely specified but highly visible account of AI research practices can be found in Weizenbaum (1976). According to Weizenbaum, AI researchers proceed in one of three modes. First, in a performance mode, researchers are purely concerned with building practical software systems which satisfy a need for some artifact capable of impressive levels of performance. As an example, Weizenbaum makes a general reference to robotics research but cites no particular study. In contrast, AI researchers working in theory mode strive towards uncovering general principles of intelligent behavior without explicit regard to implementational issues much as turn of the century aerodynamicists studied principles of flight. No exemplars are given for this view of AI. Weizenbaum's third class, simulation mode, involves the construction of computer models of human cognition which can be compared with actual human behavior as a means of validation. As an exemplar, Weizenbaum suggests Newell and Simon's (1963) work on GPS.

Feigenbaum (1977) argues for the continuing applicability of a view of AI research proposed much earlier (Feigenbaum and Feldman, 1963) which consists of two classes. In the first, researchers attempt to build useful intelligent systems and to develop a methodology which supports such construction. This view is particularily appealing to Feigenbaum, and his description of it is replete with terminology suggestive of an engineering discipline (eg. "workbench," "knowledge engineers," and "toolkit"). Artifacts as "intelligent agents" are characterized by their use of generate-and-test heuristic search guided by a considerable amount of domain knowledge. Numerous exemplars for this view are given with the heuristic DENDRAL project (Feigenbaum et.al., 1971) being distinguished both by its longevity as a research program and popularity among researchers in chemistry. Feigenbaum claims the second class of AI research closely follows Newell's (1973) view of AI as theoretical psychology. No exemplars (or discussion) are given for this second class.

Lenat (1978) proclaims a single paradigm for AI research in which human behavior is viewed as the output of a symbol processing system. Having selected some human cognitive activity, the AI researcher proposes a theory of information-processing to support that activity, operationalizing the theory in a computer program. The behavior of the running program is then examined to determine the locus of "intelligent" behavior in the hopes of uncovering a unified theory of intelligence. As an exemplar Lenat discusses his own work (1977) on automatic theory formation in mathematics, stressing a view of intelligence as heuristic search.

Hayes (1978) contrasts AI research methodology with that of general systems theory, arguing that AI defers generality in favor of working programs. According to Hayes, _applied AI_ focuses on creating practically useful artifacts in highly circumscribed task domains. As an exemplar, Hayes suggests the work of Waltz (1975) on constraint propagation in scene analysis. Hayes describes a second form, _scientific AI_, which concentrates on the construction of working programs as experimental evidence for the efficacy of theoretical explanations for intelligent behavior. Schank's (1977) claims for conceptual dependency theory in natural language understanding are given as an exemplar.

As a final set of perspectives on AI research, Ringle (1979) proposes a taxonomy consisting of four classes. The first class, _AI technology_, is quite similar to Newell's exploration, Weizenbaum's performance and Hayes's applied perspectives. The approach taken is to construct reliable, cost-effective artifacts which demonstrate intelligent functioning without regard for human behavior or processes. As an example, Ringle cites the work of Buchanan (1969) on the analysis of mass spectrogram data in the heuristic DENDRAL project. The second perspective, AI simulation, is concerned with overt human behavior but breaks into two approaches which differ in the extent to which internal human cognitive processes are considered. In the first approach (which Ringle terms "demonstrative simulation"), computer programs are constructed which produce overt human-like behavior without regard for internal cognitive processing. As exemplars, Ringle suggests many of the early game playing systems which sought a human level of performance without concern for emulating human processing. The second approach is dubbed "investigative simulation" and entails computer demonstration of human-like overt behavior followed by hypotheses about similarities between machine processing and human cognitive activities. Newell and Simon's (1963) use of protocol analyses in validating the computational mechanisms of the GPS system (specifically means-ends analysis) is given as an exemplar. A third perspective in Ringle's taxonomy is termed AI modelling and involves the construction of computer programs which are intended specifically as models of internal human cognitive representation and processing. Ringle cites the work of Hunt (1973) on human memory as an exemplar. This modelling perspective is to be considered distinct from either form of simulation in the sense that research moves from theory to object rather than constructing working programs and then advancing hypotheses of similarity between program and cognitive structures. As a final perspective, Ringle describes _AI theory_ in which general principles of intelligence are advanced without regard for particular implementations or human cognition. Ringle's claim is that this perspective amounts to a form of applied epistemology which, "when mature, will _subsume_ the theory of human intelligence" (p. 12). As an exemplar, Minsky's (1975) theory of frames is given.

Perspectives in perspective

The above review of perspectives on AI research may not be bibliographically complete. Particularily notable omissions arise when the field of AI is viewed geographically (ie. MIT, CMU, and Stanford) or in terms of well-known participants. However, we will argue that major conceptual viewpoints have been addressed in this review. Obviously, there are many similarities between the views presented thus far. In the sections which follow, an attempt will be made to extract the common features of these various perspectives in the hopes that a practical framework for describing research reports in AI may be constructed.

If the reader at this point feels somewhat ambivalent towards the research perspectives discussed above, this paper will be in some measure a success. There is considerable overlap between perspectives across each framework, and none of the perspectives clearly specify a "modus operandi" for AI researchers which one might expect from an epistemological description of a disciplinary matrix within which research occurs. Specifically, no guidelines are apparent for problem choice, practical methodology, or the critical evaluation of research reports.

The reasons for the sparseness of the foregoing descriptions are diverse and merit some consideration at this point. First, there is no consensus from within the field or among observers of the field as to what the term "artificial intelligence" means. While many might agree that artificial (particularily man-made) systems are worthy of study (Simon, 1969), there is little agreement on the extent to which intelligent systems should be "artificial" in the sense of the word that the artifact differs significantly from the original, natural object. More importantly, however, the term "intelligence" has no consensually validated meaning. Before attributing this confusion to AI specifically, one should note that the meaning of "intelligence" was a hotly contested issue many years before the emergence of AI without recognizable resolution unless one were willing to accept illusive operational definitions like intelligence is what intelligence tests measure. Hence, one should not be lulled into a sense of confidence when researchers speak of choosing a problem task which is assumed to require "intelligent" (perhaps human-like) behavior. The space of potential research problems in this process of choice is hardly well-defined. In fact the definition of intelligent (whether human or otherwise) behavior appears to be in the head of the researcher more than in any attributes of the behavior itself. For example, in Ringle's taxonomy it is unclear whether research problems (in the form of some task domain) which interest AI "technologists" would interest participants in any of the other three perspectives. Perhaps as Ringle suggests, the method and intent of the researcher overshadow choice of a particular task domain.

<u>An epistemological framework for AI research</u>

What follows will be an attempt to delineate a set of AI research perspectives with respect to problem choice, methodological approach and criteria for evaluation. No claim will be made that the resultant taxonomy is "correct," but empirical evidence for the "fit" of perspectives comprising this taxonomy will be presented by a detailed examination of a selected group of exemplary studies.

As noted earlier, most of the taxonomies reviewed in this paper have much in common. In particular, there seem to be two primary classes of research perspectives resting on either side of a fissure brought about by the level of concern for explaining natural (ie. human or animal) intelligence. Within these broad classes, further subdivisions seem reasonable on the basis of relatively coarse methodological and intentional (on the part of the researcher) grounds. For researchers not concerned with exclusively natural modes of intelligent functioning, an intentional distinction is possible between those interested solely in program performance and those interested in uncovering more general principles of intelligence. The MACSYMA system (1975) for symbolic manipulation of algebraic formulas would be an example of the <u>performance</u> perspective. For researchers interested in more general achievements, a methodological division similar to Ringle's distinction between simulation and modelling appears useful. One approach to such general principles is a bottom-up construction of human problem solving can be considered exemplary of the former, empirical approach, while Schank's (1977) claims for episodic memory organization might serve as an exemplar for the latter, speculative approach.

To foster an impression that these five perspectives describe disjoint groups of practicing researchers would be misleading. In fact many researchers in AI periodically switch back and forth between perspectives from project to project. For example, a researcher doing formal work on knowledge representation techniques might collaborate on the construction of some practical artifact, with reports of both activities reaching the literature. Furthermore, many participants seem willing to engage methodological techniques of differing perspectives even while reporting work done quite strongly within the confines of another perspective. For example, Newell (1969), while presenting what amounts to a formal description of general principles in AI problem-solving systems, utilizes the results of human protocol analyses to test the viability of hypotheses concerning generality and the existence of ill-structured problems. Since there seems to be a tendency on the part of many AI researchers to shift quite freely between perspectives and even to claim to be meeting the concerns of multiple perspectives simultaneously, the five perspectives outlined in this paper might better be considered as ideal types from which variance is to be expected.

It will be the contention of this paper, however, that published research reports can generally (although perhaps with considerable caution) be identified as subscribing to one of the five differing perspectives. To operationalize this claim, the methodological and intentional divisions depicted earlier must be sharpened somewhat by giving fairly detailed descriptions of the three components which a useful research perspective should provide: guidance in problem choice, a characteristic methodology, and a set of criteria by which "good work" can be identified.

Perspectives in particular

In this section, the three components mentioned above will be described within each of the five proposed research perspectives, and an effort will be made to "fit" exemplary studies within each of the perspectives. For each perspective, exemplars are examined first, followed by a more general summary of perspective components.

Performance AI. As an exemplar of AI research oriented entirely towards impressive levels of performance, the widely used MACSYMA system for on-line algebraic manipulation will be discussed (Moses, 1971; MACSYMA, 1975; Barr and Feigenbaum, 1982). Unlike much previous work on symbolic integration which was concerned with general AI techniques, the MACSYMA project displays a clear goal of integration performance at or beyond levels of human performance. Extensive domain specific expertise (often in the form of highly specific mathematical algorithms) is used to generate solutions without regard for human approaches to similar problems or the similarity of computational mechanisms to established AI techniques.

As a problem choice, symbolic integration represents a constrained, non-trivial task which is often difficult for humans. In addition, substantial levels of performance in this problem domain promise to be of considerable utility to individuals who regularily face difficult integration problems in their work (eg. researchers in plasma physics).

The methodology of the MACSYMA project is to develop and use whatever integration techniques seem promising for the solution of particular classes of symbolic problems. Although representational and processing issues of general interest to AI do emerge (eg. inheritance hierarchies to guide inferences over symbol types), there is little *a priori* interest in such issues. Rather, the emphasis is on what techniques might be applied so as to minimize the time/space complexity of generating solutions within the problem domain or to make the system more accessible to users.

As a criteria for success, the utility of the MACSYMA system for on-line users looms large. In addition, comparisons are drawn between successively more sophisticated system capabilities in terms of the classes of problems which can be solved and the time/space complexity characteristics of solutions to those problems. In the case of MACSYMA, program performance actually exceeds that of all but a few human experts in the problem domain.

In abstraction, the performance AI perspective selects problems oftentimes undertaken with some difficulty by naturally intelligent agents. Hence, movement towards solution of these problems would typically be of some practical significance. Methods consist of developing and applying processing and representational techniques which approach solutions with minimal computational requirements. As evidence of success, practical utility and improved performance over previous computational approaches weigh heavily.

Constructive AI. As an example of the constructive (bottom-up) approach to general principles of intelligence in AI, the heuristic DENDRAL project is particularily instructive (Buchanan, Sutherland and Feigenbaum, 1969, 1970; Feigenbaum, Buchanan and Lederberg, 1971). Taking the study of scientific hypothesis formation as a general goal, the DENDRAL project can be seen as a study of the merits of generality vs expertise with respect to the performance power of AI techniques. This effort is not to be confused with the performance perspective discussed in this paper, and interestingly enough the investigators on the DENDRAL project are careful to point out that,

> attention given to the program as an application of artificial intelligence research has tended to obscure the more general concerns of the project investigators (Feigenbaum, et.al., 1971, p.166).

As a problem choice, structure elucidation by mass spectrometry in organic chemistry serves as a complex, real world problem which is,

> complex enough and rich enough in internal structure and theory to provide many firm foundation points on which to erect a meta-level for the study of theory formation processes (Feigenbaum, et.al., 1971, p.187).

The problem is chosen to provide a "forcing function" in which domain requirements will guide system design and illuminate representational and processing issues of more general significance in AI.

Methodologically, the DENDRAL project proceeds as an iterative interplay between program design/construction and performance/experimentation, with general issues emerging in the process of accommodating demands in the task domain. For example, in attacking the amine family of chemical structures, the need for strong heuristic constraint on the space of possible structures leads to the construction of a powerful (in terms of limiting generated structures) planning mechanism. This development, combined with the incorporation of simple hypotheses concerning likely fragmentation patterns, introduces problems of consistency among multiple sources of knowledge which in turn leads to an appreciation for the desirability of separating knowledge representation from processing details. Reflecting on this chain of events initiated by an attempt to constrain search, the investigators report,

> there are a number of ways to do this, some of which were tried with success, some with failure. The failures were at least as illuminating as the successes (Feigenbaum, et.al., 1971, p.171).

As might be expected, rewriting substantial sections of code is described as a common activity in the project (Buchanan, et.al., 1970). Hence, demands originally quite specific to the task at hand can be seen to force consideration of issues of general significance in AI.

A variety of evidence is given for the success of the heuristic DENDRAL project. Comparisons of the complete space of structural candidates with generated and suggested candidates are given as evidence of efficient and correct structure determination for selected molecular classes (Feigenbaum, et.al., 1971). In addition, published reports of candidate structure spaces appearing in widely read chemistry journals are given as evidence that chemists found DENDRAL's performance interesting. Favorable comparisons with structure identification by human experts (graduate students and a post doctoral fellow) for selected molecular classes are also given (Buchanan, et.al., 1970). Finally, arguments are made for the extensibility of the DENDRAL programs to accommodate new molecular classes and rapidly accumulating theoretical knowledge of mass spectrometry.

In summary, the heuristic DENDRAL project provides considerable insight into the manner in which <u>constructive AI</u> is routinely done. Complex, real world problems are chosen to create an experimental design atmosphere in which issues of general interest to AI research are regularily forced into active consideration. Solutions to these design problems as incorporated into functioning software systems are evaluated in terms of efficiency, credibility with human experts in the domain area, and demonstrated or promised extensibility.

Formal AI. The original work of Hart, Nilsson and Raphael (1968) and subsequent descriptions (Nilsson, 1971; 1980b) of heuristic search serve as widely read exemplars of the formal perspective in AI research. Apart from any concern for naturally intelligent behavior, these reports provide an abstract framework for using domain specific information in determining minimum cost solutions for a large class of specific problems expressed more generally as graph search problems.

As a problem choice, heuristic graph search represents an abstraction of problems encountered in many applications areas (eg. navigational routing, circuit design or problem-solving). It is the intent of the authors to give a general theory of heuristic search which encompasses a variety of techniques previously reported in the AI literature.

Methodologically, these exemplary reports proceed by giving a formal problem definition of finding minimum cost paths for a restricted class of graphs which serve as a general representational medium for a variety of search problems. A generalized algorithm is developed, which, using suitably restricted evaluation functions to determine which node to consider next, can be shown always to yield a minimum cost solution path between start and goal nodes, providing such a path exists. Claims for the correctness of this general algorithm are proven, and, in later reports (1971, 1980b), performance comparisons are made between particular algorithms using differently informed evaluation functions.

Criteria for success in these reports include acceptable proofs of algorithm correctness and performance increases for more "informed" versions of the algorithm. More generally, it is shown that the proposed formalism does indeed cover a wide class of search techniques, from blind search to heuristic search in which the chosen evaluation function provides a relatively tight lower bound on actual minimum cost solution paths.

Generalizing over exemplary particulars, problem choice in the formal AI perspective appears to focus on recurring problems across multiple domains which identify the need for general techniques. Hence, general techniques should not be considered to arise in a vacuum, rather they emerge as a result of a perceived need for an encompassing formal framework for some related classes of existing problems. Methodologically, work proceeds by giving a formal problem specification, detailing some computational mechanism (eg. an algorithm) for solving the problem, and then giving some justification for the appropriateness of that mechanism. Although not appearing in the heuristic search exemplar examined above, a demonstration of the proposed general technique in a particular (typically constrained) task environment is commonly used. Successful work in this perspective requires unambiguous and adequately descriptive specifications of problem and solution mechanism for an "important" class of related specific problems. A problem class can be considered important to the extent that particular manifestations of the problem recur in the literature and various manifestations can be meaningfully viewed as members of a more general class of problems which have yet to see an encompassing solution. Demonstration of computational sufficiency must be convincing (eg. proofs or assumptions should be believable) and comparisons with alternate methods of solution should be favorable.

Speculative AI. Schank and Abelson's (1977) treatise on natural language understanding stands as a clear exemplar for AI researchers interested in naturally occuring intelligent behavior apart from the onerous task of empirical verification characteristic of empirical AI (discussed next). For Schank and Abelson, the focus of research is squarely on proposing a theory which can account for human abilities in understanding and generating routine connected discourse in a natural language. Although discussion periodically turns to a more general theory of "knowledge systems" which might encompass both human and machine performance, concern for human functioning is clearly emphasized.

As a problem choice, the authors constrain the immense domain of natural language use by focussing on what they term the "naive psychology" and "naive physics" of everyday human discourse (Schank and Abelson, 1977, p.4). Rather than strictly defining a task environment, these constraints are taken to provide a starting point with relatively simple (ie.

common-sense) forms of knowledge.

Methodologically, the approach taken by Schank and Abelson marks a sharp divergence from traditional psychological or linguistic approaches to language use which the authors find unnecessarily restrictive. In their words,

> we are willing to theorize far in advance of the usual kind of experimental validation because we need a large theory whereas experimental validation comes by tiny bits and pieces (Schank and Abelson, 1977, p.7).

This impatience with empirical demonstrations of theoretical validity is offset with what is described as a painstaking process of implementing theorized mechanisms of natural language understanding in clearly specified computer programs. Difficulties with program construction are taken as theoretical inadequacies which must be remedied before implementation will be successful. The source of theoretical speculations appears to be primarily introspection guided by "intuitive necessity" and "internal consistency" (Schank and Abelson, 1977, p.21). Occasionally, observations of the behavior of others (ie. the daughter of one of the authors) is used in discussing theoretical propositions, but these observations appear to be used more as anecdotes to motivate theoretical speculation than as empirical demonstrations of theoretical validity. Lastly, separate programming efforts (eg. the script and plan applying systems, SAM and PAM) are described as working implementations of theoretical components.

Criteria for success in this exemplar amount to the reader's sense of psychological plausibility bolstered by evidence in the form of working programs. For example, the relatively appealing proposition of script-based episodic memory for stereotypical human experiences is supported with output of the SAM (script applier mechanism) program which includes descriptions of events which are never mentioned as part of the input story, underscoring the importance of expectations in understanding natural language.

In summary, the _speculative AI_ approach to studying naturally occuring intelligent behavior selects problems which are common representatives of the system under study, unlike the rather restricted (to facilitate empirical tractability) problems chosen in the empirical perspective discussed next. Introspection guides the formation of theoretical propositions concerning intelligent behavior, which are then tested by attempting to embody those propositions in clearly specifiable computer programs. To the extent that implementational difficulties arise, theoretical propositions are reconsidered. Successful work within this perspective consists of working programs which are taken as support for the sufficiency of psychologically plausible theories.

Empirical AI. The widely-known GPS (General Problem Solver) project of Newell and Simon (1963, 1972) serves as an exemplar of the empirical approach to modelling naturally intelligent systems, in this case the behavior of humans in well defined problem solving tasks. There is little doubt that these authors intend a model of human cognition (eg. "GPS, a program that simulates human thought," Newell and Simon, 1963), and their published reports are quite strongly connected to the psychological literature.

As a problem choice, Newell and Simon suggest human tasks for which AI can provide potential representational and computational strategies (ie. data structures and means of accessing them that might have some psychological validity). Cryptarithmetic puzzles, theorem proving in logic, and chess are chosen for detailed explication in the GPS project. Extension of a similar experimental approach to a wider range of human functioning is suggested (Newell and Simon, 1972), but only after a thorough understanding of human behavior in less complex domains.

The primary methodology evident in the GPS project is a sustained interplay between program construction and a comparison of program/human performance in the current domain. This comparison proceeds as a detailed ideographic analysis of verbal problem solving protocols. Program revisions are proposed to accommodate discrepancies between program traces and human protocols. The level of detail at which this comparison is done varies, with considerable interest in the identification of subprocesses in human performance which can be shown to correspond to the actions of subcomponents within the entire program. For example, segments of the human protocol such as, "I'm looking for a way, now, to get rid of that horseshoe. Ah ... here it is, R6." (Newell and Simon, 1972, p. 461) are taken to correspond to a search through the set of operators (rules of inference in this case) during a subject's first exposure to the GPS experimental task, before a "table of connections" (used in the GPS program to index rules by difference reductions) has been acquired.

As evidence for success of the GPS project, the investigators give detailed descriptions of correspondence between program functioning and human behavior. When discrepancies do arise, to the extent that the program can be modified without major reorganization, the program is considered a valid simulation of human problem solving. Interestingly, more serious discrepancies (eg. the tendency of some subjects to backtrack and correct previous rule invocations) are viewed positively as uncovering additions which "could significantly increase the total capabilities of the program" (Newell and Simon, 1972, p.472) rather than as failures. Finally, the viability of the theory (that is humans and programs as information processing systems) is demonstrated over multiple task domains as evidence of general applicability.

More abstractly, the empirical AI perspective can be seen to choose well-defined problems which natural systems perform well. In addition, there seems to be some tendency to choose tasks for which an academic literature exists to suggest fruitful approaches. Beyond the exemplar discussed here, this tendency might be seen to extend into choosing tasks for which some empirical database exists with respect to behavior of natural systems. The methodology is essentially experimental, incrementally modifying the artificial system as a result of careful comparison with behavior of the natural system. Newell and Simon's rather strict reliance on ideographic analysis should not be generalized to the class of research undertaken in this perspective as a whole. In general, the primary methodological point is that of detailed, empirical comparison between program as model and the naturally occuring system. Criteria for success include: empirically demonstrable correspondence between program performance and natural behavior, robustness of program design with respect to incremental changes (taken as an indicator of the fidelity of the model), and (perhaps less crucial than the previous criteria) the extensibility of model concepts to varied task domains.

Conclusion

The taxonomy for methodological perspectives in AI developed in this paper is certainly not complete. Particular studies used as exemplars in the preceding section were obviously presented so as to support the taxonomic types being developed. Faced with an arbitrary research report from the AI literature, an attempt at classification according to the current taxonomy would require careful consideration of a variety of issues. In particular, methodological orientations are seldom presented with any clarity in published research reports, making the identification of what was actually done or intended as part of a research project quite difficult to determine. As was pointed out quite candidly after a recent conference (Ohlsson, 1983, p.53), researchers in AI have a pernicious tendency to write and talk about what they would like for their research to demonstrate as opposed to what is actually being demonstrated. If the accepted vocabulary of AI is, indeed, "about as precise as that of poetry and about as substantive as that of advertising copy" (Doyle, 1983), some changes in the manner in which research reports are presented would seem desirable.

This paper has been an argument not only for a taxonomy of methods in AI research, but has also been an attempt to underscore the importance of making one's methodological orientation explicit when communicating results with the rest of the AI community. It is the hope of the authors that this will mark a beginning in what should be an ongoing public discussion aimed at defining the current state of AI research. The benefits of this sort of discussion, hopefully, are obvious. Not only might adherents of divergent approaches begin to appreciate or at least understand the motivations of other researchers, but the field might become more solidly accessible to newcomers.

References

Barr A. and Feigenbaum E. (1982) The handbook of artificial intelligence. Heuristech press, Stanford, CA.

Brachman R.J. and Smith B.C. (1980) SIGART Newsletter 70, Special issue on knowledge representation.

Buchanan B., Sutherland G. and Feigenbaum E.A. (1969) Heuristic DENDRAL: a program for generating explanatory hypotheses in organic chemistry. Machine Intelligence 4, 209-254.

Buchanan B., Sutherland G.L. and Feigenbaum E.A. (1970) Rediscovering some problems of artificial intelligence in the context of organic chemistry. Machine Intelligence 5, 253-280.

Bundy A. (1982) What is the well-dressed AI educator wearing now? AI Magazine, Winter 1981-1982, 13-14.

Doyle J. (1983) What is rational psychology? Toward a modern mental philosophy. AI Magazine Fall 1983, 50-53.

Feigenbaum E.A. and Feldman J.A. (1963) Computers and thought, McGraw-Hill, New York.

Feigenbaum E.A., Buchanan B.G. and Lederberg J. (1971) On generality and problem solving: a case study using the DENDRAL program. Machine Intelligence 6, 165-190.

Feigenbaum E.A. (1977) The art of artificial intelligence - themes and case studies of knowledge engineering. IJCAI 5, 1014-1029.

Green C. (1969) Application of theorem proving to question-answering systems. IJCAI 1, 219-237.

Hart P.E., Nilsson N. and Raphael B. (1968) A formal basis for the heuristic determination of minimum cost paths. IEEE Transactions of Systems Science and Cybernetics v. SSC-4(2), 100-107.

Hayes P. (1978) Trends in artificial intelligence. Int.J. Man-Machine Studies 10, 295-299.

Hunt E. (1973) The memory we must have. In Schank and Colby (Eds.) Computer Models of Thought and Language, Freeman, San Francisco, 343-371.

Kuhn T.S. (1970) The Structure of Scientific Revolutions, Universtity of Chicago Press, Chicago, Illinois.

Lenat D.B. (1977) Automated theory formation in mathematics. IJCAI 5, 833-842.

Schank R.C. (1983) The current state of AI: one man's opinion. *AI Magazine* Winter/Spring 1983, 3-8.

Simon H.A. (1969) *The Sciences of the Artificial*. MIT Press: Cambridge, Mass.

Waltz D. (1975) Understanding line drawings of scenes with shadows. In Winston (Ed.) *The Psychology of Computer Vision*. McGraw Hill Book Company: New York, 19-91.

Weimer W.B. (1979) *Notes on the methodology of scientific research*. Lawrence Erlbaum Associates: New York.

Weizenbaum J. (1976) *Computer Power and Human Reason: From Judgement to Calculation*. W.H. Freeman and Company: San Francisco.

Lenat D.B. (1978) The ubiquity of discovery. Artificial Intelligence 9, 257-285.

MACSYMA (1975) MACSYMA Reference Manual, The Mathlab Group, Project MAC, MIT.

Masterman M. (1970) The nature of a paradigm. In Lakatos and Musgrave (Eds.) Criticism and the Growth of Knowledge. Cambridge University Press, Cambridge, 59-89.

Minsky M. (1975) A framework for representing knowledge. In Winston (Ed.) The Psychology of Computer Vision, McGraw Hill Book Co., New York, 211-277.

Moses J. (1971) Symbolic integration: the stormy decade. Comm. ACM 14(8), 548-560.

Newell A. and Simon H.A. (1963) GPS, a program that simulates human thought. In Feigenbaum and Feldman (Eds.) Computers and Thought, McGraw Hill Book Co. Inc., New York, 279-293.

Newell A. (1969) Heuristic programming: ill-structured problems. In Aronofsky (Ed.) Progress in Operations Research 3, 360-414.

Newell A. and Simon H.A. (1972) Human Problem Solving. Prentice Hall: Englewood Cliffs, N.J.

Newell A. (1973) Artificial intelligence and the concept of mind. In Schank and Colby (Eds.) Computer Models of Thought and Language, W.H. Freeman and Company: San Francisco, 1-60.

Nilsson, N.J. (1971) Problem-solving methods in artificial intelligence. McGraw Hill: New York.

Nilsson, N.J. (1980a) The interplay between experimental and theoretical methods in artificial intelligence. Cognition and Brain Theory 4(1), 69-74.

Nilsson, N.J. (1980b) Principles of Artificial Intelligence. McGraw Hill: New York.

Nilsson, N.J. (1982) Artificial intelligence: engineering, science or slogan? AI Magazine Winter 1981-1982, 2-9.

Ohlsson S. (1983) Tell me your problems: a psychologist visits AAAI82. AISBQ 46, 27-30.

Ringle M. (1979) Philosophy and artificial intelligence. In Ringle (Ed.) Philosophical Perspectives in Artificial Intelligence. Humanities Press: Atlantic Heights, N.J., 1-20.

Schank R.C. and Abelson R.P. (1977) Scripts, Plans, Goals, and Understanding. Lawrence Erlbaum: Hillsdale, N.J.