SRI International

THE INTERPLAY BETWEEN EXPERIMENTAL AND THEORETICAL METHODS IN ARTIFICIAL INTELLIGENCE

Technical Note 229

September 1980

By: Nils J. Nilsson, Director
Artificial Intelligence Center
Computer Science and Technology Division

This paper will appear in "Cognition and Brain Theory" in 1981.

SRI International

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

ABSTRACT

This note alleges that there is a dichotomy between theoretical and experimental work in Artificial Intelligence (AI). The reasons for this dichotomy are discussed, and AI is compared with other, more mature disciplines in which there is closer cooperation between experimental and theoretical branches. Some recommendations are given for achieving this needed cooperation.

I THE MISSING CONNECTION

Most scientific disciplines have rich connections between their theoretical and experimental branches. Theory guides experiment, while experiment enriches theory. In my opinion, artificial intelligence (AI) does not yet display the degree of cooperation between theory and experiment that is essential, if AI's lofty goals are to be achieved.

Of course, there are tensions between theorists and experimentalists in all disciplines. But these tensions are usually rooted at the extremes of a spectrum that extends from pure trial-and-error experimentations at one end to pure mathematical speculation at the other. In mature disciplines there is a strong central portion to this spectrum where theory-guided experiments and applied theoretical work lead to major advances. In AI, however, this middle ground appears underdeveloped--leaving the extreme theorists and experimentalists out of touch with one another.

Experiments in AI take the form of computer programs. These programs are experimental artifacts; they are tested in constrained environments, with results that enlarge our understanding of the principles of AI system design. AI theory has concerned itself largely with heuristic search and with reasoning systems using formal logics. Theory provides a language in which to represent experimentally derived knowledge and in which to explore the logical consequences of this knowledge.

AI theorists often criticize experimental efforts as ad hoc--undisciplined and shallow. AI experimentalists largely ignore what they consider to be the irrelevant and impractical speculations of the theorists. Progress in AI is being hampered by a shortage of people who can reconcile these two extremes.

In this paper I shall first make some general comments about the relationship between theory and experiment in engineering disciplines, then suggest why this relationship has not yet developed in AI. Finally, I shall make some recommendations for improving the situation.

II THEORY AND EXPERIMENT AS HEURISTIC SEARCH

Artificial Intelligence has both scientific and engineering goals. As a science, AI aspires to explore and explain the nature of intelligent processes. Here, though, I am mainly concerned with that aspect of AI that is oriented toward the engineering goal of designing and building intelligent machines. Some readers might think it inappropriate to speak of theory and experiment in connection with engineering (as contrasted with scientific) disciplines. In the sciences, theories about nature are shaped by experimental methods. But how can we speak of experiments and theories in fields of endeavor whose goals are to design and make artifacts instead of to explore and explain the natural world? Isn't design fundamentally different than natural science?

Simon (1969) analyzes questions like these in his book The Sciences of the Artificial. I adopt here his point of view that an experiment to test the interaction between a complex artifact and its environment is analogous to an experiment in the natural sciences. Theories concerning artifacts are descriptions of their possible behaviors under various circumstances. Thus we can treat engineering disciplines as sciences of the artificial that employ the scientific method of theory development and experimental test.

In another paper, Simon (1979) used the AI process of heuristic search as a metaphor to describe progress in AI itself. I shall use it in this paper as a metaphor to describe the relationship between theory and experiment in AI. I liken the process of expanding a node in heuristic search to the performance of an experiment. The goal of AI, to build versatile reasoning systems, can be attained only along the path laid down by a large number of experiments.

A policy of ad hoc, trial and error AI experimentation that is not guided by theory is tantamount to a blind search. For the same reason that blind search is ineffective such a policy will not lead to the goal of producing sophisticated AI systems; it faces a combinatorial explosion of possible experiments.

Pursuing this metaphor, I liken heuristic search representations--with their nodes, operators and goal conditions--to the theoretical knowledge of a field. I also include with "theory" the control knowledge used for efficient heuristic search. AI theory provides the very framework in which experiments should be formulated. It allows us to prune the number of plausible experiments by predicting that some will be more successful than others. AI theory is an abstraction of the knowledge gained in previous experiments. It is a repository of the results of these experiments, as well as a set of hypotheses about how proposed experiments will turn out. Well-designed experiments improve AI theory by refuting hypotheses and by yielding results that will subsequently be described by expanded and modified theories.

The current situation in AI research resembles too much a blind search process conducted in a poorly represented search space. Experimenters are often unaware of relevant theoretical conclusions. Too often they merely try out random ideas and describe their results with ad hoc constructs unrelated to existing theory. Of course, when existing theory is underdeveloped, experimenters have no other recourse. But AI has already progressed far enough to permit heuristic rather than blind search.

Note that I am not favoring theoretical over experimental methods. Experiments are necessary to give us new knowledge; theory is necessary to describe that knowledge. My point is that the AI enterprise needs a closer intermingling of the two.

III OBSTACLES TO THEORETICAL-EXPERIMENTAL COOPERATION IN AI

There are currently three major factors that, in my opinion, act to separate theoretical and experimental methods in AI. The first is the relative ease, compared with other engineering fields, in which experiments can be carried out in AI. A typical AI experiment, involving the design and testing of a large computer program, requires about one or two man-years of effort and is often performed within the context of a Ph.D. dissertation. A wide range of possible experiments has been brought within easy reach by the flexibility and universality of computers and high-level list processing languages. Using our heuristic search metaphor, the flexible components for designing and performing experiments in AI induce a high branching factor. The prospective Ph.D. candidate can quickly put together any one of a large number of interesting experiments without bothering about theory.

The relatively higher costs of experiments in other engineering fields forces experimenters to consider their experiments more carefully using existing theory as their guide. It is usually more costly to try out a new highway or air foil or bridge or antenna design than it is to write a computer program. Instead of allowing the flexibility of our computer tools to continue to overwhelm us with a surplus of riches, we should make ample use of theory to help focus these powerful resources upon the task at hand.

A second factor that acts to isolate theory from experiment is an inadequate understanding on the part of experimenters as to what constitutes a useful description of their experiments. The most helpful descriptions would employ an appropriate level of theoretical language. Instead, experimenters too often give only an excessively detailed account of their programs, using the terminology of computer programming. They describe their <u>implementations</u> by discussing the LISP functions and data structures used. Analogous descriptions in civil engineering would focus on such low-level details as the kind of reinforcing steel embedded in concrete structures.

I acknowledge that theoretical descriptions can cover an entire spectrum of levels. LISP terminology may be appropriate in lower-level theories of computational processes, but higher-level constructs should be used for theories of reasoning and representation. Detailed accounts of program implementation should be relegated to appendices at the end of papers.

Some experimenters do accept the need for providing more abstract descriptions of programs and their behavior. But too often these descriptions are not very useful; they are couched in the imprecise terminology that, unfortunately, is far too prevalent in AI. Terms like "knowledge-based," "generate-and-test," "isa hierarchies," "event-driven," and "frames" are little more than slogans or buzz words. They are more a part of the promotional or political vocabulary of AI than of its technical vocabulary.

One reason experimenters fail to pose their experimental questions and describe their results in proper technical language is that they have never been taught to do so. The training of AI scientists borders on the scandalous. In many computer science departments, AI graduate students do not receive formal training in the technical subject matter that is the foundation for good AI work; instead they are treated as apprentices who are expected to learn the tricks of the trade as best they can and then go off and do innovative research. AI curricula are still rudimentary. Thus, AI researchers too often must struggle to invent new technical vocabularies to describe their work, even though perfectly adequate ones already exist.

One example of a body of underutilized theoretical apparatus is the predicate calculus. It is shocking that many AI researchers do not have an adequate knowledge of logic. Yet logic is <u>the</u> technical language for describing mechanized reasoning. Many AI researchers have not yet accepted the fact that the predicate calculus provides the theoretical basis and language for their work. AI experiments in representation and reasoning are best described using the technical vocabulary of logic. Instead, some researchers feel compelled to invent some new terminology that usually turns out to be inadequate as well as mysterious.

The following quotation, taken from a recent paper, is a symptom of the problem I am describing. "Organizing the representation as nodes with attributes is, of course, not new and is not essential. The representation could also be expressed in a formal logic (a translation to logic would be fairly straightforward)." If something could be described easily by using the predicate calculus, it certainly should have been.

A third factor that keeps theory and experiment apart is that not enough effort has been devoted to popularizing the results of AI theorists. Many important AI theoretical results--for example, those on nonmonotonic reasoning, metatheoretic reasoning, reasoning about propositional attitudes, and the use of modal logics--are described in highly mathematical papers that are beyond the grasp or interest of most AI experimenters. Simpler explanations of these formal ideas need to be written so that experimenters can understand them.

Other engineering disciplines furnish examples in which mathematical ideas are simplified somewhat for use by experimentalists. In electrical engineering, Heaviside's operational calculus simplified the use of the LaPlace transform, and the Dirac delta function permitted certain symbolic manipulations that avoided use of the more complex Stieltjes integral. Perhaps analogous simplifications in AI will enable experimenters to use more powerful theoretical ideas.

IV RECOMMENDATIONS

To accelerate progress in AI, we must strengthen the middle ground between theory and experiment. We can and should begin to do so now. First, both hard-headed experimentalists and speculative theorists should make efforts to acknowledge their interdependence. Theorists should be prepared to admit that, in large measure, the experimentalists are the frontiersmen and explorers in the field. Their expeditions challenge established knowledge and bring back specimens for scrutiny and description. Experimentalists should realize that theorists provide indispensable bases of support. Theorists provide the vocabulary, the metaphors, and models, as well as the targets that enable experimental expeditions to focus their efforts efficiently. Theorists and experimentalists who adopt such views will constitute the core of the essential middle ground.

Probably the most important long-range step we can take to improve progress in AI is to develop an AI curriculum. The newly formed American Association for Artificial Intelligence might establish a committee charged with drafting graduate courses of study leading to M.S. and Ph.D. degrees in AI. The education of AI researchers is now deplorably unsystematic. All other engineering disciplines teach their important ideas to students, rather than merely trusting that students will somehow absorb these ideas by osmosis.

In the meantime, there are some things that can be done by people who write articles and dissertations about AI experiments. If the article is about reasoning, deduction, or planning, the work should be described using the terminology of predicate logic. The predicate calculus is the product of a century of thoughtful work; it is unlikely that a few months of a graduate student's time will produce any significant rival.

Furthermore researchers should devote some effort to creating abstractions of their systems and describing these abstractions rather than describing the systems themselves. To explain MYCIN, NOAH, HACKER, or what-have-you, first invent a totally imaginary mini-NOAH, mini-MYCIN, or mini-HACKER. Don't worry that these fictitious systems might not be as powerful as their real counterparts; worry instead that they are more elegant! Make sure that most of the important theoretical ideas to be remembered are contained in the simplified systems. Often a simplified system can be invented that will model the essence of the real system, but will be much simpler to explain. Describe the mini systems using predicate calculus terminology whenever possible.

Experimenters must realize that experiments whose descriptions are too complicated will neither be remembered nor taught. Strive for simplicity and save the details for appendices. (The very process of inventing the mini system is also quite likely to suggest further refinements of the actual system and additional experiments.)

Analogous advice regarding simplification can be given to theorists--especially to those who want to communicate with experimentalists. Temporarily sacrifice precision to gain simplicity; regain precision in the appendices. Describe the "frictionless case" first.

AI will unquestionably establish itself in time as a mature discipline--one that will build on proven theory and create a remarkable galaxy of contrivances. There is no reason to believe, however, that either experimental or theoretical methods alone can accomplish the goals of AI. Nevertheless, we have experimenters in AI today who seem almost to take pride in their ignorance of theory and counterposed to them, theorists who are blissfully unfamiliar with what experimentalists have already learned. Let both extremes reconsider the power of heuristic search, so that we can apply it, in a metaphoric sense, to our common research enterprise.

REFERENCES

Simon, H. A., 1969, The sciences of the artificial, Cambridge, Massachusetts: The MIT Press.

Simon, H. A., 1979, Artificial intelligence research strategies in the light of AI models of scientific discovery, Proc. 6th Inter. Joint Conf. Artificial Intelligence, pp. 1086-1094, Tokyo, Japan.