AI Magazine Volume 6 Number 2 (1985) (© AAAI)

A Biologist Looks At Cognitive AI

William K. Purves

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

Abstract

Although cognitive AI is not generally viewed as being "scientific" in the same, strong sense as is physics, it shares a number of the properties of the natural sciences, especially biology Certain of the special themes of biology, notably the principles of historicity and of structure-function relations, are applicable in AI research From a biologist's viewpoint, certain principles of cognitive AI research emerge

Once in a while, a new science struts onto the intellectual stage. It typically gets mixed reviews—some critics raise their hands in horror and say, "This is not how things are done. You are violating the canons of drama, and I just don't like it." Others are swept along by the excitement of the play. Some of these friendlier critics may like what they see even though it runs counter to principles they have previously espoused, but most like the new play in part because it does fit into their intellectual framework. It is in this latter spirit that I view the science of AI.

AI is exciting. It appears to cross new frontiers and bring entirely new methodologies to bear upon long-vexing philosophical, psychological, and linguistic questions. However, increasing familiarity reveals a troubled field. Some AI-ers writhe with "science envy." One leader describes AI as "in a serious state of disruption" (Schank, 1983). There is a widely perceived lack of standards, a seeming inability to determine what work is of value and what is incompetent or irrelevant. The standards problem manifests itself during the planning of meetings, the reviewing of papers and books, and the allocation of grants.

Should AI be worried about its standing among the technical disciplines? Does AI have certain distinctive principles of investigation? Is it appropriate for linguists and philosophers to reject the results of AI out of hand? Is AI a scientific discipline at all? I shall touch upon some of these questions from the viewpoint of a biologist with limited experience as a member of an AI group.

AI embraces many disparate subjects, some merging with engineering and others with mathematics. In this article, I am limiting my domain strictly to that part of AI that intersects the discipline of cognitive science. Specifically excluded are those AI subdisciplines in which no attempt is made to base models on human mechanisms. Areas thus excluded are much of robotics, some work in natural language processing, and so forth. When no attempt is made to study human behavior, and when the program or representation is its own justification, one is doing engineering or mathematics and, presumably, suffers no psychological scars. It is the cognitive AI-er whose psyche is at risk, because of both concerns internal to AI and sniping from linguists and philosophers.

Is Cognitive AI a Science?

This question is ultimately a matter of definition, rather than one of logical proof. Perhaps the most rigorous test we could propose would be that of Popper (1962), who argued that science proceeds by "conjectures and refutations." The only propositions that Popper accepts as scientific are ones that are subject to definitive, objective refutation. Such liability to refutation is the hallmark of the physical sciences—we do not generally derive much sustenance from untestable hypotheses. The biological sciences also aspire to the Popperian ideal, although there the ideal may less often be achieved.

There are areas of biology in which the amenability to objective, immediate refutation is much less than in the vast majority of propositions of the physical sciences. The propositions of molecular biology are eminently testable. Indeed, this fact may account for the explosive growth of that discipline. But what of the taxonomist? Most of us

Permanent Address: Departments of Biology and Computer Science, Harvey Mudd College, Claremont, CA 91711 I am grateful to Roger Schank and the Yale Artificial Intelligence Project for their hospitality and to Daniel Dennett, Larry Birnbaum, Margot Flowers, Larry Hunter, Stan Letovsky, Toyoaki Nishida, and Roger Schank for helpful comments on this article as it developed.

think that when a taxonomist lays out a room full of flowers flattened on herbarium sheets, assorts and reassorts the sheets, and revises the classification of those flowers, that taxonomist is probably doing something we would call "science." However, much of such "alpha-taxonomy" does not appear to be subject to refutation. It certainly was not subject to refutation a few years ago, when DNA technology was in a less developed state.

Must all scientific propositions be subject to refutation? There are cases in which changing technology has enabled subsequent generations of physical and biological scientists to attempt definitive refutations of earlier propositions. It was recently proposed that the U.S. government spend some $130 million for a test of two predictions of Einstein's general theory of relativity. With advances in technology, these conjectures have only now become truly subject to experimental refutation– by observing the behavior of a set of gyroscopes whirling at near-zero gravity in space. This example illustrates an important point: Not all scientific propositions are subject to immediate test. Or are certain proposals to be regarded as nonscientific until such time as a test becomes technologically feasible?

Other propositions may, by their very nature, be incontrovertible by experiment or observation. Here I refer to theories concerning that which is forever past, such as the origin of life on Earth. Much is now "known" about the origin of life, yet our theories are based on plausible inference and attempts to explain what we see today. It is likely that it will never be possible to choose with certainty among certain contradictory but attractive theories. The activities of paleontologists result in propositions that may never be amenable to refutation, again because they deal with events forever gone. Yet who doubts that paleontologists, cosmologists, and students of the origin of life are doing science?

Much work in histology and anatomy, until relatively recently, lay outside the bounds of the "conjectures and refutations" paradigm. In many ways, the histologist and anatomist were purely descriptive scientists, concerned with determining and describing what was there. It is difficult to see in what sense such scientists were forming conjectures, let alone how any conjectures would be subjected to possible refutation.

What we have seen is that some subdisciplines of biology, unlike most of physics, fail to meet the "refutability" test of Popper. A similar situation prevails in a substantial fraction of geology, and in virtually all of meteorology. One might say that these disciplines are, rather than sciences of conjectures and refutations, sciences of the plausible. Until such time as their conjectures may be subject to clear refutation, these conjectures are, perhaps, best evaluated in terms of their ability to account for available data, their elegance and parsimony, and, in a word, their plausibility. May not such standards be applied, as well, to parts of AI and to cognitive science and the social sciences in general?

Cognitive AI has at least some of the trappings of "science." Sometimes it involves intensive data collection, the creation of conjectures, and attempts to refute those conjectures. AI is a discipline in which a major activity is the attempt to explain a body of observations, seeking a model consistent with all the observations.

Many of the propositions of AI are not subject to Popperian refutation. Some models may be refuted by demonstrating that, when realized as computer programs, they fail to perform as desired. However, once a model can be programmed successfully, it becomes harder to test. In some cases, psychological data may be brought to bear; but this is not always so. The problem is common in the social sciences; just the fact that a model explains the data does not mean that it accurately describes the basis for actual human performance. Consider the definition of an AI result proposed by Marr (1977): "A result in AI consists of the isolation of a particular information-processing problem, the formulation of a computational theory for it, and a practical demonstration that the algorithm is successful." This definition suffices if one is concerned simply with the development of programs that do things. However, it is inadequate for cognitive AI, which is concerned with psychological validity as well as with sheer performance.

There are those who fear that AI may produce multiple, mutually contradictory models of cognitive processes. If none of the models is refutable, then we have a most unscientific state of affairs. This was the situation in psychoanalysis that first motivated Popper to consider what might be the attributes of a scientific theory, as opposed to a nonscientific one. Such fears with respect to AI may be unduly pessimistic—or, in another sense, unduly optimistic. In cognitive AI, we are still looking for a first large-scale theory that deals adequately with the data at hand. Lacking a first theory, why get upset about its relationship to others that may never be developed?

The conclusion I draw from these considerations is that cognitive AI is a science to the extent that any of the social sciences are. For that matter, it ranks with the natural sciences. Forget the "science envy."

Biological Principles Relevant to Cognitive AI

A number of important themes inform biological research, and some of them are foreign to the physical sciences I shall sketch a few of these uniquely biological themes, and then show how two of them are relevant to research in AI In so doing, I hope further to convince some AI-ers to cast off their science envy. The two themes of most apparent relevance to AI are "historicity" and the marriage of structure and function. The next three subsections deal with them.

There are other themes found in biological, but not physical, research. One is the principle of hierarchy. Biological phenomena and biological explanations have multiple levels (molecular, cellular, tissue, organ, organismal, population, community, ecosystem, biosphere). In explaining a particular biological phenomenon, the experimentalist tries to show how events at lower levels in the hierarchy result in what is observed at the level of the phenomenon being studied. This is the reductionist approach. However, reduction is not all there is to explanation in biology. While no biological phenomenon violates the laws of quantum mechanics, the phenomena of biology are not predictable from the laws of quantum mechanics. That is, there are emergent phenomena, which arise in part through the operation of chance, so that they could not have been predicted. There are phenomena of great importance at the higher levels of the hierarchy, out of reach of the lower levels. One may speculate about whether the hierarchy principle has any application in AI research. One of the primary assumptions of AI is that any phenomenon of interest has an information processing level, but it is unclear whether there is an interesting hierarchy of other levels.

Another theme of biological, but not physical, science is that its phenomena arise from a plan, and that that plan is embodied in DNA. The science of AI also studies things arising, ultimately, from information encoded in DNA molecules. This fact provides part of the rationale for basing AI models on the mechanisms used by living things to accomplish the same purposes.

Historicity

Many of the phenomena studied by physical scientists occur under conditions such that it matters not whether time is running forward or backward. Often, the history of a purely physical phenomenon is of little or no interest. It may be the case that knowledge of the current positions and momenta of objects suffices to allow inference of any past state and a prediction of any future state.

In biology, this is never the case in any deep sense. The present is understood in terms of the past. This historicity of biology has two principal forms: evolutionary (the history of a species) and developmental (the history of an organism). In a real and scientifically useful sense, all biological phenomena are bound together historically. Every living thing on this planet is related to every other one—past, present, or future—and they are bound through time by common ancestry. Further, an individual organism is a function of its biography. It is a product of complex developmental events, foreordained by its genetic heritage; but it is also a product of the history of its interaction with its environment.

The historicity of the living world is exploited by biologists in many ways. In physiology, anatomy, and biochemistry, investigators have often used the "comparative approach," clarifying a puzzling situation in the species of primary interest by comparing it with a comparable structure or function in another species. Because of the historicity of life—in this case, the evolutionary link—such comparisons are both valid and helpful.

One of the most obvious examples of time-dependence in biology is the phenomenon of development—the irreversible changes in form, size, and biochemistry that occur over time from formation to death or division of an organism. Development is a meaningful concept in the study of viruses and of all living things, however simple their body forms. Events in the juvenile stage of an organism often have effects propagated through time to all subsequent stages, and the adult form depends absolutely upon appropriate development through the earlier ones. The biologist may be able to understand a puzzling phenomenon in the adult by studying what goes on at an earlier stage. Embryological results have even solved taxonomic problems; for example, the sea squirts were recognized as members of our own phylum Chordata only after their larvae were examined.

Historicity in AI?

(Before considering the possible roles of historicity in AI research, it is necessary to remark upon the great difficulty of representing time at all in a program. This is a problem of the first magnitude in the design of a planner. Two of the more interesting approaches have been those dealing with histories (Hayes, 1978, 1979) and persistences (McDermott, 1982). A simple approach, for an extremely limited domain, allowed SHRDLU to explain sequences of actions (Winograd, 1972). Real-world historicity can be strong, urgent, and a vital area of study.)

Development is a relevant concept in two senses in AI. First, studies of cognitive processes may be enriched by taking human developmental considerations into account. This idea seems particularly appropriate in studies of natural language processing, where models of language acquisition in children may well prove to be relevant to future programming efforts. Second, true AI programs of the future must, certainly, develop significantly as they run. That is, they must learn, reorganize their memories, and change importantly during their "lifetimes."

A useful attribute of some AI programs will be their ability to explain themselves. This ability will often entail such tasks as recounting sequential events in their development. On other occasions, it may require determining whether two events occurred at the same or different times.

I have cited the biologist's comparative approach as a characteristic example of the recognition of evolutionary historicity. Can a similar approach be used in the science of AI? One area where this approach may be valid is the study of natural language processing. Natural languages are bound together in groups by an evolutionary historicity. One preliminary exploration of this sort has been done, as a side issue, with respect to multilingual output by SAM (Carbonell *et al.*, 1978). One scientific outcome of that exercise was a positive test of the adequacy of conceptual dependencies (Schank, 1972) as a base for natural language generation. I have done some preliminary work of another sort, involving a comparison of English and Swahili utterances. Swahili has a remarkable degree of semantic redundancy, provided in part by a system of concordant affixes to several parts of speech; and I am interested in determining whether this facilitates disambiguation in parsing.

Another possible application of using a comparative approach in AI would be the exploration of certain behaviors of nonhuman animals in comparison with those of our own species. While much of our cognitive activity may be "purely human," other aspects may indeed have significant homology with behaviors performed by other animals—behaviors that might prove easier to model.

Structure, Function, and Goal

One of the deepest and most productive concepts of biology is that of the intimate relationship between structure and function. To understand a structure, whether that structure is an infinitesimal protein molecule, a two-meter-tall termite hill, or the megalithic array of Stonehenge, we must understand its function. To account for the performance of a function, the biologist needs to know what structures play roles.

To view biological structure-function relations in another way, consider convergent evolution, in which groups that are only distantly related come to solve similar problems in similar ways. The wings of hummingbirds and bumblebees are only analogous to one another (not homologous like the wings of pigeons and chickens—or like bird wings, human hands, and fish fins). Evolution has solved the problem of flight (as opposed to simple gliding) only by flattened, beating structures; it never, for instance, produced a jet engine—although squid use a form of jet propulsion for swimming. A comparison of the various body forms of cacti (found only in the Western hemisphere) with those of their Old World counterparts, the distantly related euphorbs, shows how similar are the structures that arise in convergent evolution. A similar comparison may be made between the marsupials of Australia and the placental mammals of the rest of the world.

How do we draw an analogy between biological structure-function considerations and what goes on in some AI research? Pylyshyn (1979), speaking of the work of Turing (1937), makes the analogy clear: "It represents the emergence of a new level of analysis, independent of physics yet mechanistic in spirit. It makes possible a science of structure and function divorced from material substance..." What is the analogue in cognitive science, divorced from material substance, to biological structure? It is behavior that corresponds to structure. In AI, one should not just study a behavior in and for itself, but should take it in a larger context, considering its function in accomplishing goals. Natural language processing by humans relates to such goals as information transfer, and this consideration should inform research on language (compare this position with that of Dresher & Hornstein [1976], for example).

In studying and modeling behaviors, it would be foolish to ignore their purposes or functions. Just as everything in the living world has gotten there by a process of evolution, so have all the subjects of AI arisen through either biological evolution, cultural evolution, or both. Things do not pass through such a filter unless they perform some function. The cognitive AI-er must remain alert to what may seem to be quirks, oddities, and tag ends of the cognitive process under study. For example, such tag ends as long- versus short-term memory, momentary "blankouts" of memory, seemingly irrelevant memories, or the triggering of memories by physical stimuli have presumably hung on because they relate to the accomplishment of goals in memory processing.

As another example of structure-function considerations in AI, consider the various forms of "aberrant" language, such as grunts, telegraphic speech, and "clever" speech, including *Finnegan's Wake.* These may all be regarded as relatively bizarre structures; but they have, without exception, essential functions. Language with extraordinary amounts of ellipses can still mediate significant —even fluent—communication. And most of us can read "Wan sue wassail ladle guller namer Ladle Rat Rotten Hut" as the beginning of a familiar story about a little girl and her adventure with a masquerading wolf. *Finnegan's Wake* presents overwhelming problems to most readers, but we can still feel much of its beauty and derive information from it. A linguistic approach too wedded to "standard" usage and a God-given (or DNA-given) syntax will be unable to deal with such "aberrant" language. However, an AI approach focusing on these structures as having functions (or fulfilling goals) can collect nuggets from veins untouched by conventional linguistics.

Does the example of biological convergent evolution suggest anything about structure-function relations in AI? In AI it is assumed that there are only a finite number of likely ways to solve an information processing problem. Perhaps as we go on, we will find that aspects of our earlier successful models will be applicable, at least in broad principle, to other problems as well.

Some Principles of Cognitive AI

Every science operates on the basis of various principles and paradigms. It is not my purpose (nor would I have the right) to announce these for cognitive AI On the other hand, it may be appropriate for me to identify a few important principles that relate either to the biological underpinnings of cognition or to the biological paradigms I have been discussing.

- The unique, distinguishing principle of cognitive AI research is this: If a process can be carried out by the human mind, then it must be possible to produce a computer model that operates by the same algorithm This is a radical claim, and there are many who will not accept it (*e.g* , Dreyfus, 1979, Searle, 1980). However, it is this claim, above all others, that characterizes the spirit—and should characterize the method—of cognitive AI. It corresponds with the basic claims of the physical and biological sciences, which assert that there is indeed a mechanistic basis for natural phenomena As a biologist, I find the claim a credible one. As generations of philosophers will attest, some of those algorithms will be extraordinarily difficult to discover, however, the computer is a powerful tool to use in the effort

• The ability to produce a working program is not a sufficient demonstration of the validity of a theory. However, it is a necessary part of the justification of a full-fledged AI theory. (This is an obvious corollary of the first principle.)

• A cognitive AI model should be based on the algorithm used by the mind. This, of course, represents the (distant?) ideal situation, in which we actually know the brain's algorithm. However, it represents an important goal A program that simply "works," in the sense of producing the desired output, may be useful in various ways:

(a) It may help us discover the psychological algorithm,

(b) it helps us understand the space of possible models, and

(c) it may be a useful piece of engineering.

It may be difficult, temporarily impossible, or even impossible in principle to confirm that a given algorithm is the one employed by the brain; but the effort should be made to discover psychological or other appropriate data that might be brought to bear. Another way to support a working program is to present evidence that some of its characteristics are necessary in order to perform the task in question.

• A person can explain her- or himself; a good cognitive AI program may explain itself. This implementation of historicity is the first and most obvious difference between ELIZA (Weizenbaum, 1965) and an AI program. Satisfying this principle will probably mean that the domain cannot be restricted as tightly as the investigator might wish, since this will make adequate explanation difficult or impossible.

• Just as biological structure and function are inextricably linked, and as biological phenomena are understandable in terms of purposes, an AI system links behavior and cognitive processes and is understandable in terms of goals Goal-directed processing is of great importance. This assertion is in opposition to one espoused by some social scientists. (See also my comments about "aberrant" language, above)

Hard Problems, Intractable Simplicity, and a Suggestion

The early days of AI research were characterized by attempts to model "really intelligent" behavior, such as theorem proving and chess playing It is now widely recognized that, in terms of understanding the workings of the human mind, this was not the way to proceed. That is, "hard" problems for people are not the right "hard" problems for the cognitive scientist to address. There is plenty of difficulty associated with such basic human behaviors as mundane explanation, memory, learning, and speech

To model accurately these behaviors as performed by a very young child, or by a severely retarded adult, would be a triumph. Although there seems to be little interest in the AI community in studying intelligent behavior of animals other than humans (but see, for example, Dennett, 1983), a program that successfully and legitimately modeled learning or communication in nonhuman primates would also be a triumph. And we might be more impressed by a chimpanzee that understood a conversation between two human preschoolers than by one that could play chess. As a biologist, I suggest that some AI-ers might find it profitable to model cognitive processes of nonhuman primates, other vertebrates, and perhaps even invertebrates. Ethologists and physiological psychologists have laid a groundwork that should support significant AI work in this area. While there are many references that could be cited on this point, I might recommend a very recent one that may interest AI-ers (Griffin, 1984).

Some of the "simpler" phenomena studied in AI labs are ones that have engaged the attention of philosophers for more than two millennia This duration should assure us that such problems are indeed hard ones. It is also exciting to consider that the methods of AI (perhaps even applied to other species) may allow real progress to be made, at long last, in dealing with such intractable "simplicity."

Some Problems of a Young Science

At the outset, I spoke of the concern some AI investigators have about an apparent lack of standards for AI research. Such concern has been nicely expressed by McDermott (1981): "In a young field, self-discipline is not necessarily a virtue, but we are not getting any younger." It is evident that there are problems in evaluating contributions to journals and conferences. And, speaking as a biologist used to certain standards for technical publication in my own field, I have been surprised on occasion by what passes for technical communication in AI

Claims are not always spelled out, and sometimes the broader issues addressed are obscured by the technical points that are the principal focus. Most startling to me as a biologist is the frequent failure to cite appropriate literature This puzzles me all the more since the primary

AI literature is still small enough so that it is feasible for a worker in any category of cognitive AI to read all the theses and major papers and technical reports that have appeared. Relevant material from the psychological and philosophical literature seems to be too often ignored. Opportunities are missed to back assertions with evidence from these or other sources. Comparing the biological and AI literatures, it would seem that the latter could profit from the more frequent presentation of either new data or relevant evidence from the literature, in such forms as psychological observations or experiments, results of introspection, and so forth

In addressing such problems of a young, dynamic science, I would urge tolerance and patience. In a field still in its infancy, it would be most unfortunate to squeeze out true originality. AI is at a stage when it needs all the ideas it can muster, even though most of them will inevitably be discarded. If one is embarrassed or repelled by the current sins of AI, it may help to seek out the technical literature of the early centuries of research in the physical and biological sciences—fields that have somehow survived in spite of early and, undoubtedly, continuing naivete.

References

Carbonell, J , R E Cullingford, & A V Gershman (1978) Knowledge-based machine translation. Research Report 146, Computer Science Department, Yale University

Dresher, B E & N Hornstein (1976) On some supposed contributions of artificial intelligence to the scientific study of language *Cognition*, 4:321-398

Dreyfus, H. L. (1979) *What computers can't do A critique of artificial reason* 2nd edition New York: Harper & Row.

Griffin, D I (1984) *Animal thinking* Cambridge: Harvard University Press

Hayes, P J (1978) Naive physics 1: ontology for liquids Institute of Semantic and Cognitive Studies, Geneva

Hayes, P J (1979) The naive physics manifesto Expert systems in the micro-electronic age Location: Edinburgh University Press

Marr, D (1977) Artificial intelligence a personal view *Artificial Intelligence* 9:37-48.

McDermott, D (1981) Artificial intelligence meets natural stupidity *Mind Design* Cambridge: MIT Press

McDermott, D (1982) A temporal logic for reasoning about processes and plans. *Cognitive Science* 6:101-155

Popper, K. R. (1962) *Conjectures and refutations: The growth of scientific knowledge* New York: Basic Books.

Pylyshyn, Z W (1979) Complexity and the study of artificial and human intelligence *Philosophical perspectives in artificial intelligence* Atlantic Highlands, NJ: Humanities Press

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

Schank, R C (1983) The current state of AI: One man's opinion, the *AI Magazine* 4(1):3-8

Searle, J (1980) Minds, brains, and programs *The Behavioral and Brain Sciences* 3:417-457

Turing, A M (1937) *On computable numbers, with an application to the Entscheidungsproblem* Proceedings of the London Mathematical Society 42:230-265.

Weizenbaum, J (1965) ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9:36-45.

Winograd, T (1972) *Understanding Natural Language.* New York: Academic Press