Artificial Intelligence Prepares for 2001

Nils J. Nilsson
AAAI President, 1982-1983
Director, Artificial Intelligence Center
SRI International
Menlo Park, CA 94025

Abstract

Artificial Intelligence, as a maturing scientific/engineering discipline, is beginning to find its niche among the variety of subjects that are relevant to intelligent, perceptive behavior. A view of AI is presented that is based on a declarative representation of knowledge with *semantic attachments* to problem-specific procedures and data structures. Several important challenges to this view are briefly discussed It is argued that research in the field would be stimulated by a project to develop a *computer individual* that would have a *continuing existence* in time

THOSE OF US engaged in artificial intelligence research have the historically unique privilege of asking and answering the most profound scientific and engineering questions that people have ever set for themselves—questions about the nature of those processes that separate us humans from the rest of the universe—namely intelligence, reason, perception, self-awareness, and language.

It is clear—to most of us in AI, at least—that our field, perhaps together with molecular genetics, will be society's predominant scientific endeavor for the rest of this century and well into the next—just as physics and chemistry predominated during the decades before and after 1900. The preoccupation of the physical sciences was, if you like, with the *machine code* of the universe. We in AI are now moving on to the higher level processes.

AI has made an excellent beginning toward understanding these processes. Two measures of the health of the field are the many AI applications that are now being reported and the almost-weekly announcements of new AI companies being formed to develop and market these applications. AI is finally beginning to "earn its way." Although I will briefly comment later about applications, I want to concentrate on the maturation of AI as a serious scientific/engineering discipline.

In the tradition of previous addresses by AAAI Presidents, I have decided not to attempt to present a consensual or averaged view of the important matters concerning our field; instead I will deliver some personal opinions. I trust that discerning readers and the growth processes of AI as a scientific field will provide sufficient filtering of these and other views.

What is AI's Niche?

To be considered seriously as a scientific field, we need at least to have good answers to several crucial questions such as, "What are we about?" "What are our special problems?" "What sets us apart from adjacent disciplines?" and "What have been our major accomplishments?" In dealing

This article is based on the author's Presidential Address given at the Annual Meeting of the American Association for Artificial Intelligence on August 11, 1983 in Washington, D.C

with these questions, we should keep in mind that it is unproductive to attempt overprecise definitions of fields or to draw rigid boundaries between disciplines. Yet, sooner or later we must come to accept the following important insight: not *all* symbolic information processing contributing to intelligent, perceptive behavior is of a piece. There are joints at which to carve this large body of science and engineering, and there are other disciplines that have already established legitimate claims on some of the pieces. Electrical, mechanical, and control engineers are making important contributions in robotics, image and waveform analysis, pattern recognition, and specialized computational architectures. People in operations research and decision analysis are concerned with search techniques and optimal decisions under uncertainty. Cognitive psychologists claim the task of explaining human and animal intelligence. Also philosophers, linguists, and those computer scientists who do not consider themselves AI researchers are all very much involved in the enterprise of synthesizing intelligent behavior. Clearly the territory has already been partially subdivided, and AI has not emerged in control of all of it.

There are component pieces, however, for which AI has no competition. These pieces are at the core of what might be called *high-level* reasoning and perception. This core is mainly concerned with the collection, representation, and use of *propositional* or *declarative* knowledge. (Such knowledge is of the type that can be stated in sentences of some form, as contrasted, say, with knowledge that is implicit only in procedures or in ad hoc data structures.) Attempts to develop and use formalisms for the propositional representation of knowledge have centuries-old antecedents in the aspirations and work of many scientists from Leibniz through Frege and Russell to the present day. We in AI are now pursuing these core topics with a blend of practical and theoretical goals and achievements unmatched by any other scientific/engineering field. In my opinion these topics constitute AI's special niche within the overall subject of intelligent, perceiving mechanisms.

To describe more clearly what constitutes this niche, I will mention what I think are some of the major efforts and achievements of AI over the last few years:

1. An understanding that much of the knowledge that we want our programs to have can and should be represented declaratively in some sort of propositional, logic-like, formalism—and that much of the manipulation of that knowledge can and should be performed by mechanisms based on logical operations and rules of inference. We might call this point of view, the *propositional doctrine*. With only minor modifications, this style (a declarative knowledge base plus an inference mechanism) is the basis for the major AI architectures. I think there is more agreement on this point than meets the eye. Even when designers do not conceive of their systems in these terms, they can nevertheless often be coherently explained in these terms. Many expert systems, such as MYCIN (Shortliffe, 1976), for example, can be thought of as *backward-chaining* type propositional logic theorem-provers with overlay mechanisms for dealing with certainty factors. The so-called *knowledge intensive* nature of these systems does not in any way disqualify them as *theorem provers*. Such knowledge is generally achieved by a large base of axioms that expresses knowledge about the domain of application.
   Emphasis on propositional and declarative representations does not mean that knowledge represented by procedures and ad hoc data structures does not have a place in AI. But AI is a *user*—not an *inventor*—of such procedures and data structures. Their *invention* comes from other disciplines—often from the domain of application. The connection between these two types of knowledge representations is of great importance in AI. Hence:

2. A theory of how propositional and procedural knowledge are connected based on *reflection principles* and on *semantic attachment* to partial models. Richard Weyhrauch (Weyhrauch, 1980) pioneered the theory of these connections. Briefly, the theory involves thinking of specialized procedures and data structures as components of a partial model or interpretation for the language used by the propositional representation. It is said that such procedures and data structures are *semantically attached* to their corresponding linguistic entities. Semantic evaluation brings out the knowledge implicit in the special structures.

3. Demonstrations that major bodies of expert knowledge about *real problems* can be represented propositionally and used effectively according to the propositional doctrine These demonstrations have occurred in medicine, programming, geology, chemistry, business, and other subjects. They have given rise to much of what we now call *expert systems* (Hayes-Roth, Waterman & Lenat, 1983) and form the technical foundation for many of the attempts to commercialize AI. Also, the propositional formalisms being developed to capture expert knowledge will have a major impact on those fields whose knowledge is being represented—just as conventional mathematics has had a major impact on physics and other scientific subjects.

4. Achievements regarding the control of AI programs:

- Embedding procedural knowledge to control deduction efficiently Carl Hewitt proposed several ideas for controlling inference in his work on PLANNER (Hewitt, 1978). Bob Moore's master's thesis (Moore, 1975) explained the logical basis for many of these and proposed additional ideas also. A key observation involves the different procedural interpretations that can be given for logical implication. Kowalski (Kowalski, 1975), Sickel (Sickel, 1976), and Stickel (Stickel, 1982) all focused on how to make resolution procedures more efficient by precomputing unifications and storing them in a structure called a *connection graph*
- Heuristic search processes. Much work in AI began

with the realization that search was unavoidable (if not criterial) in intelligent processes and could be heuristically controlled by employing knowledge from the problem domain. Hart, Nilsson, and Raphael (Hart, Nilsson & Raphael, 1968) formalized techniques for using evaluation functions to control search, and Pearl (Pearl, 1983) has recently synthesized and extended heuristic search theory in an important monograph.

- Meta-level control processes. Genesereth and Smith (Genesereth & Smith, 1982) and others are exploring the idea that the knowledge needed for control can itself best be represented propositionally (rather than procedurally—as in early AI programs), and that a meta-level reasoning system can use this knowledge to decide what the object level reasoning system ought to do next.

5 The idea that controlled deduction (or rather a controlled search for a chain of deductions) can play the role of a computation. The conceptual basis for this idea was described by Cordell Green (Green, 1969) and developed into a serious basis for a programming language by Bob Kowalski (Kowalski, 1974). Alain Colmeraurer (Colmeraurer, 1975) first implemented a special case of such a language, called PROLOG, and David Warren and Fernando Pereira (Warren & Pereira, 1977) developed an efficient compiler that made PROLOG a serious competitor to LISP as an AI programming language. As a programming language, PROLOG is more than merely an AI development, but the general idea that propositional statements can be given a procedural interpretation has important implications for AI.

6. Elaboration of the view that the generation of sentences in natural language (such as English) by cognitive agents (such as humans) is a deliberate process planned to achieve specific changes in the cognitive state of the hearer/reader of such sentences. This view, called *speech-act theory*, is having a major impact on the way in which we look at natural-language generation and interpretation. The work builds on ideas originally proposed by philosophers Austin (Austin, 1962) and Searle (Searle, 1969). Perrault and Cohen and others (Cohen & Perrault, 1979) expanded on these ideas and have firmly established them as important AI concepts. Doug Appelt (Appelt, 1982) built the first system that used this approach in a planning system that constructed English sentences.

7. Advances in commonsense reasoning:

- Attempts to formalize commonsense domains. In addition to the knowledge that is the subject of formal, scientific disciplines such as physics, chemistry, etc., much of the knowledge that is important in human reasoning can be called *commonsense* knowledge. Commonsense knowledge concerns the everyday knowledge of the properties and behavior of physical objects that even most children know and the knowledge of culture and rules of behavior, and so on. Pat Hayes (Hayes, 1979, 1978) has been active in formalizing what he calls *naive physics*—the physics of everyday experience. Johan De Kleer and John Seeley Brown (De Kleer & Brown, 1983) and Ken Forbus (Forbus, 1982) have used the phrase *qualitative process theory* to describe a simplified model of physical processes to be used in commonsense reasoning. James Allen (Allen, 1981) and Drew McDermott (McDermott, 1982) have been concerned with the problem of representing concepts involving time and the chronology of events
Another topic that scientists have not formalized (but the best of them have mastered) is knowledge about how to discover important new concepts and heuristics. Doug Lenat (Lenat, 1982, 1983a, 1983b), inspired by the work of Polya, has gathered an impressive quantity of this knowledge together in his AI systems AM and EURISKO, although the task of formalizing it (in a language more perspicuous than LISP) remains.

- Representing and reasoning about knowledge, belief, and other propositional attitudes or cognitive states such as desires and intentions. This subject is important for several reasons. First, speech-act theory presumes the ability to represent these states so that they can be reasoned about by both speaker and hearer. Second, AI systems will have to reason about the cognitive states of their users in order to be more flexible and helpful. Third, cooperating AI systems (such as teams of cooperating robots) will have to reason about each others' cognitive states in order to coordinate their performance. Finally, AI systems will have to reason about their own cognitive states in order to display certain aspects of truly intelligent behavior. Philosophers such as Quine, Kripke, and Hintikka have all concerned themselves with problems connected with representing propositional attitudes. Several approaches to this problem have been pursued in AI. John McCarthy (McCarthy, 1979) made some early proposals for representing and reasoning about knowledge, and Bob Moore (Moore, 1979) worked out a computationally feasible approach for reasoning about knowledge and action. Kurt Konolige (Konolige, forthcoming) has investigated an approach that also accounts for limits on the ability of an agent to deduce conclusions from more primitive beliefs.

- Nonmonotonic Reasoning Reasoning is called *nonmonotonic* when the reasoning agent must withdraw a previously deduced conclusion in response to learning some new fact. In ordinary logic, the set of conclusions increases monotonically with the addition of new facts, but some human reasoning seems to be nonmonotonic. (From the statements "Tweety is a bird" and "All birds can fly," we can deduce "Tweety can fly;" yet we would probably want to withdraw that conclusion when we learn that Tweety is an ostrich.) McDermott and Doyle (McDermott & Doyle, 1980) have developed what they call a *nonmonotonic* logic that addresses this problem, and several others, including John McCarthy (McCarthy, 1980), Ray Reiter (Reiter, 1980), and Bob Moore

(Moore, 1983), have made important contributions.

These then are examples of problems and accomplishments in AI. They are topics that cannot be claimed by other disciplines. Whether we in AI like it or not, the dynamics of the development of scientific fields will exert a strong bias toward AI maturing in this direction, regardless of any claims we may make about AI encompassing a much broader range of subjects. I think there is already evidence that we are beginning to make a science of AI along these lines. Nevertheless, there are some serious challenges to this point of view, and we must look at some of these challenges next.

Challenges

The first challenge is more sociological than technical. Simply put, there is wide disagreement in the field about what AI is all about. People have come to AI from various other disciplines and bring with them different standards, traditions, and problems. There is as yet no agreed-on curriculum for training students in AI, and new researchers enter the field from the universities with quite different points of view about the field. These differences, while perhaps stimulating creative development, work against the maturation of AI as a serious science.

One of the major differences among AI researchers is whether AI is more like an empirical art (learned through apprenticeship and case studies) than like a theoretically-based technical subject. This issue is often characterized as one between the *scruffies* and the *neats*. The scruffies point to their experimental, creative work that brings new knowledge into AI, while the neats complain about the scruffies' lack of formal rigor and theory. This argument echoes similar ones in older engineering disciplines. I remember the same argument in my student days as an electrical engineer between the handbook-experimentalists and the theoretically-based, network-synthesis people. Of course, the scruffy-neat controversy is a nonissue. A dynamic field needs scruffies (informed but not overly inhibited by neat theories) at its expanding frontier of knowledge, and it needs neats to codify, clarify, and teach its core concepts. A field that is scruffy to the core has simply not yet matured as a science, and one that does not have a scruffy exterior is simply sterile.

There are also a set of potential technical challenges to the view of AI that I have been promoting. Any one of these could upset this view. First, there is what I call the *dead duck* challenge. It asserts that logical languages and other formal representational schemes having a so-called *truth-theoretic* semantics are not appropriate as representational and reasoning frameworks. In his 1982 AAAI president's address, Marvin Minsky stressed the inappropriateness of general sentences (such as "All birds can fly") for representing knowledge about the real world because the real world usually presents exceptions (such as dead ducks) to such sentences.

In its strongest form, Minsky's position challenges the power of language itself to represent the world. Humanists, of course, take such limitations for granted. Writer Edward Abbey (Abbey, 1971) expressed the problem well by saying that "Language makes a pretty loose net with which to go fishing for simple facts, when facts are infinite." But Minsky is raising the issue as a scientific one, and it deserves serious consideration.

This is the challenge that the previously mentioned work in nonmonotonic reasoning is attempting to answer. It was recognized very early in the history of AI by John McCarthy who was concerned with problems that seemed to require the withdrawal of some conclusions as certain new facts were asserted. AI researchers are indeed split about how to deal with this problem (Kolata, 1982). Some (myself included) think that formal, logical methods can be extended and employed to respond to this challenge. Minsky and others (Minsky, 1980; Winograd, 1980; Simon, 1983) feel that something beyond anything we would call logic will be required. If they are right, the way AI will develop as a science might be entirely different than the way I have described. Although it may still be a bit too early to tell how this will all work out, it seems that the people who are pursuing formal approaches to dealing with this challenge are making good progress, and I am not aware that any other approaches have been proposed that have comparable power.

Next, there is what I call the *iconic* challenge. One hears, especially from those engaged in machine vision-research, that propositional representations (even with semantically attached procedural knowledge) are not appropriate for handling many types of perceptual reasoning problems. For these, it is argued, one needs so-called *conic* representations in which, in some sense, the representational structure itself plays the major representational role. My view is that the mechanism of semantic attachment to those computations and data structures appropriate for the problem at hand is the way to achieve iconicity, but some iconocists argue that they will ultimately develop a general theory of these sorts of representations and reasoning methods for them that will put them on a par with the propositional ones. Although we are all aware of the importance of special structures, such as trees and lattices for representing hierarchies, and generalized cylinders for representing shapes, I have not yet heard of any general *iconic* language.

Another problem for AI is the *holistic* challenge. A successful (if not essential) methodology in science and engineering is to "divide and conquer" a subject by analyzing its parts. Some commentators on AI, such as Dreyfus (Dreyfus, 1977) assert that intelligent behavior is by its nature irreducible to independently analyzable components. Some of us in AI have attempted to subdivide AI's concern for *knowledge* into three areas, namely, what knowledge is pertinent to a problem, how should that knowledge be represented, and how and when should that knowledge be used. This approach is criticized by those who say that any attempt to consider these subareas separately completely misses the most important point about knowledge--its collection, representation and use must be considered simultaneously. According to this view, the way in which knowledge is to be *used* in a problem (and there are many different ways in which it could be used) dictates how it should be *represented* for that problem. If this aspect of the holistic challenge succeeds, then I think AI will not develop as a separate subject at all but will instead develop multiply and independently in the myriad of niches where intelligence might be applied. I think the very existence of more-or-less "general purpose" intelligence (occasionally!) in humans argues against this challenge.

A final challenge, the *kludge* challenge, arises out of Marvin Minsky's oft-mentioned assertion that "Intelligence is a kludge." Minksy proposes that intelligence in humans is a manifestation of a large number of complex, specialized mechanisms more-or-less haphazardly evolved in response to numerous and independent environmental pressures (Minsky, 1980). Don't look for unifying theories or simplified versions of intelligence, warns Minsky. To do so misses the important point about intelligence. If intelligence emerges only from unsimplifiable complexity, then the science of intelligence would be very much different from traditional sciences such as physics. In mechanics, for example, we gain understanding by considering first the *frictionless* case. The main principles already emerge in simplified models. We can add friction later without complete revision of the model. But, according to the kludge challenge, the phenomenon of intelligence is itself the very complexity that misguided AI theoreticians are unsuccessfully trying to simplify. In principle, intelligence has no simple models that can later be elaborated. It *is* the friction and dozens or hundreds of other unsimplifiable phenomena.

It may turn out that Minsky is right about this. If so, that would be bad news for scientists and engineers who are trying to understand and build intelligent mechanisms. I think it is far too early to concede this point. As scientists and engineers, we should continue to attempt to simplify, to organize, and to make elegant models—otherwise there are serious doubts that we would ever be able to understand enough about intelligence to design intelligent machines or to teach these design methods to students. If bridges had to be kludges, we wouldn't have a man-made bridge across the Golden Gate because complex bridge-building couldn't be understood, taught, or remembered. Successful engineering requires the frictionless case and a succession of gradually more complex models. I think AI has been reasonably successful so far in inventing understandable and useful theoretical frameworks and that it would be inappropriate for us to discontinue these attempts.

The next steps in the maturation of AI, I think, will be the successful disposal of these challenges. Their resolution will constitute an important part of the subject matter of AI.

Research Strategies

Orthogonal to the matter of what AI is all about are some questions concerning how we ought to pursue AI research. There are many facets to the topic of research strategy—too many to treat here. I want only to mention one of the interesting pairs of opposing approaches.

On the one hand, we have what might be called the *function-follows-form approach.* Here, we start with some sort of computational mechanism (the *form*) that appears attractive or relevant (for various reasons) and then see what sorts of things (the *functions*) can be done with it. There are many examples in the history of AI where this approach was followed. Work on perceptrons (Rosenblatt, 1961) was motivated largely by a certain infatuation with the idea of neural networks. Since the brain is composed of neurons, then it seemed that experimenting with networks of models of neurons might lead us to finding out some things about mental function. I think that some of the present work on so-called *connectionist* schemes (Feldman & Ballard, 1982) and active semantic networks draw on a similar motivation. The brain is a highly parallel computational mechanism, therefore we ought to be experimenting with parallel architectures. The *situation-action* rules of production systems (which can be regarded as a generalization of the *stimulus-response* pairs of psychology) provide another example of an intriguing computational mechanism that seems relevant to intelligent behavior and that invites experimentation. It must be admitted that some of the enthusiasm for the use of logic in AI might also be explained by this function-follows-form approach. Probably an historian of science would be able to point out cases in which this approach paid off by helping people create new functional ideas that they might not otherwise have thought of. A certain amount of AI research ought to have this motivation.

On the other hand, AI research might follow what might be called the *form-follows-function approach.* Here we start with a functional idea of what must be computed and then think about the forms for implementing this computation. The function itself is conceived of independently of whether or not it will be realized by serial or parallel primitive processes, by protein or silicon, by logic programming or LISP. The traditional AI heuristic search programs mainly followed this approach. Researchers first specified the kinds of functions needed for the tasks of symbolic integration, language understanding, and robot problem solving and then considered how they would implement these functions–usually in LISP. Probably most AI research follows this approach. That is why many AI researchers have a rather puzzled reaction to well-meaning lay questions like, "But how can a serial digital computer be intelligent; isn't the brain a parallel device?" Those sorts of quandaries don't arise when one is considering functions.

To predict which of these strategies will be more successful in AI requires more hubris than I care to acknowledge. I think I would commit most resources to the latter strategy but hedge my bet by encouraging some of the former.

AI and the Real World

As a science matures, its relations with the rest of the world become more interesting and complex. Here I want to comment a bit about the effects that successful AI applications are having on the science of AI and also about worldwide competition in AI.

First of all, I think that the more applications of AI we can be involved in, the better for our science, because then we'll sooner learn our weak spots and be more motivated to support the basic research needed to correct them. I don't see any obvious danger that *all* the AI researchers are going to concentrate on applications and on making money. (At SRI we have lost a large number of people to business careers, but those who are mainly interested in basic research are staying on.) As a matter of fact, the departure of applications-oriented people from the universities to businesses may be quite beneficial for AI. It brings those with applications interests into more intensive confrontation with real problems, and it leaves at the universities a higher concentration of people who are mainly interested in developing the basic science of AI

The main concern is that the funding agencies continue to appreciate the importance of long-range basic research in AI. (Let me remark in passing that, if they continue on the whole to be as enlightened as they have been during the last twenty or so years, future progress should not be hampered for lack of support.)

I expect that, in addition to those applications involving expert systems and natural language interfaces, we will soon see applications of AI planning systems (Sacerdoti, 1977; Robinson & Wilkins, 1982; Wilkins, 1983; Tate, 1976) for tasks like project planning, robot task planning, and error-recovery planning.

Most basic research in AI has occurred in the United States, although there have been some important contributions from universities in Britain and Western Europe. The American applications of AI have now begun to stimulate intense overseas interest in the subject. The Japanese have initiated a project to develop *fifth generation* computer systems (Feigenbaum & McCorduck, 1983), which will probably succeed in developing superb hardware for AI applications and will also attempt to make important conceptual contributions in natural-language processing and expert systems.

American discomfort over this Japanese challenge has prompted two major responses. First, a consortium of computer companies has formed the Microelectronics and Computer Corporation (MCC) in Austin, Texas. This new company will perform generic research and development on topics mainly related to the hardware needed for AI and other advanced computer applications. Second, the Defense Department (through its Defense Advanced Research Projects Agency—DARPA) has proposed a major new effort called *Strategic Computing and Survivability* (Cooper & Kahn, 1983). This project will support the development of AI hardware and technology and their application to important defense problems. Similar efforts have been launched in Britain and in Western Europe

These are welcome developments for AI in that they will make available needed extra resources to pursue research and applications and will stimulate productive competition among development teams. All of these projects, however, seem to be focusing on achievements that have a certain *inevitableness* about them—achievements that, although important, are almost bound to result from the strong tide of progress in memory technology, very large scale integration, and architectures for parallel computation.

So far, there has not been enough large-scale attention to those aspects of AI where progress is far from inevitable and on which the ultimate success of the fifth-generation style endeavors will really depend. Many conceptual breakthroughs are needed, for example, in commonsense reasoning and language processing, in order to meet the full goals of fifth-generation machines. One notable development, which I regard as extremely important, is the establishment at Stanford University of the Center for the Study of Language and Information—largely funded by the System Development Foundation. The Center, which includes participation by SRI International and the Xerox Palo Alto Research Center, will study, among other things, the close interrelation between language and computation. Projects like those being pursued at CSLI are the ones that will produce the conceptual developments needed to breathe intelligence into fifth-generation hardware.

The Computer-Individual Project

The future of AI as a science depends both on continued progress in basic research and on improving the flow of ideas from research settings into practical applications. I have no particular suggestions to give here about applications—although I am sure there are experts on the subject who would have something useful to say. I do have a recommendation for a research project that I think will stimulate advances in the basic science of AI.

A project should be initiated whose goal is to develop a class of new AI programs that would have a continuing existence. To use a phrase of Richard Weyhrauch's, let's call them *computer individuals*. The ground rule would be that they should never be turned off—a characteristic they would share with large time-shared computer operating systems and airline reservation systems. They would have a constantly changing model of the world and of the user(s). They should be able to engage in extended dialogs in natural language. Designing such programs would stimulate (in fact force) research in machine learning because it would be manifestly unintelligent for a computer individual existing over time not to benefit from its experiences. The problems of efficient learning, memory organization, and deciding what is important enough to learn become especially important for an individual whose storage and retrieval facilities are finite. (The SRI SHAKEY robot system (Fikes, Hart & Nilsson, 1972) "learned" new operators for achieving goals, but it never had to decide whether it ought *not* to save any of these new operators because it never "lived" long enough for memory capacity to become an issue.) Since computer individuals would exist in time, this project would also stimulate research in commonsense knowledge about time—just as robotics research has improved our ability to deal with commonsense knowledge about space.

We can think of several varieties of computer individuals, including personal assistants, meeting schedulers, expert consultants, and mobile robots. Imagine, for example, a general-purpose robot embodying one of these never-to-be-turned-off computer individuals. Or, even better, imagine a team of such robots that could communicate with each other (using an appropriate formal language) and with humans (using English, gestures, and graphics) to perform tasks. For the basic research purposes that such projects would serve, what specifically these robots would do is relatively unimportant. They could do anything that requires moving around in and sensing a real environment and manipulating that environment in some way. Aiming for roughly the same sort of sensing and manipulative abilities that people have would give us plenty of AI challenges.

In considering basic research projects of this kind, it is important to avoid the trap of insisting that the project be directed at some specific useful application. To attempt to justify robots by proposing them as useful highway constructors, for example, is misguided because general-purpose robots may be an inefficient way to solve the road-building problem—or the transportation problem—or any other specific problem. Any single application of robots alone is probably insufficient reason to justify their development. The whole reason for robots is their general-purposeness!

Projects to develop computer individuals ought to be undertaken in the same sort of spirit as would a manned landing on Mars—to see if we could do it and to learn by doing it. There is an important difference though between such projects and a Mars landing. All the basic science is in hand to achieve the Mars landing—it probably isn't yet to build the kind of systems that I am proposing. The purpose of trying to develop such systems would be to stimulate the development of the science. As basic research projects, they might fail—or they may end up doing something different. We are now only at about the same stage in being able to produce computer individuals as physicists were in relation to harnessing nuclear energy during the 1930s. We know something profound can be done, we have a few clues about how to proceed, we know that much more basic research must be done—and we want to get on with it.

Conclusions

These are truly watershed years for AI. Obviously, the applications of AI alone are ensuring that the field will change rapidly. But, more importantly, the lines along which our science is maturing should now begin to be clear to us.

Besides scientific maturity, we must also come to accept more mature social responsibilities for our technical developments. I recently attended two interesting meetings dealing with some of these matters. A panel discussion on the impact of AI on income and employment was held at the August 1983 IJCAI in Karlsruhe (Nilsson & Cook, 1983), and a workshop on the social consequences of AI was held the following week at IIASA outside Vienna. It is important for us to be concerned about the social impacts of AI. Like molecular genetics, AI can cause deep fears and confusions among our fellow citizens—some of whom might be eager to support restrictive legislation that could turn out to be extremely ill-considered. Although all citizens must participate in the decisions about how AI will be used—we AI scientists have the special responsibility for informing our fellows about the potential benefits and the limits of AI.

It is difficult to compare the magnitude of the effects of AI with those of other disciplines—some of which could destroy life on earth. At best, artificial intelligence will both liberate us from unwelcome toil and provide us with the most detailed picture we have ever had of ourselves. Possibly no science has ever posed greater challenges than those.

References

Abbey, Edward (1971) *Desert Solitaire*. New York: Ballantine Books.

Allen, J. F. (1981) An interval-based representation of temporal knowledge. *IJCAI 7* Vancouver, B.C.

Appelt, D. E. (1982) Planning natural-language utterances to satisfy multiple goals. SRI Artificial Intelligence Center Tech. Note 259.

Austin, J. L. (1962) *How to do things with words*. Cambridge Univ. Press

Cohen, P. R., & Perrault, C. R. (1979) Elements of a plan-based theory of speech acts. *Cognitive Science* 3, 177-212.

Colmeraurer, A. (1975) Les grammaires de metamorphase, Groupe d'Intelligence Artificielle, Marseille, Marseille-Luminy.

Cooper, R. S. & Kahn, R. E. (1983) Strategic computing and survivability. *Signal*, 5-28.

De Kleer, J. & Brown, J. S. (1983) The origin, form and logic of qualitative physical laws. *IJCAI 8*, 1158-1169.

Dreyfus, H. L. (1977) *What computers can't do*. Harper-Colophon Books.

Feigenbaum, E. A., & McCorduck, P. (1983) *The fifth generation: artificial intelligence and Japan's computer challenge to the world*. Reading, MA: Addison-Wesley.

Feldman, J. A., & Ballard, D. H. (1962) Connectionist models and their properties. *Cognitive Science* 6, 205-254.

Fikes, R. E., Hart, P. J., & Nilsson, N. J. (1972) Learning and executing generalized robot plans. *Artificial Intelligence* 3(4), 251-288.

Forbus, K. D. (1982) Qualitative process theory. MIT AI Lab Report AIM-664, Cambridge, Mass.

Genesereth, M. R., & Smith, D. E. (1982) Meta-level architecture. HPP-81-6, Stanford University Heuristic Programming Project Report

Green, C., (1969) Application of theorem proving to problem solving, *IJCAI 1*, 219-239.

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968) A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans. Syst. Sci. and Cybernetics* SSC-4(2), 100-107.

Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (Eds.) (1983) *Building Expert Systems.* Reading, MA: Addison-Wesley Publishing Company.

Hayes, P. J. (1978) *Naive physics: ontology for liquids* (working paper 35), Institute of Semantic and Cognitive Studies, Geneva.

Hayes, P. J. (1979) The naive physics manifesto. In D. Michie (ed.) *Expert systems in the micro-electronic age.* Edinburgh Univ. Press, 242-270.

Hewitt, C. E. (1972) Description and theoretical analysis (using schemata) of PLANNER: a language for proving theorems and manipulating models in a robot. Tech. Rept. TR-258, Artificial Intelligence Laboratory, Massachusets Institute of Technology, Cambridge, Mass.

Kolata, G. (1982) How Can Computers Get Common Sense?, *Science*, 217(4566) 1237-1238.

Konolige, K. (forthcoming, 1984) A deduction model of belief. SRI AI Center Tech. Note.

Kowalski, R. (1974) Predicate Logic as a programming language. *IFIP Congress.*

Kowalski, R. (1975) A proof procedure using connection graphs. *JACM* 22, 572-595.

Lenat, D. B. (1982) The nature of heuristics. *Artificial Intelligence* 19(2), 189-249.

Lenat, D. B. (1983a) Theory formation by heuristic search: The nature of heuristics II: background and examples. *Artificial Intelligence* 21(1,2), 31-59.

Lenat, D. B. (1983b) EURISKO: a program that learns new heuristics and domain concepts: The nature of heuristics III: program design and results. *Artificial Intelligence* 21(1,2), 61-98.

McCarthy, J. (1979) First order theories of individual concepts and propositions, in Hayes, J. E., Michie, D., & Mikulich, L. I. (Eds.) *Machine Intelligence 9.* Ellis Horwood, Chichester.

McCarthy, J. (1980) Circumscription — a form of non-monotonic reasoning. *Artificial Intelligence.* 13, 27-40.

McDermott, D. & Doyle, J. (1980) Non-monotonic logic, I. *Artificial Intelligence.* 13, 41-72.

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

Minsky, M. (1980) K-lines: a theory of memory. *Cognitive Science.* 4, 117-133.

Moore, R. C. (1975) Reasoning from incomplete knowledge in a procedural deduction system. Tech. Rep. AI-TR-347, MIT Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA.

Moore, R. C. (1979) Reasoning about knowledge and action. Tech. Note 191, SRI International, Artificial Intelligence Center, Menlo Park, CA.

Moore, R. C. (1983) Semantical considerations on non-monotonic logic. *IJCAI 8* Karlsruhe, 272-279.

Nilsson, N. J. & Cook, S. B. (1983) Artificial intelligence: its impacts on human occupations and distribution of income. *IJCAI 8* Karlsruhe. 1195-1198.

Pearl, J. (1983) *Heuristics: Intelligent Search Strategies for Computer Problem Solving.* Reading, Massachusetts: Addison-Wesley.

Reiter, R. (1980) A logic for default reasoning. *Artificial Intelligence.* 13, 81-132.

Robinson, A. & Wilkins, D. E. (1982) Man-machine cooperation for action planning. SRI Final Report on ONR Contract No. N00014-80-C-0300.

Rosenblatt, F. (1961) *Principles of Neurodynamics: perceptrons and the theory of brain mechanisms.* Washington, D. C.: Spartan Books.

Sacerdoti, E. D. (1977) *A structure for plans and behavior* New York: American Elsevier.

Searle, J. (1969) *Speech acts: an essay in the philosophy of language.* Cambridge, MA: Cambridge Univ. Press.

Shortliffe, E. H. (1976) *Computer-based medical consultation. MYCIN.* New York: American Elsevier.

Sickel, S. (1976) A search technique for clause interconnectivity graphs. *IEEE Trans. on Computers* 25(8), 823-835.

Simon, H. A. (1983) Search and reasoning in problem solving. *Artificial Intelligence* 21, 7-29.

Stickel, M. E. (1982) A nonclausal connection-graph resolution theorem-proving program. *AAAI-82 Proceedings of National Conference on Artificial Intelligence*, 229-233.

Tate, A. (1976) Project planning using a hierarchic non-linear planner. Research report no. 25, Department of Artificial Intelligence, University of Edinburgh.

Warren, D. H. D., Pereira, F. C. & Pereira, L. M. (1977) PROLOG–the language and its implementation compared with LISP. *Proceedings of the Symposium on Artificial Intelligence and Programming Languages (ACM) SIGPLAN Notices.* 12(8); and *SIGART Newsletter* no. 64, 109-115.

Weyhrauch, R. W. (1980) Prolegomena to a theory of mechanized formal reasoning. *Artificial Intelligence.* 13, 133-170.

Wilkins, D. E. (1983) Domain-independent planning: representation and plan generation. SRI AI Center Tech. Note 266R, Menlo Park, CA.

Winograd, T. (1980) What does it mean to understand language? *Cognitive Science* 4, 209-241.