EYE ON THE PRIZE

Nils J. Nilsson
Robotics Laboratory
Department of Computer Science
Stanford University
Stanford, CA 94305
e-mail: nilsson@cs.stanford.edu

January 30, 1995

Abstract

In its early stages, the field of artificial intelligence (AI) had as its main goal the invention of computer programs having the general problem solving abilities of humans. Along the way, there developed a major shift of emphasis from general-purpose programs toward "performance programs"—ones whose competence was highly specialized and limited to particular areas of expertise. In this paper I claim that AI is now at the beginning of another transition—one that will re-invigorate efforts to build programs of general, human-like competence. These programs will use specialized performance programs as tools, much like humans do.

Keywords: autonomous agents, general problem solving, habile systems

[This paper is being submitted to the AI Magazine.]

1 Diversions from the Main Goal

Over forty years ago, soon after the birth of electronic computers, people began to think that human levels of intelligence might someday be realized in computer programs. Alan Turing [Turing, 1950] was among the first to speculate that " . . . machines will eventually compete with men in all purely intellectual fields." Allen Newell and Herb Simon [Newell & Simon, 1976] made this speculation more crisp in their *physical symbol system hypothesis*: "A physical symbol system [such as a digital computer] has the necessary and sufficient means for *general* intelligent action" (emphasis mine). In its early stages, the field of artificial intelligence (AI) had as its main goal the invention of computer programs having the general problem-solving abilities of humans. One such program was the General Problem Solver, GPS [Newell, Shaw & Simon, 1960], which used what have come to be called *weak methods* to search for solutions to simple problems.

Many of the early AI programs dealt with *toy problems*—puzzles and games that humans sometimes find challenging but that they can usually solve without special training. When these early AI techniques were tried on much more difficult problems, it was found that the methods did not "scale" well. They were not sufficiently powerful to solve large problems of "real-world" consequence. In their efforts to get past the barrier separating toy problems from real ones, AI researchers became absorbed in two important diversions from their original goal of developing general, intelligent systems. One diversion was toward developing *performance programs*—ones whose competence was highly specialized and limited to particular areas of expertise. Another diversion was toward refining specialized techniques beyond those required for general-purpose intelligence. In this paper, I speculate about the reasons for these diversions and then describe growing forces that are pushing AI to resume work on its original goal of building programs of general, human-like competence.

2 The Shift to Performance Programs

Sometime during the 1970s AI changed its focus from developing general problem solving systems to developing expert programs whose performance was superior to that of any human not having specialized training, experience, and tools. A representative performance program was DENDRAL [Feigenbaum, *et al.*, 1971]). Edward Feigenbaum and colleagues [Feigenbaum, *et al.*, 1971, p. 187], who are credited with having led the way toward the development of *expert systems*, put it this way:

> ". . . general problem-solvers are too weak to be used as the basis for building high performance systems. The behavior of the best general problem-solvers we know, human problem solvers, is observed to be weak and shallow, except in the areas in which the human problem-solver is a specialist."

Observations like these resulted in a shift toward programs containing large bodies of specialized knowledge and the techniques required to deploy that knowledge. The shift was very fruitful. It is estimated that there are several thousand knowledge-based expert systems used in industry today. The American Association for Artificial Intelligence (AAAI) sponsors an annual conference on "Innovative Applications of Artificial Intelligence," and the *Proceedings* of these conferences give ample evidence of AI's successes.[1] I won't try to summarize the applications work here, but the following list taken from a recent article in *Business Week* [*Business Week*, 1992] is representative of the kinds of programs in operation:

- Shearson Lehman uses neural networks to predict the performance of stocks and bonds
- Merced County in California has an expert system that decides if applicants should receive welfare benefits
- NYNEX has a system that helps unskilled workers diagnose customer phone problems
- Arco and Texaco use neural networks to help pinpoint oil and gas deposits deep below the earth's surface

[1] Vic Reis, a former Director of the Advanced Research Projects Agency (ARPA), was quoted as saying that the DART system, used in deployment planning of operation Desert Shield, justified ARPA's entire investment in AI technology [Grosz & Davis, 1994, note 4, page 20].

- The Internal Revenue Service is testing software designed to read tax returns and spot fraud
- Spiegel uses neural networks to determine who on a vast mailing list are the most likely buyers of its products
- American Airlines has an expert system that schedules the routine maintenance of its airplanes

High-performance programs like these are all very useful; they are important and worthy projects for AI, and undoubtedly they have been excellent investments. But do they move AI closer to its original, main goal of developing a general, intelligent system? I think not. The components and knowledge needed for extreme specialization are not necessarily those that will be needed for general intelligence. Some medical diagnosis programs, for example, have expert medical knowledge comparable to that of human physicians who have had years of training and practice [Miller *et al.*, 1982]. Yet, these doctors were already far more intelligent, generally, *before* attending medical school than are the best of our AI systems. They had the ability then to acquire the knowledge that they would need in their speciality—an ability AI programs do not yet have.

3 Ever More Refined Techniques

In parallel with the move toward performance programs, AI researchers working on techniques (rather than on specific applications) began to sharpen these techniques much beyond what I think are required by general, intelligent systems. I'll give some examples. Let's look first at automatic planning. It is clear that a general, intelligent system will need to be able to plan its actions. An extensive spectrum of work on automatic planning has been done by AI researchers. Early work was done by [Newell, Shaw & Simon, 1960, McCarthy & Hayes, 1969, Green, 1969, Fikes & Nilsson, 1971]. These early programs and ideas were clearly deficient in many respects. While working on one part of a problem, they sometimes undid an already solved part; they had to do too much work to verify that their actions left most of their surroundings unchanged; and they made the unrealistic assumption that their worlds remained frozen while they made their plans. Some of the deficiencies were ameliorated by subsequent research [Waldinger, 1977, Sacerdoti 1977, Tate, 1977, Sussman, 1975]. Recent work by [Wilkins, 1988, Currie & Tate, 1991, Chapman, 1987] led to quite complex and useful planning and scheduling systems. Somewhere along this spectrum, however, we began to develop specialized planning capabilities that I think are not required of a general, intelligent system. After all, even the smartest human cannot (without the aid of special tools) plan NASA missions or lay out a factory schedule, but automatic planning programs can now do those things [Deale *et al.*, 1994, Fox, 1984].

Other examples of refinement occur in the research area dealing with reasoning under uncertainty. Elaborate probabilistic reasoning schemes have been developed, and perhaps some of these computational processes are needed by intelligent systems. But what I think is *not* needed (to give just one example) is a dynamic programming system for calculating paths of minimal expected costs between states in a Markov decision problem—and yet, some high quality AI research is devoted to that and similar problems (which do arise in special settings).

More examples exist in several other branches of AI, including automated theorem proving, intelligent database retrieval, design automation, intelligent control, and program verification and synthesis. AI research has been focussed on getting systems to solve problems beyond what humans can ordinarily do without elaborate and specialized knowledge, training, and tools.

Of course a program must be equipped with the skills and knowledge that it truly needs in its area of application. *What I am arguing for here is that those skills and knowledge bases be regarded as <u>tools</u> separate from the intelligent programs that use them.* It is time to begin to distinguish between general, intelligent programs and the special performance systems, that is, tools, that they use. AI has for many years now been working mainly on the tools—expert systems and highly refined techniques. Building the tools is important—no question. But working on the tools alone does not move us closer to AI's original goal, producing intelligent programs that are able to use tools. Such general programs do not need to have the skills and knowledge within them as refined and detailed as that in the tools they use. Instead they need to be able to find out about what knowledge and tools are available to match the problems they face and to learn how to use them.

Curiously, this view that "general intelligence" needs to be regarded as something separate from "specialist intelligence" was mentioned in the same paper that helped to move the field toward concentrating on special intelligence. [Feigenbaum, *et al.*, 1971, page 187] said:

> "The 'big switch' hypothesis holds that generality in problem solving is achieved by arraying specialists at the terminals of a big switch. The big switch is moved from specialist to specialist as the problem solver switches its attention from one problem area to another. [. . . The kinds of problem-solving processes, if any, which are involved in 'setting the switch' (selecting a specialist) is a topic that obviously deserves detailed examination in another paper.]"

Unfortunately, work on "setting the switch" (if, indeed, that's what is involved in general intelligence) has been somewhat delayed. The same authors, however, did go on to give some recommendations, which seem to me to be still quite valid:

> "The appropriate place for an attack on the problem of generality may be at the meta-levels of learning, knowledge transformation, and representation, not at the level of performance programs. Perhaps for the designer of intelligent systems what is most significant about human general problem-solving behavior is the ability to learn specialties as needed—to learn expertness in problem areas by learning problem-specific heuristics, by acquiring problem-specific information, and by transforming general knowledge and general processes into specialized forms."

4 Some Reasons for the Diversions

There are several reasons why AI has concentrated on tool-building. First, the problem of building general, intelligent systems is very hard. Some have argued that we haven't made much progress on this problem in the last forty years. Perhaps we have another forty years ahead of us before significant results will be achieved. It is natural for researchers to want to achieve specific results during their research lifetimes and to become frustrated when progress is slow and uneven.

Second, sponsors of AI research have encouraged (and have often insisted on) specialized systems. After years of supporting general AI they understandably want a return on their investment. The problem is that the people who have the dollars usually have *specific* problems they want solved. The dollars exist in niches, and those niches call forth programs to fill them.

Third, many of the systems and tools that AI has been working on have their own intrinsic, captivating interest. A community of researchers develops, and momentum carries the pursuit of techniques into areas perhaps not relevant to a general intelligent agent. Exciting whirlpools always divert some people from the main stream. Some of the work in theoretical AI (for example some non-monotonic reasoning research), may be of this character.

Fourth, some AI leaders have argued quite persuasively that the best route toward AI's main goal lies through the development of performance systems. Edward Feigenbaum, for example, has often said that he learns the most when he throws AI techniques against the wall of hard problems to see where they break. It is true that many of the early AI methods did not scale up well and that confronting hard problems in science, engineering, and medicine made our methods more robust. I believe that, but I think the hard-problem approach has now reached the point of diminishing returns. Throwing our techniques against yet more (special) hard walls is now not so likely to improve those techniques further nor lead to new and generally useful ones. (It will, of course, result in solving additional specific problems.)

Fifth, university computer science departments have increasingly shifted from "understanding-driven" to "need-driven" research. This shift has been encouraged by a number of factors, not the least of which is the alleged "new compact" between society and science in which science is supposed to be directed more toward "national needs." Also, most university computer science departments are in engineering colleges—which often have a very practical outlook. Computer science itself now seems to be more concerned with faster algorithms, better graphics, bigger databases, wider networks, and speedier chips than it is with the basic problems of artificial intelligence (or even with the basic problems of computer science). AI faculty, competing in these departments for recognition and tenure, want to be perceived as working on "real problems"—not chasing ill-defined and far-off willothewisps. The importance that is attached to being able to "evaluate" research results leads inevitably to working on projects with clear evaluation criteria, and, typically, it's easier to evaluate systems that do specific things than it is to evaluate systems whose tasks are more general.

Finally, the arguments of those who say "it can't be done" might have had some effect. People who know insufficient computer science, but yet consider themselves qualified to pronounce on what is possible and what is not, have been free with their opinions [Penrose, 1989, Penrose, 1994, Dreyfus & Dreyfus, 1985, Searle, 1980]. Out of these pronouncements has come the distinction between "strong AI" and "weak AI." In the words of [Searle, 1980]:

> "According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really *is* a mind ..."

These critics acknowledge the successes of expert systems and other AI applications—claiming them to be examples of weak AI. Strong AI is declared to be impossible (with the overtone that we shouldn't be wanting to achieve it anyway), and weak AI is embraced as appropriate, doable, and socially acceptable. Many AI researchers are willing to settle for the goals of weak AI. The weak AI agenda is also consistent with much of the rest of present-day computer science which increasingly sees its mission as providing computational tools. Paradoxically, since strong AI implies the ability to function effectively in a variety of environments, it will most probably depend on AI's so-called "weak methods," namely ones that are generally useful and unspecialized. The strong and specialized methods, on the other hand, are used by the niche systems associated with weak AI.

5 Habile Systems

Perhaps a good adjective to describe the general, intelligent systems I have in mind is "habile," which means "having general skill." What are some of the properties of a habile system? Here is my list:

- Commonsense knowledge and commonsense reasoning abilities. Wide-ranging knowledge and inference capabilities are necessary in order that the system be "generally intelligent." Unlike expert systems, we would expect habile systems (using appropriate tools) to perform reasonably, if not expertly, in a variety of situations. Of course, what we gain in breadth, we will probably have to give up in depth. This trade-off (applied to programming languages) was nicely expressed by [Stroustrup, 1994, page 201][2]:

> For every single specific question, you can construct a language or system that is a better answer than C++. C++'s strength comes from being a good answer to many questions rather than being the best answer to one specific question ... Thus, the most a general-purpose language can hope for is to be "everybody's second choice."

The fact that a habile system will be a "jack of all trades" and master of none does not diminish the value of such a system. It does make it more difficult to find funding sources for research on habile systems, however.

- Access abilities. These include whatever is needed for an agent to get information about the environment in which it operates and to affect that environment in appropriate ways. For robots, the access abilities might include perceptual processing of visual images and a suite of effectors. For software agents, the access abilities might include ability to read e-mail messages and to access databases and computer networks. The access abilities of habile systems that must deal with other agents will include facilities for receiving, understanding, and generating communications. Interaction with humans will require natural language understanding and generation programs.

- Autonomy and continuous existence. Habile systems will be "agents" that have built-in high level goals (much like the "drives") of animals. They will have an architecture that mediates between reasoning (using their commonsense knowledge) and reflexive reactions to urgent situations.

[2] I thank Ron Kohavi for bringing this citation to my attention

- Ability to learn. Agents having a continuous existence can learn from experience. New demands will create new applications, and agents must be able to learn how to solve new problems. All of the learning methods of AI will be needed here. Habile agents must be "informable" [Genesereth, 1989]. Humans will want to give advice to them that varies in precision from detailed instructions to vague hints. Since so much human knowledge exists in written form, we will want our agents to be able to get appropriate information from documents. These abilities also presuppose natural language skills.

There is reason now to think that AI will soon be placing much more emphasis on the development of habile systems. I explain why in the next section.

6 Some Forces Pushing us Toward Habile Systems

Not all of the forces affecting AI are in the direction of niche systems. There have always been good reasons to build habile systems, but now I think there are some new needs—just now becoming more pressing. These new forces arise from the rapid development of the information "superhighway," multi-media for entertainment, education, and simulation, and the growing demand for more flexible robots. I'll make a few comments about each of these influences.

6.1 The Information Superhighway

The exploding access to databases, programs, media, and other information provided by computer networks will create a huge demand for programs that can aid the consumers and producers of this information. In the words of a *Wall Street Journal* article about electronic agents [Hill, 1994]: "The bigger the network and the more services on it, the greater the potential power of agents." All kinds of special "softbot" agents (sometimes called "spiders" when they inhabit the World Wide Web) have been proposed—personal assistants, database browsers, e-mail handlers, purchasing agents, and so forth. Several people are working on prototypes that aim toward such agents [Etzioni & Weld, 1994, Ball & Ling, 1993, Maes, 1994]. Even though a variety of very specialized "niche" agents will be built to service these demands, the casual user will want a general-purpose personal assistant to act as an intermediary between him or her and all the specialized agents and the rest of the World Wide Web. Such a personal assistant should have many of the features of habile agents: general common-sense knowledge, wide-ranging natural language ability, and continuous existence. As a step in that direction, the architecture being explored for CommerceNet uses an agent called a "facilitator" that has quite general capabilities [Genesereth, 1992]. Demand for habile personal assistants will be unceasing and growing as services available on the Internet continue to expand.

6.2 Entertainment, Education, and Simulation

Interactive, multi-media video art and entertainment require characters that are "believable" in their emotions and actions [Bates, 1994]. The human participants in these interactions want characters that act and think very much like humans do. So long as such characters are perceived to be "simply mechanical" and easily predictable, there will be competitive pressure to do better. Similar needs exist as we develop more sophisticated computer educational systems. On-the-job training in an environment with customers, co-workers, and even adversaries is an important style of education for many occupations. To provide real environments and their inhabitants for purposes of training is expensive and perhaps dangerous, and therefore simulations and simulated inhabitants are being used increasingly. This need for realistic simulated agents exerts continuing pressure to develop ones with wide-ranging, human-like capabilities.

6.3 The Requirement for More Flexible Robots

A recent article in *The New York Times* [Holusha, 1994] said " ... sales are booming for robots, which are cheaper, stronger, faster and smarter than their predecessors." One reason for the sales increase is that robots are gradually becoming more flexible—in action and in perception. I expect that there will be increasing demand for flexibile mobile robots in manufacturing, construction, and in service industries. Some possible applications include delivery vehicles, carpenters' assistants, in-orbit space-station constructors, robots that work in hazardous environments, household robots, sentry robots, and underwater robots. Although there will be many niche systems (just as there are in the biological world), cost considerations will favor habile robot architectures that can be applied to a variety of different tasks. I think the main challenge in developing flexible robots (in addition to providing those features of habile systems already mentioned) is to integrate perception, reasoning, and action in an architecture designed especially with such integration in mind. There are several such general-purpose robot architectures being explored—including one I am currently working on [Benson & Nilsson, 1995].

These factors will combine with those that have existed for quite some time. To name just a few of the latter, there is still a need for more versatile natural language processing systems, for more robust expert systems, and for computational models of human and animal intelligence.

6.4 Natural Language Processing

Several important applications require more general and competent natural language (NL) abilities. These include systems for dictation, automated voice services using the telephone system, translation between different natural languages, interfaces to certain applications programs for casual users, agents for filtering voice mail, electronic mail and other messages, automatic abstracting, optical character recognition, and information retrieval programs. Both NL understanding and generation are required. The demand for these abilities will exert an unceasing and growing pressure to create the knowledge bases and programs required for general, wide-domain (we might say habile) NL systems. The desire for better NL processing systems will not disappear, even though the technical problems involved are very difficult and progress on solving them is slow.

6.5 The Brittleness of Expert Systems

AI applications specialists acknowledge that the main defect of most expert systems is that they are very brittle. Within their specialized areas these systems contain much more expertise than is needed by a general, intelligent system, but once off the high mesa of their specialized knowledge, they fall to the flat plain of complete ignorance. Worse, they don't even know when they are off their mesa. These expert systems need what John

McCarthy [McCarthy, 1990] calls *commonsense*—without it they are *idiot savants*. There is growing insistence that these programs be less brittle. Making their "knowledge cliff" less steep means extending their competence at least to semi-hability in the areas surrounding their field of expertise. Several projects have as their goal making expert systems more flexible. One that is attempting to do so by giving such systems more general knowledge surrounding their specialized area is the "How Things Work" project at Stanford [Iwasaki & Low, 1993], which is producing a knowledge base of general physical and electro-mechanical laws that would be useful to a wide variety of different expert systems.

6.6 The Scientific Interest to Understand How the Brain Works

One of the motivations for AI research all along has been to gain insight into mental processes. The efforts of neuro-scientists, psychologists, ethologists, cognitive scientists, and AI researchers are all contributing their own results and points of view to the integrated, multi-level picture appropriate for this most difficult scientific quest. Just as knowledge of transistor physics alone is not adequate for an understanding of the computer, so also neuro-science must be combined with higher level concepts, such as those being investigated by AI researchers, in order to fill out our picture of mental functioning. The steadily accumulating body of knowledge about neural processes will add to the urgency of understanding how the higher level processes combine with the others to form a *mind*.

Even within AI, there are several approaches being followed by people whose main interest is the scientific study of mental functioning. There is what might be called the *animat* approach [Wilson, 1991], which holds that AI should concern itself first with building simple, insect-like, artifacts and gradually work its way up the evolutionary scale [Brooks, 1991]. Whatever one might believe about the long-range potential for this work, it is contributing significantly to our understanding of building autonomous systems that must function in a variety of complex, real environments and thus reinforces the trend toward habile systems. Such work also provides a base that arguably may be necessary to support higher cognitive functions.

At a distinctly higher level is the work on Soar [Laird, *et al.*, 1987], an architecture for general intelligence, which is aimed at modeling various cognitive and learning abilities of humans. It is interesting to note that even with these general goals, the Soar architecture can be specialized to function as an expert system for configuring computer systems as well as for a number of other specialized tasks [Rosenbloom, *et al.*, 1985, Pearson, *et al.*, 1993]. At a similarly high level is an attempt to duplicate in computer agents some of the stages of Piagetian learning [Drescher, 1991].

All of these efforts are directed at understanding the common mechanisms in naturally occurring, biological individuals. The scientific quest to understand them will never cease, and will thus always exert a pull on the development of habile systems.

In summary, I think all of these factors, old and new, suggest the strong possibility that AI will once again direct a substantial portion of its research energies toward the development of general, intelligent systems.

7 Some Important Research Projects

In addition to the research efforts already mentioned, several others are quite relevant to habile systems. I'll remark on just three of the ones I know the most about.

One is the CYC project led by Douglas Lenat [Guha & Lenat, 1990]. It has as its goal the building of a commonsense knowledge base containing millions of facts and their interrelationships. It is striving to encompass the knowledge that is seldom written down—knowledge, for example, that the reader of an encyclopedia is assumed to possess *before* reading the encyclopedia and which, indeed, is required in order to understand what he reads. It seems clear to many of us that this kind of knowledge, in some form, will be required by habile systems—in particular by any systems that are expected to use more-or-less unconstrained natural language. I think projects of this sort are very important to AI's long-range goals, and I agree with Marvin Minsky who said "I find it heartbreaking [that] there still are not a dozen other such projects [like CYC] in the world, ... " [Riecken & Minsky, 1994].

Another project of general importance is the attempt to build an "interlingua" for knowledge representation such as the *Knowledge Interchange Format (KIF)* [Genesereth & Fikes, *et al.*, 1992]. For efficiency, niche applications will want their specialized knowledge in customized formats, but some of this knowledge, at least, will be the same as the knowledge needed by other niche systems. To permit "knowledge-sharing" among different systems, knowledge must be translatable from one system's format into another's, and a common inter-lingua, such as KIF, greatly facilitates the translation process. Although, as some argue, it may be too early to codify standards for such an inter-lingua, it is not too early to begin to consider the research issues involved.

Agents that are part of communities of agents will need knowledge of each other's "cognitive structure" and how to affect the beliefs and goals in such structures through communication. Yoav Shoham's *Agent-Oriented Programming (AOP)* formalism [Shoham, 1993] is one attempt to facilitate the construction of communicating agents.

8 Summary and Conclusions

AI's founding fathers, Marvin Minsky, John McCarthy, and Allen Newell, always kept their eyes on the prize—even though they pursued different paths toward it. McCarthy's work on commonsense reasoning [McCarthy, 1958, McCarthy, 1986] has been directly aimed at general, intelligent systems. The same can be said for Minsky's work on structuring knowledge in "frames" [Minsky, 1975] and on his "society of mind" [Minsky, 1986]. Newell's work on production systems and Soar [Newell, 1990] focussed on the same prize. Now it appears that there are strong and insistent reasons for many others also to resume work on AI's original goal of building systems with human-like capabilities. Even though this prize may still be distant, the ultimate benefits of practical, re-targetable, tool-using systems will more than repay the long-term investments.

I think there is no reason to be discouraged by the present pressures to concentrate on "mission-specific" research. There are now people whose very missions *require* the development of habile systems, and much basic research needs to be done before their needs can be satisfied. Several different architectures need to be explored. There are still many unresolved questions, such as "Is general intelligence dependent on just a few weak methods (some still to be discovered) plus lots and lots of commonsense knowledge?" or "Does it depend on perhaps hundreds or thousands of specialized mini-competences in a heterarchical society of mind?" No one knows the answers to questions like these, and only experiments and trials will provide these answers. We need, as Minsky recommends, ten more CYC projects. We also need support for young investigators and post-docs, graduate fellowships, individual investigator-initiated grant programs, and research equipment and facilities.

With the right sort of research support, AI will now proceed along two parallel paths—specialized systems and habile systems. Niche systems will continue to be developed because there are so many niches where computation is cost-effective. Newell foresaw this path when he charmingly predicted [Newell, 1992] that there would someday be "brakes that know how to stop on wet pavement, instruments that can converse with their users, bridges that watch out for the safety of those who cross them, streetlights that care about those who stand under them—who know the way, so no one need get lost, [and] little boxes that make out your income tax for you." He might also have mentioned vacuum cleaners that know how to vacuum rooms, garden hoses that know how to unroll themselves when needed and roll themselves back up for storage, automobiles that know where you want to go and drive you there, and thousands of other fanciful and economically important agents. Society's real world and its invented virtual worlds together will have even more niches for computational systems than the physical world does for biological ones. AI and computer science have already set about trying to fill some of these niches, and that is a worthy, if never-ending, pursuit. But the biggest prize, I think, is for the creation of an artificial intelligence as flexible as the biological ones that will win it. Ignore the naysayers; go for it!

Acknowledgements

This paper is based on invited talks given at the Iberamia '94 Conference in Caracas, Venezuela (October 25, 1994) and at the Department of Computer Science, University of Washington (May 4, 1995). I thank my hosts at those venues, Professors José Ramirez and Hector Geffner; and Professor Steve Hanks. I received many valuable comments (if not complete agreement) from Peter Hart, Barbara Hayes-Roth, Andrew Kosoresow, and Ron Kohavi.

Biographical Sketch

Nils J. Nilsson, Kumagai Professor of Engineering in the Department of Computer Science at Stanford University, received his PhD degree in electrical engineering from Stanford in 1958. He spent twenty-three years at the Artificial Intelligence Center of SRI International working on statistical and neural-network approaches to pattern recognition, co-inventing the A* heuristic search algorithm and the STRIPS automatic planning system, directing work on the integrated mobile robot, SHAKEY, and collaborating in the development of the PROSPECTOR expert system. He has published four textbooks on artificial intelligence. Professor Nilsson returned to Stanford in 1985 as the Chairman of the Department of Computer Science, a position he held until August 1990. Besides teaching courses on artificial intelligence and on machine learning, he is conducting research on flexible robots that are able to react to dynamic worlds, plan courses of action, and learn from experience. Professor Nilsson served on the editorial board of the journal Artificial Intelligence, was an Area Editor for the Journal of the Association for Computing Machinery, and is currently on the editorial board of the Journal of Artificial Intelligence Research. He is a past-president and Fellow of the American Association for Artificial Intelligence and is also a Fellow of the American Association for the Advancement of Science. He is a founding director of Morgan Kaufmann Publishers, Inc. In 1993, he was elected a foreign member of the Royal Swedish Academy of Engineering Sciences.

References

[Ball & Ling, 1993] Ball, J. E., and Ling, D., "Natural Language Processing for a Conversational Assistant," Microsoft Research Technical Report MSR-TR-93-13, October, 1993.

[Bates, 1994] Bates, J., "The Role of Emotion in Believable Agents," *Communications of the ACM*, 37(7): 122-125, July, 1994.

[Benson & Nilsson, 1995] Benson, S., and Nilsson, N., "Reacting, Planning and Learning in an Autonomous Agent," in *Machine Intelligence 14*, (eds. K. Furukawa, D. Michie, and S. Muggleton), Oxford: the Clarendon Press, 1995.

[Brooks, 1991] Brooks, R. A., "Intelligence Without Representation," *Artificial Intelligence*, 47(1-3), pp. 139-159, January, 1991.

[*Business Week*, 1992] "Smart Programs Go to Work," *Business Week*, pp. 97ff, March 2, 1992.

[Chapman, 1987] Chapman, D., "Planning for Conjunctive Goals," *Artificial Intelligence*, 32: 333-377, 1987.

[Currie & Tate, 1991] Currie, K. W., and Tate, A., "O-Plan: The Open Planning Architecture," *Artificial Intelligence*, 51(1), 1991.

[Deale *et al.*, 1994] Deale, M., *et al.*, "The Space Shuttle Ground Processing Scheduling System," in Zweben, M, and Fox, M., *Intelligent Scheduling*, San Francisco: Morgan Kaufmann, 1994

[Drescher, 1991] Drescher, G., *Made-Up Minds: A Constructivist Approach to Artificial Intelligence*, Cambridge, MA: MIT Press, 1991.

[Dreyfus & Dreyfus, 1985] Dreyfus, H., and Dreyfus, S., *Mind Over Machine*, New York: MacMillan/The Free Press, 1985.

[Etzioni & Weld, 1994] Etzioni, O., and Weld, D., "A Softbot-Based Interface to the Internet," *Communications of the ACM*, 37(7): 72-76, July, 1994.

[Feigenbaum, *et al.*, 1971] Feigenbaum, E., *et al.*, "On Generality and Problem Solving: A Case Study Using the DENDRAL Program," in Meltzer, B., and Michie, D., (eds.), *Machine Intelligence 6*, pp. 165-190, Edinburgh: Edinburgh University Press, 1971.

[Fikes & Nilsson, 1971] Fikes, R. E. and Nilsson, N. J., "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving," *Artificial Intelligence*, 2(3–4): 189–208, 1971.

[Fox, 1984] Fox, M., and Smith, S., "ISIS—A Knowledge-Based System for Factory Scheduling," *Expert Systems*, 1(1), 1984.

[Genesereth, 1989] Genesereth, M., "A Proposal for Research on Informable Agents," Logic-89-9, Stanford University Computer Science Logic Group Report, June 1989.

[Genesereth, 1992] Genesereth, M., "An Agent-Based Approach to Software Interoperability," in *Proceedings of the DARPA Software Technology Conference*, 1992.

[Genesereth & Fikes, *et al.*, 1992] Genesereth, M., and Fikes, R., *et al.*, *Knowledge Interchange Format Version 3 Reference Manual*, Logic-92-1, Stanford University Computer Science Logic Group Report, 1992.

[Green, 1969] Green, C., "Application of Theorem Proving to Problem Solving," *Proceedings of the First International Joint Conference on Artificial Intelligence*, Washington, DC, 1969.

[Grosz & Davis, 1994] Grosz, B., and Davis, R. (eds.), "A Report to ARPA on Twenty-First Century Intelligent Systems," *AI Magazine*, pp. 10-20, Fall 1994.

[Guha & Lenat, 1990] Guha, R., and Lenat, D., "Cyc: A Mid-Term Report," *The AI Magazine*, Vol. 11, no. 3, pp. 32-59, Fall, 1990.

[Hill, 1994] Hill, G., "Cyber Servants," *The Wall Street Journal*, page 1, September 27, 1994.

[Holusha, 1994] Holusha, J., "Industrial Robots Make the Grade," *The New York Times*, September 7, 1994.

[Iwasaki & Low, 1993] Iwasaki, Y., and Low, C. M., "Model Generation and Simulation of Device Behavior with Continuous and Discrete Change," *Intelligent Systems Engineering*, 1(2), 1993.

[Laird, *et al.*, 1987] Laird, J., *et al.*, "Soar: An Architecture for General Intelligence," *Artificial Intelligence*, 33:1-64, 1987.

[McCarthy & Hayes, 1969] McCarthy, J., and Hayes, P. J., "Some Philosophical Problems from the Standpoint of Artificial Intelligence," in Meltzer, B., and Michie, D. (Eds.), *Machine Intelligence 4*, pp. 463-502, Edinburgh: Edinburgh University Press, 1969.

[McCarthy, 1990] McCarthy, J., "Some Expert Systems Need Commonsense," in Lifschitz, V. (ed.), *Formalizing Common Sense: Papers by John McCarthy*, pp. 189-197, Norwood, NJ: Ablex, 1990.

[McCarthy, 1958] McCarthy, J., "Programs with Common Sense," *Mechanisation of Thought Processes, Proceedings of the Symposium of the National Physics Laboratory*, Vol. I, pp. 77–84, London, UK: Her Majesty's Stationary Office, 1958.

[McCarthy, 1986] McCarthy, J., "Applications of Circumscription to Formalizing Commonsense Knowledge," *Artificial Intelligence*, 28(1): 89–116, 1986.

[Maes, 1994] Maes, P., "Agents that Reduce Work and Information Overload," *Communications of the ACM*, 37(7): 31-40, July, 1994.

[Miller *et al.*, 1982] Miller, R. *et al.*, "INTERNIST-1: An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine," *New England Journal of Medicine*, 307:468-476, 1982.

[Minsky, 1975] Minsky, M., "A Framework for Representing Knowledge," in Winston, P. H. (ed.), *The Psychology of Computer Vision*, pp. 211–277, New York: McGraw-Hill, 1975.

[Minsky, 1986] Minsky, M., *The Society of Mind*, New York: Simon and Schuster, 1986.

[Newell, 1990] Newell, A., *Unified Theories of Cognition*, Cambridge, MA: Harvard University Press, 1990.

[Newell, 1992] Newell, A., "Fairy Tales," *The AI Magazine*, vol. 13, no. 4, pp. 46-48, Winter, 1992.

[Newell, Shaw & Simon, 1960] Newell, A., Shaw, J. C., and Simon, H. A., "Report on a General Problem-Solving Program for a Computer," *Information Processing: Proc. of the Int. Conf. on Information Processing*, UNESCO, Paris, pp. 256-264, 1960.

[Newell & Simon, 1976] Newell, A. and Simon, H. A., "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the Association for Computing Machinery*, 19(3): 113–126, 1976.

[Pearson, *et al.*, 1993] Pearson, D., *et al.*, "Intelligent Multi-Level Control in a Highly Reactive Domain," in Groen, F., Hirose, S., and Thorpe, C., *Intelligent Autonomous Systems, IAS-3*, pp. 449-458, Washington: IOS Press, 1993.

[Penrose, 1989] Penrose, R., *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford: The Oxford University Press, 1989.

[Penrose, 1994] Penrose, R., *Shadows of the Mind: Search for the Missing Science of Consciousness*, Oxford: The Oxford University Press, 1994.

[Riecken & Minsky, 1994] Riecken, D., and Minsky, M., "A Conversation wilth Marvin Minsky About Agents," *Communications of the ACM*, 37(7): 23-29, July, 1994.

[Rosenbloom, *et al.*, 1985] Rosenbloom, P., "R1-Soar: An Experiment in Knowledge-Intensive Programming in a Problem-Solving Architecture," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7:561-569, 1985.

[Sacerdoti 1977] Sacerdoti, E. D., *A Structure for Plans and Behavior*, New York: Elsevier, 1977.

[Searle, 1980] Searle, J., "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, vol 3, 1980.

[Shoham, 1993] Shoham, Y., "Agent Oriented Programming," *Artificial Intelligence*, **60**, 1:51-92, 1993.

[Stroustrup, 1994] Stroustrup, B., *The Design and Evolution of C++*, Reading, MA: Addison-Wesley, 1994.

[Sussman, 1975] Sussman, G. J., *A Computer Model of Skill Acquisition*, New York: American Elsevier, 1975.

[Tate, 1977] Tate, A., "Generating Project Networks," *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pp. 888-893, San Francisco, CA: Morgan Kaufmann, 1977.

[Turing, 1950] Turing, A., "Computing Machinery and Intelligence," *Mind*, 59:433-460, October, 1950.

[Waldinger, 1977] Waldinger, R. J., "Achieving Several Goals Simultaneously," in Elcock, E. and Michie, D. (eds.), *Machine Intelligence 8: Machine Representations of Knowledge*, pp. 94–136, Chichester, UK: Ellis Horwood, 1977.

[Wilkins, 1988] Wilkins, D. E., *Practical Planning: Extending the Classical AI Planning Paradigm*, San Francisco: Morgan Kaufmann, 1988.

[Wilson, 1991] Wilson, S., "The Animat Path to AI," in Meyer, J. A., and Wilson, S. (eds.), *From Animals to Animats; Proceedings of the First International Conference on the Simulation of Adaptive Behavior*, The MIT Press/Bradford Books, Cambridge, MA, 1991.