AI Magazine Volume 10 Number 1 (1989) (© AAAI)

Part One

Expert Systems: How Far Can They Go?

Randall Davis, Editor

A panel session at the 1985 International Joint Conference on artificial intelligence in Los Angeles dealt with the subject of knowledge-based systems; the session was entitled "Expert Systems: How Far Can They Go?" The panelists included Randall Davis (Massachusetts Institute of Technology); Stuart Dreyfus (University of California at Berkeley); Brian Smith (Xerox Palo Alto Research Center); and Terry Winograd (Stanford University), chairman. The article begins with Winograd's original charge to the panel, followed by lightly edited transcripts of the panel's remarks. Part 1 includes presentations from Winograd and Dreyfus. Part 2, which will appear in the Summer 1989 issue, includes presentations from Smith and Davis and concludes with the panel discussion. Although three years have passed since this session took place, the issues raised and the points discussed are no less relevant today.

Charge to the Panel

We are in the midst of a great wave of enthusiasm about the potential for expert systems in every area of human life and work. There is no agreement, however, as to just how much they can do, and where they will run into fundamental limits. The intent of this panel is to present and discuss some basic questions as to what expert systems can really be expected to do:

- What is the nature of the problem domains in which expert systems are likely to succeed and those in which they will not? Are there domains in which their use might be dangerous?
- How will their performance compare with that of human experts in the domain? Are there different facets of expertise that are not amenable to programming? How can human and machine expertise best be combined?
- To what extent can we count on rule-based systems for "flexibility" in dealing with unexpected situations? How reliable will such systems be in cases where the programmers (or knowledge engineers) did not anticipate significant possibilities?
- How can a knowledge base be subjected to standards of accountability? Who is responsible for what an expert system contains and what it does?

Winograd's Presentation

I ran into some people today who asked "What is this panel really going to be about? What are you going to do? I hope you're not going to just raise the old religious issues."

I got to thinking about some similarities between what gets said about expert systems and what gets said about religion. On the one side you have the atheists who say that it's impossible, doesn't exist, and can't be done. In the middle there are a variety of agnostics and partial believers who aren't sure, or are interested and want to discuss the questions but don't have a firm view. Then, on the other side you have the TV evangelists who say, "Send in your money and you shall receive miracles."

I think there's another similarity sometimes, which is that the discussions tend to get couched in terms of what it's going to be like in the sweet bye and bye. That is: "Can you imagine that at any day in the future it might be possible that a computer could do this or couldn't do that?" It is easy to lose touch with the reality of what people are actually doing—with what kind of sense it makes now and what kind of sense it is likely to make in the foreseeable future.

So, my response to those comments was "Yes, indeed, I don't want to make this a religious discussion. I think that it will be more useful to everybody to focus on a more specific set of questions that are more practical." In this panel, we will take on questions about what expert systems really are doing and can do. By that I don't mean focusing on just the current systems. I don't think it's particularly useful to pick apart the weaknesses of a particular knowledge-based system or a particular rule formalism. We need to take a broader view, covering the body of techniques, of which particular systems are examples, and anticipating to some degree where they're going to go.

Looking at where things stand now, it is interesting to see how they are different from what they were two to three years ago, which was again different from a few years before that. There has been a huge wave of enthusiasm for expert systems and in the last couple of conferences like this one we have seen the beginnings of a reaction—an attempt to go beyond the enthusiasm and to better understand what they really are. If you look at AI people—not when they're writing their marketing plan, or their proposal to the Pentagon, but when they're really seriously discussing with their colleagues—you see new signs of caution. People are really asking themselves "What have we promised? What does the public believe about AI and expert systems? What's really going to get delivered? Is there going to be an 'AI winter'—a dark age? Are we going to see the result of an oversell?"

There is also a fair amount of recognition within the field that it isn't just a matter of shoveling lots more knowledge into the systems we have. We have developed particular technologies that are useful within certain ranges but there is no sense that we just need to hire lots of people to put in encyclopedias and everything will be solved. There is still a lot of uncertainty about what can and can't be done. This uncertainty is reflected to some extent within the profession, in meetings like this, and is certainly reflected to a larger extent in the public.

If you read the public press on what computers are doing and will be doing, you find wildly discrepant themes—from "They can't really do anything" to "They'll be wonderful little friends you'll tuck under your arm and they'll solve all your problems." This confusion has become particularly relevant as we see large organizations sponsor research with emphasis on AI components and expert systems (for example in projects like the Strategic Computing Initiative and the Strategic Defense Initiative). People who have a lot of money to put into big projects are putting high expectations on the future results and don't really know what it's going to come to.

So what? Why should we worry? Well, first, there's been worry about a backlash. If you oversell, then you pay the consequences. I think that's an interesting point, but not the one I want to focus on. I think there's a more important problem with misplaced confidence.

Once systems are actually proposed and built and they are being used, people don't have a sense of what they can count on and what they can't count on. This is true to some extent for the researchers, and to an even larger extent for the public. That is, they don't recognize the restricted range within which systems work. They project onto them broader capabilities, broader abilities than they really have.

There is also a mystique about the ability of AI programs to be flexible in unanticipated situations. The myth is that when an ordinary program hits an unanticipated situation, it will do a core dump, but an AI program will somehow put together some rules that you hadn't thought of combining and will come up with the right answer. This misplaced optimism has serious effects in attempts to put AI systems into critical areas of any kind and especially in areas like weapons systems.

There is also a problem of overexpectation that obscures responsibility. Once again, there's a kind of mystique—an objectification—that if the expert system says so, it must be right. "It's in the 'knowledge base'." If the knowledge base were called the "opinion base" because it's full of opinions that various people put into it, you might have a somewhat different view, especially in areas like medicine where there really is a significant difference of opinion about lots of things.

The essential issues that need to be raised are not peculiar to expert systems. All of AI has to deal with many of these same things. My own background is from natural language and the way I got into thinking about these issues was by trying to figure out why it was so hard to get a computer to understand ordinary English. The problems that come up there are the same as the ones that come up for expert systems.

I've written a book on these kinds of topics.[1] We discuss what we call the "rationalistic tradition"—a broad tradition of thought that is directly reflected in computer science and also in management theory and a lot of other fields (which suffer from many of these same problems). The problems have to do with basic issues of representation and problem solving, what those are, and how they relate to human situations in which computer systems are used. The basic issue is that in creating a representation for use in a program—whether it be for diagnosis of medical problems or for prospecting for oil—we create an artificial formal domain. This domain embodies a collection of distinctions—atoms, tokens, whatever structures are in the system—with the expectation that this artificial domain will correspond in a reasonable and substantial way with the world we care about. In creating a representation we are carving out from the reality that we experience a collection of formalized structures. We thereby create a blindness to everything that is not expressible within those structures.

Now, "blindness" has a kind of negative connotation that I want to dispel. I don't think it's avoidable and I don't think it's something bad. It's just inevitable. Whenever you create a precise structure within an open-ended context, that structure by necessity focuses on certain things and thereby drops out others. What we are able to do with expert systems, with the kind of technology we have, depends on creating a very narrow focus and structure.

By that I don't mean tackling very narrow problems. We operate by reducing any domain of concern to a collection of elements that can be related in reasonably definable ways—among which you can make clear distinctions. If you couldn't do that, people would not understand what was in the knowledge base or be able to add knowledge to it. You must articulate an understanding that is derived from an intuitively understood or partly articulated domain. This works very well in areas where that articulation captures most of what's going on—in narrowly technical kinds of areas where you can safely ignore all but a few critical factors and their interactions. As you move into more open-ended things (which includes almost anything dealing with human behavior, though it may not

In creating a representation, we are carving out from the reality we experience a particular collection of formalized structures. We thereby create a blindness to everything that is not expressible with those structures.

include things like choosing the best routing for phone calls through a network), you are creating a kind of blindness in a system that has no way to step back from it. A person, when faced with a situation in which the previous articulation is not applicable, uses what we call "common sense" or "intuition" to say "Wait a minute! The formula says such and such, but it doesn't work. It doesn't deal with the thing I should be looking at here."

That kind of stepping back is based on a larger embedding of ourselves in our world, which cannot be programmed using the techniques that work for the narrow domains. Part of the appeal (or mystique) of rule-based systems is the hope that you can get around the limitations of a particular formulation by having lots of little disjointed pieces rather than insisting on putting together a whole coherent picture. When you put in a bunch of rules you can always say "We may be leaving things out but we can always add more rules to cover it." I think that's wishful thinking. When you carve out the structure of the domain—before you even write the rules—you have already created blindness by the set of relations, objects and properties of which the rules are composed. Once you decide there is something called a "secondary infection," regardless of how many rules you put in for particular secondary infections, you have already set up a kind of rigid structure. Also, of course the hope that things put in independently will actually do what you want when they interact is only at best wishful thinking. It does happen sometimes, and it's very pleasing when it does, but you certainly can't count on it.

I think there's also a more fundamental question about the notion of problem-solving. Problem-solving takes for granted some account of a "problem space" in which you are searching for a solution. It doesn't take into account the process by which the problem itself comes to be formulated. Within the AI literature in general (and certainly within expert systems) only lip service has been paid to that issue. There has been a tremendous emphasis on the solving techniques and an understandable non-emphasis on the formulation techniques, because they aren't amenable in any easy way to the same kinds of devices.

Where does all that lead us?

I'll give my own personal view on that by quoting from Randy Davis. Randy gave a talk four years ago at this same conference entitled "Expert Systems—Where Are We and Where Do We Go from Here?"[2] He described what he calls the state of the art. I'll just quote the items: "narrow domain of expertise," "fragile behavior at the boundaries," "limited knowledge representation language," "limited explanation," and "one expert as knowledge base czar." Those are not just limitations of the state of the art four years ago. I think everybody would pretty much have to agree that they are still state of the art limitations now, four years later, and they are more fundamental. Randy's observations touch on some fundamental limitations of the ability to articulate artificial domains in a way that corresponds to reality.

The "narrow domain," of course, fits into that; the "fragile behavior" comes precisely from not being able to articulate all the possible interactions. You get unexpected interactions which do totally crazy things because there's no common sense to back them up.

"Limited representation language" and "limited explanation" are a function of keeping the articulation of the domain in a form that can be executed, which is not the same as a form that knowledge-base builders can understand and use. If somebody asks why you did something and you say "Because factor A, factor B, and factor C made me conclude that D," most of the time, in a human context, they will think it is either a silly answer or a purposely deceitful one. Typically you will more usefully answer with something like "Well, I didn't want to do such and such else." You don't just say that three particular factors led me to that conclusion because those factors appeared within a huge background of possibilities. What you were doing took into account the background in picking something out to focus on, and an explanation is not the same as a trace of particular rules.

Finally, the observation that in existing systems there is usually one expert as knowledge base czar also reflects a fundamental limitation. The limitation isn't that there must be just one expert, but that designing a knowledge base is not a matter of picking out of experts' brains something that's already there. It is the creation of a systematic domain—a new construct that reflects what is important in the situation of interest. In some fields, much of that work has already been done and all you need to do is program it. But, in most fields, what's in the textbooks is only a tiny fraction of what really goes on. The creation and articulation of this new domain has to be done by consensus, and involves the full knowledge and discussion of experts.

To conclude, where does all that lead us? I'm not claiming that expert systems are useless or impractical or that we shouldn't be doing them. Rather, it is critical to understand the nature of the limitations that result from the kinds of systematic articulated domains in which they work. They will work much better in areas satisfying obvious criteria of simplicity—having well distinguished factors with clearly describable interactions. These may include many scientific and technical areas, factory operations, and so forth. When you move farther afield, including some of the most popular development areas such as medicine and computer-aided instruction, you go beyond, and the road is much harder.

A second key point is that people who are using an expert system need to understand what its domain really is rather than being enticed by what the domain seems to be about. A system may be broadly described as doing medical diagnosis, but what it really does is figure out the connections between certain kinds of bacteria and certain kinds of blood infections. That may be very useful, but if consumers think it's giving a diagnosis in a broad sense, they are being misled and potentially very dangerously misled.

So what's important is for the people who use the systems to actually understand what the computer is doing, in terms of its explicit domain. This goes against the common wisdom because it suggests that the people who are going to use expert systems the most and the best are themselves experts. The public is not going to have an understanding that allows them to really work with limitations and constraints in a system. Putting an expert system in place of the expert for general use is going to lead to serious problems of misplaced confidence. On the other hand, you can say to an expert "Here's a system that handles this particular formalized part of your job, and you don't need to worry about looking all these things up in the textbooks or figuring all the drug interactions or the stresses or whatever." The expert knows what aspect of the problem the computer is really dealing with, and can put it into proper perspective. In that lies the real potential for expert systems.

Dreyfus's Presentation

In his introduction, Terry Winograd correctly said that I am sometimes confused with my brother, Hubert Dreyfus. I should explain why, in the AI context, that is not an unreasonable error. The statements that I'll be making tonight and the positions that I'll be taking are totally in agreement with those of Bert. So if you hear me say something that you agree with, don't say "Ah ha. That's sensible. He's not crazy like his brother" or if I say something that you think is wrong don't say "Well, at least Bert Dreyfus isn't that foolish" because he probably believes it too. We agree completely on the proper uses and the limitations of AI and of mathematical modeling although we come from different backgrounds; he's a philosopher and I'm a mathematician. For about half of our academic careers I didn't understand anything that he was saying and he didn't understand what I was doing, but it turned out, when we finally understood each other, that we were thinking very similar things. For the last half of our careers we've actually been working together in this area.

I'd like to begin with a few disclaimers to dispel some misunderstandings of mainly my brother (since he speaks out on this subject more than I do) and of me. First, we do not want to be seen as "soulists" or mystics or people who believe that it is impossible for a man-made physical device to be as intelligent as a human being in isolated areas of skill where human bodily capacities and social conventions can be ignored. The basic issue for us isn't whether the brain is something that goes beyond the physical materials composing it. We don't claim that there is anything in the brain transcending its electro-chemical activities.

The issue as we see it is that the attempt to simulate intelligence in the field of AI in general, and the expert system area in particular, is based on using isolated facts, and rules relating those facts, as a means of duplicating the abilities of the human mind. We believe that research using information processing devices or logic machines or whatever you call machines that process facts by using complex reasoning based on rules in order to produce so-called intelligence has provided no convincing evidence that such machines can indeed exhibit a high level of intelligence in any but structured combinatorial domains. (Structured domains are those where what constitutes the relevant facts and how these facts can be changed by decisions is known objectively.) It's an empirical question. We don't claim to have an in principle proof that some superior logic, maybe beyond what anybody has thought of so far in AI, can't produce intelligence. However, the burden of proof is upon those people who believe that it can. So far it hasn't. I don't need to tell you, who are the experts in this field, that AI has not yet captured the common sense or the ability to distinguish relevant from irrelevant information that even a small child can exhibit. We believe that facts, rules, and logic won't get you to full intelligence. We claim that the evidence produced by observing human skill acquisition and by the difficulties of AI justifies this belief. We do not hold that devices created by man cannot be intelligent. But we ought to back off from this obsession with rules, logic, facts, and the like and look at some other approach, perhaps that of the so-called "new connectionists."

Here is a short, but rather typical, definition of an expert system: a system that draws conclusions from facts, beliefs, and rules, some of which are heuristic, using complex reasoning based on inference rules. I take it that an expert system is a rule-based system. When I am critical of rule-based systems I am criticizing expert systems. Someday there may be systems that behave like experts that are not complex reasoning systems.

I'd like to begin by defining what I mean by "intuition." It is the ability, effortlessly and rapidly to associate with one's present situation an action or decision which experience has shown to be appropriate. We think that this ability is achieved by the modification during learning of synapses that determine the strength of connections between neurons of the brain. No manipulation of meaningful physical symbols takes place during the association of learned outputs with inputs similar to those previously experienced. Intuition is what people have and complex reasoning systems do not have and, I believe, will never have. It's a kind of glorified pattern discrimination ability. Intuition isn't something mystical, it isn't something that some people have and others don't have or that women have more or less of than men. It's something that everybody has and uses every day in order to function in the world. One cannot have intuition about a situation if one has not faced similar situations in the past. It's not something, I believe, that survives in the air as vibes to be picked up by someone else.

For example, by my definition of intuition, the recognition of a friend's face is an intuitive act. I don't believe that one does this by rules and some kind of complicated analysis. Rather, our brains' synapses have been modified on the basis of a lot of faces we've seen before, maybe lots of typical examples of each of these faces. As a result we have the ability, when we see a familiar face (even though it doesn't have exactly the same expression and is not seen from exactly the same angle or in exactly the same light as we've ever seen before) to associate with it a name or some other response. I would call this an intuitive ability. Since an infant has this capacity, I find it hard to believe that it is analytical.

In my research I have found that world-class chess players have this same intuitive ability to discriminate and associate. To such a player, but not to most of us amateurs who try to think ahead and reason things out, chess is largely an associational or intuitive game despite the fact that, as a structured combinatorial domain, chess is also amenable to complex enumerative reasoning of the sort that computers can do much better than people. A master chess player looks at a position he has never seen before but which is a legitimate chess position similar to those seen before, and just like we would associate a response with a familiar face, after two seconds he associates with that position a certain move. I claim, and my research has given some support to this (although research in psychology never definitively proves anything), that association is the essence of human skill, at the highest level, in chess; it is also the essence of skill in driving a car. It's the essence of skill in almost everything we do in the world—walking, conversing, and so forth. It's also the essence of most professional expertise.

We don't claim there is anything in the brain transcending its electrochemical activities.

Given this idea of associative intuition, I assert that it is mistaken to believe (and I hope no one here does) that expert systems are repositories of human expertise. In no sense can one capture human expertise and store it in the form of a complex reasoning system if human expertise is an intuitive associative ability not based on processing facts by means of rules.

I further claim that expert systems do not currently perform as well as experts in areas where the most skilled human beings are intuitive decision makers. Whether expert systems, using methods entirely different from experts, can ultimately, in some area, perform as well as intuitive experts is an empirical question. I don't believe that they have so far or that in any but structured areas they ever will. Perhaps in some unstructured area, some day, somebody will invent wonderful rules that go beyond what anybody ever has used or could use, that will result in performance at or above the quality of intuitive experts. I don't claim that I have an in principle argument against the possibility. But the evidence so far is that, except in structured, combinatorial domains, expert systems do not perform as well as the intuitive experts consulted by knowledge engineers when building them and there is no reason to be optimistic about their future.

Let me be specific about our view of skill acquisition. I can't go into detail here, but, as Terry said, Bert and I have written a book that goes into great detail on this matter.[3] In our studies of skill acquisition we found that people pass through at least five distinguishable stages in which they look at their problems and respond to their situations in significantly different ways. There are many qualifications concerning this model; too many to enumerate all of them here. One is that in a skill domain a person is not at just one of these levels with respect to all situations in that domain. He can be a beginner with respect to certain situations different from what he has seen before while he is at a very high level of skill in certain other familiar situations. While the five-stage progression I shall describe characterizes most people in most skill domains, there are exceptions. Anyone learning to ride a bicycle generally learns by trial-and-error, repeating what works, rather than by passing through the initial analytical stages that I shall describe. And while the best performers in most domains rely on intuition, it is possible to intuitively respond to familiar patterns in learned ways and still perform badly—just see all patterns as similar to each other and always respond the same way no matter what the situation. Some skill domains deal with situations that are so combinatorial and jig-saw puzzle–like that no amount of experience allows one to learn associations that work for similar patterns, so even the best performers apply rules and use principles, although they usually use intuition about which rules to apply and when to break them.

Let me briefly outline our model of skill acquisition since it allows one to see the place of rule-based expert systems. At the lowest level of skill, novice, a person uses context-free facts and context-free rules to produce actions based on these facts. Such facts and rules are what beginners are taught in most skill domains. A beginning chess player might be taught a point-value system for counting values of pieces and the rule "if you can capture pieces that add up to more points than you lose in the process, always do it." This gets him off the ground so he can begin to play chess and to gain concrete experience. It's what computers do when they are programmed in the most simplistic way. They use context-free facts and process these by precise rules.

An advanced beginner learns to recognize elements of a situation that are not defined in terms of context-free facts. They are recognized as similar to what one has already seen. For instance, in chess, with enough experience with examples, a player will recognize an over-extended position or weakness on the king's side. This is not recognition of a whole position but recognition of what we call a situational aspect of the position. The rules that an advanced beginner employs when selecting a move can now refer to these aspects that are recognized based on experience as well as on context-free features.

A competent performer, at the third level of our five stages, learns to organize thought in terms of plans and goals. This is the subject of much of the sophisticated research on decision making. Researchers recognize that problem solvers don't just apply rules to facts, but do so with goals in mind. The goal determines what are seen as salient facts and what rules to apply. Sophisticated problem solvers use a hierarchy of goals and subgoals. By my definition, a competent performer is someone who consciously and deliberately decides upon a goal and perhaps subgoals. Then he selects what are the important facts and what rules to apply to determine what to do. This sounds like the description of a very sophisticated AI program or an expert system. However such systems can only use situational aspects if they are provided by a human being with the capacity to recognize them. Expert systems, used interactively, are capable of competent performance. They are capable of performing in the sophisticated, detached analytical way that I have just described.

The last two stages of skill acquisition pass, I believe, beyond the capabilities of rule-based computers, for example logic machines. The proficient performer, at the fourth level of skill, intuitively recognizes the sense of the situation. When faced with a familiar situation, what kind of situation it is or what the goal should be leaps out at him due to similarity with situations previously experienced as well as due to recent past events which determine what appears as salient in the situation. He doesn't stand back and ask "How should I see this situation?" It is simply there in front of him and generally he has no choice in the matter. This lack of choice is a form of blindness, as Terry said. When you are proficient you are also, in a sense, blind. It's a danger, but also a human strength that takes one beyond the level of reason and beyond merely competent performance. While proficient performers know intuitively what is going on without having to figure it out, they still have to calculate what to do based on what is going on. A proficient chess player, for example, will know from experience that one must attack in a certain position, but will do analysis in order to decide how to best go about it.

The expert, at the highest stage of skill, has the ability not only to intuitively know the sense of the situation but he also knows from prior experience what to do. Not only what is going on but also what needs to be done leaps to mind in the one or two seconds that it takes an intuitive world-class chess player to arrive at a move. The expert martial artist responds in even less time.

I don't want to give the impression that I think that all skill is ultimately, at its highest level, purely automatic response to situations. Truly skilled individuals, if they have time, will deliberate about their intuitions. I've identified a form of detached deliberation, and I'm sure I'm not the first to do so, that involves a kind of rationality, but not the kind of calculative rationality that programmers know how to put into a computer. It's deliberative thinking about one's intuitions. Do I have a right to be intuitive in this situation—have I had enough experience? Have things changed since I got that experience? Am I being blind—is there perhaps another way of looking at this situation where other things might stand out? Should I rethink the history that led me to see things the way I do and question whether, consistent with that history, I might be able to see the same situation in a different light? These are the kinds of questions that an involved, intuitive decision maker will ask if he has time for detached, rational deliberations.

While good decision making is more that just knee-jerk intuitive response to situations, my experiments have shown that if you take a world-class chess player and restrict him to virtually instantaneous intuitive responses, he still plays near master caliber chess. That's why I claim that the essence of expertise is intuition. I attempted to remove from a chess player any ability to do analytical thinking by first requiring that he add heard numbers as fast as he could so that I could determine his speed. Then I said "I want you to continue doing addition just as fast and now play chess at the same time." The idea was to demonstrate to him and to me and hopefully to the field of experimental psychology with its obsession with problem solving behavior, that there is a kind of expertise that does not require figuring things out or using any kind of analysis. I attempted to eliminate any analysis by occupying the analytical mind with the addition task. At first my subjects would turn pale at the prospect of chess without analysis. It turned out, however, that a world-class player, International Master Julio Kaplan, could play chess at the rate of a few seconds a move, simultaneously add as fast as he could when not playing chess, and still beat a good master player.

This gives you a perspective from which to think about expert systems and expertise. When considering building an expert system to deal with a certain type of problem, ask yourself "Is this an unstructured domain where the best performers can and do become intuitive? Is this a domain where an expert has a gut feeling almost instantaneously about what's going on?" If it is, stay away from that domain as far as making promises about capturing that expert's understanding and skill in the form of a complex reasoning system using facts, rules and inferences. Medicine, prospecting for minerals and all policy-level decision making, to name a few, are domains where the best performers seem to develop and use intuition. These are also domains where expert systems, contrary to what you may have heard, do not measure up to experts. Likewise, an experienced automobile mechanic, repairing a familiar type of car with a familiar malfunction, proceeds intuitively.

If, on the other hand, you find a domain where no one, even after considerable experience, acquires intuition and there are people who apply rules in that domain who do as well as anyone, then you've got a domain where there is hope that an expert system will perform as well as the experienced specialist and indeed, to some extent, capture his expertise. Such, apparently, is the case with the well-known R1 or XCON VAX configuring expert system. As I understand it, a human expert VAX configurer doesn't just look at an order and immediately have a strong intuitive feeling about the optimal configuration. He may have an intuitive feeling about what goals and subgoals to use to figure out the best configuration, but he still looks at his problem in a rule-based way. So there is an area where I would have predicted some success for an expert system.

So-called expert systems can, in almost all isolated domains cut off from everyday human affairs, perform as well as competent human beings. They can perform as well or better than anyone on novel problems where no one has had enough experience to become an intuitive expert. They an approximate the skill of the best human rule-users in highly combinatorial situations and in structured domains they may someday outperform the best intuitive human beings. But in most domains, since they are unstructured, expert systems can neither imitate nor perform as well as mature and practiced intuitive experts.

Notes

1. Winograd, T., and Flores, F. 1986. *Understanding Computers and Cognition.* Norwood, N.J.: Ablex.

2. A version of the talk, entitled "Expert Systems: Where Are We? And Where Do We Go from Here?" appeared in *AI Magazine* 3(2).

3. Dreyfus, H., and Dreyfus, S. 1986. *Mind over Machine.* New York: The Free Press.

Terry Winograd is associate professor of Computer Science and Linguistics at Stanford University. He received his Ph.D. from the Massachusetts Institute of Technology. in Applied Mathematics (AI) in 1970 and joined the computer science department at Stanford in 1973. His research on natural language understanding by computers is often cited as a major milestone in AI. It was the basis for the book *Understanding Natural Language* (Academic Press, 1972) and the textbook *Language as a Cognitive Process* (Addison-Wesley, 1983). He coauthored, with Fernando Flores, *Understanding Computers and Cognition: A New Foundation for Design* (Ablex, 1986) which was honored as the 1987 Information Sciences Book of the Year by the American Society for Information Sciences.

Stuart E. Dreyfus is professor of Engineering Science in the Department of Industrial Engineering and Operations Research at the University of California at Berkeley. He is author of several books on dynamic programming and optimal control theory. Dreyfus is coauthor, with his brother, Hubert L. Dreyfus, of *Mind over Machine* (The Free Press, revised, 1988). Dreyfus's current research interest is in the theory of artificial neural networks (connectionism) and its relevance to cognitive science.

Randall Davis is at the Massachusetts Institute of Technology, where he is a professor of Information Science in the School of Management and associate director of the Artificial Intelligence Laboratory. He received his Ph.D. from Stanford University in artificial intelligence in 1976 and joined the MIT faculty in 1978, where he held an Esther and Harold Edgerton endowed chair from 1979 through 1981. Davis's current research focuses on model-based systems—programs that work from descriptions of structure and function and that reason from first principles. He and his group at MIT have developed model-based systems for troubleshooting, generating diagnostics, design for testability, innovative design, and other functions.

Dr. Davis has published widely in the field and serves on several editorial boards including *Artificial Intelligence* and *AI in Engineering.* In 1984 he was selected as one of America's top 100 scientists under the age of 40 by *Science Digest.*