

Panel: Expert systems: How far can they go?

Terry Winograd, Randall Davis, Stuart Dreyfus, Brian Smith

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?

Expert systems:

What to do until the theory arrives

Randall Davis

MIT

In reading a newspaper recently I was struck by the profusion of expert advice available. There were three different expert opinions on the future course of the economy, several compelling (and contradictory) opinions about the likely course of events in the mideast, and a number of suggestions about avoiding heart disease, as well as claims about long term weather patterns, advice from Ms. Manners on behavior and guidelines from a therapist on drinking and sex.

All of which made me begin to wonder:

Experts: How far can they go?

We are in the midst of a great wave of enthusiasm about the potential for experts 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:

What is the nature of the problem domains in which experts are likely to succeed and those in which they will not? Are there domains in which their use might be dangerous?

To what extent can we count on carbon-based systems for 'flexibility' in dealing with unexpected situations? How reliable will such systems be in cases where their teachers did not anticipate significant possibilities?

How can a person's knowledge be subjected to standards of accountability? Who is responsible for what an expert contains and what that person does?

Hardly an original satire, but it does serve several purposes. First, it demonstrates that the questions are neither unfamiliar nor inherently mysterious. The answers for people may not be well established, but we do know something of how to proceed and we do believe there is no magic here: *some* form of knowledge accounts for an expert's competence. The nature and source of it may be far from understood, but that doesn't make the question unanswerable.

Next, it sets the argument out on what I believe to be an important direction: the questions we are asking are first about knowledge and only then about technology. That is, the first question should not be *What can expert systems do?* but rather *What do we know?* Only then we can address the technology issue and ask *And how easily can we encode that knowledge?* Both matter but the order is important.

Finally, the comparison is valid and provides an interesting way to proceed. Asking the same questions about people provides a useful, non-threatening way of examining the topics. Our answers may differ for people

and programs, but even those disparities will prove interesting. We explore people's understanding of a subject with a test that examines only a limited sample of their knowledge, and then extrapolate, saying that people who pass 'understand' the material, meaning by that something more than that they can do exactly the problems chosen for the exam. If a program passed the same exam, would we be willing to do the same extrapolation? If not, why not? If we can determine what it is that makes us hesitate in the case of the program, we have the beginnings of an intriguing research agenda.

In developing these themes I will argue that there are two attributes of expert systems (and much of AI) that are central to this discussion:

It is a weak technology

It is a technology for dealing with incompletely understood ideas.

I will suggest that the first of these is a temporary vice that will be remedied in time (though not soon) and that the second is a permanent virtue.

Both of these have interesting implications for the use of expert systems (and indeed much of AI). Perhaps the most important implication is that most traditional rule-based expert systems will never have all the knowledge they need, and as a consequence they are guaranteed to fail occasionally (though perhaps infrequently) during all of their operational lives.

I will suggest ways of proceeding that take these issues into account, allowing us to employ the technology while reducing the potential difficulties that can arise.

The nature of expertise

Stuart E. Dreyfus

University of California, Berkeley

All AI work with the exception of a few 'connectionist' theories assumes that knowledge must be represented in the mind as symbolic descriptions. Expert-system builders further assume that the expert possesses a particular kind of symbolic description: a knowledge base of facts, beliefs and 'if-then' rules that allow the drawing of inferences.

I will argue that expert-system builders fail to recognize the real character of *expert* human understanding. Expertise is acquired in a five-stage process. The *beginner* applies rules to context-free features as would an expert system deprived of situational knowledge appropriate to the particular case. The *advanced beginner* learns from experience to recognize aspects of a situation without requiring a definition of them in terms of context-free features.

Aspects are recognized after seeing several examples, apparently because of their similarity to already experienced prototypical cases. An interactive expert system could use aspects if they were identified for it by a human user. At the next stage, the *competent* performer organizes behavior by selecting plans, goals, or perspectives which determine hierarchically what facts to consider and what rules to apply. Expert systems can do likewise and, if they accept human situational assessments, could appropriately be called 'competent systems.' This, however, is the best they can do. The fourth stage, *proficiency*, is achieved when the performer no longer uses his knowledge to select a perspective or goal, but simply recognizes the appropriate one based on prior experience in similar situations in which goals were chosen and events either confirmed the wisdom of the choice or showed it to be mistaken. As with the recognition of situational aspects by the advanced beginner, the involved, intuitive recognition of similarity of whole situations is apparently not produced by rules operating on features but seems to be effortless and holistic. While the proficient performer still analytically figures out what to do once the situation is intuitively understood, at the highest level of skill the *expert* has experienced so many situations that he associates with each prototypical situation in his memory the decision, action or strategy that he has found to work. He reasons out neither strategy nor action. Intuitively responding to situational patterns as experience has shown appropriate, his skill depends neither on problem solving nor planning. Expert systems can neither recognize situations holistically without analysis into components nor know what to do without applying rules to decomposed knowledge, so they can be neither proficient nor expert.

If time permits, an expert will deliberate about his intuitive understanding. To fine-tune his responses, he will attempt to take account of subtle differences between his current situation and similar prior ones, he will ask himself whether there might be another quite different way of intuitively viewing his circumstance, and he will consider whether he has had enough experience in the particular kind of situation to trust his intuition. But he will rarely regress to competent detached problem solving.

While 'competent systems' have their useful place, there is no reason to expect them to perform as well as experts who have passed beyond the use of facts, beliefs, and rules of inference and who rely on memories of thousands of concrete experiences and what has worked in each. An examination of the performance of various expert systems supports the above analysis. In domains where human beings pass from reasoning to recognition as they become experts, expert systems, even when experts participate in their development, never perform as well as experts.