Book Review

John McCarthy
Computer Science Department
Stanford University
Stanford, CA 94305
Internet: jmc@cs.stanford.edu

June 13, 2000

Hubert Dreyfus *What Computers Still Can't Do* (The MIT Press, Cambridge, MASSACHUSETTS, 1992);

1 Progress in Logic-Based AI

Hubert Dreyfus claims that "symbolic AI" is a "degenerating research program", i.e. is not making progress. It's hard to see how he would know, since he makes no claim to have read much of the recent literature.

In defending "symbolic AI", I shall concentrate on just one part of symbolic AI—the logic-based approach. It was first proposed in [McCarthy, 1959], attracted only intermittent following at first, but has had an increasing number of workers since 1970. [1] I think other approaches to AI will also eventually succeed, perhaps even connectionism. To contradict an earlier Dreyfus metaphor "AI at the Crossroads Again", it isn't a crossroads but a race including logic based AI, SOAR, connectionism and several other contenders.

How goes the logic-based runner? In fits and starts, as problems are identified and progress made.

Within logic-based AI, I shall emphasize one development—formalized nonmono-

[1] The logic-based approach doesn't include Newell's [Newell, 1992] approach based on Soar, which is also "symbolic AI". A reasonable interpretation of logic-based AI is that it works directly at what Newell, [Newell, 1982] and [Newell, 1992], calls the logic level.

tonic reasoning, because it illustrates intermittent but definite progress. It was first proposed in [McCarthy, 1977], gained momentum with the 1980 special issue of *Artificial Intelligence*, and summarized in the collection [Ginsberg, 1987]. It has continued to develop, see e.g. [Lifschitz, 1993].

Minsky [Minsky, 1975] mentioned the need for something like nonmonotonic reasoning, but used this fact as evidence for the inadequacy of logic-based approaches to AI and the need for approaches not based on logic. This isn't how things have gone. Nonmonotonic reasoning has developed as a branch of mathematical logic, using the concepts, such as that of interpretation and model, that logicians have been developing since the 1930s.

The circumscription method of nonmonotonic reasoning would have been entirely comprehensible to Hilbert and probably even to Frege. However, no formalized nonmonotonic logic was developed until the 1970s. This is typical of the slow progress in developing mathematical logical formalisms. Just consider the sequence, Aristotle, Leibniz, Boole, Frege, Hilbert, Gödel. Each step would have been comprehensible to the predecessor, yet it took a long time for the new ideas to appear. Formalized nonmonotonic reasoning is surely not the last step in the chain aimed at formalizing useful reasoning. Formalizing nonmonotonic reasoning required realizing that there is proper reasoning that is not conclusive and that is often not the same as probabilistic reasoning.

For this reason, neither Dreyfus or anyone else is entitled to conclude that if the logic oriented AI problem hasn't been solved in 40 years, it won't ever be solved. To do that it would be necessary to prove a theorem about limitations of logic. Not even showing that there has been no progress at all would be conclusive. However, Dreyfus makes no reference to nonmonotonic reasoning in his book. That's about 1,000 papers he doesn't know about. However, in answer to a question at a book-selling talk, he said that claiming progress in nonmonotonic reasoning is progress towards AI is like claiming that climbing a tree is progress towards reaching the moon—thus recycling a metaphor from the book.

Although formalized nonmonotonic reasoning was discovered in connection with AI, many logicians pursue it as a purely mathematical study, independent of applications to AI or logic programming (another non-entry in Dreyfus's index).

2 The Future of Logic Based AI

The review editors asked me to say what I think the obstacles are to human-level AI by the logic route and why I think they can be overcome. If anyone could make a complete list of the obstacles, this would be a major step towards overcoming them. What I can actually do is much more tentative.

Workers in logic-based AI hope to reach human-level in a logic based system. Such a system would, as proposed in [McCarthy, 1959], represent what it knew about the world in general, about the particular situation and about its goals by sentences in logic. Other data structures, e.g. for representing pictures, would be present together with programs for creating them, manipulating them and for getting sentences for describing them. The program would perform actions that it inferred were appropriate for achieving its goals.

Logic-based AI is the most ambitious approach to AI, because it proposes to understand the common sense world well enough to express what is required for successful action in formulas. Other approaches to AI do not require this. Anything based on neural nets, for example, hopes that a net can be made to learn human-level capability without the people who design the original net knowing much about the world in which their creation learns. Maybe this will work, but then they may have an intelligent machine and still not understand how it works. This prospect seems to appeal to some people.

Common sense knowledge and reasoning is at the core of AI, because a human or an intelligent machine always starts from a situation in which the information available to it has a common sense character. Mathematical models of the traditional kind are imbedded in common sense. This was not obvious, and many scientists supposed that the development of mathematical theories would obviate the need for common sense terminology in scientific work. Here are two quotations that express this attitude.

> One service mathematics has rendered to the human race. It has put common sense back where it belongs, on the topmost shelf next to the dusty canister labelled 'discarded nonsense'. —*E. T. Bell*

> All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word 'cause' never occurs ... The law of causality, I believe, like much that passes muster among philoso-

> phers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm ....—*B. Russell*, "On the Notion of Cause", *Proceedings of the Aristotelian Society*, 13 (1913), pp. 1-26.

The "Nemesis" theory of the mass extinctions holds that our sun has a companion star that every 13 million years comes close enough to disrupt the Oort cloud of comets, some of which then come into the inner solar system and bombard the earth causing extinctions. The Nemesis theory involves gravitational astronomy, but it doesn't propose a precise orbit for the star Nemesis and still less proposes orbits for the comets in the Oort cloud. Therefore, the theory is formulated in terms of the common sense notion of causality.

It was natural for Russell and Bell to be pleased that mathematical laws were available for certain phenomena that had previously been treated only informally. However, they were interested in a hypothetical information situation in which the scientist has a full knowledge of an initial configuration, e.g. in celestial mechanics, and needs to predict the future. It is only when people began to work on AI that it became clear that general intelligence requires machines that can handle the common sense information situation in which concepts like "causes" are appropriate. Even after that it took 20 years before it was apparent that nonmonotonic reasoning could be and had to be formalized.

Making a logic-based human-level program requires enough progress on at least the following problems:

extensions of mathematical logic
Besides nonmonotonic reasoning other problems in the logic of AI are beginning to take a definite form including formalization of contexts as objects. This can provide a logical way of matching the human ability to use language in different ways depending on context. [McCarthy, 1987], [Guha, 1991], [McCarthy, 1993], [Buvac and Mason, 1993].

elaboration tolerance Formalisms need to be elaboratable without a human having to start the formalism over from the beginning. There are ideas but no articles as yet.

concurrent events
[Gelfond, Lifschitz and Rabinov, 1991] treats this using the situation calculus, and I have some recent and still unpublished results aimed at a simpler treatment.

intentionality The treatment of mental objects such as beliefs (much discussed in the philsophical literature) and the corresponding term concept, e.g. "what he thinks electrons are" (hardly at all discussed in the formal literature).

reification We need a better understanding of what are to be considered objects in the logic. For example, a full treatment of the missionaries-and-cannibals problem together with reasonable elaborations must allow us to say, "There are just two things wrong with the boat."

introspection and transcendence
Human intelligence has the ability to survey, in some sense, the whole of its activity, and to consider any assumption, however built-in, as subject to question. Humans aren't really very good at this, and it is only needed for some very high level problems. Nevertheless, we want it, and there are some ideas about how to get it. What may work is to use the context mechanism as discussed in [McCarthy, 1993] to go beyond the outermost context considered so far.

Unfortunately, too many people concentrated on self-referential sentences. It's a cute subject, but not relevant to human introspection or to the kinds of introspection we will have to make computers do.

levels of description If one is asked how an event occurred, one can often answer by giving a sequence of lower level events that answer the question for the particular occurrence. Once I bought some stamps by going to the stamp selling machine in the airport and putting in six dollars, etc. Each of these subevents has a how, but I didn't plan them, and cannot recall them. A stamp buying coach would have analyzed them to a lower level than I could and would be able to teach me how to buy stamps more effectively. For AI we therefore need a more flexible notion than the computer science theories of how programs are built up from elementary operations.

Dreyfus asks why anyone should believe all this can be done. It seems as good a bet as any other difficult scientific problem. Recently progress has become more rapid, and many people have entered the field of logical AI in the last 15 years. Besides those whose papers I referenced, these include Raymond Reiter, Leora Morgenstern, Donald Perlis, Ernest Davis, Murray Shanahan, David Etherington, Yoav Shoham, Fangzhen Lin, Sarit Kraus, Matthew Ginsberg, Douglas Lenat, R. V. Guha, Hector Levesque, Jack Minker, Tom Costello, Erik Sandewall, Kurt Konolige and many others. There aren't just a few "die-hards".

However, reaching human level AI is not a problem that is within engineering range of solution. Very likely, fundamental scientific discoveries are still to come.

3 Common Sense in Lenat's Work

Douglas Lenat is one of the few workers in AI at whose recent work Dreyfus has taken a peek.

Dreyfus, p. xvii and xviii, writes:

> When, instead of developing philosophical theories of the transcendental conditions that must hold if the mind is to represent the world, or proposing psychological models of how the storage and retrieval of propositional representations works, researchers in AI actually tried to formulate and organize everyday consensus knowledge, they ran into what has come to be called the commonsense-knowledge problem. There are really at least three problems grouped under this rubric:
>
> 1. How everyday knowledge must be organized so that one can make inferences from it.

> 2. How skills or know-how can be represented as knowing-that.
> 3. How relevant knowledge can be brought to bear in particular situations.
>
> While representationalists have written programs that attempt to deal with each of these problems, there is no generally accepted solution, nor is there a proof that these problems cannot be solved. What is clear is that all attempts to solve them have run into unexpected difficulties, and this in turn suggests that there may well be in-principle limitations on representationalism. At the very least these difficulties lead us to question why anyone would expect the representationalist project to succeed.

That's not too bad a summary except for the rhetorical question at the end. Why should one expect it to be easy, and why should one expect it not to succeed eventually in reaching human level intelligence? Most of the people who have pursued the approach have seen enough of what they regard as progress to expect eventual success. I have referred to some of this progress in my account of the invention and development of formalized nonmonotonic reasoning.

Mostly I agree with what Lenat said (as Dreyfus quotes him in the book), and I don't find much support for Dreyfus's assertions that empathy rather than just verbalizable understanding is required in order to understand human action. I think the example on p. xix of what "it" means in

Mary saw a dog in the window. She wanted it.

is within the capability of some current parsers that use semantic and pragmatic information.

However, I think the following assertion of Lenat's [Lenat and Guha, 1990] quoted by Dreyfus on p. xxv is an oversimplification.

> These layers of analogy and metaphor eventually 'bottom out' at physical-*somatic*-primitives: up, down, forward, back, pain, sssssscold, inside, seeing, sleeping, tasting, growing, containing, moving, making noise, hearing, birth, death, strain, exhaustion, ...

The contact of humans (and future robots) with the common sense world is on many levels, and our concepts are on many levels. Events that might bottom out *physically*—as informing someone of something may physically bottom out in making a noise—often don't bottom out epistemologically. We may assert that A informed B of something without our being able to describe the act in terms of making noise or typing on a keyboard.

While I don't agree with Lenat's formulation, the success of Cyc doesn't depend on its correctness. Cyc perfectly well can (and indeed does) store information obtained on several levels of organization and used by programs interacting with the world on several levels.

All this doesn't guarantee that Cyc will succeed as a database of common sense knowledge. There may be to big a conceptual gap in the AI community's ideas of what are the usefully stored elements of common sense knowledge.

4 The Degenerating Research Program

In the first edition of Dreyfus's book there were some challenges to AI. Dreyfus said computers couldn't exhibit "ambiguity tolerance", "fringe consciousness" and "zeroing in". These were left so imprecise that most readers couldn't see any definite problem at all. In the succeeding 30 years Dreyfus has neither made these challenges more precise nor proposed any new challenges, however imprecise. It's a pity, because AI could use a critic saying, "Here's the easiest thing I don't see how you can do".

That part of Dreyfus's research program has certainly degenerated.

However, I can give a definite meaning to the phrase "ambiguity tolerance" that may not be too far from Dreyfus's vague idea, and with which formalized nonmonotonic reasoning can deal. The idea is that a concept that may be ambiguous in general is to be taken by default as unambiguous in a particular case unless there is reason to do otherwise.

Here's an example.

Suppose that some knowledge engineer has the job of making an adviser for assistant district attorneys. The prosecutor answers some questions about the facts of the case, and the program suggests asking for indictments for certain crimes. We suppose that attempting to bribe a public official is one of these crimes.

We ask whether the knowledge engineer must have anticipated the following three possible defenses against the charge, i.e. have decided whether the following circumstances still justify an indictment.

1. The defendant's lawyer claims that his client did not know the person he offered money to fix his drunk driving convictions was the commisioner of motor vehicles. His client thought he was just an influential

lawyer.

2. The defendant's lawyer claims that while his client may have thought he was bribing the commissioner of motor vehicles, he really wasn't, because the Governor had never properly signed the commission.

3. The defendant put an advertisement in the *Criminal Gazette* offering $5,000 to any public official who would fix his conviction. Must the prosecution exhibit a specific public official the defendant was attempting to bribe in order to get a conviction for "attempting to bribe *a* public official".

There may be further potential ambiguities in the statute. If we demand that the knowledge engineer have resolved all of them before he can write his expert system, we are asking for the impossible. Legislators, lawyers and judges don't see all the ambiguities in advance.

Notice that in most cases of bribing a public official, there was a specific individual, and he really was a public official and this was really known to the defendant. Very likely, the legislators had not thought of any other possibilities. The nonmonotonic reasoning approach to ambiguity tolerance says that by default the statute is unambiguous in a particular case. Indeed this is how the law works.

The courts will not invalidate a law because of a general amiguity; it has to be ambiguous in a significant way in the particular case.

Since the expert system writer cannot anticipate all the possible ambiguities, he must make his system *ambiguity tolerant*.

When an ambiguity is actually pointed out to the expert system, it would be best if it advised looking at cases to see which way the statute had been interpreted by judges. I don't know whether to be a useful adviser in statutory criminal law, the expert system would have to have a library of cases and the ability to reason from them.

I have not written logical formulas for *ambiguity tolerance*, i.e. expressing the default that a concept, possibly ambiguous in general, is to be considered unambiguous in particular cases unless there is evidence to the contrary. However, I would be strongly motivated to give it high priority if Dreyfus were to offer to bet money that I can't.

To conclude: Dreyfus has posed various challenges to AI from time to time, but he doesn't seem to make any of them precise. Here is my challenge to Dreyfus, whereby he might rescue his research program from degeneration.

What is the least complex intellectual behavior that you think humans can do and computers can't? It would be nice to have more details than were given in connection with *"ambiguity tolerance"* and *"zeroing in"*.

References

[Buvac and Mason, 1993] **Saša Buvač and Ian A. Mason.**: "Propositional logic of context", in *Proceedings of the Eleventh National Conference on Artificial Intelligence*, 1993, to appear.

[Gelfond, Lifschitz and Rabinov, 1991] **Gelfond, M., Lifschitz, V. and A. Rabinov.**: "What Are the Limitations of the Situation Calculus?", in *Automated Reasoning: Essays in Honor of Woody Bledsoe*, Robert S. Boyer, (ed.), Kluwer Academic Publishers, Dordrecht, Boston and London.

[Ginsberg, 1987] **Ginsberg, Matthew**. *Readings in Non-Monotonic Reasoning.* Los Altos, CA: Morgan Kaufmann.

[Guha, 1991] **Guha, R. V.**: *Contexts: A Formalization and Some Applications*, Stanford PhD Thesis, 1991.

[Lenat and Guha, 1990] **Lenat, Douglas and R. V. Guha**: ""Building Large Knowledge Based Systems: Representation and Inference in the Cyc Project", Addison-Wesley, 1990.

[Lifschitz, 1993] **Lifschitz, Vladimir** "Circumscription", in *Handbook of Logic for Artificial Intelligence and Logic Programming*, Vol. 3, Oxford University Press, 1993.

[McCarthy, 1959] **McCarthy, John (1959)**: "Programs with Common Sense", in *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, Her Majesty's Stationery Office, London.

[McCarthy, 1977] **McCarthy, John (1977)**: "Epistemological Problems of Artificial Intelligence", in *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, M.I.T., Cambridge, Mass.

[McCarthy, 1987] **McCarthy, John (1987)**: "Generality in Artificial Intelligence", *Communications of the ACM.* Vol. 30, No. 12, pp. 1030-1035. Also in *ACM Turing Award Lectures, The First Twenty Years*, ACM Press, 1987. Reprinted in [McCarthy, 1990].

[McCarthy, 1990] **McCarthy, John**: *Formalizing Common Sense*, Ablex, Norwood, New Jersey, 1990.

[McCarthy, 1993] **McCarthy, John**: "Notes on the Formalization of Context",*Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, Los Altos, CA, Morgan-Kaufmann.

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

[Newell, 1982] **Newell, Allen**: "The Knowledge Level", *Artificial Intelligence* **18** pp. 87-127.

[Newell, 1992] **Newell, Allen**: *Unified Theories of Cognition*, Harvard 1992.

[Newell, 1992] **Newell, Allen**: "Reflections on the Knowledge Level", *Artificial Intelligence* **59** pp. 31-38.

/@sail.stanford.edu:/u/ftp/pub/jmc/reviews/dreyfus.tex: begun 1993 Mar 3, LaTeXed 2000 Jun 13 at 1:04 a.m.