

Why Heideggerian AI failed and how fixing it would require making it more Heideggerian[☆]

Hubert L. Dreyfus

University of California, Department of Philosophy, 314 Moses Hall 2390, Berkeley, CA 94720-2390, USA

Available online 10 October 2007

1. The convergence of computers and philosophy

When I was teaching at MIT in the early sixties, students from the Artificial Intelligence Laboratory would come to my Heidegger course and say in effect: “You philosophers have been reflecting in your armchairs for over 2000 years and you still don’t understand how the mind works. We in the AI Lab have taken over and are succeeding where you philosophers have failed. We are now programming computers to exhibit human intelligence: to solve problems, to understand natural language, to perceive, and to learn.”¹ In 1968 Marvin Minsky, head of the AI lab, proclaimed: “Within a generation we will have intelligent computers like HAL in the film, *2001*.”²

As luck would have it, in 1963, I was invited by the RAND Corporation to evaluate the pioneering work of Alan Newell and Herbert Simon in a new field called Cognitive Simulation (CS). Newell and Simon claimed that both digital computers and the human mind could be understood as *physical symbol systems*, using strings of bits or streams of neuron pulses as symbols representing the external world. Intelligence, they claimed, merely required making the appropriate inferences from these internal representations. As they put it: “A physical symbol system has the necessary and sufficient means for general intelligent action.”³

As I studied the RAND papers and memos, I found to my surprise that, far from replacing philosophy, the pioneers in CS had learned a lot, directly and indirectly from the philosophers. They had taken over Hobbes’ claim that reasoning was calculating, Descartes’ mental representations, Leibniz’s idea of a “universal characteristic”—a set of primitives in which all knowledge could be expressed,—Kant’s claim that concepts were rules, Frege’s formalization of such rules, and Russell’s postulation of logical atoms as the building blocks of reality. In short, without realizing it, AI researchers were hard at work turning rationalist philosophy into a research program.

At the same time, I began to suspect that the critical insights formulated in existentialist armchairs, especially Heidegger’s and Merleau-Ponty’s, were bad news for those working in AI laboratories—that, by combining rational-

[☆] The essence of this article will appear in the book “The Mechanization of Mind in History”, MIT Press.
E-mail address: dreyfus@cogsci.berkeley.edu.

¹ This isn’t just my impression. Philip Agre, a PhD student at the AI Lab at that time, later wrote:

I have heard expressed many versions of the propositions . . . that philosophy is a matter of mere thinking whereas technology is a matter of real doing, and that philosophy consequently can be understood only as deficient.

Philip E. Agre, *Computation and Human Experience* (Cambridge: Cambridge University Press, 1997), 239.

² Marvin Minsky as quoted in a 1968 MGM press release for Stanley Kubrick’s *2001: A Space Odyssey*.

³ Newell, A. and Simon, H.A., *Computer Science as Empirical Inquiry: Symbols and Search*, *Mind Design*, John Haugeland, Edt. (Cambridge, MA, MIT Press), 1988.

ism, representationalism, conceptualism, formalism, and logical atomism into a research program, AI researchers had condemned their enterprise to reenact a failure.

2. Symbolic AI as a degenerating research program

Using Heidegger as a guide, I began to look for signs that the whole AI research program was degenerating. I was particularly struck by the fact that, among other troubles, researchers were running up against the problem of representing significance and relevance—a problem that Heidegger saw was implicit in Descartes' understanding of the world as a set of meaningless facts to which the mind assigned what Descartes called values, and John Searle now calls functions.⁴

But, Heidegger warned, values are just more meaningless facts. To say a hammer has the function of being for hammering leaves out the defining relation of hammers to nails and other equipment, to the point of building things, and to the skills required when actually using the hammer—all of which reveal the way of being of the hammer which Heidegger called *readiness-to-hand*. Merely assigning formal function predicates to brute facts such as hammers couldn't capture the hammer's way of being nor the meaningful organization of the everyday world in which hammering has its place. “[B]y taking refuge in ‘value’-characteristics,” Heidegger said, “we are . . . far from even catching a glimpse of being as readiness-to-hand.”⁵

Minsky, unaware of Heidegger's critique, was convinced that representing a few million facts about objects including their functions, would solve what had come to be called the commonsense knowledge problem. It seemed to me, however, that the deep problem wasn't storing millions of facts; it was knowing which facts were relevant in any given situation. One version of this relevance problem was called “the frame problem.” If the computer is running a representation of the current state of the world and something in the world changes, how does the program determine which of its represented facts can be assumed to have stayed the same, and which would have to be updated?

As Michael Wheeler in his recent book, *Reconstructing the Cognitive World*, puts it:

[G]iven a dynamically changing world, how is a nonmagical system . . . to take account of those state changes in that world . . . that matter, and those unchanged states in that world that matter, while ignoring those that do not? And how is that system to retrieve and (if necessary) to revise, out of all the beliefs that it possesses, just those beliefs that are relevant in some particular context of action?⁶

Minsky suggested that, to avoid the frame problem, AI programmers could use what he called frames—descriptions of typical situations like going to a birthday party—to list and organize those, and only those, facts that were normally relevant. Perhaps influenced by a computer science student who had taken my phenomenology course, Minsky suggested a structure of essential features and default assignments—a structure Husserl had already proposed and already called a frame.⁷

But a system of frames isn't *in* a situation, so in order to select the possibly relevant facts in the current situation one would need frames for recognizing situations like birthday parties, and for telling them from other situations such as ordering in a restaurant. But how, I wondered, could the computer select from the supposed millions of frames in its memory the relevant frame for selecting the birthday party frame as the relevant frame, so as to see the current relevance of, say, an exchange of gifts rather than money? It seemed to me obvious that any AI program using frames to organize millions of meaningless facts so as to retrieve the currently relevant ones was going to be caught in a

⁴ John R. Searle, *The Construction of Social Reality* (New York: The Free Press, 1995).

⁵ Martin Heidegger, *Being and Time*, J. Macquarrie & E. Robinson, Trans. (New York: Harper & Row, 1962), 132–133.

⁶ Michael Wheeler, *Reconstructing the Cognitive World: The Next Step* (Cambridge, MA: A Bradford Book, The MIT Press, 2007), 179.

⁷ Edmund Husserl, *Experience and Judgment* (Evanston: Northwestern University Press, 1973), 38.

To do the same job, Roger Schank proposed what he called *scripts* such as a restaurant script. “A script,” he wrote, “is a structure that describes appropriate sequences of events in a particular context. A script is made up of slots and requirements about what can fill those slots. The structure is an interconnected whole, and what is in one slot affects what can be in another. A script is a predetermined, stereotyped sequence of actions that defines a well-known situation.” R.C. Schank and R.P. Abelson, *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures* (Hillsdale, NJ: Lawrence Erlbaum, 1977) 41. Quoted in: *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, John Preston and Mark Bishop, Eds (Oxford: Clarendon Press, 2002).

Download English Version:

<https://daneshyari.com/en/article/377427>

Download Persian Version:

<https://daneshyari.com/article/377427>

[Daneshyari.com](https://daneshyari.com)