

Dennett on Artificial Intelligence (forward by John Donovan)

Philosophers are only now beginning to understand how much of biology isn't physics, but more like engineering. I think that in some ways, like the tendency to go with what works, the practice of science is also much more like engineering than is realized. Not always pretty, but it gets results. Which is probably the biggest difference it has with philosophy and metaphysics which tends to be very pretty, but not very useful.

Philosophers often ask how natural, mechanistic and materialistic processes could possibly produce "truth" or "understanding" (as though the supernatural must be invoked for a reasonable answer). Strange as it may sound, I suspect it happens in a similar way to how science does it- by trial and error (plus some guessing based on previous experience) and testing against reality (as opposed to our intuitions). If "fine dining" can evolve from digestion, and "true love" can evolve from hormones, then I suspect that " $1 + 1 = 2$ " can evolve from a self-interest in making sure that the gathered nuts are equitably shared within the group.

Here is a recent quote on a related aspect of this question that I find most enlightening and still pointedly relevant even though it was written by Dan Dennett ("When Philosophers Encounter AI" from his book "Brainchildren") almost 20 years ago:

"How is it possible for a physical thing--a person, an animal, a robot--to extract knowledge of the world from perception and then exploit that knowledge in the guidance of successful action? That is a question with which philosophers have grappled for generations, but it could also be taken to be one of the defining questions of Artificial Intelligence. AI is, in large measure, philosophy. It is often directly concerned with instantly recognizable philosophical questions: What is mind? What is meaning? What is reasoning, and rationality? What are the necessary conditions for the recognition of objects in perception? How are decisions made and justified?

Some philosophers have appreciated this, and a few have even cheerfully switched fields, pursuing their philosophical quarries through thickets of Lisp. In general, however, philosophers have not welcomed this new style of philosophy with much enthusiasm. One might suppose that was because they had seen through it. Some philosophers have indeed concluded, after cursory inspection of the field, that in spite of the breathtaking pretension of some of its publicists, artificial intelligence has nothing new to offer philosophers beyond the spectacle of ancient, well-drubbed errors replayed in a glitzy new medium. And other philosophers are so sure this must be so that they haven't bothered conducting the cursory inspection. They are sure the field is dismissable on "general principles."

Philosophers have been dreaming about AI for centuries. Hobbes and Leibniz, in very different ways, tried to explore the implications of the idea of breaking down the mind into small, ultimately mechanical, operations. Descartes even anticipated the Turing Test (Alan Turing's much-discussed proposal of an audition of sorts for computers, in which the computer's task is to convince the judges that they are conversing with a human being) , and did not hesitate to issue a confident prediction of its inevitable result:

It is indeed conceivable that a machine could be made so that it would utter words, and even words appropriate to the presence of physical acts or objects which cause some change in its organs; as, for example, if it was touched in some spot that it would ask what you wanted to say to it; if in another, that it would cry that it was hurt, and so on for similar things. But it could never modify its phrases to reply to the sense of whatever was said in its presence, as even the most stupid men can do.¹

1. Rene' Descartes, Discourse on Method (1637), translated by Lawrence LaFleur (New York: Bobbs Merrill, 1960).

Descartes' appreciation of the powers of mechanism was colored by his acquaintance with the marvelous clockwork automata of his day. He could see very clearly and distinctly, no doubt, the limitations of that technology. Even a thousand tiny gears--even ten thousand!--would never permit an automaton to respond

gracefully and rationally! Perhaps Hobbes or Leibniz would have been less confident of this point, but surely none of them would have bothered wondering about the a priori limits on a million tiny gears spinning millions of times a second. That was simply not a thinkable thought for them. It was unthinkable then, not in the familiar philosophical sense of appearing self-contradictory ("repugnant to reason"), or entirely outside their conceptual scheme (like the concept of a neutrino), but in the more workaday but equally limiting sense of being an idea they would have had no way to take seriously. When philosophers set out to scout large conceptual domains, they are as inhibited in the paths they take by their sense of silliness as by their insights into logical necessity. And there is something about AI that many philosophers find off-putting--if not repugnant to reason, then repugnant to their aesthetic sense.

This clash of vision was memorably displayed in a historic debate at Tufts University in March of 1978, staged, appropriately, by the Society for Philosophy and Psychology. Nominally a panel discussion on the foundations and prospects of Artificial Intelligence, it turned into a tag-team rhetorical wrestling match between four heavyweight ideologues: Noam Chomsky and Jerry Fodor attacking AI, and Roger Schank and Terry Winograd defending. Schank was working at the time on programs for natural language comprehension, and the critics focused on his scheme for representing (in a computer) the higgledy-piggledy collection of trivia we all know and somehow rely on when deciphering ordinary speech acts, allusive and truncated as they are. Chomsky and Fodor heaped scorn on this enterprise, but the grounds of their attack gradually shifted in the course of the match. It began as a straightforward, "first principles" condemnation of conceptual error--Schank was on one fool's errand or another--but it ended with a striking concession from Chomsky: it just might turn out, as Schank thought, that the human capacity to comprehend conversation (and more generally, to think) was to be explained in terms of the interaction of hundreds or thousands of jerry-built gizmos--pseudo-representations, one might call them--but that would be a shame, for then psychology would prove in the end not to be "interesting." There were only two interesting possibilities, in Chomsky's mind: psychology could turn out to be "like physics"--its regularities explainable as the consequences of a few deep, elegant, inexorable laws--or psychology could turn out to be utterly lacking in laws, in which case the only way to study or expound psychology would be the novelist's way (and he much preferred Jane Austen to Roger Schank, if that were the enterprise).

A vigorous debate ensued among the panelists and audience, capped by an observation from Chomsky's MIT colleague, Marvin Minsky, one of the founding fathers of AI, and founder of MIT's AI Lab: "I think only a humanities professor at MIT could be so oblivious to the third interesting possibility: psychology could turn out to be like engineering."

Minsky had put his finger on it. There is something about the prospect of an engineering approach to the mind that is deeply repugnant to a certain sort of humanist, and it has little or nothing to do with a distaste for materialism or science. Witness Chomsky's physics-worship, an attitude he shares with many philosophers. The days of Berkeleyan idealism and Cartesian dualism are over (to judge from the current materialistic consensus among philosophers and scientists), but in their place there is a widespread acceptance of what we might call Chomsky's fork: there are only two appealing ("interesting") alternatives.

On the one hand, there is the dignity and purity of the Crystalline Mind. Recall Aristotle's prejudice against extending earthly physics to the heavens, which ought, he thought, to be bound by a higher and purer order. This was his one pernicious legacy, but now that the heavens have been stormed, we appreciate the beauty of universal physics, and can hope that the Mind will be among its chosen, "natural kinds", not a mere gerrymandering of bits and pieces.

On the other hand, there is the dignity of ultimate mystery, the Inexplicable Mind. If our minds can't be Fundamental, then let them be Anomalous. A very influential view among philosophers in recent years has been Donald Davidson's "anomalous monism", the view that while the mind is the brain, there are no lawlike regularities aligning the mental facts with the physical facts. His Berkeley colleague, John Searle, has made a different sort of mystery of the mind: the brain, thanks to some unspecified feature of its biochemistry, has some terribly important--but unspecified--"bottom-up causal powers" that are entirely distinct from the mere "control powers" studied by AI.

One feature shared by these otherwise drastically different forms of mind-body materialism is a resistance

to Minsky's tertium quid: in between the Mind as Crystal and the Mind as Chaos lies the Mind as Gadget, an object which one should not expect to be governed by "deep," mathematical laws, but nevertheless a designed object, analyzable in functional terms: ends and means, costs and benefits, elegant "solutions" on the one hand, and on the other, shortcuts, jury-rigs, and cheap ad hoc fixes.

This vision of the mind is resisted by many philosophers despite being a straightforward implication of the current received view (among scientists and science-minded humanists) of Our Place in Nature: we are biological entities, designed by natural selection, which is a tinker, not an ideal engineer. Computer programmers call an ad hoc fix a "kludge"--to rhyme with Scrooge--and the mixture of disdain and begrudged admiration reserved for kludges parallels the biologists' bemusement with the panda's thumb and other fascinating examples of bricolage, to use Francois Jacob's term. "