

This work originally appeared in:

Dennett, Daniel. 1982. The Myth of the Computer: An Exchange. *New York Review of Books* XXIX: 56.

It is available online at:

<http://www.nybooks.com/articles/6542>

This is Daniel C. Dennett's final draft before publication. It has been modified to reflect the pagination of the published version of the work.

# The Myth of the Computer: An Exchange

*To the Editors:*

In *The Mind's I*, Douglas Hofstadter and I reprint (correctly) John Searle's much-discussed article, "Minds, Brains, and Programs," and follow it with a "Reflection" that is meant to refute his position, as he notes in his review (NYR, April 29). Searle charges that in that Reflection we "fabricate a direct quotation" which moreover "runs dead opposite" to what he in fact says. The Pocket OED says "fabricate" means "invent (Ile., etc.)" forge(document)" so Searle is suggesting (at some length) that this is a deliberate misquotation--a very serious charge which we categorically deny.

Here are the facts. We do misquote him in the Reflection, alas; we have him saying "a few slips of paper" where he in fact says "bits of paper." This misquotation was entirely inadvertent: we apologize to him for it: we have arranged for the error to be corrected in any future printings of the book.

Now, does the error make a difference worth mentioning? Searle claims it does. He claims that the misquotation is "the basis" of our argument, which could not proceed without it, since it "runs dead opposite" to his meaning. We do, as he says, repeat the error five times (In effect. we got off on the wrong foot and then quoted our own error four times.) But so little does our case depend on the misquotation, that once it is corrected no further revision-- not so much as a word or comma-of our Reflection is called for or contemplated.

How could Searle think "a few slips of paper" differs so dramatically from "bits of paper"? We had better look at the context from which we have (mis)taken the fatal phrase. Here is what Searle says, as printed correctly on p. 359 of *The Mind's I*:

The Idea is that while a person doesn't understand Chinese, somehow the conjunction of that person and bits of paper might understand Chinese. It is not easy for me to imagine how someone who was not in the grip of an ideology would find that idea at all plausible.

Here Searle is ridiculing what he calls "the systems reply" to his view, and as he admits, he has a hard time taking it seriously. That is one of the points we were trying to make. He also says, in his review: "The mental gymnastics that partisans of strong AI have performed in their attempts to refute this rather simple argument (his "Chinese Room" thought experiment) are truly extraordinary."

Here we have the spectacle of an eminent philosopher going around the country trotting out a "rather simple argument" and then marveling at the obtuseness of his audiences, who keep trying to show him what's wrong with it. He apparently cannot bring himself to contemplate the possibility that he might be missing a point or two, or underestimating the opposition. As he notes in his review, no less than twenty-seven rather eminent people responded to his article when it first appeared in *Behavioral and Brain Sciences*, but since he repeats its claims almost verbatim in the review, it seems that the only lesson he has learned from that response was that there are several dozen fools in the world. (Several dualists, including Sir John Eccles, the Nobel laureate neurophysiologist, sided with Searle.)

We claim that he has frankly misunderstood the systems reply and that his remark about "bits of paper" betrays this—and has "blinded him to the realities of the situation." Sometimes it even seems as if he deliberately misrepresents the systems reply, as when he says in his review: "Adherents of this view believe, to my constant amazement, that thoughts man fails to understand, the room understands Chinese." Searle's amazement stops just short of inspiring any doubt in his mind about the fidelity of his interpretation, but perhaps this is to be explained by a certain exegetical carelessness rather than willful caricature.

What is the heart of the system reply? It is a distinction of levels that is not at all mysterious, or new, though Searle's diminutive "bits of paper" act to minimize(or obfuscate) the point. "The conjunction of a person and bits of paper" doesn't sound like a very different system from a person alone, does it? How about "the conjunction of a person and the library of Congress with its attendant staff"? Does that sound like a supersystem that just might have some interesting powers or properties lacked by any of its proper parts or subsystems? The latter comparison should suit Searle just fine, if (as he now claims) he meant his "bits of paper" to carry no diminutive implications. And it is fairer, since Searle is supposedly talking about an imagined super-program that passes any and all Turing tests, a program many orders of magnitude grander than anything yet written.

Searle, in a letter to me (which he has kindly permitted me to quote), says:

In any case you and Hofstadter still miss the point. No matter how big the program, the conjunction of man and bits of paper is no different from man alone. All of the bits of paper in the world add nothing to the neurophysiological powers of the man's brain. The whole point of reminding the reader that these are just "bits of paper" is that they are not in any way an addition to the specific neurophysiological powers of the man's brain.

Here Searle manifestly misunderstands the systems reply. No one claims the supersystem gives the subsystem by itself special new powers or properties. Rather, we (and many others) claim that the supersystem itself—the whole supersystem—has these powers. Searle's persistent deaf ear to this point puzzles me, particularly since it is really just a "category mistake" claim of the sort that was all the rage during Searle's graduate student days at Oxford. In his reply to my earlier commentary on his paper (In *Behavioral and Brain Sciences*)he objects to my rather Oxonian claim that I understand English-my brain doesn't-with the retort: "I find his claim as implausible as insisting, 'I digest pizza: my stomach and my digestive tract don't.'" How important a single word can be! The verb "digest" is nicely chosen, for note how radically the image shifts if we switch to "eat" or "enjoy." Does Searle find it quite all right to say that his stomach eats pizza? Can his mouth eat pizza? Which proper part of him could be said to enjoy the pizza? Levels do make a difference. Anyone who hunts for a pizza-enjoying subsystem in a human being is on a fool's errand, and anyone who denies that a supersystem understands Chinese on the grounds that none of its subsystems do is making the same error moving in the other direction.

This error is hidden in the flurry (or is it a mountain) of bits of paper. Searle's original article abounds in misdirection of this sort. Is it deliberate or inadvertent? Searle objects to our giving him the benefit of the doubt and calling his phrase "casual" and "offhand."

Would he prefer us to all it deliberate miscorrections? In my earlier commentary in *Behavioral and Brain Sciences* I described his article as "sophistry," but Hofstadter and I took a more charitable line in our volume. We, unlike Searle, do not pretend to be able to divine intention in the slips of our opponents.

We are sorry we slipped over "a few slips," but if Searle actually thinks this was a deliberate "fabrication"-or that our case against his view depends on misquotation—he has deluded himself.

As for the rest of Searle's review, it contains much to which we object, but we have pre-refuted virtually all of it, point by point, in the book he was reviewing. Indeed, Searle's review is, with perhaps one novelty, simply a telescoped version of his article. Searle may think that "Say it again, faster, in the pages of *The New York Review*" is a sound tactic of persuasion, but we don't. So for the most part we are content to refer readers who want to figure out what is wrong with Searle's view to our book. The one somewhat new element in the review is the enlargement on his unusual idea that we ignore the "causal power of the brain," and since one can easily misread Searle on this point, a little clarification is in order.

Searle stresses that a computer program, being "purely formal," has no causal power of its own. True, but of course when a program is physically realized in some hardware, and attacked by "transducers and effectors" to relevant portions of the rest of the world, that physically realized program can have lots of causal powers: such a program can control an oil refinery, make out payroll checks or-terrible to say-guide nuclear missiles to their targets. Let's call such causal powers control powers. Such powers are not simulated but real; the computer doesn't simulate controlling the refinery; it really does control the refinery. (The distinction between simulating and duplicating is not as unproblematic as Searle supposes, but we will give him the distinction here for the sake of argument.)

Now Searle has admitted (in conversation on several occasions) that in his view a computer program, physically realized on a silicon chip (or for that matter a beer-can contraption suitably sped up and hooked up) could in principle duplicate--not merely simulate--the control powers of the human brain. That is, such a computer program (somehow realized) could control a human body in all its activities. Would such a body have a mind? We on the outside would find its behavior indistinguishable from that of a normal human being but whether or not it really had a mind would depend, Searle insists on whether the hardware realization of the control program shared with the missing brain not only all its control powers (granted ex hypothesis) but also some other "causal powers" entirely undetectable by others in behavior, including the behaviors of introspective speech, emotional reaction, and so forth.

What powers could these be? Where would the physical effects of the, neurophysiological powers show up? Searle answers that they would show up in the individual subject's consciousness of his own intentionality. But would these be physical effects? If so, they must be detectable (in principle) by outsiders. Would they register on the instruments of neuroscientists(if not "behaviorists")? Searle does not say, but since he insists that the effects are introspectible (only?) it is tempting to conclude that the effects are presumed to be nonphysical, and that Searle is some sort of dualist. He adamantly denies it: he insists the causal powers he is discussing are physical, so they must have physical, publicly observable effects. Where, if not in the subject's behavior? Just in the brain? What would these effects do?

These are mysterious causal powers indeed, despite their scientific-sounding name. We frankly disbelieve in them--which is the extent of our "behaviorism." Surely we all agree that anything that has all the relevant causal powers of food--it saves one from starving, sustains growth and repair, tastes good, etc.-- is food. And anything that has all the causal power of oxygen is oxygen. We think that you could in principle give a body an artificial brain by giving it something that duplicated all the brain's control powers. And any creature so equipped would "have a mind" in the only sense that makes any sense: it would have a well-functioning (prosthetic) brain. Now perhaps we are wrong: perhaps there are some other causal powers that matter. Searle thinks so; he thinks organic brains "produce intentionality." It sometimes seems as if he thinks intentionality is some marvelous fluid secreted by the brain-but we shrink from imputing such a silly view to him, and await his further clarification of his position.

Searle paints us as taken in by the "mythology" of computers. We see ourselves as demythologizers, and Searle as the victim of several superannuated myths, but perhaps we have misinterpreted his view.

**Daniel C. Dennett**

*Tufts University*

*Medford, Massachusetts*