

BOOK REVIEWS

The Philosophical Review, XCVI, No. 1 (January 1987)

MINDS, BRAINS AND SCIENCE. By JOHN SEARLE. Cambridge, Mass., Harvard University Press, 1984. Pp. 107.

In his intellectual autobiography, David Armstrong paraphrases the bittersweet judgment of Australian poet James McAuley: "Despite all that can be said against the philosophy of our age, what a time it has been to be a philosopher!"¹ What a time indeed. Recent years have been witness to extraordinary developments in American philosophy, among which perhaps the most encouraging has been the maturation of the seeds planted by Dewey more than a half century ago. Dewey advocated the wholesale rejection of boundaries separating philosophy from the sciences, and in one domain after another those boundaries have indeed begun to dissolve. Nowhere is this development more pronounced than in contemporary cognitive science, where psychologists, neuro-scientists, linguists and computer scientists have increasingly accepted philosophers as full partners in their venture. The surest sign of this acceptance is the fact that even the philosophical critics of cognitive science are taken seriously, not only by fellow philosophers, but also by large numbers of cognitive scientists whose training is in other disciplines.

John Searle has been the most persistent of these critics, the most plain spoken, and by far the most influential. In this slim volume, drawn from his Reith lectures on the BBC, Searle sets out his critique in a format intended to be "completely accessible to an interested and alert audience most of whose members have no familiarity whatever with the subject matter, with its terminology, or with the special preoccupations of its practitioners" (p. 7). On this score, the volume is an unqualified success. Searle's six brief chapters are models of straightforward, vigorous, non-technical argument. Yet accessibility, for Searle, is not bought at the price of pulling his punches. All the boldness of his controversial critique is here.

The first three chapters of the volume are devoted to setting out Searle's view of the mind-body problem, and developing his critiques of functionalism (or "strong AI") and cognitive science. In the fourth chapter he sketches his account of the structure of human action, an account which entails that "actions differ from other natural events in the world" (p. 67). The fifth chapter is devoted to a discussion of the "prospects for the social sciences." According to Searle, there is a "radical discontinuity between the

¹D. M. Armstrong, "Self-Profile," in *D. M. Armstrong*, Radu Bogdan, ed. (Dordrecht, Holland: D. Reidel Publishing Co., 1984), p. 49.

social and the natural sciences,” (p. 84) since “the social sciences in general are about various aspects of intentionality” (p. 82). Searle’s last chapter offers a rather pessimistic perspective on the problem of free will. Contemporary science, he maintains, cannot be reconciled with “the doctrine of human freedom.” In order for us to have free will, “it looks as if we would have to postulate that inside each of us was a self that was capable of interfering with the causal order of nature. That is, it looks as if we would have to contain some entity that was capable of making molecules swerve from their paths” (p. 92). Such a view, Searle maintains, is “certainly not consistent with what we know about how the world works from physics. And there is not the slightest evidence to suppose that we should abandon physical theory in favor of such a view” (p. 92). But Searle also argues that we cannot “give up the conviction of freedom because that conviction is built into every normal, conscious intentional action” (p. 97). We are thus condemned to a world view that is in conflict with itself.

All of this heady and provocative stuff makes Searle’s book an exciting read, and a splendid choice for the instructor who wants to enliven a course in cognitive science or the philosophy of mind. The publisher has made it more attractive still by issuing a reasonably priced paperback edition. Yet having said so many nice things about this admirable little book, I hasten to add a pair of caveats: On my view, there is hardly a single controversial claim in the book that is not seriously mistaken, and there is hardly a single major argument that is sound.

Since much of the volume echoes positions Searle has elaborated elsewhere, there is already a substantial critical literature taking on both his arguments and his conclusions. In the remaining pages of the current review, I will focus on a point that has received relatively little attention in the literature. What I shall argue is that, taken together, Searle’s central views about the nature of the mind are incoherent. More specifically, I shall argue that his view about the nature of mental states, if true, would undermine his own argument against strong AI and in favor of the thesis that “biology matters.” For if he were right about mental states, then he could not possibly know that the central claim in his argument against strong AI is true.

Let me begin with Searle’s view about the nature of mental states. His first chapter lists “four features of mental phenomena which have made them seem impossible to fit into our ‘scientific’ conception of the world” (p. 15). Despite these difficulties, Searle insists, the four “are real features of our mental lives. . . . If your theory ends up by denying any one of them, you know you must have made a mistake somewhere” (p. 17). The four features on Searle’s list are *consciousness*, *intentionality*, *subjectivity*, and *mental causation* (the capacity of thoughts and feelings to have some causal effect

on the physical world). Just prior to setting out this list, Searle suggests that conscious states are not only real, they are *irreducible*—“irreducible as anything else in the universe” (p. 15). It is the third item on this list, subjectivity, that will be the focus of my argument. What subjectivity amounts to, according to Searle, is a kind of privileged access. “My present state of consciousness is a feature of my brain, but its conscious aspects are accessible to me in a way that they are not accessible to you” (p. 25).

A second component of Searle’s view about mental states is that “biology matters” (p. 40). “Mental phenomena, all mental phenomena . . . indeed all of our mental life, are caused by processes going on in the brain” (p. 18). He rejects the functionalist or “strong artificial intelligence” view that “instantiating the right computer program with the right inputs and outputs [is] sufficient for, or constitutive of, thinking” (p. 36). Searle’s point is not that a normal, conscious, thinking human brain cannot be viewed as a computer. For, as he rightly points out, “anything whatever can be described *as if* it were a digital computer” (p. 36). Rather, his point is that instantiating the right program is not *sufficient* for consciousness or thought. “No computer program by itself is sufficient to give a system a mind” (p. 39). Unlike his Berkeley colleague, Hubert Dreyfus, Searle quite sensibly does not claim that electronic computers could not even simulate such intelligent activities as playing world class chess, writing good poetry or doing decent science. What he does claim is that such simulations of intelligent behavior, if they are indeed possible, would be *only* simulations; they would not be duplications. An electronic computer that simulated a person involved in a conversation would not really understand the sentences it produced, nor would it be consciously aware of the sentences it heard. Searle is careful to note that mentality need not be restricted exclusively to creatures with brains like our own. Martians with “green slime” in their heads might perfectly well have conscious mental lives quite like our own, so long as the green slime had “causal powers equal to those of the human brain” (p. 41).

The trouble with all this is that if Searle is right about the subjective nature of mental states, then it is hard to see how we could possibly know whether the Martians really had mental lives. How could we possibly know that the green slime was duplicating subjective mental processes, rather than just simulating it? It is equally hard to see how Searle could possibly know that instantiating the right computer program is *not* sufficient for real, subjective mentality. For recall that Searle is willing to concede that instantiating the right program might well be sufficient for *simulating* mentality. And a computer driven robot simulating a well-taught Berkeley graduate student might well insist that the “conscious aspects” of its mental life “are accessible to me in a way that they are not accessible to you” (p.

25). If the simulation is flawless, and if consciousness is indeed subjective and directly accessible only to the conscious agent himself, then how could Searle possibly know whether or not the robot has a mental life?

The point can be made a bit more vividly with the help of a philosophical shaggy dog story.² Suppose that after your annual check-up your doctor tells you that there is bad news and good news. The bad news is that a microscopic tumor has been detected in your brain which, if not removed, will kill you. The good news is that there is a new process enabling the surgical team to connect up a tiny micro-chip that mimics the nerve impulse transmission functions of the nerve cells they remove. Like all shaggy dog stories, this one iterates. One tumor after another is discovered and replaced. Finally there is nothing left in your head except micro-chips, though *ex hypothesis* those around you detect no change in your behavior. Searle, as we have seen, is committed to the view that by the time your biological brain has been completely replaced by an electronic one, your mental lights may have gone out. The mere fact that the electronic brain mimics the program of your biological brain is not causally sufficient to guarantee that it will produce the unique, subjective states and processes that are fundamental to mentality. But just when does Searle suppose the lights go out? Do they go out gradually, or is there some crucial bit of biological circuitry on which subjective consciousness depends? Among the many different logically possible electronic computers that might simulate my brain and my behavior, are there some that, like green slime, actually are conscious because they have “causal powers at least equivalent to those of the brain?” If so, which ones are these? And, more important, since mental states are subjective, *how could we possibly know?*

Searle’s only serious defense for the claim that “no computer program by itself is sufficient to give a system a mind” (p. 39) is the by now well known “Chinese Room” example in which a person who understands no Chinese is put inside a closed room where he runs through all the steps that a computer program would run through in simulating a fluent Chinese speaker holding a conversation in Chinese. *Ex hypothesis*, the system which consists of the person plus the rules plus the facilities for accepting input and issuing output, is an excellent simulation of a Chinese speaker. But, Searle claims, there is no real understanding here, no intentionality. To convince us of this he relies on the stipulation that the person in the room understands no Chinese, along with the intuition that nothing else here

²I have been using this example in my Introduction to Philosophy lectures for a dozen years or more, and I doubt it is original with me. Where I first heard it, and from whom, are past recalling.

does either—not the room or the rules or the “system.” But this argument comes to grief if we take Searle’s account of the mind seriously. For, on that account mental states, including intentional states, are the “real” and “irreducible” causal products of certain complex physical systems. My brain produces real consciousness and real intentionality; I know they exist because I have subjective access to them. Other brains, including yours, which is physically similar to mine, and the Martian’s, which is made of green slime, *might* produce real mental states as well. Whether or not your brain or the Martian’s really does produce mental states is, on Searle’s view, a brute matter of fact, not a question that can be settled by intuition. The problem I am belaboring is that it is not clear how any evidence whatever that I might have could bear on the question of whether you and the Martian are merely simulating conscious mentality or actually experiencing it. And, of course, everything that has been said about your brain and the Martian’s can be said about the Chinese Room “system” as well. On Searle’s view of mental states, we have no way at all of finding out whether or not this system is really producing genuine intentional states or merely simulating them.

In his first chapter, Searle tells us that one of his aims is to try to break out of the “tired old categories” of the philosophy of mind (p. 14). But in this effort Searle has failed rather badly. According to him, mental states are subjective, irreducible states of the brain, and that is enough to place him squarely within one of those “tired old categories.” Searle is a property dualist. Like any dualist, he is plagued by the problem of other minds—the problem of how I could possibly know that anyone or anything else in the universe has mental states. But unlike most other dualists, Searle refuses to address or even acknowledge the problem. He seems to think that if he mentions brains and biology often enough, somehow the problem will just go away.

Searle suggests that progress on the mind-body problem has been inhibited by “an inherited cultural resistance to treating the conscious mind as a biological phenomenon like any other” (p. 10). And I agree. But if we are to take this idea seriously we must renounce just those claims about the mind that condemn Searle to dualism. No other biological phenomenon is either subjective or irreducible. If we are to resolve the philosophical puzzles about the mind we will have to recognize that thinking about the mind in those terms is a mistake.

STEPHEN P. STICH

University of California, San Diego