

One of the Ten Best Books of the Year—New York Times

CONSCIOUSNESS EXPLAINED



DANIEL C. DENNETT

Author of Brainstorms and coauthor of The Mind's I

stuff. The conscious mind, it seems, cannot just be the brain, or any proper part of it, because nothing in the brain could

- (1) be the medium in which the purple cow is rendered;
- (2) be the thinking thing, the I in "I think, therefore I am";
- (3) appreciate wine, hate racism, love someone, be a source of mattering;
- (4) act with moral responsibility.

An acceptable theory of human consciousness must account for these four compelling grounds for thinking that there must be mind stuff.

4. WHY DUALISM IS FORLORN

The idea of mind as distinct in this way from the brain, composed not of ordinary matter but of some other, special kind of stuff, is dualism, and it is deservedly in disrepute today, in spite of the persuasive themes just canvassed. Ever since Gilbert Ryle's classic attack (1949) on what he called Descartes's "dogma of the ghost in the machine," dualists have been on the defensive.² The prevailing wisdom, variously expressed and argued for, is materialism: there is only one sort of stuff, namely matter — the physical stuff of physics, chemistry, and physiology — and the mind is somehow nothing but a physical phenomenon. In short, the mind is the brain. According to the materialists, we can (in principle!) account for every mental phenomenon using the same physical principles, laws, and raw materials that suffice to explain radioactivity, continental drift, photosynthesis, reproduction, nutrition, and growth. It is one of the main burdens of this book to explain consciousness without ever giving in to the siren song of dualism. What, then, is so wrong with dualism? Why is it in such disfavor?

The standard objection to dualism was all too familiar to Descartes himself in the seventeenth century, and it is fair to say that neither he nor any subsequent dualist has ever overcome it convincingly. If mind

2. A few brave souls (and they surely cannot object to being so categorized!) have bucked the tide: Arthur Koestler's defiantly titled *The Ghost in the Machine* (1967) and Popper and Eccles's *The Self and Its Brain* (1977) are by unquestionably eminent authors, and two other iconoclastic and quirkily insightful defenses of dualism are Zeno Vendler's *Res Cogitans* (1972) and *The Matter of Minds* (1984).

and body are distinct things or substances, they nevertheless must interact; the bodily sense organs, via the brain, must *inform* the mind, must send to it or present it with perceptions or ideas or data of some sort, and then the mind, having thought things over, must *direct* the body in appropriate action (including speech). Hence the view is often called Cartesian interactionism or interactionist dualism. In Descartes's formulation, the locus of interaction in the brain was the pineal gland, or *epiphysis*. It appears in Descartes's own schematic diagram as the much-enlarged pointed oval in the middle of the head.

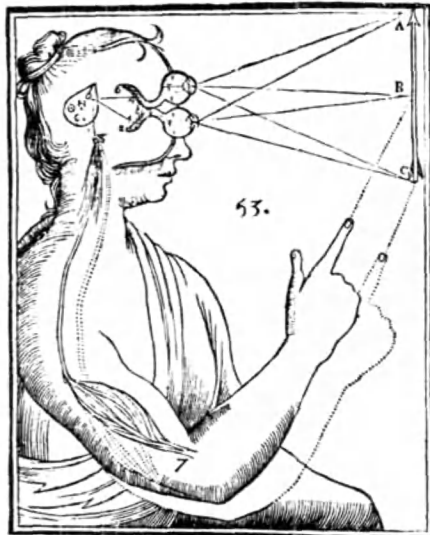


Figure 2.1

We can make the problem with interactionism clear by superimposing a sketch of the rest of Descartes's theory on his diagram (Figure 2.2).

The conscious perception of the arrow occurs only after the brain has somehow transmitted its message to the mind, and the person's finger can point to the arrow only after the mind commands the body. How, precisely, does the information get transmitted from pineal gland to mind? Since we don't have the faintest idea (yet) what properties mind stuff has, we can't even guess (yet) how it might be affected by physical processes emanating somehow from the brain, so let's ignore those upbound signals for the time being, and concentrate on the return signals, the directives from mind to brain. These, *ex hypothesi*, are not physical; they are not light waves or sound waves or cosmic rays or



Figure 2.2

streams of subatomic particles. No physical energy or mass is associated with them. How, then, do they get to make a difference to what happens in the brain cells they must affect, if the mind is to have any influence over the body? A fundamental principle of physics is that any change in the trajectory of any physical entity is an acceleration requiring the expenditure of energy, and where is this energy to come from? It is this principle of the conservation of energy that accounts for the physical impossibility of “perpetual motion machines,” and the same principle is apparently violated by dualism. This confrontation between quite standard physics and dualism has been endlessly discussed since Descartes’s own day, and is widely regarded as the inescapable and fatal flaw of dualism.

Just as one would expect, ingenious technical exemptions based on sophisticated readings of the relevant physics have been explored and expounded, but without attracting many conversions. Dualism’s embarrassment here is really simpler than the citation of presumed laws of physics suggests. It is the same incoherence that children notice — but tolerate happily in fantasy — in such fare as *Casper the Friendly Ghost* (Figure 2.3, page 36). How can Casper both glide through walls and grab a falling towel? How can mind stuff both elude all physical measurement and control the body? A ghost in the machine is of no help in our theories unless it is a ghost that can move things around — like a noisy poltergeist who can tip over a lamp or slam a door — but anything that can move a physical thing is itself a physical thing (although perhaps a strange and heretofore unstudied kind of physical thing).

What about the option, then, of concluding that mind stuff is

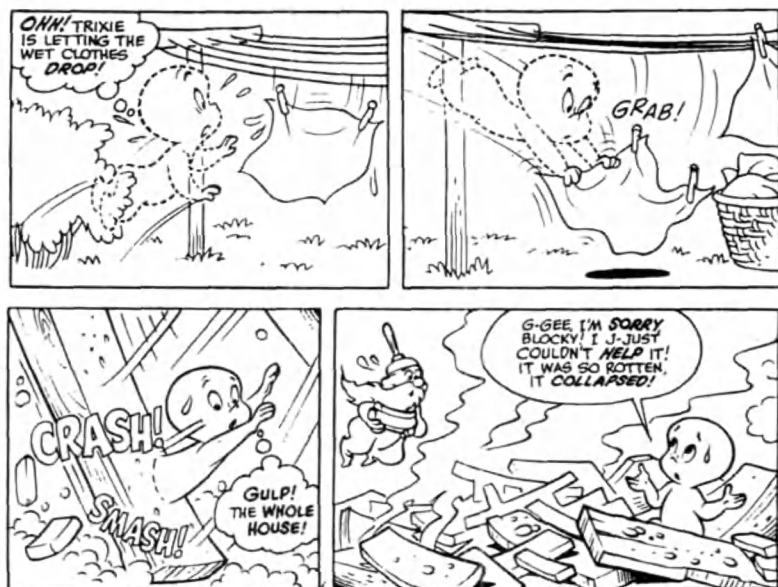


Figure 2.3

actually a special kind of matter? In Victorian séances, the mediums often produced out of thin air something they called "ectoplasm," a strange gooeey substance that was supposedly the basic material of the spirit world, but which could be trapped in a glass jar, and which oozed and moistened and reflected light just like everyday matter. Those fraudulent trappings should not dissuade us from asking, more soberly, whether mind stuff might indeed be something above and beyond the atoms and molecules that compose the brain, but still a scientifically investigatable kind of matter. The ontology of a theory is the catalogue of things and types of things the theory deems to exist. The ontology of the physical sciences used to include "caloric" (the stuff heat was made of, in effect) and "the ether" (the stuff that pervaded space and was the medium of light vibrations in the same way air or water can be the medium of sound vibrations). These things are no longer taken seriously, while neutrinos and antimatter and black holes are now included in the standard scientific ontology. Perhaps some basic enlargement of the ontology of the physical sciences is called for in order to account for the phenomena of consciousness.

Just such a revolution of physics has recently been proposed by the physicist and mathematician Roger Penrose, in *The Emperor's New Mind* (1989). While I myself do not think he has succeeded in making

his case for revolution,³ it is important to notice that he has been careful not to fall into the trap of dualism. What is the difference? Penrose makes it clear that he intends his proposed revolution to make the conscious mind more accessible to scientific investigation, not less. It is surely no accident that the few dualists to avow their views openly have all candidly and comfortably announced that they have no theory whatever of how the mind works — something, they insist, that is quite beyond human ken.⁴ There is the lurking suspicion that the most attractive feature of mind stuff is its promise of being so mysterious that it keeps science at bay forever.

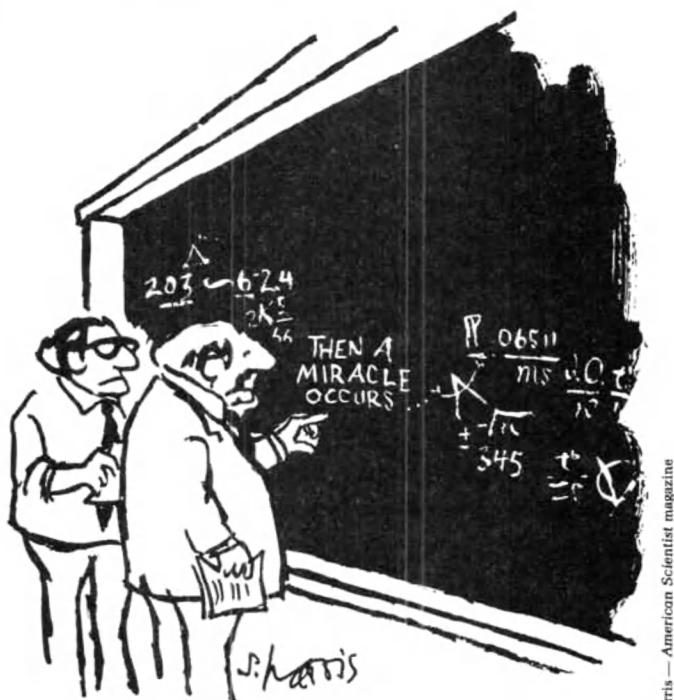
This fundamentally antiscientific stance of dualism is, to my mind, its most disqualifying feature, and is the reason why in this book I adopt the apparently dogmatic rule that dualism is to be avoided at all costs. It is not that I think I can give a knock-down proof that dualism, in all its forms, is false or incoherent, but that, given the way dualism wallows in mystery, *accepting dualism is giving up* (as in Figure 2.4, page 38).

There is widespread agreement about this, but it is as shallow as it is wide, papering over some troublesome cracks in the materialist wall. Scientists and philosophers may have achieved a consensus of sorts in favor of materialism, but as we shall see, getting rid of the old dualistic visions is harder than contemporary materialists have thought. Finding suitable replacements for the traditional dualistic images will require some rather startling adjustments to our habitual ways of thinking, adjustments that will be just as counterintuitive at first to scientists as to laypeople.

I don't view it as ominous that my theory seems at first to be strongly at odds with common wisdom. On the contrary, we shouldn't expect a good theory of consciousness to make for comfortable reading — the sort that immediately "rings bells," that makes us exclaim to ourselves, with something like secret pride: "Of course! I knew that all along! It's obvious, once it's been pointed out!" If there were any such theory to be had, we would surely have hit upon it by now. The mysteries of the mind have been around for so long, and we have made

3. See "Murmurs in the Cathedral" (Dennett, 1989c), my review of his book.

4. Eccles has proposed that the nonphysical mind is composed of millions of "psychons," which interact with millions of "dendrons" (tracts of pyramidal cells) in the cortex: each psychon corresponds roughly to what Descartes or Hume would call an idea — such as the idea of red, or the idea of round, or hot — but aside from this minimal decomposition, Eccles has nothing to say about the parts, activities, principles of action, or other properties of the nonphysical mind.



"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Figure 2.4

© 1975 Sidney Harris — American Scientist magazine

so little progress on them, that the likelihood is high that some things we all tend to agree to be obvious are just not so. I will soon be introducing my candidates.

Some brain researchers today — perhaps even a stolid majority of them — continue to pretend that, for them, the brain is just another organ, like the kidney or pancreas, which should be described and explained only in the most secure terms of the physical and biological sciences. They would never dream of mentioning the mind or anything "mental" in the course of their professional duties. For other, more theoretically daring researchers, there is a new object of study, the mind/brain (Churchland, 1986). This newly popular coinage nicely expresses the prevailing materialism of these researchers, who happily admit to

the world and themselves that what makes the brain particularly fascinating and baffling is that somehow or other it is the mind. But even among these researchers there is a reluctance to confront the Big Issues, a desire to postpone until some later date the embarrassing questions about the nature of consciousness.

But while this attitude is entirely reasonable, a modest recognition of the value of the divide-and-conquer strategy, it has the effect of distorting some of the new concepts that have arisen in what is now called cognitive science. Almost all researchers in cognitive science, whether they consider themselves neuroscientists or psychologists or artificial intelligence researchers, tend to postpone questions about consciousness by restricting their attention to the "peripheral" and "subordinate" systems of the mind/brain, which are deemed to feed and service some dimly imagined "center" where "conscious thought" and "experience" take place. This tends to have the effect of leaving too much of the mind's work to be done "in the center," and this leads theorists to underestimate the "amount of understanding" that must be accomplished by the relatively peripheral systems of the brain (Dennett, 1984b).

For instance, theorists tend to think of perceptual systems as providing "input" to some central thinking arena, which in turn provides "control" or "direction" to some relatively peripheral systems governing bodily motion. This central arena is also thought to avail itself of material held in various relatively subservient systems of memory. But the very idea that there are important theoretical divisions between such presumed subsystems as "long-term memory" and "reasoning" (or "planning") is more an artifact of the divide-and-conquer strategy than anything found in nature. As we shall soon see, the exclusive attention to specific subsystems of the mind/brain often causes a sort of theoretical myopia that prevents theorists from seeing that their models still presuppose that somewhere, conveniently hidden in the obscure "center" of the mind/brain, there is a Cartesian Theater, a place where "it all comes together" and consciousness happens. This may seem like a good idea, an inevitable idea, but until we see, in some detail, why it is not, the Cartesian Theater will continue to attract crowds of theorists transfixed by an illusion.

5. THE CHALLENGE

In the preceding section, I noted that if dualism is the best we can do, then we can't understand human consciousness. Some people are