

What Neuroscience Cannot Tell Us About Ourselves

Raymond Tallis

There has been much breathless talk of late about all the varied mysteries of human existence that have been or soon will be solved by neuroscience. As a clinical neuroscientist, I could easily expatiate on the wonders of a discipline that I believe has a better claim than mathematics to being Queen of the Sciences. For a start, it is a science in which many other sciences converge: physics, biology, chemistry, biophysics, biochemistry, pharmacology, and psychology, among others. In addition, its object of study is the one material object that, of all the material objects in the universe, bears most closely on our lives: the brain, and more generally, the nervous system. So let us begin by giving all proper respect to what neuroscience can tell us about ourselves: it reveals some of the most important conditions that are *necessary* for behavior and awareness.

What neuroscience does not do, however, is provide a satisfactory account of the conditions that are *sufficient* for behavior and awareness. Its descriptions of what these phenomena are and of how they arise are incomplete in several crucial respects, as we will see. The pervasive yet mistaken idea that neuroscience does fully account for awareness and behavior is *neuroscientism*, an exercise in science-based faith. While to live a human life requires having a brain in some kind of working order, it does not follow from this fact that to live a human life is to *be* a brain in some kind of working order. This confusion between necessary and sufficient conditions lies behind the encroachment of “neuroscientific” discourse on academic work in the humanities, and the present epidemic of such neuro-prefixed pseudo-disciplines as neuroaesthetics, neuroeconomics, neurosociology, neuropolitics, neurotheology, neurophilosophy, and so on.

The failure to distinguish consciousness from neural activity corrodes our self-understanding in two significant ways. If we are just our brains, and our brains are just evolved organs designed to optimize our odds of

Raymond Tallis, emeritus professor of geriatric medicine at the University of Manchester, United Kingdom, is the author, most recently, of Michelangelo's Finger: An Exploration of Everyday Transcendence (Yale, 2010) and Aping Mankind: Neuromania, Darwinitis and the Misrepresentation of Humanity (Acumen, forthcoming in 2011). This essay has been adapted from a lecture delivered in February 2010 at the American Enterprise Institute.

FALL 2010 ~ 3

survival—or, more precisely, to maximize the replicative potential of the genetic material for which we are the vehicle—then we are merely beasts like any other, equally beholden as apes and centipedes to biological drives. Similarly, if we are just our brains, and our brains are just material objects, then we, and our lives, are merely way stations in the great causal net that is the universe, stretching from the Big Bang to the Big Crunch.

Most of those who subscribe to such “neuroevolutionary” accounts of humanity don’t recognize these consequences. Or, if they do recognize them, then they don’t subscribe to these accounts sincerely. When John Gray appeals, in his 2002 book *Straw Dogs*, to a belief that human beings are merely animals and so “human life has no more meaning than the life of slime mold,” he doesn’t really believe that the life of John Gray, erstwhile School Professor of European Thought at the London School of Economics, has no more meaning than that of a slime mold—else why would he have aspired to the life of a distinguished professor rather than something closer to that of a slime mold?

Wrong ideas about what human beings are and how we work, especially if they are endlessly repeated, keep us from thinking about ourselves in ways that may genuinely advance our self-understanding. Indeed, proponents of the neuroscientific account of human behavior hope that it will someday supplant our traditional understandings of mind, behavior, and consciousness, which they dismiss as mere “folk psychology.” According to a 2007 *New Yorker* profile of professors Paul and Patricia Churchland, two leading “neurophilosophers,” they like “to speculate about a day when whole chunks of English, especially the bits that constitute folk psychology, are replaced by scientific words that call a thing by its proper name rather than some outworn metaphor.” The article recounts the occasion Patricia Churchland came home from a vexing day at work and told her husband, “Paul, don’t speak to me, my serotonin levels have hit bottom, my brain is awash in glucocorticoids, my blood vessels are full of adrenaline, and if it weren’t for my endogenous opiates I’d have driven the car into a tree on the way home. My dopamine levels need lifting. Pour me a Chardonnay, and I’ll be down in a minute.” Such awkward chemical conversation is unlikely to replace “folk psychology” anytime soon, despite the Churchlands’ fervent wishes, if only because it misses the actual human reasons for the reported neurochemical impairments—such as, for example, failing to get one’s favored candidate appointed to a post.

Moreover, there is strong reason to believe that the failure to provide a neuroscientific account of the sufficient conditions of consciousness and conscious behavior is not a temporary state of affairs. It is unlikely

that the gap between neuroscientific stories of human behavior and the standard humanistic or common-sense narratives will be closed, even as neuroscience advances and as our tools for observing neural activity grow more sophisticated.

In outlining the case that neuroscience will always have little to say about most aspects of human consciousness, we must not rely on weak mysterian arguments, such as Colin McGinn's claim, in his famous 1989 *Mind* paper "Can We Solve the Mind-Body Problem?," that there may be a neuroscientific answer but we are biologically incapable of figuring it out. Nor is there much use in appealing to arguments about category errors, such as considering thoughts to be "kinds of things," which were mobilized against mind-body identity theories when philosophy was most linguistically turned, in the middle of the last century. No, the aim of this essay is to give principled reasons, based on examining the nature of human consciousness, for asserting that we are not now and never will be able to account for the mind in terms of neural activity. It will focus on human consciousness—so as to avoid the futility of arguments about where we draw the line between sentient and insentient creatures, because there are more negative consequences to misrepresenting human consciousness than animal, and because it is human consciousness that underlines the difficulty of fitting consciousness into the natural world as understood through strictly materialist science.

This critique of the neural theory of consciousness will begin by taking seriously its own declared account of what actually exists in the world. On this, I appeal to no less an authority than the philosophy professor Daniel Dennett, one of the most prominent spokesmen for the neuroevolutionary reduction of human beings and their minds. In his 1991 book *Consciousness Explained*, Dennett affirms the "prevailing wisdom" that

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.... 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.

So when we are talking about the brain, we are talking about nothing more than a piece of matter. If we keep this in mind, we will have enough ammunition to demonstrate the necessary failure of neuroscientific accounts of consciousness and conscious behavior.

It is a pure dedication to materialism that lies behind another common neuroscientific claim, one that arises in response to the criticism that there are characteristics of consciousness that neuroscience cannot explain. The response is a strangely triumphant declaration that that which neuroscience cannot grasp does not exist. This declaration is particularly liable to be directed at the self and at free will, those two most persistent “illusions.” But even neuroscientists themselves don’t apply this argument consistently: they don’t doubt that they *think* they are selves, or that they have the illusion that they act freely—and yet, as we will see, there is no conceivable neural explanation of these phenomena. We are therefore justified in rejecting the presumption that if neuroscience cannot see it, then it does not exist.

The Outward Gaze

A good place to begin understanding why consciousness is not strictly reducible to the material is in looking at consciousness *of* material objects—that is, straightforward perception. Perception as it is experienced by human beings is the explicit sense of being aware of something material *other than oneself*. Consider your awareness of a glass sitting on a table near you. Light reflects from the glass, enters your eyes, and triggers activity in your visual pathways. The standard neuroscientific account says that your perception of the glass is the result of, or just *is*, this neural activity. There is a chain of causes and effects connecting the glass with the neural activity in your brain that is entirely compatible with, as in Dennett’s words, “the same physical principles, laws, and raw materials that suffice” to explain everything else in the material universe.

Unfortunately for neuroscientism, the inward causal path explains how the light gets into your brain but not how it results in a gaze that looks *out*. The inward causal path does not deliver your awareness of the glass as an item explicitly separate from you—as *over there* with respect to yourself, who is *over here*. This aspect of consciousness is known as *intentionality* (which is not to be confused with intentions). Intentionality designates the way that we are conscious *of* something, and that the contents of our consciousness are thus *about* something; and, in the case of human consciousness, that we are conscious of it *as* something other than ourselves. But there is nothing in the activity of the visual cortex, consisting of nerve impulses that are no more than material events in a material object, which could make that activity be *about* the things that you see. In

other words, in intentionality we have something fundamental about consciousness that is left unexplained by the neurological account.

This claim refers to fully developed intentionality and not the kind of putative proto-intentionality that may be ascribed to non-human sentient creatures. Intentionality is utterly mysterious from a material standpoint. This is apparent first because intentionality *points in the direction opposite* to that of causality: the causal chain has a directionality in space-time pointing from the light wave bouncing off the object to the light wave hitting your visual cortex, whereas your perception *of* the object refers or points from you *back* to the object. The referential “pointing back” or “bounce back” is not “feedback” or reverse causation, since the causal arrow is located in physical space and time, whereas the intentional arrow is located in a field of concepts and awareness, a field which is not independent of but stands aside from physical space and time.

Ironically, by locating consciousness in particular parts of the material of the brain, neuroscientism actually underlines this mystery of intentionality, opening up a literal, physical space between conscious experiences and that which they are about. This physical space is, paradoxically, both underlined and annulled: The gap between the glass of which you are aware and the neural impulses that are supposed to be your awareness of it is both a spatial gap and a non-spatial gap. The nerve impulses inside your cranium are six feet away from the glass, and yet, if the nerve impulses *reach out* or *refer* to the glass, as it were, they do so by having the glass “inside” them. The task of attempting to express the conceptual space of intentionality in purely physical terms is a dizzying one. The perception of the glass *inherently* is *of* the glass, whereas the associated neural activity exists apart from the cause of the light bouncing off the glass. This also means, incidentally, that the neural activity could exist due to a different cause. For example, you could have the same experience of the glass, even if the glass were not present, by tickling the relevant neurons. The resulting perception will be mistaken, because it is of an object that is not in fact physically present before you. But it would be ludicrous to talk of the associated neural activity as itself mistaken; neural activity is not *about* anything and so can be neither correct nor mistaken.

Let us tease out the mystery of intentionality a bit more, if only to anticipate the usual materialist trick of burying intentionality in causation by brushing past perception to its behavioral consequences. If perceptions really are material effects (in one place—the brain) of material causes (in another place—the object), then intentionality seems to run in the contrary direction to and hence to lie outside causation. That your perception

of the glass requires the neural activity in your visual cortex to reach causally upstream to the events that caused it is, again, utterly mysterious. Moreover, it immediately raises two questions. First, why does the backward glance of a set of effects to some of their causes stop *at a particular point* in the causal chain—in this case, at the glass? And, second, how does this reaching backward create a solid, stable object out of something as unstable as an interference with the light? The ordinary inference implicit in everyday perception is that the events which cause nerve impulses are manifestations of something that transcends those events—namely, an “object” that is the relatively permanent locus of possibility for many future events—making intentionality even more mysterious.

The bounce back is necessary to mark the point at which sense experiences are, as it were, “received”; the same point where, via a variety of intermediate steps, they can trigger behavioral outputs. This is a crucial point of demarcation within the causal nexus between perceptual input and behavioral output. And yet there is nothing within the nervous system that marks this point of arrival, or the point at which arrival passes over into departure (perceptual input into behavioral output). Nor is there anything to distinguish, on the one hand, those parts of the nervous system that are supposed to be the point of arrival of neural activity as a component of conscious experience from, on the other, those parts that are mere unconscious way stations en route to some other point of arrival.

In any event, identifying experiences with neural activity requires that intentionality, which has no place in the material world—since no material object is *about* any other material object—nevertheless fastens us into the material world. Examination of neural activity reveals only an unbroken causal chain passing from sensory inputs to motor outputs. Intentionality is significant because it is that which opens up the otherwise causally closed physical world. It lies at the root of our being a point of departure in the world, a site at which events originate—that is, of our being *actors*. And the weaving together of individual intentional spaces creates the human world—that shared, public, temporally deep sphere of possibilities, that outside-of-nature which makes our individual and collective human lives possible. It lies at the origin of everything that distances us from the material world. Without intentionality, there is no point of arrival of perceptions, no point of departure for actions, no input and output, no person located in a world. It is intentionality that opens up the present to the absent, the actual to the possible, and the now to the past and the future, so that we are able to live in a world that is an infinitely elaborated space of possibilities, rather than being simply “wired in” to

what *is*. These are large claims, some of which I have already elaborated in these pages (see “How Can I Possibly Be Free?,” Summer 2010). But the aspects presented here will be enough to wrest ourselves back from being assimilated into our brains.

It should also be noted that looking at the difficulties intentionality poses to materialism relieves us too of the need for the problematic views of (intermittently quite sensible) philosophers such as John Searle, who argues in his famous 1980 paper “Minds, Brains, and Programs” that intentionality “is a biological phenomenon, and it is as likely to be as causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomena.” Searle says this to undermine computational and functional theories of mind; but he still remains inside the biological frame of reference. And this requires him to think of intentionality—that in virtue of which an effect reaches back to its cause—to be itself the effect of another cause or set of causes. (The functionalism that Searle was rebutting claims that, just as a computer program is defined by how it transforms input to output, a piece of consciousness is defined by the particular causal transformation it effects between an organism’s perceptual input and its behavioral output. But since functionalism tries to assimilate perception into causation by arguing that the contents of consciousness are identical with their causal relations, it is just as easily disposed of by looking carefully at the counter-causal nature of intentionality, and the need for a point of arrival and departure, for input and output, without resorting to Searle’s argument from biological causes.)

Focusing on intentionality and placing it in the context of a materialistic, neuroscientific theory underlines what an extraordinary phenomenon perception is. It is that in virtue of which an object is revealed to a subject; or, rather, that in virtue of which the experiences of a subject are the revelation of an object. And this brings us to the heart of the trouble that the neural theory of perception is in: its central claim is that the interaction between two material objects—either directly, such as by touch, or indirectly, such as by vision—will cause one to *appear* to the other. The counter-causal direction of intentionality not only shows that this cannot be accommodated in physical science (of which neuroscience is a part) but that appearance is not something that the material world, a nexus of causation, affords. Indeed, we could go further and argue that the progressive enclosure of the world within the framework of physical science, its being construed as a *material* world, tends towards *the elimination of appearance*.

Making Appearance Disappear

Physical science begins when we escape from our subjective, first-person experiences into objective measurement, and thereby start to aspire towards Thomas Nagel's "view from nowhere." You think the table over there is large; I think it is small. We make a measurement and discover that it is two feet by two feet. We now characterize the table in a way that is less beholden to our own, or anyone else's, personal experience. Or we terminate an argument about whether the table is light or dark brown by translating its color into a mixture of frequencies of electromagnetic radiation. The table has lost contact with its phenomenal appearance to me, to you, or to anyone, as being characteristic of what it is.

As science progresses, measurement takes us further from actual experience, and the phenomena of subjective consciousness, to a realm in which things are described in abstract, general quantitative terms. The most obvious symptom of this is the way physical science has to discard what it regards as "secondary qualities"—such as color experiences, feelings of warmth and cold, and tastes. They are regarded as somehow unreal, or at least as falling short of describing what the furniture of the world is "in itself." For the physicist, light is not in itself bright or colorful; rather, it is a mixture of vibrations of different frequencies in an electromagnetic field. The material world, far from being the colorful, noisy, smelly place we experience, is purportedly instead composed only of colorless, silent, odorless atoms or quarks, or other basic particles and waves best described mathematically.

Physical science is thus about the marginalization, and ultimately the elimination, of phenomenal appearance. But consciousness is centrally about appearances. The basic contents of consciousness are these mere "secondary qualities." They are what fill our every conscious moment. As science advances, it retreats from appearances towards quantifying items that do not in themselves have the kinds of manifestation that constitute our experiences. A biophysical account of consciousness, which sees consciousness in terms of nerve impulses that are the passage of ions through semi-permeable membranes, must be a contradiction in terms. For such an account must ultimately be a physical account, and physical science does not admit the existence of anything that would show why a physical object such as a brain should find, uncover, create, produce, result in, or cause the emergence of appearances and, in particular, secondary qualities in the world. Galileo's famous assertions that the book of nature "is written in the language of mathematics" and that "tastes, odors, colors...

reside only in consciousness,” and would be “wiped out and annihilated” in a world devoid of conscious creatures, underline the connection, going back to the very earliest days of modern physical science, between quantification and the disappearance of appearance.

Any explanation of consciousness that admits the existence of appearances but is rooted in materialist science will fail because, on its own account, matter and energy do not intrinsically have appearances, never mind those corresponding to secondary qualities. We could, of course, by all means change our notion of matter; but if we do not, and the brain is a piece of matter, then it cannot explain the experience of things. Those who imagine that consciousness of material objects could arise from the effect of some material objects on another particular material object don't seem to take the notion of matter seriously.

Some neurophilosophers might respond that science does not eliminate appearance; rather, it replaces one appearance with another—fickle immediate and conscious appearance with one that is more true to the reality of the objects it attends to. But this is not what science does—least of all physical science, which is supposed to give us the final report on what there is in the universe, for which matter (or mass-energy) is the ultimate reality, and equations linking quantities are the best way of revealing the inner essence of this reality. For, again, it is of the very nature of mass-energy, as it is envisaged in physics, *not* to have any kind of appearance in itself.

This lack of appearance to mass-energy may still seem counter-intuitive, but it will become clearer when we examine a well-known defense, again made by John Searle, of the theory that mind and brain are identical—or specifically, that experiences can be found in neural impulses because they are the same thing. In his 1983 book *Intentionality*, Searle—who, as already noted, is committed to a neural account of consciousness—addresses the most obvious problem associated with the claim that experiences are identical with neural activity: experiences are nothing like neural activity, and the least one might expect of something is that it should be like itself.

Searle denies that this is a problem by arguing that neural activity and experiences are *different aspects* of the same stuff; more precisely, that they are the same stuff seen at “different levels.” The immediate problem with this claim is in knowing what it means. Clearly, neural activity and experiences are not two aspects of the same thing in the way that the front and back of a house are two aspects of the same house. Searle tries to clarify what he means using an analogy: experience is related to neural

activity, he says, as “liquid properties of water” are related to “the behavior of the individual molecules” of H₂O. They are the same stuff even though molecules of H₂O are nothing like water. Water is wet, he argues, while individual molecules are not.

Wetness is the one specific “liquid” property of water he cites at the outset, and the only others he mentions are that “it pours, you can drink it, you can wash in it, etc.” Because of this, it may seem at first that all Searle has accomplished is isolating the experiential qualities of water from the non-experiential. That is, one interpretation of Searle’s supposed explanation is that neural activity is related to experience in the same way water is related to experiences of water. This explanation, of course, is completely inadequate, because it simply sets us at a further regress from the answer.

But it turns out that this interpretation of Searle’s argument is the charitable one. We can see why in a section where Searle responds to this famous argument made by Leibniz in *The Monadology* (1714):

And supposing that there were a machine so constructed as to think, feel, and have perception, we could conceive of it as enlarged and yet preserving the same proportions, so that we might enter it as into a mill. And this granted, we should only find on visiting it, pieces which push against another, but never anything by which to explain perception. This must be sought for, therefore, in the simple substance and not in the composite or in the machine.

Searle’s response:

An exactly parallel argument to Leibniz’s would be that the behavior of H₂O molecules can never explain the liquidity of water, because if we entered into the system of molecules “as into a mill we should only find on visiting it pieces which push one against another, but never anything by which to explain” liquidity. But in both cases we would be looking at the system at the wrong level. The liquidity of water is not to be found at the level of the individual molecule, nor [is] the visual perception... to be found at the level of the individual neuron or synapse.

The key to understanding Searle’s argument and its fatal flaw is in the words, “But in both cases we would be looking at the system.” It turns out that in this water/H₂O analogy, it is not just the water but both levels that are already levels of experience or of observation. Searle in fact requires experience, observation, description—in short, consciousness—to generate the two levels of his water analogy, which are meant to sustain his

argument that two stuffs can be the same stuff even if they do not look like one another. This supposed explanation evades the question of experience even more than does the first. For what Searle is in effect arguing, though he does not seem to notice it, is that the relationship between neural activity and experience is like the relationship between two kinds of experience of the same stuff. And this is unsatisfactory because the problem he is supposedly solving is that neural impulses are not like experiences at all. (This rebuttal also applies—even more obviously, in fact—to another, very popular analogy, between dots of newsprint and a picture in the newspaper as neural activity and experiences. The dots/picture analogy also has the benefit of making clear another vulnerability of such analogies: the suggestion that neural activity is “micro” while experiences are “macro,” when it is not at all evident why that should be the case.)

Some will object to this experiential characterization of the “levels” argument, and will formulate it instead in terms of levels of *organization* or *complexity*: for example, the Earth’s climate and weather system is organized into many different levels of complexity, each exhibiting distinct behavior and distinct sorts of phenomena, from the interplays causing cycles of temperature over the centuries, down to the behavior of storms, down to the interactions of molecules. The implicit idea is that each level of complexity is governed by its own distinct set of laws. But one cannot take the distinction between these sets of laws to be inherent in the climate/weather system without in effect saying that when enough matter gets together in the same vicinity, it becomes another kind of matter which falls under the scope of another kind of law (at the same time that it remains the more basic kind of matter under the scope of the more basic kind of law). That flies in the face of reductive materialism—not to mention raises some very difficult questions about the identity of these different kinds of matter. What is more, the “getting together” that makes, say, a storm a whole made out of parts, is itself description-dependent and hence perception-dependent. The very term “complexity” refers to a description-dependent property. A pebble may be seen as something very simple—one pebble—or something infinitely complex—a system of a trillion trillion sub-atomic particles interacting in such a way as to sustain a static equilibrium.

The persistent materialist may launch a final defense of the argument, to the effect that the particular descriptions of water and H₂O molecules Searle mentions do not really depend on experience at all. He writes, “In its liquid state water is wet, it pours, you can drink it, you can wash in it, etc.... When we describe the stuff as liquid we are just describing those

very molecules at a higher level of description than that of the individual molecule.” One might argue that these enumerated qualities of water are all physical facts, that they are true even when there is no one present to observe them. But to the extent that this reinterpretation of Searle’s argument helps it to hold water (so to speak), it is only due to the original argument’s imprecision. For if we take it to be truly independent of experience that water “is wet, it pours, you can drink it, you can wash in it,” then these facts are *equally true* of a collection of molecules of H₂O, because of course the physical stuff known as water *just is* a collection of molecules of H₂O. Water and H₂O molecules, considered solely as physical things, are identical, and have all of the same properties. No appeal to “levels of description” should even be necessary. The reason it *is* necessary hinges on Searle’s description of one level as that of “*the individual molecule*.” But an individual molecule is not at all the same thing as water—which is a collection of many individual molecules—and so of course we should not expect the two to have the same properties. If we remove from the analogy the differing appearances to us of water and H₂O molecules as sources of their un-likeness, then all Searle has demonstrated is how a thing can be unlike a *part* of itself, rather than unlike itself. This is trivially true, and does not apply in any event to the question at hand if neural impulses are taken to be identical to experiences.

Searle makes his position even more vulnerable by arguing that not only are neural activity and the experience of perception the same but that the former *causes* the latter just as water is “caused” by H₂O. This is desperate stuff: one could hardly expect some thing A to cause some thing B with which it is identical, because nothing can cause itself. In any event, the bottom line is that the molecules of H₂O and the wet stuff that is water are two *appearances* of the same thing—two conscious takes on the same stuff. They cannot be analogous to, respectively, that which supposedly *causes* conscious experiences (neural impulses) and conscious experiences *themselves*.

What Physical Science is Blind To

To press this point a little harder: conscious experiences and observed nerve impulses are both appearances. But nerve impulses do not have any appearance in themselves; they require a conscious subject observing them to appear—and it is irrelevant that the observation is highly mediated through instrumentation. Like all material items, nerve impulses lack appearances absent an observer. And given that they are material events

lacking appearances in themselves, there is no reason why they should bring about the appearances of things other than themselves. It is magical thinking to imagine that material events in a material object should be appearances of objects other than themselves. Material objects *require* consciousness in order to appear.

All Searle has explained, again, is how two different appearances of the same thing can be unlike each other; but the problem he means to solve in the first place—or should mean to solve—is how something that *itself has no appearance* can give rise to, in fact be identical to, appearances. That is, Searle's task is to show how something that *itself has no appearance* can *be* an appearance—and without someone else observing the thing so as to *give* it the appearance. The question becomes pointed in Leibniz's thought experiment: how, from looking at the raw material of neurons in someone's brain, are we, as outside observers, supposed to get the appearances these neurons are meant to be causing, or generating, or identical to? Searle is forced into this conclusion: "If one knew the principles on which the system of H₂O molecules worked, one could infer that it was in a liquid state by observing the movement of the molecules, but similarly if one knew the principles on which the brain worked one could infer that it was in a state of thirst or having a visual experience." In other words, just by looking at neural impulses and "translating" them into the other "level of description," we can get at the corresponding experiences. This sounds fine until we consider just what "getting at" those experiences means. For what Searle does not account for is how *knowing that* a particular brain is having a particular experience is supposed to be enough to deliver *actually having* that experience yourself. To fully accept Searle's conclusion, we would have to believe that having the experience is the same as knowing that it exists—that it arises for the one person experiencing it, perhaps, from some implicit act of translation or "inference"—and so, among other things, that just by *looking* at someone else's brain in the proper way, we could *have* the same experiences they are having. These are absurd conclusions, as we will see.

From a more practical standpoint, we can see why it will never be enough to dismiss the problem of appearances out of hand by appealing either to the idea that perceptions are just brain activity like any other brain activity, or the idea that consciousness (and so all appearance) is an illusion. For in either case, while appearances are "nothing but" neural activity, we still must be able to explain why *some* neural activity leads to the sensation (or illusion) of appearance while *other* neural activity does not; and we must be able to distinguish between the two by looking only

at the material neurons. Neurophilosophers should be able to recognize this problem, since they acknowledge that the vast majority of neural impulses are *not* associated with appearances or consciousness of any sort. The search for neural correlates of consciousness has in fact turned up clusters, patterns, and locations of activity that are not in any significant respect different from neural activity that is not so correlated. What is more, “clusters,” “patterns,” and so forth also require an observer, to bring them together into a unity and to see that unity *as* a unity. That which requires an observer cannot be the basis of an observation.

The fact that intentionality does not fit into the materialist world picture has often been noted, but it is important to emphasize its anomalous nature because it lies at the root of pretty well everything that distances us from the material world, including other animals. The nature of intentionality becomes most clear when we see that the perceiver is an embodied subject—when the object is related to an “I” who perceives it, and who experiences himself as being located in the same experiential field as the object. The requirement of admitting the existence of a perceiving self is, of course, enough to make neurophilosophers hostile to the notion of the subject. But if they deny the existence of a self, they still have to account for how it is that matter can be arranged around a *viewpoint* as “near,” “far,” and so forth—for the construction of what Bertrand Russell called “egocentric space,” in which indexical words like “this,” “that,” and “here” find their meanings.

There are no appearances without viewpoints: for example, there are no appearances of a rock that are neither from the front of it nor from the back of it nor from any other angle. But there is nothing in the material brain, as we have seen, that could make it anyone’s *own* brain, or that could locate it at the center of anyone’s sensory field as the foundation of a viewpoint. We cannot appeal to the objective fact that the brain is located in a particular body to install it as *someone’s* brain, because ownership does not reside in bodies absent consciousness, or indeed self-consciousness—but consciousness is just what we are unable to find by looking at the material of the brain. Nor is the fact that the brain is located at a particular point in space sufficient for making it the center of a particular someone’s personal space any more than that fact is sufficient for a rock to have its own personal space.

The “view from nowhere” of physical science does not accommodate viewpoints. And since the material world has of itself no viewpoints, it does not, of itself, have centers—or, for that matter, peripheries. The equation $E=mc^2$, like all laws of physics, captures an ultimate, all-encompassing

scientific truth about the world, a viewless view of material reality, and has nothing to say about the experience of the world. Absent from it is that which forms the basic contents of consciousness: the phenomenal appearances of the world.

Mysteries of the Subjective Self

The loss of appearances is not an accidental mislaying. It is an inevitable consequence of the materialist conception of matter as we have it today. The brain, being a piece of matter, must be person-free. This is true not only by definition but also in other, specific senses. Persons—or selves—have two additional features which cannot be captured in neural terms.

The first is *unity in multiplicity*. At any given moment, I am aware of a multitude of experiences: sensations, perceptions, memories, thoughts, emotions. I am *co-conscious* of them—that is, I am aware of each of them at once, so that they are integrated into a unity of sorts. Moreover, co-consciousness includes consciousness of things I cannot currently see or touch: it includes consciousness of the absent past, of the absent elsewhere of the present, and of the possible future.

It is difficult to see how this integration is possible in neural terms, since neurophysiology assigns these experiences to spatially different parts of the brain. Aspects of consciousness are supposedly kept very tidily apart: the pathways for perception are separate from those for emotion, which are separate from those for memory, which are separate from those for motivation, which are separate from those for judgment, and so on. Within perception, each of the senses of vision, hearing, smell, and so forth has different pathways and destinations. And within, say, visual perception, different parts of the brain are supposed to be responsible for receiving the color, shape, distance, classification, purpose, and emotional significance of seen objects. When, however, I see my red hat on the table, over there, and see that it is squashed, and feel cross about it, while I hear you laughing, and I recognize the laughter as yours, and I am upset, and I note that the taxi I have ordered has arrived so that I can catch the train that I am aware I must not miss—when all of these things occur in my consciousness at once, many things that are kept apart must somehow be brought together. There is no model of such synthesis in the brain. This is the so-called “binding” problem.

Converging neural pathways might seem to offer a means by which things are all brought together—this is the standard neurophysiological

account of “integration.” However, it solves nothing. If all those components of the moment of consciousness came together in the same spot, if their activity converged, they would lose their separate identity and the distinct elements would be lost in a meaningless mush. When I look at my hat, I see that it is red, and squashed, and over there, and a hat, and all of the rest. Here is the challenge presented to neuroscience by the experienced unity *and* multiplicity of the conscious moment: that which is brought together has also to be kept apart. Consciousness is a *unification that retains multiplicity*.

Neurophilosophers have tried to deal with this problem by arguing that, while the components of experience retain their individual locations in the brain, the activity that occurs in those different locations is bound together. The mechanism for this binding is supposed to be either rhythmic mass neural activity or emergent physical forces which transcend the boundaries of individual neurons, such as electromagnetic fields or quantum coherence arising out of the properties of nerve membranes. This way of imagining the unity of consciousness assumes, without any reason, that linked activity across large sections of the brain—say, a coherent pattern of rhythmic activity, made visible as such to an observer by instrumentation—will be translated, or more precisely will *translate itself*, from an objective fact to a subjective unity. We are required to accept that something that is observed as an internal whole—via instrumentation—will be experienced as a whole, or itself be the experience of a whole, such that it will deliver the wholeness of a subset of items in the world while *at the very same time* retaining the separateness of those items.

The other distinctive feature of subjectivity is *temporal depth*. The human subject is aware of a past (his own and the shared past of communities and cultures) and reaches into a future (his own and the shared future). For simplicity’s sake, let us just focus on the past. There are many neurophilosophical accounts of memory, but they have one thing in common: they see memory as, in the slightly scornful phrase of the philosopher Henri Bergson, “a cerebral deposit.” Memory is, to use the slippery term, “stored” as an *effect* on the brain, expressed in its altered reactivity. This theory has been demonstrated, to the satisfaction of many neurophysiologists and cognitive neuroscientists, in creatures as disparate as apes and fruit flies. Some of the most lauded studies on “memory,” such as those that won Columbia University neuroscientist Eric Kandel his 2000 Nobel Prize, have been on the sea slug.

In reality, Kandel did not examine anything that should really be called *memory*—it was actually *altered behavior* in response to training by

means of an electric shock—essentially a conditioned reflex. A sea slug does not, so far as anybody knows, have semantic memory of facts—that is, memory of facts *as* facts, laden with concepts. It does not have explicit episodic memories of events—that is, events remembered as located in the past. Nor does it have autobiographical memories—that is, events remembered as located in its *own* past. It does not even have an explicit sense of the past or of time in general, and even less of a collective past where shared history is located. Nor can one seriously imagine an elderly sea slug actively trying to remember earlier events, racking its meager allocation of twenty thousand neurons to recall something, any more than one can think of it feeling nostalgic for its youth when it believed that it still had a marvelous life ahead of it.

Of course, neurophilosophers are not impressed by the objection that the sea slug or any other animal model does not possess anything like the kind of memory that we possess. It is, they say, simply a matter of different degrees of complexity of the nervous system in question: explicit memories involve more elaborate circuitry, with more intermediate connections, than the kinds of conditioned reflexes observed in sea slugs. To dissect this response, we have to examine critically the very idea that memories are identical with altered states of a nervous system. Let us look first of all at how the fallacy commanded acceptance. Kandel, like many other researchers, seems to assimilate all memory into *habit* memory, and habit memory in turn into altered behavior, or altered reactivity of the organism. And altered reactivity can be correlated with the altered properties of the excitable tissue in the organism, which may be understood in biophysical, biochemical, or neurochemical terms—the kinds of chemical changes one can see in the contents of a Petri dish. But these changes have nothing to do with memory as we experience and value it, though they have everything to do with overlooking the true nature of memory.

This is because habit memory is merely implicit, while human memory is also explicit: the former sort of “memory” is merely altered behavior, while the latter is something one is *aware of as a memory*. Those who think this a false elevation of the human must address not only the fact that there are two broad categories of memory to be found among animal species, but that both of these types of memory coexist in human beings. We not only have uniquely explicit memory, but also have the same sort of implicit memory as Kandel’s sea slugs. Moreover, we can have both of these types of memory about the same event: After a spark from a doorknob shocks my hand as I close the door during the winter, I will instinctively flinch from touching it again, and will then stop and explic-

itly *remember that* I had previously received an electrostatic shock. This time, I will explicitly plan to shut the door with my foot, an act that will itself after a few repetitions become instinctive or implicit, until I again stop to recall the explicit memory of the event that led to the habit. The neurophysiological account fails to address these distinctions.

To get to the bottom of the fallacies that underlie the very idea of a “neurophysiology of memory,” we need to remind ourselves that the nervous system is a material object and that material objects are identical with their present states. A broken cup is a broken cup. It is not in itself a record of its previous states—of a cup that was once whole—except to an outside observer who previously saw the cup in its unbroken state and now remembers it, so that he or she can compare the past and present states of the cup. The broken cup has an altered reactivity—it moves differently in response to stimuli—but this altered reactivity is not a *memory* of its previous state or of the event that caused its altered reactivity, namely its having been dropped. Likewise, although the altered state of the sea slug is, as it were, a “record” of what has happened to it, it is a record only to an external consciousness that has observed it in both its past and present states and is aware of both. And this is equally true of the altered reactivity of neurons exposed to previous stimuli in higher organisms.

Indeed, just as a conscious observer is required for the present state of the broken cup to be regarded as a “record” or “memory” of its having been dropped, so it must be a consciousness that identifies the particular piece of matter of the cup as a single object distinct from its surroundings, having its own distinct causal history, of which there is one special event among all others of which the cup is a “record”: its being dropped. From a consciousness-free material standpoint, the cup is but an arbitrary subset of all matter, and its present state owes equally to every prior state of the matter that composes it. The cup would have to be at once a “memory” of the moment it was dropped and of the innumerable moments when it sat motionless in the cupboard—with the former in no way privileged. In fact, if you believe that the present state of an object is a record or memory of all the events that brought it to its present state, you are committed to believing that, at any given moment, the universe is a memory of all its previous states. This need not be so, of course, if an object is encountered by a conscious individual who can see its present state as a sign of its past state, and so can focus on salient causes of salient aspects—for example, the event that led to the cup being broken. The conscious individual alone can see the present state as a sign of a past state and pick out one present state as a sign of one of the events that brought it from its past state to its present state.

This final point illustrates how the effect of an experienced event is a *record* of this event only to an observer. But the brain, being a material object, cannot be its own observer, comparing its past and present states. More precisely, the present state of a portion of the brain cannot reach out or *refer*, by the temporal equivalent of intentionality, to those salient events that changed it from an earlier state. And yet this is what memory does. Memories, that is to say, have an even more mysterious and counter-causal *about-ness* than perceptions of present events: they reach back to previous experiences, which themselves, through perception, reached out to that which, according to orthodox neuroscience, caused the experience. Memories supposedly therefore reach back to the mental causes of their physical causes. What is more, just as in vision I see the object as separate from myself, in memory I see the remembered object as different from the present, from the totality of what is here—I see it as *absent*. The memory explicitly locates its intentional object in the *past*. To borrow a phrase that Roger Scruton used in relation to music, memories have a *double intentionality*.

The failure of neuroscientism to deal with this last twist of the knife is illustrated by a recent paper in *Science* which some regarded triumphantly as having nailed memory. The authors found that the same neurons were active when an individual watched a TV scene (from, of all things, *The Simpsons*) as when they were asked to remember it. Memory, they concluded, is simply the replication of the neural activity that was provoked by the event that is remembered. This fails to distinguish, and so leaves unexplained, how it is that an individual experiences a memory *as a memory* rather than as something present—or, actually, a hallucination of something present.

A putative neural account of memory cannot deal with the difference between perceptions and memories because there is no past tense—or indeed future tense—in the material world. Consciousness, with its implicit sense of “now,” is required to locate events in one panel or other of the triptych past, present, and future; it is the conscious subject that provides the reference point. This is why Einstein said that physicists “know that the distinction between past, present, and future is only a stubbornly persistent illusion.” A consistent materialism should not allow for the possibility of memory, of the sense of the past. It only manages to seem to do so because observers, viewpoint, and consciousness are smuggled into the descriptions of the successive states of the brain, making it seem that later states can be *about* earlier states.

As if the unity of the self or subject or “I” at a *particular* time were not sufficiently resistant to neurological explanation, the unity of the self

over time is even further beyond its reach. The objective endurance of the brain does not generate the sensed co-presence of successive states of the self, even less the sense that one has temporal depth. Even if the self were reduced to a series of experiences, as in the accounts of David Hume or Oxford philosopher Derek Parfit, it would not be possible to see how the series was explicit *as* a series, with different moments explicitly related to each other, where one part accessed another and saw that it belonged to the same self. Indeed, starving the self down into a mere implicit thread linking successive experiences renders it *less*, not more, amenable to neural reduction, since the question of why some particular set of successive experiences rather than another should be linked together as a single series becomes even more glaring when the experiences are seen as but some arbitrary physical events among other physical events occurring in many different locations in physical space.

And the problem is by no means absolved if the sense of self is—as claimed by some neuroscientists, like so many other things they are unable to explain—an illusion. My feeling that I am the same person as the person who married my wife in 1970 is just as impossible to explain neurologically if it is an illusion as if it is true. Neural activity does not have the wherewithal to create the sense that we have of feeling that we are the same individual at different times—just as little if the sense is illusory as if it is true. The notion that the material brain can produce the illusion of the self but not be the basis of the real thing seems, to put it mildly, rather odd. And what is it to which the illusion is presented? Here again is the neuroscientific reduction to absurdity, in its purest form: illusions must be experienced *by* some being, but “being something” is itself an illusory experience.

An Insincere Materialism

The belief among neurophilosophers that the brain, a material object, can generate *tensed* time is one among many manifestations of the insincerity of their materialism. As we have seen, under cover of hard-line materialism, they borrow consciousness from elsewhere, smuggling it into, or presupposing it in, their descriptions of brain activity. This ploy is facilitated by a mode of speaking which I call “thinking by transferred epithet,” in which mental properties are ascribed to the brain or to parts of the brain (frequently very tiny parts, even individual neurons), which are credited with “signaling,” and often very complex acts such as “rewarding,” “informing,” and so forth. The use of transferred epithets is

the linguistic symptom of what Oxford philosopher P. M. S. Hacker and University of Sydney neuroscientist M. R. Bennett described, in their 2003 book *Philosophical Foundations of Neuroscience*, as the “mereological fallacy”: ascribing to parts properties which truly belong to wholes. This fallacy bids fair to be described as the Original Sin of much neurotalk, and it certainly allows the mind-brain barrier to be trespassed with ease.

This ease is in turn concealed by the ubiquity of transferred epithets outside brain science in everyday life. We are so used to talking about machines (particularly computers) “detecting,” “signaling,” “recording,” “remembering,” “warning,” and so forth, that we hardly notice, even less object, when this talk is applied to brains. Indeed, given that the brain is often billed as the most sophisticated of all machines, the computer to end all computers, it hardly needs to demonstrate its entitlement to being credited with such activities. While the homunculus is out of fashion, and ghosts have been exorcised from the machine, there are apparently billions of micro-homunculi haunting the cerebral cortex. The exiled homunculus has crept back in the form of a million billion angels bearing messages from one part of the brain to another, chattering endlessly across synapses.

This absurdity is concealed yet more deeply by a mode of speech that populates even the material environment that *surrounds* the brain with “signals” and “messages” and “information.” All the nervous system has to do is to extract and transmit those signals and messages and information. The Princeton psychologist Philip Johnson-Laird, a leading figure of the school of thought that held that the brain-mind is a computer, stated in his 1988 book *The Computer and the Mind* that “light reflected from surfaces and focused on the retinae contains a large amount of information.” (Gossipy stuff, light.) He admitted, however, that there were no entirely free gifts:

No matter how much information is in the light falling on the retinae, there must be a mental mechanism for recovering the identities of the things in a scene and those of their properties that vision makes explicit to consciousness.

Nevertheless, stipulating that there is information in the energy tickling up the brain is a flying start, and gets you across the brain-consciousness barrier without any scientific or indeed conceptual work being done. The otherwise inexplicable miracle by which the brain is supposed to support intelligent consciousness is made rather easier to understand when the energy that impinges on it is billed as *information*—information “about” the brain’s surroundings.

This trend, incidentally, is the top of a slippery slope at the bottom of which much lunacy lies. Information, once freed from the confinement of conscious human beings offering information to other human beings requiring to be informed, is everywhere. It is in the light; it is in DNA and other structures of the body. It is even in the material transactions of the non-living universe, as has been suggested by the advocates of “digital physics”—the idea that the universe *is* computation. By such misuse of language, matter becomes consciousness, or the energy in the material world comes to know itself, as has been suggested by the advocates of “panpsychism”—the idea that all matter is at least partially conscious.

The promotion of energy to information is the inverse of the demotion of consciousness to material transactions. In one direction, consciousness is in nothing; in the other, it is in everything. It gets right to the heart of how inherently absurd and paradoxical is neuroscientism to recognize that it naturally splits into these two wholly and fundamentally opposed modes of thinking, yet relies simultaneously on them both.

Finding Ourselves

We can see more clearly now the wide gap between brain function and consciousness—really, between people and their brains. This gap is seemingly crossed by linguistic legerdemain: people can be “brainified” if the brain is personified. But we have seen reasons why this gap should be unbridgeable. This, however, only throws into greater relief the magnitude of what remains to be answered, and so we must ask where we go from here. The failure to explain consciousness in terms of the brain—which follows from the failure of matter as understood in the most rigorous scientific manner to be able to house consciousness—raises two immediate questions.

The first and most obvious question is: Why, if the brain is not the basis of consciousness, is it so intimately bound up with it? Even those of us who object to the reduction of persons to brains have to explain why, of all the objects in the world, the brain is so relevant to our lives as persons. Nor can we overlook the extraordinary advances that have come from neuroscience in our ability to understand and treat diseases that damage voluntary action, consciousness, and mood—something that has been central to my entire professional career as a clinical neuroscientist. If consciousness, mind, volition, and so forth are not deeply connected with brain activity, then what are we to make of the genuine advances that neuroscience has contributed to our management of conditions that affect these central underpinnings of ordinary life?

The second question is whether, having shown the difficulty—no, the impossibility—of trying to get from brains alone to persons, we should abandon the very notion of the brain as a starting point for our thoughts about human consciousness. This question, however, brings us back to the first. If we say “I wouldn’t start from here,” then what do we do with the facts of neuroscience? Where does the brain fit into a metaphysics, an epistemology, and an ontology of mind that deny the brain a place at their center? If we are thinking of a new ontology, an account of the kinds of things there are in the universe that goes beyond the traditional division into mental and physical things; or if we are to go beyond an interactive epistemology that begins with sensations arising out of the impingement of energy on our brains and ascends to our knowledge of the laws of nature; then how shall we make sense of the things neuroscience tells us? How shall we deal with the fact that we are evolved organisms as well as persons?

These questions are posed because the case outlined here has been, necessarily, quite negative. It has merely been meant to clear the decks so we can set sail on the real work of finding a *positive* description of our nature, of the place of mind in nature, and, possibly, of the nature of nature itself. We need to start again thinking about our hybrid status: as pieces of matter subject to the laws of physics, as organisms subject to the laws of biology, and as people who have a complex sense of themselves, who narrate and lead their lives, and who are capable of thinking thoughts like these.