

My Brain and I

Roger Scruton

In her highly influential 1986 book *Neurophilosophy*, Patricia Churchland recommends that we ask ourselves just what philosophy has contributed to our understanding of human mental processes—just what, that is, compared to the extensive findings of neuroscience. The answer is not much, or even nothing at all, depending on your level of exasperation. Churchland, a professor at the University of California, San Diego, is of the view that philosophical arguments about our existing concepts—the concepts of “folk psychology,” as she calls them—are of no real significance. These philosophical arguments ignore the fact that our folk concepts belong to a theory—a useful theory, and one that gives us a handle on human language and human behavior—but a theory nevertheless. And theories get replaced by better ones. That, Churchland tells us, is what is happening, as neuroscience takes over from folk psychology, providing better explanations of human behavior than could ever be obtained from our old-fashioned language of belief, perception, emotion, and desire. In the wake of Churchland’s book, whole disciplines have sprung into being, proudly sporting the prefix “neuro-” by way of attaching themselves to Churchland’s banner. We have entered a new period in which philosophy, once the handmaiden of theology, is seen by a substantial community of its practitioners as the handmaiden of neuroscience, whose role is to remove the obstacles that have been laid in the path of scientific advance by popular prejudice and superstitious ways of thinking.

On the other hand, the concept of the person, which has been a central concern of philosophy at least since the Middle Ages, resists translation into the idiom of neuroscience, being associated with ways of understanding and interpreting human beings that bypass causal laws and theory-building categories. We evaluate human conduct in terms of free choice and responsibility. Persons are singled out from the rest of our environment as recipients of love, affection, anger, and forgiveness. We face them eye to eye and I to I, believing each person to be a center of self-conscious reflection who responds to reasons, who makes decisions, and whose life

*Roger Scruton, a New Atlantis contributing editor, is a visiting professor of philosophy at Oxford University and a senior fellow at the Ethics and Public Policy Center. This essay is adapted with permission from his most recent book, *The Soul of the World*, published by Princeton University Press in 2014.*

SPRING 2014 ~ 35

forms a continuous narrative in which individual identity is maintained from moment to moment and from year to year. All those aspects of our interpersonal understanding are assumed in moral judgment, in the law, in religion, politics, and the arts. And all sit uneasily with the picture of our condition that is advocated by many students of neuroscience, who describe the supposedly distinctive features of humanity as adaptations, more sophisticated than the social skills to be observed in the other animals, but not fundamentally different in their origin or function. These adaptations, they suggest, are “hardwired” in the human brain, to be understood in terms of their function in the cognitive processing that occurs when sensory inputs lead to behavioral outputs that have served the cause of reproduction. Moreover, the neuroscientists are likely to insist, the brain processes that are represented in our conscious awareness are only a tiny fragment of what is going on inside our heads. In the charming simile David Eagleman offers in his book *Incognito*, the “I” is like a passenger pacing the deck of a vast oceangoing liner while persuading himself that he moves it with his feet.

Converging Disciplines

In developing their account of human beings, the neurophilosophers draw on and combine the findings and theories of three disciplines. First, they lean heavily on neurophysiology. Brain-imaging techniques have been used to cast doubt on the reality of human freedom, to revise the description of reason and its place in human nature, and to question the validity of the old distinction of kind that separated person from animal, and the free agent from the conditioned organism. And the more we learn about the brain and its functions, the more do people wonder whether our old ways of managing our lives and resolving our conflicts—the ways of moral judgment, legal process, and the imparting of virtue—are the best ways, and whether there might be more direct forms of intervention that would take us more speedily, more reliably, and maybe more kindly to the right result.

Second, the neurophilosophers rely on cognitive science, a discipline concerned with understanding the kind of link that is established between a creature and its environment by the various “cognitive” processes, such as learning and perception. This field has been shaped by decades of theorizing about artificial intelligence. The conviction seems to be gaining ground that the nervous system is a network of yes/no switches that function as “logic gates,” the brain being a kind of digital computer, which operates by carrying out computations on information received through the many receptors

located around the body and delivering appropriate responses. This conviction is reinforced by research into artificial neural networks, in which the gates are connected so as to mimic some of the brain's capacities.

In a creature with a mind, there is no direct law-like connection between sensory input and behavioral output. How the creature responds depends on what it perceives, what it desires, what it believes, and so on. Those states of mind involve truth-claims and reference claims, which are not explicable in mechanistic terms. Cognitive science must therefore show how truth-claims and reference-claims arise, and how they can be causally efficacious. Much of the resulting theory arises from *a priori* reflection, and without recourse to experiment. For instance, Rutgers philosopher Jerry Fodor's well-known modular theory of the mind identifies discrete functions by reflecting on the nature of thought and on the connections between thought and action, and between thought and its objects in the world. It says little about the brain, although it has been an inspiration to neuroscientists, many of whom have been guided by it in their search for discrete neural pathways and allocated areas of the cortex.

Third, the neurophilosophers draw from the field of evolutionary psychology, which tells us to view the brain as the outcome of a process of adaptation. To understand what the brain is doing, we should ask how the genes of its owner would gain a competitive advantage by its doing just *this*, in the environment that originally shaped our species. For example, in what way did organisms gain a genetic advantage in those long hard Pleistocene years, by reacting not to changes in their environment, but to changes in their own *thoughts* about the environment? In what way did they benefit genetically from a sense of beauty? And so on. Much has been said by evolutionary psychologists concerning altruism, and how it could be explained as an "evolutionarily stable strategy." Neuroscientists might wish to take up the point, arguing that altruism must therefore be "hard-wired" in the brain, and that we should expect to find dedicated pathways and centers to which it can be assigned. Suppose you have proved that something called altruism is an evolutionarily stable strategy for organisms like us; and suppose that you have found the web of neurons that fire in your brain whenever you perform some altruistic act or gesture. Does that not say *something* at least about the workings of our moral emotions, and doesn't it also place severe limits on what a philosopher might say?

You can see from that example how the three disciplines of neurophysiology, cognitive science, and evolutionary psychology might converge, each taking a share in defining the questions and each taking a share in answering them. I want to urge a few doubts as to whether it is right to

run these disciplines together, and also whether it is right to think that, by doing so, we cast the kind of light on the human condition that would entitle us to rebrand ourselves as neurophilosophers.

The Problem of Explaining Morality

Consider the evolutionary psychologist's explanation of altruism as we find it delicately and passionately expounded by Matt Ridley in his 1997 book *The Origins of Virtue*. Ridley plausibly suggests that moral virtue and the habit of obedience to what Kant called the moral law is an adaptation, his evidence being that any other form of conduct would have set an organism's genes at a distinct disadvantage in the game of life. To use the language of game theory, in the circumstances that have prevailed during the course of evolution, altruism is a dominant strategy. This was first shown by John Maynard Smith in a paper published in 1964, and taken up by Robert Axelrod in his 1984 book *The Evolution of Cooperation*.

But what exactly do those writers mean by "altruism"?

An organism acts altruistically, they tell us, if it benefits another organism at a cost to itself. The concept applies equally to the soldier ant that marches into the flames that threaten the anthill, and to the officer who throws himself onto the live grenade that threatens his platoon. The concept of altruism, so understood, cannot explain, or even recognize, the distinction between those two cases. Yet surely there is all the difference in the world between the ant that marches instinctively toward the flames, unable either to understand what it is doing or to fear the results of it, and the officer who consciously lays down his life for his troops.

If Kant is right, a rational being has a motive to obey the moral law, regardless of genetic advantage. This motive would arise, even if the normal result of following it were that which the Greeks observed with awe at Thermopylae, or the Anglo-Saxons at the Battle of Maldon—instances in which an entire community is observed to embrace death, in full consciousness of what it is doing, because death is the honorable option. Even if you do not think Kant's account of this is the right one, the fact is that this motive is universally observed in human beings, and is entirely distinct from that of the soldier ant, in being founded on a consciousness of the predicament, of the cost of doing right, and of the call to renounce life for the sake of others who depend on you or to whom your life is owed.

To put it in another way, on the approach of the evolutionary psychologists, the conduct of the Spartans at Thermopylae is overdetermined. The "dominant reproductive strategy" explanation and the "honorable

sacrifice” explanation are both sufficient to account for their conduct. So which is the real explanation? Or is the “honorable sacrifice” explanation just a story that we tell ourselves, in order to pin medals on the chest of the ruined “survival machine” that died in obedience to its genes?

But suppose that the moral explanation is genuine and sufficient. It would follow that the genetic explanation is trivial. If rational beings are motivated to act in accordance with moral laws regardless of any genetic strategy, then that is sufficient to explain the fact that they do behave in this way.

Moral thinking unfolds before us a view of the world that transcends the deliverances of the senses, and which it is hard to explain as the byproduct of evolutionary competition. Moral judgments are framed in the language of necessity, and no corner of our universe escapes their jurisdiction. And morality makes sense only if there are reasons for action that are normative and binding. It is hard to accept this while resisting the conclusion drawn by Thomas Nagel, that the universe is ordered by teleological laws.

What Information Is About

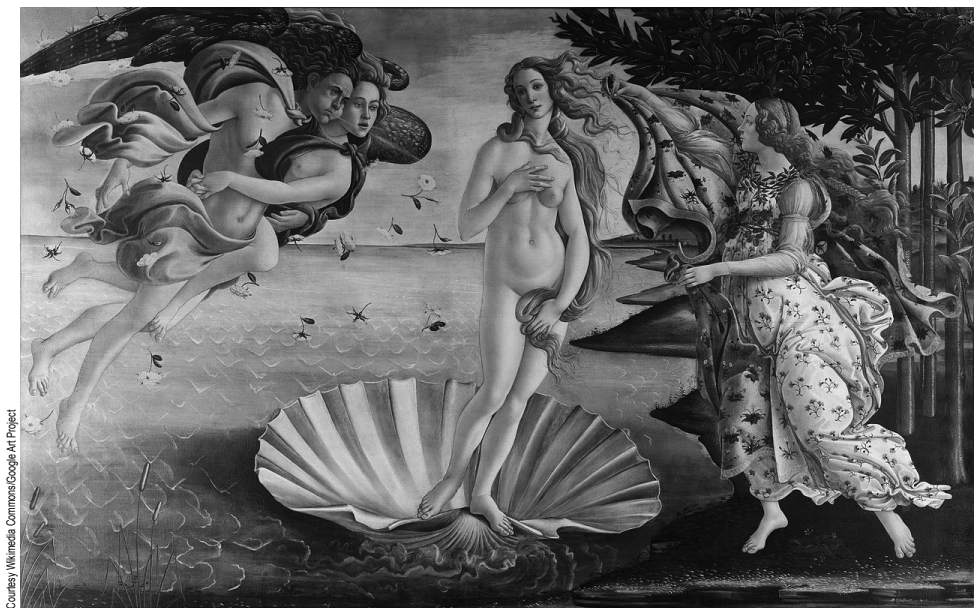
Another difficulty for the neurophilosophers arises from their reliance on cognitive science, which concerns the way in which information is processed by truth-directed creatures and which aims to explain perception, belief, and decision in terms of information-processing functions. However, is there a single notion of information at work here? When I inform you of something, I also inform you *that* something: I say, for example, that the plane carrying your wife has landed. Information, in this sense, is an intentional concept, which describes states that can be identified only through their content. Intentionality—the focusing on representations, on the “aboutness” of information—is a well-known obstacle in the way of all stimulus-response accounts of cognitive states; but why is it not an obstacle in the way of cognitive science?

It is surely obvious that the concept of information as “information *that*” is not the concept that has evolved in computer science or in the cybernetic models of human mental processes. In these models information generally means the accumulated instructions for taking this or that exit from a binary pathway. Information is delivered by algorithms, relating inputs to outputs within a digital system. These algorithms express no opinions; they do not commit the computer to living up to them or incorporating them into its decisions, for it does not make decisions or have opinions.

For example, suppose a computer is programmed to “read,” as we say, a digitally encoded input, which it translates into pixels, causing it to display the picture of a woman on its screen. In order to describe this process, we do not need to refer to the woman in the picture. The entire process can be completely described in terms of the hardware that translates digital data into pixels, and the software, or algorithm, which contains the instructions for doing this. There is neither the need nor the right, in this case, to use concepts like those of seeing, thinking, observing, in describing what the computer is doing; nor do we have either the need or the right to describe the thing observed in the picture as playing any causal role, or any role at all, in the operation of the computer. Of course, *we* see the woman in the picture. And to us the picture contains information of quite another kind from that encoded in the digitalized instructions for producing it. It conveys information about a woman and how she looks. To describe this kind of information is impossible without the use of intentional language—language that describes the content of certain thoughts rather than the object to which those thoughts refer. (I previously touched on this theme in these pages in “Scientism in the Arts and Humanities,” Fall 2013.)

Consider Botticelli’s famous painting *The Birth of Venus* (shown opposite). You are well aware that there is no such scene in reality as the one depicted, that there is no such goddess as Venus, and that this indelible image is an image of nothing real. But there she is all the same. Actually there was a real woman who served as Botticelli’s model—Simonetta Vespucci, mistress of Lorenzo de’ Medici. But the painting is not *of* or *about* Simonetta. In looking at this picture you are looking at a fiction, and that is something you know, and something that conditions any interpretation you might offer of its meaning. This is the goddess of erotic love—but in Plato’s version of the erotic, according to which desire calls us away from the world of sensual attachments to the ideal form of the beautiful (which is, incidentally, what Simonetta was for Botticelli). This painting helps to make Plato’s theory both clear and believable—it is a work of concentrated thought that changes, or ought to change, the worldview of anyone who looks at it for long. There is a world of information contained in this image—but it is information *about* something, information *that*, which is not captured by the algorithm that a computer might use to translate it, pixel for pixel, onto the screen.

The question is how do we move from the one concept of information to the other? How do we explain the emergence of thoughts *about* something from processes that are entirely explained by the transformation of visually



Courtesy Wikimedia Commons/Google Art Project

*Sandro Botticelli, The Birth of Venus (ca. 1485)
Florence, Galleria degli Uffizi*

encoded data? Cognitive science does not tell us. And computer models of the brain will not tell us either. They might show how images get encoded and transmitted by neural pathways to the center where they are “interpreted.” But that center does not in fact *interpret*—interpreting is a process that *we* do, in drawing conclusions, retrieving information *that*, and seeing what is there before us. And also what is *not* there, like the goddess Venus in Botticelli’s picture. A skeptic about intentionality might say that this simply shows that, in the last analysis, there is only *one* scientifically respectable concept of information—that there is, in reality, no such thing as *aboutness*, and therefore no question as to how we proceed from one concept of information to the other. But we would need a strong independent argument before drawing this conclusion. After all, isn’t science *about* the world, and does it not consist precisely in information of the kind that the skeptic denies to exist?

Parts and Wholes

In their controversial 2003 book *The Philosophical Foundations of Neuroscience*, Max Bennett and Peter Hacker describe something that they call the “mereological fallacy,” from *meros*, a part, and “mereology,” the branch of logic that studies the part–whole relation. This is the fallacy, as

they suppose it to be, of explaining some feature of a thing by ascribing that *very same feature* to a part of the thing. One familiar case of this is the well-known homunculus fallacy in the philosophy of mind, sometimes associated with Descartes, who tried to explain the consciousness of a human being by the presence of an inner soul, the “real me” inside. Clearly, that was no explanation, but merely a transferal of the problem.

Bennett and Hacker believe that many cognitive scientists commit this fallacy when they write of the brain “forming images,” “interpreting data,” being conscious or aware of things, making choices, and so on. And it certainly would be a fallacy to think that you could explain something like consciousness in this way, by showing how the brain is conscious of this or that—the explanation would be self-evidently circular.

The philosophy professor Daniel Dennett, among others, objects that no cognitive scientist ever intended to produce explanations of that kind. For Dennett there is no reason why we should not use intentional idioms to describe the behavior of entities that are not persons, and not conscious—just as we can use them, he believes, to describe simple feedback mechanisms like thermostats. When the thermostat acts to switch on the cooling system, because the room has reached a certain temperature, it is responding to information from the outside world. Sometimes it makes mistakes and heats the room instead of cooling it. You can “trick” it by blowing hot air over it. And in more complex devices like a computer it becomes yet more obvious that the mental language we apply to each other provides a useful way of describing, predicting, and making general sense of what computers do. In Dennett’s view there is nothing anthropomorphic when I refer to what my computer thinks, what it wants me to do next, and so on; for Dennett I am simply taking up what he calls the “intentional stance” toward something that thereby yields to my attempts to explain it. Its behavior can be successfully predicted using intentional idioms. Likewise with the brain: I can apply intentional idioms to a brain or its parts without thereby implying the existence of another conscious individual—other, that is, than the person whose brain this is. And in doing this I might in fact be explaining the connections between input and output that we observe in the complete human organism, and in that sense be giving a theory of consciousness.

There is an element of justice in Dennett’s riposte. There is no reason to suppose, when we use intentional idioms in describing brain processes and cognitive systems generally, that we are thereby committing ourselves to a homunculus-like *center* of consciousness, which is being described exactly as a person might be described in attributing thought, feeling, or intention. The problem, however, is that we know how to eliminate the intentional

idioms from our description of the thermostat. We can also eliminate them, with a little more difficulty, from our description of the computer. But the cognitive-science descriptions of the digital brain only seem to help us explain *consciousness* if we retain the intentional idioms and refuse to replace them. Suppose one day we can give a complete account of the brain in terms of the processing of digitalized information between input and output. We can then relinquish the intentional stance when describing the workings of this thing, just as we can with the thermostat. But we will not, then, be describing the consciousness of a person. We will be describing something that goes on when people think, and which is necessary to their thinking. But we won't be describing their thought, any more than we will be describing either the birth of Venus or Plato's theory of erotic love when we specify all the pixels in a screened version of Botticelli's painting.

Some philosophers—notably John Searle—have argued that the brain is the seat of consciousness, and that there is no *a priori* obstacle to discovering the neural networks in which, as Searle puts it, consciousness is “realized.” This seems to me to be a fudge. I don't exactly know what is intended by the term “realized”: could consciousness be realized in neural pathways in one kind of thing, in silicon chips in another, and in strings and levers in another? Or is it necessarily connected with networks that deliver a particular kind of connection between input and output? If so, what kind? The kind that we witness in animals and people, when we describe them as conscious? In that case, we have not advanced from the position that consciousness is a property of the whole animal, and the whole person, and not of any particular part of it. All of that is intricate, and pursuing it will lead into territory that has been worked over again and again by recent philosophers, to the point of becoming utterly sterile.

Who Am I?

Yet the real problem for cognitive science is not the problem of consciousness. Indeed, I am not sure that it is even a problem. Consciousness is a feature that we share with the higher animals, and I see it as an “emergent” feature, which is in place just as soon as behavior and the functional relations that govern it reach a certain level of complexity. The real problem, as I see it, is self-consciousness—the first-person awareness that distinguishes us from the other animals, and which enables us to identify ourselves, and to attribute mental predicates to ourselves, in the first-person case—the very same predicates that others attribute to us in the second- and third-person case.

Some people think we can make sense of this if we can identify a *self*, an inner monitor, which records the mental states that occur within its field of awareness, and marks them up, so to speak, on an inner screen. Sometimes it seems as though the neuroscientist Antonio Damasio is arguing in this way, but of course it exposes him immediately to Bennett and Hacker's argument, and again duplicates the problem that it is meant to solve. How does the monitor find out about the mental goings-on within its field of view? Could it make mistakes? Could it make a misattribution of a mental state, picking out the wrong subject of consciousness? Some people have argued that this is what goes on in the phenomenon of "inserted thought," which characterizes certain forms of mental illness. But there is no reason to think that a person who receives such "inserted thoughts" could ever be mistaken in thinking that they are going on in *him*.

The monitor theory is an attempt to read self-consciousness into the mechanism that supposedly explains it. We are familiar with a puzzling feature of persons, namely, that, through mastering the first-person use of mental predicates, they are able to make certain statements about themselves with a kind of epistemological privilege. People are immune not only to "error through misidentification," as Sydney Shoemaker puts it, but also (in the case of certain mental states like pain) to "error through misascription." It is one of the deep truths about our condition, that we enjoy these first-person immunities. If we did not do so, we could not in fact enter into dialogue with each other: we would be always describing ourselves as though we were someone else. However, first-person privilege is a feature of people, construed as language-using creatures: it is a condition connected in some way with their mastery of self-reference and self-predication. Internal monitors, whether situated in the brain or elsewhere, are not "language-using creatures": they do not enter into their competence through participating in the linguistic network. Hence they cannot possibly have competences that derive from language use and which are enshrined in and dependent upon the deep grammar of self-reference.

This is not to deny that there is, at some level, a neurophysiological explanation of self-knowledge. There must be, just as there must be a neurophysiological explanation of any other mental capacity revealed in behavior—although not necessarily a *complete* explanation, since the nervous system is, after all, only one part of the human being, who is himself incomplete as a person until brought into developing relation with the surrounding world and with others of his kind (for personhood is relational). But this explanation will not be framed in terms taken directly from the language of mind. It will not be describing persons at all, but ganglia and

neurons, digitally organized synapses and processes that are only figuratively described by the use of mental language in just the way my computer is only figuratively described when I refer to it as unhappy or angry. There is no doubt a neurophysiological cause of “inserted thoughts.” But it will not be framed in terms of thoughts in the brain that are in some way wrongly connected up by the inner monitor. There is probably no way in which we can, by peering hard, so to speak, at the phenomenology of inserted thoughts, make a guess at the neural disorder that gives rise to them.

Words like “I,” “choose,” “responsible,” and so on have no part in neuroscience, which can explain why an organism utters those words, but can give no material content to them. Indeed, one of the recurrent mistakes in neuroscience is the mistake of looking for the referents of such words—seeking the place in the brain where the “self” resides, or the material correlate of human freedom. The recent excitement over mirror neurons has its origins here—in the belief that these brain cells associated with imitative behavior might be the neural basis of our concept of self and of our ability to see others too as selves. But all such ideas disappear from the science of human behavior once we see human behavior as the product of a digitally organized nervous system.

On the other hand, ideas of the self and freedom cannot disappear from the minds of the human subjects themselves. Their behavior toward each other is mediated by the belief in freedom, in selfhood, in the knowledge that I am I and you are you and that each of us is a center of free and responsible thought and action. Out of these beliefs arises the world of interpersonal responses, and it is from the relations established between us that our own self-conception derives. It would seem to follow that we need to clarify the concepts of the self, of free choice, of responsibility and the rest, if we are to have a clear conception of what we are, and that no amount of neuroscience is going to help us in this task. We live on the surface, and what matters to us are not the invisible nervous systems that explain how people work, but the visible appearances to which we respond when we respond to them as people. It is these appearances that we interpret, and upon our interpretation we build responses that must in turn be interpreted by those to whom they are directed.

Again we have a useful parallel in the study of pictures. There is no way in which we could, by peering hard at the face in Botticelli’s Venus, recuperate a chemical breakdown of the pigments used to compose it. Of course, if we peer hard at the canvas and the substances smeared on it, we can reach an understanding of its chemistry. But then we are not peering at the face—not even *seeing* it.

We can understand the problem more clearly if we ask ourselves what would happen if we had a *complete* brain science, one that enabled us to fulfill Patricia Churchland's dream of replacing folk psychology with the supposedly true theory of the mind. What happens then to first-person awareness? In order to know that I am in a certain state, would I have to submit to a brain scan? Surely, if the *true* theory of the mind is a theory of what goes on in the neural pathways, I would have to find out about my own mental states as I find out about yours, achieving certainty only when I have tracked them down to their neural essence. My best observations would be of the form "It seems as though such-and-such is going on...." But then the first-person case has vanished, and only the third-person case remains. Or has it? If we look a little closer, we see that in fact the "I" remains in this picture. For the expression "It seems as though," in the report that I just imagined, in fact means "It seems *to me* as though," which means "I am having an experience of this kind," and to that statement first-person privilege attaches. It does not report something that I have to find out, or which could be mistaken. So what *does* it report? Somehow the "I" is still there, on the edge of things, and the neuroscience has merely shifted that edge. The conscious state is not that which is being described in terms of activity in the nervous system, but that which is being *expressed* in the statement "It seems to me as though...."

Moreover, I could not eliminate this "I," this first-person viewpoint, and still retain the things on which human life and community have been built. The I-You relation is fundamental to the human condition. We are accountable to each other, and this accountability depends on our ability to give and take reasons, which in turn depends upon first-person awareness. But the concepts involved in this process—concepts of responsibility, intention, guilt, and so on—have no place in brain science. They bring with them a rival conceptual scheme, which is in inevitable tension with *any* biological science of the human condition.

A Cognitive Dualist Approach

So how should a philosopher approach the findings of neuroscience? The best way of proceeding, it seems to me, is through a kind of cognitive dualism (not to be confused with the kind of ontological dualism famously postulated by Descartes). Such a cognitive dualism allows us to understand the world in two incommensurable ways, the way of science and the way of interpersonal understanding. It lets us grasp the idea that there can be *one* reality understood in more than one way. In describing a

sequence of sounds as a melody, I am situating the sequence in the human world: the world of our responses, intentions, and self-knowledge. I am lifting the sounds out of the physical realm, and repositioning them in what Husserl called the *Lebenswelt*, the world of freedom, action, reason, and interpersonal being. But I am not describing something *other* than the sounds, or implying that there is something hiding behind the sounds, some inner “self” or essence that reveals itself to itself in some way inaccessible to me. I am describing what I hear *in* the sounds, when I respond to them as music. In something like that way I situate the human organism in the *Lebenswelt*; and in doing so I use another language, and with other intentions, than those that apply in the biological sciences.

The analogy is imperfect, of course, like all analogies. But it points to the way out of the neuroscientist’s dilemma. Instead of taking the high road of theory that Patricia Churchland recommends, attempting to provide a neurological account of what we mean when we talk of persons and their states of mind, we should take the low road of common sense, and recognize that neuroscience describes one *aspect* of people, in language that cannot capture what we mean when we describe what we are thinking, feeling, or intending. Personhood is an “emergent” feature of the human being in the way that music is an emergent feature of sounds: not something over and above the life and behavior in which we observe it, but not reducible to them either. Once personhood has emerged, it is possible to relate to an organism in a new way—the way of personal relations. (In like manner we can relate to music in ways in which we cannot relate to something that we hear merely as a sequence of sounds—for example, we can dance to it.) With this new order of relation comes a new order of understanding, in which reasons and meanings, rather than causes, are sought in answer to the question “why?” With persons we are in dialogue: we call upon them to justify their conduct in our eyes, as we must justify our conduct in theirs. Central to this dialogue is the feature of self-awareness. This does not mean that people are really “selves” that hide within their bodies. It means that their *own way of describing* themselves is privileged, and cannot be dismissed as mere “folk psychology” that will give way in time to a proper neuroscience.

The *Lebenswelt* stands to the order of nature in a relation of emergence. But this relation is not reducible to a one-to-one relation between particulars. We cannot say of any individual identified in one way that it is the *same individual* as one identified in the other way. We cannot say “one thing, two conceptions,” since that raises the question what thing, which raises the question, under which conception is the thing identified?

You can see the problem very clearly in the case of persons, since here, if I say one thing, two conceptions, and then ask *what* thing, the answer will depend upon which “cognitive scheme” I am “within” at the time. The answer could be: this animal; or it could be: this person. And we know these are different answers, from all the literature on the problem of personal identity. In other words, each scheme provides its own way of dividing up the world, and the schemes are incommensurable. Because the concept of personhood is so difficult, we are tempted to duck out of the problem, and say that there is just *one* way of identifying what we are talking about, using the concept “human being” to cross back and forth from scientific to interpersonal ways of seeing things. But seeing things in that way, it seems to me, we underplay the difference that is made by the first-person case. My self-awareness enables me to identify myself without reference to my physical coordinates. I am assured of my identity through time without consulting the spatiotemporal history of my body—and this fact gives a peculiar metaphysical resonance to the question “where in the world am I?”

Light is cast on this issue by Aristotle’s theory of “hylomorphism.” Aristotle believed that the relation between body and soul is one of matter and form—the soul being the organizing principle, the body the matter from which the human being is composed. The suggestion is obscure and the analogies Aristotle gives are unpersuasive. But the theory becomes clearer when expressed in terms of the relation between a whole and its parts. Thus Mark Johnston has defended, in the name of hylomorphism, the view that the essential nature of an individual thing is given by the concept under which its parts are gathered together in a unity.

If we accept that approach, then we should conclude that in the case of human beings there are two such unifying concepts—that of the human organism, and that of the person, each embedded within a conceptual scheme that sets out to explain or to understand its subject matter. Humans are organized from their material constituents in two distinct and incommensurable ways—as animal and as person. Each human being is indeed two things, but not two separable things, since those two things reside in the same place at the same time, and all the parts of the one are also parts of the other.

The project of neurophilosophy will likely be confounded by the resiliency of the ideas of “I,” “you,” and “why?” Cognitive dualism of the sort I have begun to sketch out here allows us to appreciate and learn from what empirical science can teach about what we are, while acknowledging the objective reality of melodies as well as sounds, faces as well as physiognomies, meanings as well as causes.
