Neurophilosophy and Its Discontents

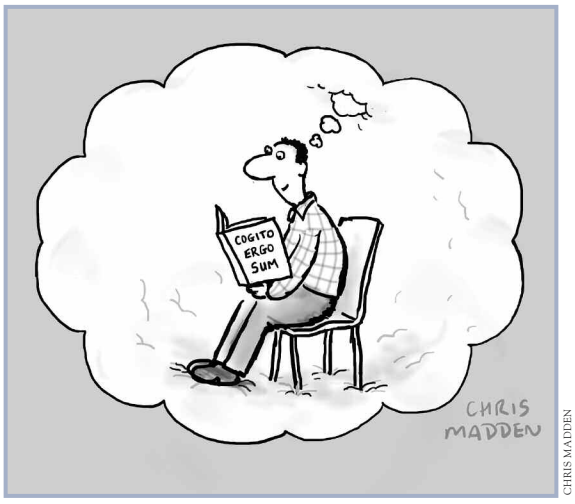
How Do We Understand Consciousness Without Becoming Complicit in that Understanding?

BY GABRIELLE BENETTE JACKSON

What is consciousness? "It is being awake," "being responsive," "acting," "being aware," "being self-aware," "paying attention," "perceiving," "feeling emotions," "feeling feelings," "having thoughts," "thinking about thoughts," "it is like this!"

Who is conscious? "We humans, surely!" Well, maybe not all the time. "Animals!" Debatable. "Computers?" No—at least, not yet. "Other machines?" Only in fiction. "Plants?" Absolutely not, right?

Nearly twenty-five years ago, we lived through "the project of the decade of the brain," a governmental initiative set forth by President George H. W. Bush.[1] Presidential Proclamation 6158 begins, "The human brain, a three-pound mass of interwoven nerve cells that controls our activity, is one of the most magnificent—and mysterious—wonders of creation. The seat of human intelligence, interpreter of senses, and controller of movement, this incredible organ continues to intrigue scientists and laymen alike. Over the years, our understanding of the brain—how it works, what goes wrong when it is injured or diseased—has increased dramatically. However, we still have much more to learn." And it concludes, "Now, Therefore, I, George Bush, President of the United States of America, do hereby proclaim the decade beginning January 1, 1990, as the Decade of the Brain. I call upon all public officials and the people of the United States to observe that decade with appropriate programs, ceremonies, and activities."



CHRIS MADDEN

What the former President did not say—what is perhaps understood by his readers—is that the brain is quite different from other body parts that have come under scientific investigation. We might be grateful to receive a donated kidney, or to have an artificial heart. But unlike every other body part, without *my* brain, there may be no I. Our sense of self, of awareness, of life—are profoundly connected to a working brain. A philosopher once said, "in a brain transplant, one wants to be the donor not the recipient." Indeed, saying that the brain is the seat of mentality is like saying that the sun is a source of light.

The decade of the brain is now over. No longer are the questions on the order of "What region of the brain is associated with facial recognition?" But rather, "Which particular neuron fires before a picture of Halle Berry's face?" We have learned that the different frequency and synchronization with which neurons fire is associated with different states of conscious awareness. There is optimism that diseases such as Alzheimer's could be treatable with therapies implemented at the neural level. We have entered the era of the neuron.

SAYING THAT THE BRAIN IS THE SEAT OF MENTALITY IS LIKE SAYING THAT THE SUN IS A SOURCE OF LIGHT.

But for all that this trajectory has and will accomplish, we seem no closer to answering basic (actually, quite old) questions about the relationship between the mind and the body—between consciousness and the physical substrates that realize it.

These questions come in two general forms. First, a metaphysical point: why should this particular physical matter (the neurochemical, the nerve cell, the neural network) give rise to consciousness? It seems we can imagine creatures who have brains just like ours, but who don't feel pain. So why do we feel it? Why does activation of group C nerve fibers in my brain give rise to pain, rather than some other feeling, or nothing at all? Second, an epistemological point: even if we were to know everything about this particular physical matter (the neurochemical, the nerve cell, the neural network), what does this tell us about consciousness? Tell me all there is to know about the chemistry of H_20, and I might know what water is. Tell me all there is to know about the neural biological basis of pain, and I still surely won't know what pain is. Unless I have experienced it myself, a truly essential aspect—*how pain feels*—has been left out.

These metaphysical and epistemological questions together form what philosophers call, respectively, "the hard problem" and "the knowledge argument." We can combine them, limit the jargon, and talk about "the problem of consciousness."

In what is perhaps the best known articulation of the problem of consciousness, in Thomas Nagel's essay "What is it like to be a bat?" (1974), he stipulates that no matter what form consciousness might take, "there is something it like to be" conscious, there is "something it is like for" a conscious being, and no objective fact will ever explain this subjective fact.[2] The best we can hope is to establish correlations between the two. Posit a connection stronger than correlation, and we overstep.

If this were all there was to say, we would have to learn to live with the problem of consciousness. But in the final paragraphs of the self-same article, Nagel offered an alternative. The problem of consciousness "should be regarded as a challenge to form new concepts and devise a new method—an objective phenomenology not dependent on empathy or imagination. Though presumably it would not capture everything, its goal would be to describe, at least in part, the subjective character of experience in a form comprehensible to beings incapable of having those experiences. […] It should be possible to devise a method of expressing in objective terms much more than we can at present, and with much greater precision." The proposal, simply put, was to develop a language to describe subjectivity in non-subjective terms. And although it is definitely not the case that all theorists pushing past the problem of consciousness consider themselves to be implementing Nagel's plan, it does help to understand a particular set of accumulated answers. Two fundamental approaches have been *neurophilosophy* and *neurophenomenology*, each emphasizing one aspect of Nagel's suggestion—either the objective part (*viz.* neurophilosophy) or the phenomenology part (*viz.* neurophenomenology).

Despite the similarity of nomenclature, neurophilosophy and neurophenomenology are very different approaches emerging from different traditions.

Suppose we start, though, with what the neurophilosopher and the neurophenomenologist share. Both hold in common the belief that the problem of consciousness is a pseudoproblem created by our inability to move beyond the conceptual binarism of mind versus body—an error Gilbert Ryle identified in his famous critique of "the dogma of the ghost in the machine." Both the neurophilosopher and the neurophenomenologist agree that the problem of consciousness is generated by some combination of false dichotomies and faulty concepts. However, they each have different ways of solving it.

Neurophilosophy develops in the "analytic" philosophical tradition in the late twentieth century. Its early formulation can be found in Patricia Churchland's 1986 book *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. But it is also manifest in the work of many other theorists (e.g., Paul Churchland, Antonio Damasio, Christof Koch). Neurophilosophy is a reductionist theory of consciousness, one that aspires to the Quinean goal of eliminating all things that cannot be reduced to physical (or functional) processes, within which a general method emerges. First, it identifies ideas about consciousness derived from common sense, folk psychology, or introspection. Second, it reduces these "soft" concepts to "hard" neuroscientific data. Third, if they no longer are practically useful, it eliminates the original ideas about consciousness in favor of their neurobiological counterparts. To give an embarrassingly oversimplified example, take the conventional idea of the love one feels for one's child. The neurophilosopher takes this subjective idea and, informed by the best neuroscience available, translates it into an objective account—imagine the neurophilosopher saying, "Love is nothing more than oxytocin release." In the future, the neurophilosopher will replace the word "love" with the more perspicuous word "oxytocin" in everyday conversation. About the possibility that the feeling of parental love is just neural chemistry, Patricia Churchland herself said, "well, actually, yes, it is. But that doesn't bother me."[3]

Thus, the neurophilosopher strives to convert mind into matter, parsing a singular subjective phenomenon in a shared objective language. But eliminating consciousness in our favor of the neuron may give away more than is necessary or useful. The price of characterizing consciousness in a more scientific language need not be the abandonment of consciousness itself.

If neuroscience hopes to do more than describe arbitrary processes at the neural level, it will always need conscious experience to direct where to look in the brain. We are not mere bystanders in the investigation of consciousness. Our own consciousness is both essential and unavoidable in this endeavor. Neuroscience sometimes forgets that it begins with what interests us about our own conscious experience. And while this certainly involves the fascination with our own subjectivity, it also involves our personal histories, our embodied and embedded situations, and our social values. Strictly speaking, the feeling of love for one's child is not oxytocin release. More precisely, parental love is *disclosed to us* through oxytocin release, as a situated normative phenomenon. An (imagined) culture that doesn't value parental love will not care one lick to discover what its neural correlate happens to be. Oxytocin release is important to us here and now because it is tied to the feeling of love for one's child, a subjective phenomenon that we already recognize and value. For this reason alone, neuroscience needs the first-person point of view.

There is a deeper problem to consider, however, one that insinuates itself into all investigations of consciousness. Technically, we never establish identity statements linking neurochemical processes directly to consciousness. What we do get are equivalences linking our conception of neurochemical processes to our conception

(Continued on page 6)

of consciousness. We then have to wonder how accurate and stable our concepts are. To what extent do the concepts we use *transform* the explananda? This is particularly relevant when what we are trying to explain is consciousness itself.

How do we understand consciousness without becoming complicit in that understanding, wrongly attaching the properties of our (conscious) inquiry to the properties of the inquired into (consciousness)? We can never completely avoid our own contribution, however accidental, to the discovery process. This is true for the philosopher in her armchair *as well as* the scientist in her lab. When investigating consciousness, there emerges, for lack of a better phrase, a kind of "observer effect." The fact that consciousness is both the tool for investigation and the thing to be investigated leads to a lot of mischief. To take a classic example: when we talk about the visual experiences of a red apple, a red stop sign, and a red sweater, we can isolate their red quality, their *redness*. But is this abstracted quality—the color distinct from its object—a property of our *visual experience* of the object, as is generally assumed, or is it rather a property of our *reflection* on the visual experience of the object? What if "a color is never simply a color, but rather the color of a certain object, and the blue of a rug would not be the same blue if it were not a wooly blue?" (Maurice Merleau-Ponty[4]). Simply put, in visual experience, prior to reflection, what if there is no such thing as uninstantiated redness?

WHEN INVESTIGATING CONSCIOUSNESS, THERE EMERGES, FOR LACK OF A BETTER PHRASE, A KIND OF "OBSERVER EFFECT."

The neurophilosopher in search of the neural correlate of redness has already assumed an answer to these questions. But this assumption may be, at best, unwarranted and, at worst, wrong. The neurophenomenologist, on the other hand, takes such concerns effectively as her starting point.

Neurophenomenology emerges out of "Continental" philosophy in the late twentieth century. At its inception, we find Francisco Varela, who articulated the approach in his 1996 article "Neurophenomenology: A Methodological Remedy for the Hard Problem." Since then there have been many collaborators in the development of this movement (e.g., Evan Thompson, Shaun Gallagher, Vittorio Gallese, Giacomo Rizzolatti). Growing out of the phenomenological tradition initiated by Edmund Husserl, neurophenomenology is primarily a method that attempts to naturalize consciousness. First, it identifies a multiplicity of cases, both observed (scientific, empirical) and introspected (described, imagined), in which consciousness is operative. Second, setting aside questions of the physical (functional) reality of consciousness, it identifies the invariant structures that all these cases have in common. Third, it uses these invariant structures to furnish an idea of consciousness that is consonant with the natural sciences.

To give an example of how this works, consider two accounts—the phenomenological and the neuroscientific—of how we come to understand the actions of others. That is, why do we experience the observed bodily movements of other people as genuine actions rather than as mere automation? I do not see a sequence of movements, take a moment to assess the situation, and then make an inference to the best explanation of what a person is doing. As is often the case with my own movements, I know immediately, directly, and implicitly what action is underway. But those are my actions to which I have privileged access. How is it, then, that I seem to have the same kind of access to the actions of others?

Gabrielle Benette Jackson, Visitor in the School of Social Science (2013–14), is an Assistant Professor at the State University of New York, Stony Brook. Her research concerns the areas of overlap among philosophy of mind, cognitive science, and phenomenology.

Phenomenologists have long argued that our access to the intentional goal-directed actions of others is actually the reverse side of our access to our own intentional actions (e.g., Edmund Husserl, Maurice Merleau-Ponty, Dan Zahavi, Natalie Depraz). For instance, when my partner lifts a heavy box, I understand what is happening instantaneously, without any inference or intervening judgment. I might even, also without thinking, reach out and offer my assistance. We live in a shared world that we experience together, the phenomenologists claim, through a kind of "bodily reciprocity"—when we see another person acting intentionally, we experience the same intention in our own bodies, and we transpose our motor intentions into that person.

Neuroscientists discovered that activity in the premotor cortex of the brain was correlated with the preparation of certain physical movements in response to sensory stimulus (e.g., Giacomo Rizzolatti, Vitorrio Gallese). These physical movements were not reflexive responses, but rather intentional goal-directed actions—such as reaching, tearing, grasping. These same neuroscientists also discovered that a subset of neurons in the premotor cortex not only activate during the execution of intentional actions, but also discharge when observing similar intentional actions performed by another individual. These neurons do not activate, however, when one's body moves or is moved unintentionally in a physically analogous manner. For an example from the original experiments, whether a monkey reaches for a raisin or watches another monkey reach for a raisin, the same sets of neurons are activated. But if the monkey's arm is moved passively towards a raisin, those neurons do not discharge. These neurons appear to mirror the active movements of others, particularly, conspecifics—hence the nomenclature, "mirror neurons."

The task of the neurophenomenologist now becomes to integrate all this data, finding the invariant structures they share. A first striking commonality is that, on both accounts, to observe an action is also to simulate it through transpositional (*viz.* bodily reciprocity) or mirroring (*viz.* mirror neurons) processes. Another point of convergence is that these processes are not reflective, linguistic or intellectual. Instead, they appear to be prereflective, nonverbal, and practical. A third parallel is that these interactions occur specifically among conspecifics. This suggests some kind of intersubjectivity at work—in order to simulate the other, we have first to identify with it. There may be other homologies, too, but even with just these three, a single unified explanation is already taking shape. Our understanding of the movements of others as genuine actions is fundamentally a bodily understanding, one that is experienced through shared empathetic connections with other like beings, whereby we simulate in ourselves their intentional goal-directed actions, transposing into them our motor intentions, a capacity realized by dedicated neural processes in the brain.

Neurophenomenology, as an approach to understanding consciousness, is not in competition with phenomenological description or scientific data. It is an intriguing place where we are allowed to surpass the alternative of subjectivity and objectivity, interpolating a conceptual space between them, in which a deeper understanding of both can emerge, a place that we already knew could be inhabited, in a way, because our very existence proves mind and matter compatible. ■

1 George H. W. Bush. Presidential Proclamation 6158. July 17, 1990. www.loc.gov/loc/brain/proclaim.html

2 Thomas Nagel. "What is it like to be a bat," *Mortal Questions* (Cambridge University Press, 1979).

3 "The benefits to realizing you are just your brain," Graham Lawton interviews Patricia Churchland for *New Scientist* 2945 (November 29, 2013).

4 Maurice Merleau-Ponty (1945). *Phenomenology of Perception*, trans. Donald A. Landes (Routledge, 2012).

From B-Mode Cosmology to the Fate of Spacetime

The Institute's thirteenth annual Prospects in Theoretical Physics (PiTP) summer program for graduate students and postdoctoral scholars, which focused on string theory, was truly extraordinary in that it overlapped with Strings 2014. This is one of the field's most important gatherings, which the Institute hosted with Princeton University, convening international experts and researchers to discuss string theory and its most recent developments. Six hundred attendees gathered for Strings 2014, which made it one of the largest Strings conferences since their inception in 1995.

Strings 2014 talks, which covered topics from B-mode cosmology and the theory of inflation to quantum entanglement, the amplituhedron, and the fate of spacetime, may be viewed at https://physics.princeton.edu/strings2014/Talk_titles.shtml.

The program for PiTP and videos of its string theory talks may be viewed at https://pitp2014.ias.edu/schedule.html. As part of the PiTP program, the Institute showed a screening of "Particle Fever," a new film that follows six scientists, including the Institute's Nima Arkani-Hamed, during the launch of the Large Hadron Collider and fortuitously captures the discovery of the Higgs particle. Peter Higgs, who predicted the existence of the particle fifty years ago, gave one of his first seminars on the topic at the Institute in 1966.