4 *Neurophilosophy*

PATRICIA CHURCHLAND

Introduction: What Is Neurophilosophy?

"Neurophilosophy" explores the impact of discoveries in neuroscience on a range of traditional philosophical questions about the nature of the mind. This subfield aims to move forward on questions such as the nature of knowledge and learning, decision making and choice, and self-control and habits by drawing on data from the relevant sciences – not only neuroscience and clinical neurology but also evolutionary biology, experimental psychology, behavioral economics, anthropology, and genetics. It draws also on lessons from the history of philosophy and the history of science, which saw mysteries about the nature of the blood or fire or infectious disease become less mysterious as experimental science began to provide new observations and tested explanations (Thagard 2014).

The massive accumulation of neurobiological data from many levels of brain organization and many species of nervous systems is a recent development because neuroscience did not really reach full steam until about the 1970s. Why was the development of neuroscience delayed until recently?

Although clinical observations had long implicated the brain in mental functions, understanding exactly *why* lesions affected mental functions remained out of reach. This was so because essentially until very recently nothing was known about the microstructure of brains – about neurons and how neurons worked, about how the brain was organized into networks and systems, and about how neurochemicals mediated interactions between neurons. Notice that detailed drawings of nerve cells were produced by Camillo Golgi and Ramón y Cajal only in the latter part of the nineteenth century. How neurons *interacted*

Particular thanks to Paul Churchland and Joshua Brown for clear-headed discussion and to David Livingstone Smith for wise advice.

with each other to yield effects such as a behavior was still a profound mystery.

Chemistry, by contrast, was a vastly more mature science in the early nineteenth century, strengthened by basic organizing principles of atomic theory, as outlined by Dalton in 1805 and a clear appreciation of the fundamental elements – no longer deemed to be earth, air, fire, and water. Instead, the elements were characterized by Mendeleyev in the 1880s in the periodic table – things such as oxygen, hydrogen, tin, and gold. As for neuroscience, it is perhaps surprising to realize that the existence of *inhibitory* connections between nerve cells was demonstrated by John Eccles and colleagues only in the 1950s. Physics, far more mature in terms of theory and explanation by that time, had begun to investigate the inner structure of the atom.

To get a perspective here, note that effective brain imaging techniques came into their own only in the last two decades of the twentieth century. At the micro level, many details regarding the synapse and how neurons communicate are not completely understood even now, nor are the functions and dynamics of neural *networks*. Neuroscience is a young science.

Because the brain's basic units work by changes in voltage across the cell membrane and by chemicals that regulate such changes, and because the units are not visible to the naked eye, development has depended on a theoretically and experimentally rich physics and chemistry. Specifically, neuroscience depends on tools and devices that exploit the knowledge of physics and chemistry, for example, the electron microscope, microelectrodes, nuclear magnetic resonance, monoclonal antibodies, and most recently, optogenetics. It is noteworthy that understanding how neurons work required knowledge of electricity, and that knowledge was not in hand until Michael Faraday's discoveries in the first half of the nineteenth century.

Some philosophers take it as dead obvious that the enduring existence of many puzzles in neuroscience entails that neuroscience can never, *ever* discover much in the way of mechanisms of cognitive function. One major reason for this conclusion is that they have generally failed to appreciate the clear historical point that the sciences of the nervous system are very young indeed.

The Relation Between Mind and Brain

The words "mind" and "brain" are distinct. Even so, that linguistic fact leaves it open whether mental processes are in fact processes of the physical brain. (Remember: water and H_2O are different words, but they do name the very same stuff.) A favored theory in philosophical thought, championed by Plato, developed by Descartes, and even now defended by Thomas Nagel (2012), holds that just as the words are distinct, so too are the processes. This approach is known as "dualism" – a "two stuffs" theory embracing physical stuff and the utterly different soul stuff. Thinking, seeing, and choosing, according to dualism, are processes of the nonphysical mind or soul. For dualists, the mind/body problem is the problem of how a physical state of the brain can interact with a totally nonphysical state of the soul. By contrast, according to an equally venerable if less popular tradition, there is only the brain; mental processes are processes of the physical brain whose exact nature remains to be discovered. This is known as "physicalism" and found adherents in Hippocrates, Hobbes, Hume, and Helmholtz. Physicalists realize that there is no problem about how the mind and body *interact* inasmuch as there are not two things, but only one thing: the brain. The mind is what the brain does. For them, the important problem concerns how the brain learns and remembers, how the brain enables us to see and hear and think, and how it enables us to move our eyes, legs, and whole body. Their problem concerns the nature of the brain mechanisms that support mental phenomena. Interestingly, dualists also have a closely related set of problems: how does soul stuff work such that we learn and remember, see and hear and think, and so forth. Whereas in neuroscience physicalists have a vibrant research program to address their questions, dualists have no comparable program. No one has the slightest idea how soul stuff does *anything*.

Neurophilosophy as a research program has poor prospects unless mental processes such as remembering and attending are processes of the brain. Otherwise, we should just study the stuff that *does* perform attending and remembering and find out how *that* works, stuff such as the "soul stuff" postulated by Descartes. At this stage in the sciences, the evidence overwhelmingly indicates that all mental events and processes, including visual or auditory perception, learning, memory,

language use, and decision making, are in fact events and processes of the physical brain. It is not that there is one single experiment that decisively shows this. Rather, the evidence has steadily accumulated over countless observations and experiments, and no counterevidence raises doubts.[1] Even though we may not understand in detail the mechanisms whereby we recall an event that occurred in childhood, we are reasonably sure that such a recollection is a brain process. This is not unlike Michael Faraday's realization that electricity was not an occult phenomenon but rather a natural physical phenomenon, even though he did not understand in precise detail the nature of electromagnetism.

One of the most dramatic observations of mind/brain dependency came from the split-brain studies published in the late 1960s. These studies involved patients whose cerebral hemispheres were surgically separated in order to treat drug-resistant epilepsy. The nerve sheet connecting the two hemispheres – the corpus callosum – was the structure that was cut, thereby disconnecting the cortex of the right and left hemispheres. The aim was to aid the patient by preventing a seizure from traveling from its origin in one hemisphere to the other hemisphere. Astonishingly, tests of "split brain" subjects showed that the mental life of the two hemispheres was also disconnected: the right hemisphere might have knowledge the left did not or see something or decide something that the left did not, for example (Gazzaniga and LeDoux 1978). The implications for the mind/body problem were obvious: if mental states were *not* brain states, why would cutting the corpus callosum allow knowledge and experience to be confined to activity in *one* hemisphere? Although a defiant dualist might invent some story to accommodate the facts (and a diehard few did this), the best and most reasonable explanation for the disconnection effects was simply that a *physical* pathway was interrupted, a pathway essential for mental unity, and that soul stuff was just not in the game. As Michael Gazzaniga (2015), one of the leading split-brain researchers puts it, consciousness can be split.

The many observations made by clinical neurologists of patients who suffered focal brain damage also weighed in. Focal brain damage could result in highly specific losses of cognitive function, such as the loss of

[1] See P. M. Churchland (1996a), Frith (2007), and P. S. Churchland (2002) and excellent textbooks such as Baars and Gage (2007).

the capacity to recognize familiar faces, loss of recognition of a limb as one's own, and loss of the capacity to perform an action on command, such as saluting or waving hello. The Damasios, Hanna and Antonio, launched a huge project at the University of Iowa Medical College to systematically document as many cases as possible involving similarly located lesions to test whether there were similar functional effects. This important project elevated brain lesion studies beyond the single case study to a more systematic understanding of the outcome of focal brain lesions and their effect on capacities.[2]

Studies of a few patients who had suffered bilateral damage to the hippocampus (a small curved structure beneath the cerebral cortex) showed them to be severely impaired in learning new things (anterograde amnesia). This finding initiated a massive research program to understand the relation between learning and memory and the hippocampal structures (Squire, Stark, and Clark 2004). Memory losses associated with dementing diseases also linked memory with neural loss and further suggested the tight link between the mental and the neural. Important also are studies of attention using brain imaging along with single neuron physiology. These varied studies suggest that at least three anatomic networks, connected but somewhat independent of the other, are involved in different aspects of attention: alerting, orienting, and executive control. Moreover, each of these functions has been the target of detailed further study, indicating, for example, that there are strong associations between these functions and awareness, especially between detection of a target (consequent on orienting) and awareness (Petersen and Posner 2012).

Developments in psychology, especially visual psychology, also implicated neural networks in mental functions, and this work tended to dovetail well with the neuroscientific findings on the visual system. Explanations of color vision, for example, depended on the retina's three cone types and on opponent processing by neurons in cortical areas. It was well appreciated that much in the world – such as ultraviolet and radio waves – could not be detected by our visual system because of its physical organization.[3] Perception of visual motion was linked to the behavior of single neurons in a visually sensitive area of

[2] For a simple account, see Grens (2014).

[3] See Solomon and Lennie (2007), pp. 276–86, and also chapters 9 and 10 in P. M. Churchland (2007).

cortex known as MT (middle temporal). Visual hallucinations were known to be caused by physical substances such as LSD or ketamine, and consciousness could be obliterated by drugs such as ether, as well as by other substances employed by anesthesiologists, such as propofol. No evidence linked these drugs to soul stuff. On the contrary, many anesthetics appear to work by altering the normal balance of excitation and inhibition of neurons in circuits.

Short-term memory can be transiently blocked by a blow to the head or by a drug such as scopolamine; emotions and moods can be affected by Prozac and by alcohol; decision making can be affected by hunger, fear, sleeplessness, and cocaine; elevated levels of cortisol cause anxiety. Very specific changes in whole-brain activity corresponding to periods of sleep versus dreaming versus being awake have been documented, and explanations for the neuronal signature typifying these three states have made considerable progress (Pace-Schott and Hobson 2002). In aggregate, these findings weighed in favor of the hypothesis that mental functions are a subset of functions of the physical brain, not of some spooky "soul stuff."

Evolutionary biology encouraged us to dwell on the fact that nervous systems are the product of evolution and that the human nervous system is no exception. Comparisons of anatomy, between human and nonhuman nervous systems, have revealed that the functional organization, at both macro and micro levels, has been highly conserved over hundreds of millions of years (Allman 1999). Although human brains are larger than the brains of other land mammals, we share all the same structures, pathways, innervation patterns, neuronal types, and neurochemicals. Neurons in a fruit fly work essentially the same way as neurons in the human brain. Molecular biology revealed that the genetic differences between humans and our nearest relatives, chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*), are very small (Striedter et al. 2014).

These evolutionary relationships imply that either no mammals have nonphysical souls or all do. Now questions flood in: if humans *alone* do have a soul, where do human souls come from, and why does the soul suddenly appear, some 4 million years after the *Homo* species branched off from our common ancestor with chimpanzees? Did extinct *Homo* species such as *Homo erectus* and *Homo neanderthalensis* have souls too? Based on cranial measurements, anthropologists believe that the brains of *Homo neanderthalensis* were typically larger than our brains.

Neanderthals probably had some form of acoustic communication even though they may not have been able to make all the vocalizations of which humans are capable (Lieberman 2013). Moreover, genetic data reveal that they did interbreed with *Homo sapiens* (Pääbo 2014). What about *their* souls? Still other questions challenge the idea that the human soul, not the human brain, is the repository of all that makes us clever. How can ravens and rats and monkeys solve complex problems – how can they sleep, dream, pay attention, and so forth – if a soul is needed for such functions?

By the 1980s, there was impressive, if cautious, agreement among scientists as well as philosophers that the existence of a nonphysical soul that feels, decides, sees, and reasons was improbable. Where disagreement flourished unabated, however, concerned whether neuroscience could *explain* those functions, physical though they may be. Neuroscientists tended to expect that with new techniques and more experiments, progress would continue to be made. How far we shall get, time and research effort will tell.

Some philosophers, by contrast, confidently predicted that neuroscience would never explain cognitive functions, a view particularly associated with Jerry Fodor (1975, 1980, 1998) and his colleagues but widely espoused within the subdiscipline of philosophy of mind. This view tended to be known as the "autonomy of psychology" – autonomous with respect to other sciences, especially neuroscience. It is important to understand that this claim about the limits of neuroscience was just a *prediction*, and it was supported by philosophical speculation, not scientific evidence. Although highly popular until about 1990, the idea has slowly and systematically been undercut by actual progress in the neurosciences, especially by increasingly suggestive links between data at the behavioral, whole-brain, and neural levels. Embarrassingly for the philosophical prediction, convergent studies on functions such as decision making (Glimcher and Fehr 2013), attention (Petersen and Posner, 2012), and spatial representation (Moser et al. 2014), for example, have revealed much more about mechanisms than some skeptical philosophers thought was remotely conceivable (Fodor 2000).

One further reason for ignoring much of neuroscience arose from a misguided analogy. The idea was that cognition is like running software on a computer, where the brain is analogous the computer

hardware.[4] Just as you need not know anything about a computer's hardware to understand an application such as PowerPoint, so you need not understand anything about the brain to understand cognition, or so the argument went. To anyone who looks at all closely at the brain, the disanalogies between brains and conventional computers are so numerous and so profound that the brain/hardware analogy was not taken seriously in neuroscience or bioengineering. Not least among the differences are that brains are parallel not serial processors, that storage and processing in brains are not done by separate modules but by the same structures, and that brains change their structure as they develop from gestation to adulthood and at all stages as they learn (Churchland and Sejnowski 1992). The actual nature of the brain's anatomy and physiology became an inspiration for developing unconventional computers that are more brainlike (Hinton 2013; Yu et al. 2013).

The point where influential philosophers are still confident that the mysteries permanently have the upper hand concerns conscious experience. Typically, there are two distinct arguments to support this conviction. The first argument makes a straightforward prediction about where science will go in the future. It is based on current intuitions about the tractability of the problem of explaining consciousness in neurobiological terms. With great confidence it will be claimed that consciousness is so completely and utterly and thoroughly mysterious, it will *never* be explained at all, period (McGinn 2012, 2014). By way of illustration, it may be suggested that expecting any science to explain how conscious experience emerges from the activity of neurons is like expecting a rat to understand differential equations. Despite its chest-pounding confidence, this prediction should be taken with ample doses of caution because predicting where science will go and what will be discovered is really a rather risky business, to put it politely.

The second and more influential argument rests on the dualist's belief that although nonconscious events such as memory consolidation and preprocessing in vision are brain events, *conscious* events such as feeling nauseous are not brain events. Hence neuroscience cannot explain them. Thus, when I am aware of a pain in my tooth or a decision to kick off my shoes, some philosophers, such as

[4] Dennett (1987) was especially fond of this analogy and appears still reluctant to abandon it.

David Chalmers (1996) and Thomas Nagel (2012), consider those conscious events to be extraphysical, merely running parallel to the physical events.

A methodologic point may be pertinent in regard to the dualist's argument: however large and systematic the mass of empirical evidence supporting the empirical hypothesis that consciousness is a brain function, it is always a logically *consistent* option to be stubborn and to insist otherwise, as do Chalmers and Nagel. Here is the way to think of this: identities – such as that temperature really *is* mean molecular kinetic energy, for example – are not directly observable. They are underwritten by inferences that best account for the mass of data and the appreciation that no explanatory competitor is as successful. One could, if determined, dig in one's heels and say, "temperature is *not* mean molecular kinetic KE, but rather an occult phenomenon that merely runs parallel to KE" (Churchland 1996b). It is a logically consistent position, even if it is not a reasonable position.

In a similar vein, causality, as Scottish philosopher David Hume famously noticed, is not directly observable. It involves an inference to the best explanation available.[5] I cannot literally observe the causal relation between a mosquito on my arm and the itch that follows its departure. But my causal inference is based on strong background knowledge. For another example, despite the powerful evidence that human immunodeficiency virus (HIV) is the major cause of AIDS, some still insist, without contradiction, though perhaps with much mischief, that the cause of AIDS lies elsewhere, such as God's punishment for bad behavior.

To be sure, caution concerning accepted theory does sometimes facilitate the emergence of new causal hypotheses that surpass the prevailing theory in predictive and explanatory power. Scientists, if they are not foolish, then upgrade their causal explanations. For example, it was widely believed that anxiety and poor diet were the major causal factors behind gastritis (inflammation of the stomach lining) until Barry Marshall and Robin Warren in the 1980s challenged that hypothesis experimentally. They discovered the more fundamental cause – a bacterium known as *Helicobacter pylori*. They did not merely vaguely wave in the direction of a *conceivable* different causal claim,

[5] For a new and quite possibly correct account of how causality is represented in the brain, see Danks (2014).

however. They showed experimentally that they had *discovered* a more powerful causal explanation. In the case of conscious experience, although philosophers such as Chalmers and Nagel express their reservations about the brain, the only thing they really do have are reservations. Moreover, their reservations are based on intuitions about how different experience seems to be from states occurring in the physical brain. They have neither competing experiments nor a competing hypothesis with any power or detail; in particular, they have no hypothesis that surpasses let alone competes seriously with the neuroscientific hypothesis.[6] For example, there is nothing that even begins to approach the richness of the neuroscientific literature on attentional mechanisms, for example, that alerting is different from orienting, which, in turn, is different from detection and from executive control. Surprisingly perhaps, with the appropriate intervention, these functions are dissociable, and they are supported by different neural networks.[7]

How do the dualists address the dependencies – the causal dependencies that suggest identification – between consciousness and brain activities? A favored strategy is to propose that conscious states just run parallel to brain states. This proposal may be embellished, perhaps by the idea that conscious states neither cause nor are caused by brain states – the two streams are causally isolated. A variation of this opts instead for a one-way causal street – brain states cause conscious states, but conscious states do not cause brain states. Traditionally, the view that mental states do not cause brain states is called "epiphenomenalism." Actual evidence is lacking for both hypotheses – both are merely empty denials of the idea that consciousness is a biological phenomenon.

Historically, the most renowned defender of two-way causal isolation was Gottfried Leibniz. Leibniz held this view because he thought that it was inconceivable that completely different substances could interact causally. If they shared no properties – not even spatial properties – how could they affect each other? Moreover, with the benefit of contemporary physics, we can see that the causal interaction between *nonphysical* stuff such as a soul with *physical* stuff such as electrons

[6] For discussion of a brain-based hypothesis, see P. S. Churchland (2013a) and Graziano (2013).

[7] See again Petersen and Posner (2012).

would be an anomaly relative to the current and rather well-established laws of physics. More exactly, it would affect the law of conservation of energy. If brains can cause changes external to the physical domain, there should be an anomaly with respect to conservation of energy. No such anomaly has ever been seen or measured. The absence of anomalous data suggests either that the hypothesis of a nonphysical conscious stream of states lacks credibility or that the conscious stream of conscious states does not interact with brain states at all.

When the neuroscientist Josef Parvizi used a tiny electrical stimulus to activate a very specific part of the brain (the middle cingulate gyrus) as part of the preparation of his human patient for surgery, his patient described the emergence of a conscious state consisting of the determination to muster courage to deal with a problem. When the stimulus was off, the feeling vanished (Parvisi et al. 2013; P. S. Churchland 2013b).[8] This experiential event was repeatable in that patient. Moreover, a very similar state was also reproducible in yet another patient stimulated in the same region. The reasonable conclusion is that the stimulus caused the change in conscious state. Some naysayers may wish to take the option that the brain events and the experienced event happen synchronously without causation: the experience stream and the brain stream are separate.

What keeps the two streams synchronized? That is the stunning puzzle that emerges from the epiphenomenal hypothesis. Here is how Leibniz dealt with the puzzle: God sets up and maintains a "pre-established harmony" to keep mental and physical states properly aligned. Needless to say, Leibniz' solution is completely ad hoc, cobbled together to in order to fill an embarrassing silence. Chalmers' does not appeal to God, but he does advert to a future physics that allegedly will explain the alignment between noninteracting streams of mental and brain events. A revolutionary new physics, according to Chalmers' (1996) conjecture, ultimately will explain the nature of consciousness as a nonbrain phenomenon. I have been unable to escape the feeling that this is really the old Leibniz solution suited up in the duds of a future physics instead of theology.

Granting that there are uncertainties in physics, is there a rationale within physics for claiming that a revolution provoked by the mysteries

[8] For a review article on drug-resistant surgery for epilepsy, see Ryvlin, Cross, and Rheims (2014).

of consciousness is in the cards? According to Chalmers, there will be, because nothing less will explain consciousness. Consciousness is so extraordinarily mysterious that only a revolution in physics will account for it.

My small sampling of physicists indicates that they do not wish to rush into investing heavily in a new physics just to address *consciousness*, especially when neuroscience has not by any means been stopped dead in its tracks. And especially when neuroscience has not yielded anomalies that challenge *particle* physics, but only puzzles that might possibly challenge neuroscience. Physicists acknowledge puzzles concerning the possibility of a new theory at the subatomic level to link strong forces, weak forces, and gravity, but these are phenomena in the range of 10^{17}, not in the range of milliseconds and micrometers (10^{-3}), where neurons exist and function. As physicist Steven Weinberg said, the puzzles in physics that motivate a possible revision to the standard model are at the wrong spatial and temporal scale to offer even the barest hint of a solution to the matter of explaining consciousness.[9] Have the philosophers themselves proposed anything substantive by way of a new physics to replace existing physical theory? No. There is nothing substantive – nothing even weakly semisubstantive.

If you are a dualist, either you can pretend that the huge accumulation of dependency evidence in neuroscience is not really there (not a realistic option), or you can say something substantial to address them. Rationally, something must be done insofar as this accumulation appears strongly to favor the hypothesis that conscious states are brain states. A novel strategy, tendered by Chalmers, claims that neuroscientific data are actually *neutral*, as between his parallel-stream hypothesis and the hypothesis that mental states are states of the physical brain.[10]

To assess the figures of merit of this "neural data neutral" strategy, try it elsewhere in science and see what results. Consider the nature of light as understood within contemporary physics: light is electromagnetic radiation (EMR) – light visible by humans is just one part of

[9] This was Weinberg's answer to a question at Gustavus Adolphus College, October 8, 2014. See also Weinberg (2015).

[10] This is a view Chalmers has made explicit only in conversation, though he acknowledges that it is implicit in his earlier writing, even in *The Conscious Mind*.

a larger spectrum that includes x-rays, microwaves, and so forth. Here is what the "neutral strategy" could say about light: "actually, the physical evidence is neutral between the hypothesis that light *is* EMR and that light is not EMR but a spooky thing. That is, light and EMR run in parallel streams, whose synchrony will be explained by a revolution in physics."

Here is what the "neutral strategy" says about life: "all of cell biology is neutral between the hypothesis that life is an occult force (vitalism) and the hypothesis that life is the outcome of the biological structure and organization – cells, membranes, genes, ribosomes, mitochondria, and so forth."

Scientifically, these "data neutral" proposals look counterproductive and more elaborate that the facts require. Silly though they may be, they are not, however, internally incoherent hypotheses. One bizarre claim that oddly appeals to various philosophers of mind is that if the "parallel stream" hypotheses are not internally contradictory, they are as reasonable as established scientific theories. Notice that it is not internally contradictory to say that the Earth is only one hour old, but it would be strange to say that this is as reasonable as saying it is about 5 billion years old.

The twin predictions regarding mind and brain – that neuroscience will never account for conscious experience and that a revolution in physics will explain why – are generally motivated by emphasizing the difference between a neuron, on the one hand, and a feeling of tooth pain, on the other, for example. On reflection, it is argued, the differences appear to be so profound and so complete that surely, *surely* it is inconceivable that the pain in my tooth might really be the activity of neurons in the brain.

Striking though the touted differences are, it is sobering to recall that the history of science is full of discoveries in which seemingly very different phenomena turn out to be one and the same but were viewed from different perspectives (Thagard 2014; Churchland 1989). Breathlessly dramatizing the striking differences lacks the scientific heft to make the dual streams hypothesis compelling.

One problem with relying on what seems inconceivable is this: what is and is not conceivable is, after all, merely a psychological fact about us – about what we can and cannot imagine given our current beliefs and our capacity for imagination. It is not a metaphysical fact about the nature of the universe. In the opinion

of some philosophers, however, trained philosophical intuition has special status and must be taken as revealing deep, "necessary" truths unavailable to untrained others – in particular, unavailable to those with only a scientifically educated intuition (McGinn 2014).[11]

An issue that spells trouble for a nonbrain theory of consciousness concerns the fact that the division between awareness and lack of awareness is typically blurry and often fluid. One place this really shows up is in the automatization of behavior as a skill is acquired, a commonplace phenomenon. As a child learns to read, she ceases to be aware of a word's individual letters; this is also demonstrated in the "word superiority" effect, whereby it is easier for an accomplished reader to read a *word* than to read individual *letters*, as measured by reaction time and errors. Another simple case: I can ride a bike without being aware of my feet working the pedals as I zoom along and think about my upcoming swim. Not so at the beginning of learning to ride a bike, where I had to pay attention to every aspect of riding. Here is the issue: are the many behavioral decisions of which I am unaware just mental brain events that blink out of the mental experience stream until an emergency arises and I must pay attention? Ditto for skating, driving a car, lots of speech and conversation, and, in my case, recently learning to be proficient at standing on my head. And here is related issue: are you aware of body position when you are concentrating on pitching a tent? Sort of and sort of not. Moreover, the neurobiological research on attention helps us to see why the answer is not simple. Apart from automatization of skills, what about shifts of attention, for example, where I cease to hear the speaker and reflect on what I will order for dinner? When I lose awareness of what the speaker is saying, does that just snap out of the consciousness stream and then snap back in? How does that work? What orchestrates and coordinates the snapping? And what *is* snapping?

This raises a second issue. Are our short-lived conscious experiences properties of a "substance"? Or are they just events, properties of "no thing" in the experience "stream"? What maintains the stream as *one* stream? Compared to the serious research in neuroscience on the mechanisms of sleep, attention, visual perception, coma, anesthesia,

[11] See my reply to McGinn (2014) in the *New York Review of Books*, June 19, 2014, p. 65.

and so forth, the naysayers seem to have a totally threadbare alternative, with very little in the way of a substantive explanatory framework.

Why do some philosophers of mind oppose so strenuously the two hypotheses: (1) mental states are states of the brain and (2) probably neuroscience can at least outline the mechanisms of cognitive functions? A range of reasons contributes, but as the frontiers of the behavioral and brain sciences push ever forward into what seems like a thicket of unapproachable mysteries, questions about turf and territory inevitably emerge. A strong assumption in the philosophy of mind is that philosophers are uniquely equipped to set the boundaries of what we can know and to outline the essential and enduring features of concepts that scientists might apply. Philosophical intuition, in this view, is a special trained capacity that can home in on those necessary properties of a phenomenon that science must respect and not challenge. In this way, philosophy sets the foundations for the science. And if philosophers characterize necessary properties of the mind that intuition and logic show cannot be explained by properties of the brain, then that is the contribution of philosophy that science needs to honor.

Thus some philosophers of mind believe that they own a problem space that is concerned with conceptual necessities – necessary truths about psychological states and processes, discovered by conceptual analysis and so-called thought experiments.[12] A necessary truth cannot, according to this approach, be falsified by scientific data. Intuitions trump data. Scientists, not surprisingly, are puzzled by where such *a priori* knowledge might really come from, and they do not want to be bamboozled by philosophical flimflam. After all, intuitions appear to be just strongly held beliefs that are likely grounded in education and reinforcement learning. Intuitions are not, by anyone's account, special reports from Plato's heaven concerning Absolute Truths.

Philosophers are apt to defend their intuitions as supported by thought experiments about what could obtain in any possible world. Supposedly, the outcome of the "thought experiments" will identify the *necessary* truths about, for example, the nature of knowledge. This is a suspect strategy. Recall that Kant thought that he had shown by

[12] This view is not limited to a small minority but is widely espoused and widely taught in philosophy courses. This is readily seen in entries in the online *Stanford Encyclopedia of Philosophy*, which presumably represents the mainstream in the field. See, for example, the entry under "Analysis of Knowledge."

thought experiments that space – the space our Earth and solar system inhabit – is necessarily Euclidean. Alas, the Euclidean claim is not even true, let alone necessarily true. Space is non-Euclidean. Thought experiments, for all the homage paid to them by philosophers, are not real experiments in any sense. Starting an inquiry with intuitions is fine if that is all you have to go on, but then experiment and observation should subject those intuitions to test, and other hypotheses should be considered. In this well-known fashion, experimental psychology and neuroscience have illuminated the nature of our knowledge of the world and the nature of learning, along with the broader question concerning the nature of how nervous systems of all mammals represent the external world (Squire et al. 2012).

How could our intuitions be misguided? Here is how: complex nervous systems are not mere reflex machines or simple conditioning machines; they build models of the external world that are deployed in navigating the world. But not all models are equally accurate to the world itself. A mouse's model of the spatial world may be sufficient to get it around its environs given its limited goals, but it will not be as accurate as *my* model of the spatial world or indeed that of a wolf. Brains also build models of the *causal* world – for example, that fire is hot and can burn us, that red raspberries are tasty, and so on. Regarding causality, too, models have different degrees of accuracy – my general causal model of the world is more accurate than that of my great grandmother or my dog, for example. Finally, the brain builds models of the *inner* world – the world of brain events, including processes we call emotions, drives, and attention. Here again, there are varying degrees of accuracy, and in particular, according to Michael Graziano (2013), the brain's ongoing model of attention can be inaccurate. In particular, it *will* be inaccurate if it embodies the idea that attention is a nonphysical, spooky phenomenon and hence that consciousness is also. Can this sense of "spookiness" be easily shed?

Probably not. By and large, our brains update our world models for us, but the control we have on the updating is limited. I might successfully update my causal model of the world as I come to realize that cholera is caused not by "bad air" but by bacteria. Somehow that information will modify and reshape my causal model of the world. However, a rainbow will still *look* like it has a location in space, even though I know full well that it does not. What about the model of attention and mental states generally? The model of mentality may

persist in *seeming* to be spooky, even when I know "cognitively" that spooky is not accurate to the facts. This may be owed to deep biological features of the way the neural model works.

Here is a comparison: it is a deep biological feature of brains that we extend touch sensations to the end of the pencil or scalpel, to the digger end of the backhoe, and so forth. It seems that we can feel the end of the tool. We all know full well that we have no sensors at the end of the backhoe bucket, but our brain's model finds it very efficient to work that way anyhow – an evolutionary adaptation, no doubt. The point is that as we learn more about the brain, our scientific understanding of our model of attention may become more accurate, but the brain's model of conscious states we use on a moment-to-moment basis may itself be largely unmodified by such neuroscientific knowledge. Thus we may understand more about why it is so easy ("intuitive") to think that consciousness is a spooky phenomenon, even when we appreciate scientifically that consciousness is not spooky but brainy.[13] What is really interesting to me is that we can simultaneously hold both ideas – "spooky" and "brainy" – in our minds, albeit in different ways.

How Did Neurophilosophy Get Started?

Neurophilosophy was more or less inevitable, given the progress in neuroscience and the many links between higher functions and neural activities. Because I happened to be the first to publish using the word "neurophilosophy" (the title of my 1986 book bore that name), I will say a little about my own history.

In about 1978, I came to think that the arguments for an autonomous psychology – a science of the mind autonomous with respect to neuroscience – were too flimsy and self-serving to be taken seriously (e.g. Fodor 2000). If, as seems probable, there is no nonphysical soul but only the physical brain, then surely what is known in neuroscience cannot help but be relevant to understanding the nature of psychological phenomena, including vision, decision making, memory, and learning. Although I have always emphasized that understanding neuroscience was necessary to understand the mind, some philosophers read me as saying neuroscience is both necessary *and* sufficient. This

[13] I owe this point to Michael Graziano in conversation. But see also Graziano (2013).

was a poorly disguised straw man designed to make the project look extreme and unproductive (see McGinn 2014; Churchland 2014).

To appreciate more exactly the contribution neuroscience might make, I recognized that I needed to know as much as I could about neuroanatomy (structure) as well as about the developments in neurophysiology (function). I went to the head of the Neuroanatomy Department at the University of Manitoba Medical College and explained my need. To my everlasting gratitude, he warmly welcomed me and encouraged me to take courses alongside the medical students. The arrangement was informal because I was not enrolled as a medical student – I was, after all, still being paid to teach philosophy to undergraduates. Soon thereafter, I was invited to attend neurology rounds and neurosurgical rounds with the clinicians, a weekly event in which patients with neurologic conditions were presented, following which their cases were discussed in detail. After finishing all available courses, I then became associated with the spinal cord laboratory of Dr. Larry Jordan, which was focused on the neural circuitry that maintained rhythmic walking motions. In the lab, I began to dig much deeper into basic neuroscience.

Among other things, the experience in the Jordan lab taught me that understanding the available techniques is essential to evaluating an experimental article. Data will be unreliable if the technique is unreliable. It also taught me to remember that nervous systems, including our own, are the products of evolution. One of the deepest insights I learned from visiting neuroscientist Rodolfo Llinas was this: the fundamental function of nervous systems is to move the body so that the animal may survive and reproduce. Perception, emotions, and cognition are functions whose features were selected for insofar as they served behavior in the business of survival and reproduction. More exactly, perception and cognition serve prediction, and the capacity to make good predictions is a major driver of brain evolution. Commonplace thirty years later, Llinas's insight provoked me to see everything about cognition and perception in a fresh way.

Of course, my husband and philosophical colleague, Paul Churchland, was as fascinated as I by the adventures in the lab, and he too began to participate in experiments. He readily saw how his own ideas about weaknesses in parts of folk psychology fit with emerging data in the behavioral and brain sciences. Among my colleagues, Jeff Foss and Michael Stack also became hooked, and our daily lunches

were effectively seminars batting around what we were all eagerly learning.

After Paul and I moved to the University of California San Diego, we encouraged our graduate students to have some laboratory exposure while engaged in philosophical research. Many of them did, and some, such as Elizabeth Buffalo, Adina Roskies, and Eric Thomson, eventually left philosophy to find their professional home in neuroscience. Others, such as Rick Grush and Brain Keeley, successfully straddled the two fields. In San Diego, the main neuroscience lab that I was associated with was run by Terry Sejnowski, whose lab was located in the Salk Institute. Francis Crick was also an associate of the lab and was an active participant on a daily basis. Terry's lab focused on a range of topics, including reinforcement learning and the question of what kinds of computations neurons and networks might be using.[14] We also frequently discussed the problem of consciousness and what experiments might help us to understand it as a brain-based phenomenon. Some of the most productive, broad-based, large-scale (one might say "philosophical") conversations took place over tea at that lab. Lab meetings and teatime continue even now to be a source of inspiration and reflection for me.

By and large, the reception of philosophers to the publication of *Neurophilosophy* in 1986 was anything but welcoming. Neuroscientists, by contrast, gave it a much warmer reception, something that seemed to further exasperate philosophers of mind.[15] Owing largely to the blossoming of the brain sciences, the book apparently facilitated the decision of many philosophy undergraduates to do graduate work in neuroscience rather than philosophy.

The hostility from philosophers that greeted neurophilosophy in its early days has mostly abated, and a small but enterprising cohort of younger philosophers has eagerly embraced its general intellectual attitude. They tend to be comfortably immersed in the neuroscience of psychology and philosophy with no sign of metaphysical angst.

[14] What emerged early on was the collaboration that resulted in Churchland and Sejnowski (1992).

[15] John Marshall, a well-known neuroscientist and frequent reviewer of books in Oxford, told me he was asked by the *New York Review of Books* in 1986 to review *Neurophilosophy*. Several years after he submitted his review, he gave me a typewritten copy of the gratifyingly positive review he had written. He explained that the *New York Review of Books* had declined to publish it. He vowed never to write for them again and did not.

Washing their hands of conceptual necessities seems to have left their creativity undiminished. Washington University in St. Louis was the first to set up a graduate program called "Philosophy, Neuroscience, and Psychology" (PNP), which has truly flourished, as has the coordinated undergraduate program. It set the benchmark for other similar programs. Duke University also saw a future in linking with psychology and neuroscience programs, and its programs also have flourished.

No one would call the shift to recognizing the relevance of scientific data a philosophical stampede, however. A quick look at the current graduate courses and syllabi from high-ranking schools in the United States reveals that conceptual analysis tends even now to dominate the philosophical agendas. Mainstream philosophical research on the mind/brain prides itself in being mainly about words, not things. Philosophers in other countries may be moving ahead more quickly. For example, Poland's prestigious Copernicus Center is at the forefront of research on such difficult problems as norms – what norms are; how they are learned, expressed and changed; and what data from psychology and neuroscience reveal about how they guide behavior.[16] Moscow's Center for Consciousness Studies likewise has a cohort of young researchers who are aiming to make progress on traditional problems about the nature of consciousness, knowledge, and representation by integrating data from many labs.[17]

Quine and the Conceptual Analysis Dogma

A powerful but oft-ignored lesson of Quine's (1960) discussion concerning naturalizing philosophical inquiry[18] is simple: clarifying a concept used to categorize the world can be very helpful in avoiding confusion in a seminar, but that clarification cannot itself tell us whether that concept truly applies to phenomena in the world, whether it should be revised in the light of facts, or even whether it possibly should be ditched altogether.

The applicability of a concept to phenomena in the actual world depends on science (broadly speaking) and discovery of the facts. This is obvious in the case of a concept such as "caloric," where we can be

[16] See, for example, Brozek (2013) and Heller, Brozek, and Kurek (2013).
[17] See the well-informed interviewers, Vadim Vasiliyev and Dmitry Volkov discuss neurophilosophy with me at: https://youtu.be/GP8o-yjZePc.
[18] See also my Preface to the second edition (2013).

reasonably clear about what were believed to be the properties of caloric fluid, were it to exist; for example, it moves from hot things to cold things, hot things have more of it than cold things, it has no mass, and so forth. All that clarity notwithstanding, there is no such thing as caloric fluid. Differences in temperature are a matter of differences in mean molecular kinetic energy, not in volume of caloric fluid.

Consider now the case of a concept such as "soul," where we might have something like Descartes' idea of what we mean by the concept. A philosophical analysis of that concept tells us precisely nothing whatever about whether souls really exist or even whether they have the properties outlined in its analysis. The meaning of a word merely reflects current beliefs, and those beliefs may be misguided. Think of wholesale revisions to the concept of an "element," originally believed to comprise *earth, air, fire* and *water*, not one of which is now considered an element. The point extends more generally. In particular, it extends to words such as "knows," "believes," "rational," and "decides."

To elaborate, Quine's point was that what is *meant* by a word reflects what is *believed to be true* about the *things* the word denotes. Thus, meaning changes as knowledge expands. This point has been stoutly resisted by scientifically naive philosophers who supposed that if something is considered part of the very meaning of a word, then it is a necessary feature of the stuff denoted by that word. That the phenomenon has that meaning-linked feature is, allegedly, a necessary truth, and necessary truths are, needless to say, necessarily true regardless of what science discovers. Thus, these philosophers convince themselves that they can dope out the deep – *necessary* – features of the world by conceptual analysis.[19]

The argument sinks into the fallacious when it shifts from saying that something is part of the meaning of the *word* to saying what is a necessary feature of things in the *world*. Hence, even if, for nineteenth-century physicists, "is indivisible" is part of the very meaning of the word "atom," this does not make it necessarily true – or even true at all – that atoms are indivisible, that they have no substructure. Nevertheless, philosophers have been prone to make claims about what must be true

[19] See, for example, the entry in the online *Stanford Encyclopedia of Philosophy* under "Analysis of Knowledge." The authors, Jonathan Jenkins Ichikawa and Matthias Steup, state that a proper analysis of knowledge "should at least be a necessary truth."

about the mind based on their analyses of the meaning of words, words such as "knows" and "believes" and "conscious."

One quick further point about conceptual analysis: typically what is marketed under the banner of "conceptual analysis" is not actually a reflection of what a word means in its everyday use by ordinary folks (Schooler et al. 2014). Rather, it is a *theory*, albeit a camouflaged theory, about the nature of some phenomenon, such as consciousness or choice or knowledge. Consider, for example, the idea that beliefs require language because beliefs are states of mind standing in relation to a sentence. This idea is not based on what ordinary speakers of the language mean or even on what is implied by what they mean. Such claims go well beyond meaning. These are actually empirical hypotheses, *disguised* and *sold* as conceptual truths, based on scanty, or even no, empirical evidence.

Theorizing is an important undertaking in the effort to advance knowledge and understanding of the world, including the world of the mind/brain. Philosophers are as welcome into the theorizing tent as anyone else, and certainly some philosophers have made important contributions in this domain.[20] Clinging to outdated ideas concerning conceptual analysis and necessary truths impedes the progress that philosophers might otherwise make. In general, it is more rewarding to take account of existing data when trying to generate an explanatory theory of a phenomenon than to troll one's intuitions for "necessary truths," something the witty biologist Sir Peter Medawar (1979) suggested is the philosophical equivalent of "psychokinetic" spoon bending.

Concluding Remarks

As more is discovered about brain organization and the dynamics of neural networks and whole systems, our knowledge of mental functions also will expand, undoubtedly in unpredictable ways. Whether unsurmountable obstacles will be encountered is not known – certainly not known even by philosophers who insist that their well-trained intuitions have already spied such obstacles.

[20] For example, Eliasmith (2013), Craver (2009), Silva, Landreth, and Bickle (2014), Smith (2011), Danks (2014), Bickle (2013), Arstila and Lloyd (2014), P. M. Churchland (2013), and Glymour (2001).

In science, we typically cannot tell whether a problem is just not yet solved or absolutely unsolvable. You cannot tell just by looking – or just by using your intuition. Just as the Straits of Gibralter were once thought to mark the outer limits of the world, so it may be tempting to think that what we cannot now imagine marks the limits of what science can discover. This is a mistake, one that is rooted in philosophical complacency and a failure of intellectual courage. Of course, some problems are not problems for neuroscience or for philosophy – such as the problem of making a vaccine against the Ebola virus or sequencing the genome of an extinct species of humans such as *Homo erectus*. Some problems, as Sir Peter Medawar wisely reminded us, are political problems concerning the more effective way to address terrorism or whether to allow doctor-assisted suicide for the terminally ill. Some problems are personal problems about whether to change jobs.[21] But some problems *are* problems for science, and it is highly likely that the nature of consciousness is one of those problems. Whether we do actually solve it remains to be seen.

Young philosophers need to ask themselves a basic question: what is it that I really want to understand? Is it just what other philosophers *say* about a problem and how I might figure out a clever response within their framework of assumptions? Is it something about current English usage, such as what the word that names the problem usually *means*? Or is it the *nature* of the thing – how it works? These are quite different questions, using very different methods, and leading a researcher in very different directions.

[21] I too make this point, for example, in Brain-Wise (2002), yet philosophers such as Roger Scruton (2014) continue to wag their finger and warn that science cannot solve all problems.