Philosophy and the Brain Sciences

PETER MACHAMER and JUSTIN SYTSMA

Abstract: What are the differences between philosophy and science, or between the methods of philosophy and the methods of science? Unlike some philosophers we do not find philosophy and the methods of philosophy to be sui generis. Science, and in particular neuroscience, has much to tell us about the nature of the world and the concepts that we must use to understand and explain it. Yet science cannot function well without reflective analysis of the concepts, methods, and practices that constitute it. For example, experimental methods and their resulting empirical data are essential for understanding the world, yet such data is not a-conceptual. Understanding how and what theoretical assumptions, conceptual assumptions, and practical knowledge guide the use of experimental methods is relevant to understanding the results of that use. In this way, philosophy — whether done by philosophers or scientists — has a role to play in understanding the world. Neuroscience is typically individualistic in focus; nonetheless, the mechanisms of learning and linguistic ability that some neuroscientists study also have a role to play in understanding communication. Philosophy cannot offer adequate understanding, even conceptual clarity, in isolation from empirical investigation. Yet, this does not require that science or scientific concepts will replace or reduce philosophical concepts, let alone those of ordinary language.

The excitement surrounding advances in the brain sciences over the past several decades has worked its way into philosophy. Philosophy of neuroscience is a growing field. However, boundary disputes over the nature of philosophy and science have arisen, often under the guise of trying to establish the nature of the mind/brain relation. In this essay we will explore such disputes by contrasting two recent books. In Philosophical Foundations of Neuroscience M. R. Bennett and P. M. S. Hacker argue, in neo-Wittgensteinian manner, that philosophy and philosophical problems are independent from neuroscience, while Patricia Churchland, Brain-Wise: Studies in Neurophilosophy, claims that neuroscience is our best bet for answering some perennial philosophical questions about the mind. We are not attempting to review these books, but to use them to show that this tension is not necessary, that philosophy has a role

1 The authors wish to thank Edouard Machery and Lisa Osbeck for their helpful comments.
to play in scientific work just as science has a role to play in explaining phenomena that concern us. That is, we hope to redirect the divisive questions about the nature of mind and world found in these works to a more inclusive treatment of the nature of explanations.

There is something slightly misleading in the title of Bennett and Hacker’s *Philosophical Foundations of Neuroscience*. They do not deal with many important aspects of neuroscience or much of what falls under philosophy of neuroscience, and the foundational claims they investigate have strict philosophical biases. They write that, “neuroscientists are aware of the fact that the problems with which they are struggling are not unrelated to reflections of philosophers on the nature of mind and mental phenomena.” But, the bulk of the problems that neuroscientists are struggling with are not undertaken with an eye toward such connections; most neuroscience, even cognitive neuroscience, does not directly deal with the nature of mind or persons.

What Bennett and Hacker discuss is the relevance of neuroscience to traditional problems in philosophy of mind with an emphasis on debunking those neuroscientists who claim, they say, that it is “time for philosophy to step aside and let science have its turn.” As such, their critique can be readily applied to neurophilosophy as articulated by Patricia Churchland. Against such an approach, Bennett and Hacker repeatedly suggest that the problems of relating mind and brain are conceptual (i.e. philosophical), not empirical (i.e. scientific). For them, philosophy (and more specifically their version of analytic linguistic philosophy) is “above all a conceptual investigation”:

Its province is not the domain of empirical truth or falsehood, but the domain of sense and nonsense. It investigates and describes the bounds of sense: that is, the limits of what can coherently be thought and said. Its destructive task is the criticism of transgressions of the bounds of sense.

From this perspective they argue that various statements in neuroscience/philosophy of mind transgress the bounds of sense. Some primary themes are

---

4 For example, neuroscience, as a science, raises all of the problems about explanation, evidence, role of experiment, discovery of mechanisms, etc. that all sciences have. That is, the basic issues in philosophy of science may be raised specifically about neuroscience. In fact, we will argue that much of the debate about the relation of neuroscience to philosophy of mind depends upon how one views philosophy. See, e.g., many of the articles in P. Machamer, P. McLaughlin, and R. Grush (eds.), *Theory and Method in the Neurosciences*, Pittsburgh: University of Pittsburgh Press, 2001.

5 Bennett and Hacker, *Philosophical Foundations of Neuroscience*, p. 396.

6 Ibid., p. 397


what Bennett and Hacker term “the mereological fallacy in neuroscience” (i.e., that brain parts are not mind or person parts), the misapplication of person-level psychological predicates to the brain or its parts, and how to talk sensibly about “conscious experience, mental states and qualia.”

Patricia Churchland nicely stands as a contrasting exemplar since she looks to neuroscience for answers to questions in the philosophy of mind. In her textbook on neurophilosophy she points in this direction although she shies away from the extreme position of science no-philosophy-necessary that Bennett and Hacker target. Nonetheless, she holds up neuroscience as the route that progress in philosophy of mind will take. Churchland writes:

Neurophilosophy arises out of the recognition that at long last, the brain sciences and their adjunct technology are sufficiently advanced that real progress can be made in understanding the mind–brain. More brashly, it predicts that philosophy of mind conducted with no understanding of neurons and the brain is likely to be sterile. Neurophilosophy, as a result, focuses on problems at the intersection of a greening neuroscience and a graying philosophy.10

Where Bennett and Hacker see philosophers as being in the business of clearing up conceptual confusions, including those that have crept into the sciences, Churchland suggests that scientific results will provide, or minimally must be used to inform, answers to traditional philosophical questions about consciousness, selfhood, and free will.11

While Churchland treats these as real problems in need of empirical data for resolution, Bennett and Hacker take them to reflect conceptual confusions that are not illuminated by experimental methods.12 Not surprisingly, Bennett and Hacker’s method follows Wittgenstein’s linguistic turn. Reflecting on their methodology, they write: “Believing that methods are vindicated by their products, we have preferred to apply the methods of connective analysis to the conceptual problems that characterize neuroscience at the point where it abuts on psychology.”13

The basic problem that needs to be addressed, and we shall comment upon it throughout this essay, is what are the differences between philosophy and science, or between the methods of philosophy and the methods of science. Unlike Bennett and Hacker we do not find philosophy and the meth-

9 Churchland, Brain-wise.
10 Ibid., p. 3.
11 Ibid., p. vii.
13 Bennett and Hacker, Philosophical Foundations, p. 378.
of philosophy to be *sui generis*. Science, and in particular neuroscience, has much to tell us about the nature of the world and the concepts we must use to understand, explain, and function in the world. Yet science cannot function well without reflective analysis of the concepts, methods and practices that constitute it. For example, experimental methods and their resulting empirical data are essential for understanding the world, yet such data is not a-conceptual. Understanding how and what theoretical and conceptual assumptions and practical knowledge guide the use of experimental methods is relevant to understanding their results and assessing their validity. In this way, philosophy, whether done by philosophers or scientists, has a role to play in understanding the world.

While Bennett and Hacker acknowledge the relevance of philosophy to science, noting, for example, that “when empirical problems are addressed without adequate conceptual clarity, misconceived questions are bound to be raised,” they are far less generous concerning the reverse direction of influence:

> Conceptual questions antecede matters of truth and falsehood [...] They determine not what is empirically true or false, but rather what does and does not make sense. Hence conceptual questions are not amenable to scientific investigation and experimentation or to scientific theorizing.\(^\text{15}\)

This view of philosophy, however, cannot be fulfilled by Bennett and Hacker’s faith in what they call “connective analysis.” In fact, even their guiding principle – that “the meanings of words are determined by their rule-governed use, and they are given by what are accepted as correct explanations of meaning by the community of speakers”\(^\text{16}\) – belies such a clear philosophy-science separation. Demarcating communities of speakers and determining what they accept as correct explanations of meaning would seem to require some empirical investigation, unless there is some mysterious *a priori* way of determining who constitute the relevant communities, what uses are accepted, and what the normative criteria of acceptable usage are. Such questions are related to sociology, socio-linguistics and social psychology most directly. Communication is obviously interpersonal and social; but, this does not mean that it is not illuminated by empirical investigation.\(^\text{17}\) Neuroscience is typi-

---

\(^{14}\) Ibid., p. 2.

\(^{15}\) Ibid., p. 2.

\(^{16}\) Ibid., p. 382.

cally individualistic in focus; nonetheless, the mechanisms of learning and linguistic ability that some neuroscientists study also have a role to play in understanding communication. Philosophy cannot offer adequate understanding, even conceptual clarity, in isolation from empirical investigation. Yet, this does not require that science or scientific concepts will replace or reduce philosophy, let alone de-throne ordinary language.

We begin by considering Bennett and Hacker’s use of their philosophical dogma, comparing it to that of Wittgenstein in his *Philosophical Investigations*, which they take as the basis for their treatment of the purported mereological fallacy in neuroscience and the so-called problem of consciousness. We will then briefly investigate the contrasting treatment given by Patricia Churchland. We conclude with a comparison of their respective views on reduction in neuroscience and some general remarks concerning the interplay between the brain sciences and philosophy.

1. The So-called Mereological Fallacy

Bennett and Hacker’s main claim concerns the logic of psychological predicates. They assert:

The organs of an animal are parts of the animal, and psychological predicates are ascribable to the whole animal not to its constituent parts. Mereology is the logic of part/whole relations. The neuroscientists’ mistake of ascribing to the constituent parts of an animal attributes that logically apply only to the whole animal we shall call “the mereological fallacy” in neuroscience.\(^\text{18}\)

This supposed mistake is amply illustrated with examples of psychological predicates being applied to the brain. Certainly some of this usage seems, at the very least, incautious. Bennett and Hacker, however, treat it as especially grievous, charging that “this application of psychological predicates to the brain makes no sense.”\(^\text{19}\) They continue: “The brain is not a logically appropriate subject for psychological predicates. Only a human being and what behaves like one can intelligibly and literally be said to see or be blind, hear or be deaf, ask...

---

\(^{18}\) Bennett and Hacker, *Philosophical Foundations*, p. 73.

\(^{19}\) Ibid., p. 72.
questions or refrain from asking.”\textsuperscript{20} The question is why such usage is so untoward and whether it is in principle—not in any specific instance—nonsense?

Unfortunately, little in the way of actual argument is given for this restriction of psychological predicates to whole animals. That the meanings of psychological predicates are sensible when applied to wholes, and that, therefore, applying such terms to parts is to speak nonsense, supposedly follows from connective analysis and the principle that “the meanings of words are determined by their rule-governed use.”\textsuperscript{21} Assuming there to be such rules of use, connective analysis then extrapolates to what combinations of words do or do not make sense. Thus, “humans see” makes sense, while “brains see” does not. On the face of it, however, a statement such as “brains see” does not appear to be nonsensical. Further, it is not clear why “whole animals see” is sensible. We are not told what seeing consists in such that some combinations are illicit. Surely in some cases metonymy, whether intended or not, is a perfectly good trope.

Bennett and Hacker’s primary support is drawn from Wittgenstein’s *Philosophical Investigations*\textsuperscript{22} and is misleadingly isolated from its context. They write:

Wittgenstein made a profound remark that bears directly on our concerns. “Only of a human being can one say: It has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious.” This epitomizes the conclusion we shall reach in our investigation.\textsuperscript{23}

Despite the strength of the language of §281, however, Wittgenstein did not hold that secondary applications of psychological predicates to things that do not behave like human beings are nonsensical. Rather, his point was that our primary use of such terms is applied to human behavior. As such, in §282, he offered a contrast class for his statement that was rather different from the part/whole distinction:

We do indeed say of an inanimate thing that it is in pain: when playing with dolls for example. But this use of the concept of pain is a secondary one. Imagine a case in which people ascribed pain only to inanimate things; pitied only dolls!

Such pain ascriptions to dolls are not nonsensical. While the secondary extension presupposes the primary usage, it does not follow that the extension is

\textsuperscript{20} Ibid., p. 72.
\textsuperscript{21} Ibid., p. 382.
\textsuperscript{22} Wittgenstein, *Philosophical Investigations*, §281.
\textsuperscript{23} Bennett and Hacker, *Philosophical Foundations*, p. 71.
meaningless. It appears doubtful that psychological attributions to brains would be nonsensical, but the question remains as to whether they are informative.\textsuperscript{24}

Wittgenstein emphasized behavior and drew a contrast with inanimate things that do not behave. Brains, however, are not inanimate, not clearly without behavior. Roughly, brains receive inputs that lead to series of neuronal firings, which activate further outputs. Such neuronal activity may warrant treating such firings as a behavior. Thus, one way to make sense of attribution of psychological predicates to brains would be to consider their application to brain behavior. If psychological predicates may be used to describe brain behavior, then such uses are not nonsensical, and may be informative in a way that the attribution of the same predicates to dolls is not.\textsuperscript{25}

For example, if we are concerned with edge-detecting cells in the visual cortex, we might say that a cell saw an edge in the presented visual stimulus when it fires, and that it did not see the edge if it did not fire. The cell exhibiting the appropriate response is enough to make the secondary application of the psychological predicate understandable. Bennett and Hacker write that in our primary application “we say of an animal or a human being that they perceive something in their field of perception if, for example, they respond to what is visible (or audible, etc.) in appropriate ways.”\textsuperscript{26} The application of “sees” to the edge-detector cell follows this directly, taking into account the types of response open to cells and how they are related to causes of behavior. While it is unclear exactly what Bennett and Hacker mean by “appropriate,” here, the edge-detector cell can be said to behave \textit{appropriately} in that it reliably responds to edges and in doing so to behave \textit{appropriately} with respect to the larger system of which it is a functional part. As J. J. Gibson\textsuperscript{27} showed, edge-detection is related to humans’ ability to perform visually guided movement in a physical environment.

\textsuperscript{24} Another general point made repeatedly by Wittgenstein, but running contrary to Bennett and Hacker’s work, is that our usage is seldom clearly and strictly bounded. Where Bennett and Hacker hold the meanings of terms to be determined by their rule-governed use, Wittgenstein challenged this, arguing that meanings are not typically fixed and, as he put in §69, that a term’s use is “not everywhere circumscribed by rules.” Bluntly, we find with Wittgenstein, \textit{pace} Bennett and Hacker, that sense and nonsense are not so cut and dry in natural language. Their approach, in Wittgenstein’s words in §81, is “to think that if anyone utters a sentence and means or understands it he is operating a calculus according to definite rules.” Wittgenstein’s later work, including the \textit{Investigations}, denied the validity of such a conception, despite his earlier attraction to it.

\textsuperscript{25} We are aware that there is a distinction between behavior and action that may be drawn. But usually the distinction depends on attribution of intentions, and we are not sure that intentions may not be ascribed to brain parts as representations. So for our purposes, behavior and actions are what things do, the activities they engage in.

\textsuperscript{26} Bennett and Hacker, \textit{Philosophical Foundations}, p. 81.

In this way, psychological ascriptions to neurons or the brain also can be informative on a deeper level than as mere descriptions of their immediate behavior. Perhaps the most basic foundation of neuroscience is that the behaviors of neurons and brain systems cause overt animal behavior. Most neuroscientists seek to explain the phenomena they are interested in by discovering and elaborating the neural and systemic mechanisms responsible for some cognitive behavior that is validly related to the behavior experimentally tested.\textsuperscript{28} Many applications of psychological predicates to the brain simply isolate the neural mechanism thought to explain animal behaviors. The \textit{parts} that neuroscientists deal with (neurons, cells, systems, etc.) exhibit behaviors or activities (firing, polarizing, activating, etc.) that are causally related to the behavior of \textit{whole} organisms.

It is conventional, in both broad and restricted communities of speakers, to apply predicates for the whole to the parts responsible for the abilities being predicated. This is a form of metonymy. It is not at all incomprehensible or nonsensical, for example, to say that my car engine is powerful, even though we mean that the car as a whole is powerful. The engine is powerful because it is functionally relevant to the ability of the car to achieve high speeds. When properties of the component are part of the explanation of the relevant ability of the whole, it is often meaningful to apply the predicate to the component. Likewise, the application of psychological predicates to the brain can play a role in explaining the abilities that our customary application to humans reflects by drawing links between the activities of the brain and the abilities of the human that they enable.

Such linking, or bridging, of explanations will be discussed in the final section of this essay. Here it is sufficient to note that if the link is provided by a mechanism (if the use of “brains seeing” falls within an articulation of neural mechanisms by which humans see), then nothing illicit appears to be going on. It is quite legitimate to make inferences such as, e.g., “He can’t see, therefore there might be something wrong with the optic nerve.” Nonetheless, the application of psychological predicates to parts of a human touches on a number of philosophical issues. While a series of neurons might be causally and functionally relevant to our ability to see, there remains the temptation to say that the neurons do not truly see. Even when a neuron systematically responds to visual cues, and its doing so is causally related to the ability of the

\textsuperscript{28} See P. Machamer, L. Darden, and C. Craver, “Thinking about Mechanisms,” \textit{Philosophy of Science}, 67 (2000), pp.1-25. It has been suggested that we are equivocating on the word “behavior” here. There is no doubt that brain behavior, cognitive behavior, and animal behavior may be distinguished. But the point is that they all may be studied and cognitive neuroscience attempts to elicit the mechanisms which relate them.
human it is part of to see, there is an inclination to deny any real seeing to it. Likewise, while we might describe metaphorically an automatic door with a light sensor as seeing when it exhibits the appropriate behavior of opening, it nonetheless seems odd to take such speech too seriously. We want to say that the door does not *truly* see anything.

There are at least two reasons we might have for this reticence. First, one might hesitate to say that the door sees because unlike human seeing, the door has no understanding of what it sees. While the door might exhibit (a very limited range of) seeing behavior, it is not conscious of what it is seeing and is unable to report any contents of its awareness. (We often say things like, “I didn’t see you,” in this sense to mean that we did not recognize the person in front of us as someone that we know or were not aware of the identity of the person we were seeing.) In this sense, we also want to deny that neurons as isolated parts can see because they are not conscious of anything, let alone visible objects.

Second, one might hesitate to say that the door truly sees if you hold that true seeing involves “phenomenal consciousness,” “conscious experience,” a “qualitative aspect,” “qualia,” etc. – phrases that are often used to indicate a belief that the entity has something more than is suggested by its possession of the relevant abilities and that is not expressed in its behaviors. If “seeing” is understood in this way, then ascription of seeing to an electronic door, a neuron, or a brain is unlikely to be explanatorily useful, as will be discussed in the following section. If we demand *explanations* of something more than an entity’s ability to see – its ability to exhibit sight-behavior (whether micro or macro, and including both simple responses and those indicative of awareness such as first-person reports) – then we are likely to encounter problems whether we are talking about a human, its parts, or another type of entity all together; but neuroscientists are not generally occupied with this problem of consciousness.

2. Conscious Experience, Mental States, and Qualia

As noted above, one foundation of neuroscience is that the behaviors of neurons and brain systems are crucial causes involved in producing overt animal behavior. Part of this is the belief that overt animal behavior is open to (but not exhausted by) mechanistic explanation. In this sense it is a philosophical foundation of neuroscience that animals (including humans) are mere physical systems. The contrast, here, is with those (vitalists, dualists) who would argue that some non-physical entities are needed because without them human behavior is *in principle* inexplicable. The usually also hold some very strong views about what may count as legitimate physical terms. Put another way, neuroscientists generally deny that there is a fundamental difference in kind between physical,
mechanical systems and mammals, although there is a huge gulf in complexity and in how they may interact with other systems. This is a complex claim and we shall return and elaborate it somewhat in the final section.

Many of the points we raised above can be found in Bennett and Hacker’s discussion of consciousness. They argue for a multifaceted conception of consciousness that emphasizes powers or abilities but restricts these to sentient creatures (or certain whole animals). They write:

First, sentient creatures who are conscious beings do not contain consciousness, they are conscious (or unconscious), and conscious of various things. They feel pain, perceive objects in their environment, feel fear or anger, take pleasure in various activities and conditions, desire things and pursue what they want. They have various active and passive powers, including the passive power of having their attention caught and held by something they perceive: that is, the power to become and then to be conscious of something they perceive.29

Accepting this, we hold that neuroscience is essential to explaining these powers, to explaining how it is that “sentient creatures” do the things that they do. In doing so, neuroscientists will inevitably treat sentient creatures as physical systems because their interest is in uncovering the physical mechanisms involved in producing the capacities at issue.

In contrast, Bennett and Hacker hold that “it is deeply misleading to refer to sentient beings as ‘physical systems’. “30 They claim:

[S]entient beings are not mere physical systems. The atmosphere (weather system) might be said to be a “mere physical system,” a volcano can be said to be a mere physical system, and so too might a pocket calculator or a computer. But animals and human beings are not mere physical systems, but living, sentient “systems.” Sentient beings are precisely what we contrast with mere physical systems. And it is obvious why we do so, for they have capacities that mere physical systems lack.31

This is not the only contrast to be drawn, however: Most neuroscientists hold that sentient creatures are physical systems in that they are not, even in part, non-physical – they have no extra vital spirit, immaterial soul, or Cartesian mind. In other words, it is held that despite their immense complexity and their extraordinary abilities, sentient creatures are still just physical beings whose (quite sophisticated) capacities can be explained in physical terms. We will return to this issue in Section 4, where we argue that such attempts to reductively explain the behavioral capacities of animals do not

30 Ibid., p. 299.
31 Ibid., p. 298.
commit neuroscientists to the type of ontological reductionism that Bennett and Hacker fear when they write that the “properties and powers [of sentient creatures] ultimately depend on, but are not reducible to, the physical and microphysical processes that characterize their physical constitution.” We argue that while it is true that these powers are not ontologically reducible to their generating mechanisms, they are explained by those mechanisms – and producing such explanations is a core goal of the brain sciences.

Explanatory vs. ontological reductionism of the “properties and powers” of sentient beings is not the issue that Bennett and Hacker’s discussion of conscious experience focuses on, however. Instead, they look at the broad way in which the term “consciousness” is meant when it occurs, for example, in phrases like “conscious experience” and “conscious mental states”:

[N]euroscientists, cognitive scientists and philosophers are inclined to equate consciousness with sentience in general, or indeed to extend it dramatically to almost the whole range of the mental. Our first task is to investigate this extension. Subsequently, we shall probe one of its roots: namely, the conception of qualia that is currently embraced by many neuroscientists and philosophers.

This usage – the sense of “consciousness” tied to the philosophical concepts of phenomenal consciousness and qualia – is Bennett and Hacker’s primary target. There is much to like in their analysis of these problematic concepts. Our concern here, however, is with the interplay between philosophy and the brain sciences. Focusing on the latter, it is misleading to emphasize those neuroscientists who embrace qualia, or “those who characterize ‘the problem of consciousness’ as the leading problem of neuroscience,” in elucidating the foundations of neuroscience. Most neuroscientists do not consider the problem of consciousness to be a problem for serious research.

Bennett and Hacker characterize the problem as claiming that consciousness seems to be beyond the physical. They write:

Puzzlement is generated by the thought that a merely physical description of the world would omit experience. To be sure, the behaviour of animate bodies would be included in it. Experiences, however, are not behaviour. They are something that underlies behaviour, something essentially subjective. The behavior that exhibits sorrow, hope, joy, fear, affection, etc. is merely the outer husk of the inner psychological reality with which each subject is intimately acquainted.

32 Ibid., p. 301.
33 Ibid., p. 263.
34 Ibid., p. 261.
However, this supposed problem is not a scientific problem for mainstream neuroscientists, rather it is a meta-problem that challenges the scope of their research or adequacy of their methods. This problem of consciousness is essentially an argument that there is something (conscious experience or phenomenal consciousness) that neuroscience should be able to explain, but which is thought to resist the types of physical and behavioral explanations that mainstream neuroscience employs. By stipulation, this problem is not and cannot be a central problem of brain neuroscience. It is by fiat ruled out as a problem for the brain sciences to tackle empirically, and by this exclusion claimed as a philosophical problem that questions the solely reductivist, physicalist ontology often read into the explanatory work done by neuroscientists.

Not surprisingly, even a cursory review of current neuroscience journals reveals that this problem of consciousness is not the central problem of neuroscience. In fact, the bulk of the work dealing with consciousness over the past two decades concerns the various active and passive powers — such as attention and awareness — that Bennett and Hacker treat as philosophically unproblematic. This fact is easily missed, however, in the swirl of quotes that Bennett and Hacker reproduce from a relatively small number of researchers who have been actively pushing for scientific engagement with the problem of consciousness; authors such as Benjamin Libet, Bernard Baars, Francis Crick, and Gerald Edelman. Baars, in particular, has championed what he calls “the new consciousness science,” noting that it is an ongoing challenge to interest mainstream brain scientists in treating “consciousness as an object of scientific scrutiny.” He writes: “For years it was common to hear scientists say that human consciousness was unlike any other scientific problem, in that it was not at all clear what evidence was relevant to it; as for theory, it seemed so far beyond our comprehension that it was hardly worth talking about.”

While Bennett and Hacker’s characterization of the centrality of the problem of consciousness is misleading, the enthusiasm and optimism displayed by researchers like Baars indicates that there is a need for philosophical analysis of the concept of consciousness. Redirecting Bennett and Hacker’s analysis from mainstream neuroscience to the growing minority of researchers engaged with the new consciousness science, much of their critique is appropriate. These scientists are engaged with an essentially ontological problem that owes much of its prominence to a backlash against certain materialist or physicalist

38 Ibid., p. 34.
39 Ibid., p. 13.
It is striking (and perhaps should be disturbing) that the most fundamental reason [for this puzzle] is philosophical, indeed Cartesian. Seen from one particular perspective, it can appear deeply puzzling how causal transactions in the material world can give rise to anything as categorically distinct from matter as experience. How can the impact of radiation upon the cones and rods of the retina generate the experience of (consciously) seeing something? How can events in the cortex “give rise” to conscious mental states?  

This puzzle is classified as philosophical because it is ontological in nature; but, one can deny that neuroscience is (or should be) in the business of generating ontologies without denying that the findings of neuroscientists are relevant to the philosophical debates.

The puzzlement that Bennett and Hacker point out has been commonly seen in the philosophical debates on consciousness of the past 30 years. This is not a failing of neuroscience to be rectified by an everyday perspective, however. The basic problem, here, is not in talking about experiences or even conscious experiences and their properties, but in taking all descriptions to refer to kinds of ontological entities and assuming that physical or material descriptions can only refer to fundamental particles and forces. The problem lies in generating incommensurable ontological pictures of the world from our descriptions of it. But there is no obvious reason why physiological explanation must be incompatible with first-person descriptions of one’s experiences, despite the fact that they have different grammars. Different types of criteria are used in scientific and phenomenological assertions. As Wittgenstein noted, my recognition that I am in pain operates differently from my judgment that someone else is in pain: “it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself.”41 In the first-person case, “what I do is not, of course, to identify my sensation by criteria.”42 The point is that we usually do not observe ourselves and then decide whether we are experiencing a given sensation or not based on applying a criterion to some inner-experience. Scientific third-person attribution functions differently. It is based on applying criteria in order to give observation reports, and they are subject to justification by reason giving. Despite this, a merging of first- and third-person approaches is not only possible, but

40 Bennett and Hacker, _Philosophical Foundations_, p. 262.
41 Wittgenstein, _Philosophical Investigations, §246._
42 Ibid., §290.
is commonplace in medicine where first-person descriptions of the patient’s pain are used by the physician as a basis for diagnosis. Treatment is guided both by physiological understanding and phenomenological description. And sometimes it is difficult for the patient to report what it is that is felt.

We agree with Bennett and Hacker that a problem of consciousness arises when language goes astray. However, it is not primarily the neuroscientists who have misused language. As J. L. Austin held, it is very often the philosophers who have created an unnecessary technical language. Nonetheless, to meaningfully explore issues of consciousness requires more than just cleaning up the terminology; we also need knowledge of what neuroscience shows us about the workings of the brain if we are to integrate our mechanistic understanding of the world with our sensory experience of it. While Bennett and Hacker’s almost exclusive focus is on conceptual issues, and thereby largely denies the relevance of neuroscience to issues of consciousness, Patricia Churchland holds that there is a real philosophical problem of consciousness that neuroscience could solve by way of reduction. Against this, we argue in the next section, that her, sometimes, strictly reductive physicalist ontology in fact closes off neuroscience from actual relevance to the problem.

3. Churchland and Consciousness

Patricia Churchland’s demarcation of the positions on the problem of consciousness follows her physicalist assumptions. She pejoratively describes two basic attitudes that one could have with regard to phenomenal consciousness: “Whereas pragmatists tend to emphasize that consciousness is a natural phenomena of the brain, mysterians favor the idea that it is a supernatural phenomenon, or at least is beyond the physical in some sense or other.” In this division a physicalist ontology is equated with the natural, and results in an either/or forced choice task: consciousness is either physical or it is mysteriously beyond the physical. Her next step is to argue that the physical choice is possible and to articulate scientific approaches to the problem:

It is possible, for all we can tell now, that consciousness, or at least the sensory component of consciousness, may be subserved by a physical substrate with a distinctive signature. In the hope that there is some distinct and discernible physical marker of the substrate, the direct strategy aims first to identify the

44 Churchland, Brain-wise, p. 128.
substrate as a correlate of phenomenological awareness, then eventually to get a reductive explanation of conscious states in neurobiological terms.\textsuperscript{45}

The suggestion is that consciousness itself is the primary problem. The first goal is to find a physical substrate that correlates with consciousness and which might serve as a mechanism for consciousness. Thus, Churchland notes that:

Discovering some one or more of the neural correlates of consciousness would not on its own yield an explanation of consciousness. Nevertheless, in biology the discovery of which mechanism supports a specific function often means that the next step – determining precisely how the function is performed – suddenly becomes a whole lot easier.\textsuperscript{46}

One question that should be raised here, however, is whether consciousness is best conceived of as the function of some neural mechanism? Following Bennett and Hacker, if we think of consciousness in terms of the various active and passive powers that are distinctive of sentient creatures, then minimally neuroscientists should be investigating a host of functions and looking for a variety of mechanisms that enable those diverse powers.

Churchland’s suggestion for a direct scientific approach to consciousness is somewhat odd. It suggests that there is a big problem out there (a hard problem of consciousness); but, this makes it seem like the brain sciences have not made any progress in explaining how we consciously do some of the things that we do. Thus, Churchland’s concern with the sensory component of consciousness cannot be a concern with how we perceive objects in the environment, for physical explanations of perceiving objects are not merely possible, but actual (in mechanism schemata, if not all the details). We can explain, for example, why a certain wine tastes bitter in terms of tannins and phenolic properties of those tannins, and how these activate the bitter taste receptors in the tongue, etc. Such examples can be produced \textit{ad nauseam}. Churchland, however, seems to want to explain something more; she wants to explain consciousness itself – treating it as something had by a conscious being, or in Bennett and Hacker’s terms, as something contained in such a being – by reducing it to a set of working entities and activities within her physicalist ontology.

The basic problem is that Churchland allows her ontological concerns to shape her assessment of the scientific questions that need to be asked and her commitment to a very specific ontological form of physical reduction. It is also unclear what kind of things are allowed to count as physical in her reduction. Churchland’s concern with offering a physicalist response to the prob-

\textsuperscript{45} Ibid., p. 134.
\textsuperscript{46} Ibid., p. 135.
lem of consciousness leads her to pose a big ontologically-loaded question, rather than asking a host of smaller questions about the things we are able to do as conscious beings. Instead of asking what consciousness itself is, we should start by asking why we are aware of some of the sensory distinctions we make (but not others). Why is information about the bitter taste of a wine accessible to some subjects, but not accessible in the same way to other people? Why does the information that provides the content of our first-person reports have the mode of access that it does, and not function in some other, more mechanical or unconscious way?

Churchland’s engagement with consciousness as a potential problem for ontological physicalism is in step with the philosophical literature. Consider, for example, Ned Block’s distinction between access consciousness and phenomenal consciousness. It is phenomenal consciousness that raises the big problems, but it does so because it is taken to conceptually irrelevant for the explanatory practices in the brain sciences. Block’s concern is that in answering the various informational questions about our sensory abilities, we leave something significant out (phenomenal consciousness). This sense that something is left out, however, again relates to treating scientific explanations as ontological descriptions. Exactly how committed Churchland is to explaining what is left out is not always clear, because it is far from clear what she is looking for in her search physical substrates or mechanisms for consciousness.

The point we wish to stress is that insofar as consciousness is understood as a substantive thing whose very existence is problematic, the possibility for scientific explanation is undercut. Instead we should focus on developing explanations of what is done consciously, which will include detailing the processes by which information is provided and used by some animal systems. These explanations, in turn, need to include how such information is accessed. Knowing consciously is adverbial, as opposed to knowing unconsciously or procedurally. A relevant distinction is between conscious and unconscious activities, and the different forms these take and the uses they have, not between the nature of consciousness versus the nature of physical things. That is, the goal should be to offer explanations of what we do, not to make unneeded ontological claims. Bennett and Hacker seem to agree, and do not make ontological claims, though they have little to say about the status of the referents of the terms related to consciousness in ordinary language.

4. Science, Reduction, and Consciousness

Churchland writes that “when I step on a thorn, it still hurts in the same way, whether I know that the pain is really an activity in neurons or not.” This expresses a traditional philosophical form of reduction, or ontological reduction. Churchland holds that, as it turns out, pain is really (is nothing but) brain activity. This is an ontological claim about the nature of pain. Churchland asserts, for example, that the pain is not in the foot but in the brain. At the same time, Churchland also focuses her description of reduction on explanation. She stresses that “the possibility that mental phenomena might be understood in a neuroscientific framework is associated with reductive explanation in science generally.” She continues:

Most simply, a reduction has been achieved when the causal powers of the macrophenomenon are explained as a function of the physical structure and causal powers of the microphenomenon. That is, the macroproperties are discovered to be the entirely natural outcome of the nature of the elements at the microlevel, together with their dynamics and interactions.

What is not clear is that in giving a reductive explanation of a creature’s abilities, for example, to recognize, conceptualize, and respond to pains in terms of neuronal activity we will have shown that pain is really nothing but neuronal activity.

More broadly, Churchland writes that there are three hypotheses that underpin her book. The first hypothesis is: “Mental activity is brain activity. It is susceptible to scientific methods of investigation.” The issue is what connects these two sentences. The first suggests an ontological reduction, but the second indicates that she is concerned with scientific explanation. Further, the second sentence is not a hypothesis; mental activity (understood in terms of the powers and abilities evidenced in conscious behavior) has been and is being investigated scientifically. This does not require or imply that mental activity is nothing but brain activity, however.

While we argued above that a neuron seeing is not nonsensical, what was meant is that (1) some neuronal behavior can be reasonably and informatively described as seeing behavior and (2) that such neuronal behavior is part of an explanation of the organism’s seeing. That we explain behavior in a certain way does not imply an ontological reduction – that seeing is really (nothing but)

49 Ibid., p. 20.
50 Ibid., p. 21.
51 Ibid., p. 30.
brain activity. The explanation does not imply elimination, but phrases such as “is really,” “is nothing but” do. Such a move indicates confusion about different senses of the term “reduction.” Churchland writes that, “an understanding of mental phenomena – such as memory, pains, dreaming, and reasoning – in terms of neurobiological phenomena is a candidate case of reduction, inasmuch as it looks reasonable to expect that they are brain functions.”\textsuperscript{52} Rather, we find that it is reasonable to investigate our cognitive abilities in terms of neurobiological mechanisms. This investigation need not be, and we argue should not be, seen as reduction in this strong philosophical sense of identity.

Bennett and Hacker are clearer on the distinction between ontological reductionism and explanatory reductionism.\textsuperscript{53} They correctly note that the “is really” type of reduction found (at points) in Churchland is “not science but metaphysics.”\textsuperscript{54} At the same time, they go on to play fast and loose with the distinction – points made against ontological reduction are shifted to explanatory reduction, including an emphasis on laws that reifies the entities in explanations.

Bennett and Hacker’s initial point is to deny ontological reduction of psychological attributes to the interactions of neurons. Thus, they strike the wedge that drives the split between neuroscience and psychological attributes in their discussion of the so-called mereological fallacy. They write:

So, if there is no sense to literally identifying neural states and configurations with psychological attributes, there cannot be general bridge principles linking the reducing entities (neural configurations) with the entities that are to be reduced (psychological attributes). But if there can be no bridge principles, then there is no hope for any form of reduction that will allow one to derive the laws governing phenomena at the higher level of psychology from the laws governing phenomena at the neural level.\textsuperscript{55}

Thus, even in explanatory reduction, Bennett and Hacker see the denial of bridge principles linking these two realms of entities to hinge on the possibility of nomological reduction. The suggestion is that explanatory reduction must be nomological reduction; they then argue that the latter fails because there are no laws of human behavior: “Not only are there no bridge principles allowing any form of ontological reduction of psychological attributes to neural configurations, but it is far from evident that there is anything that can be dignified by the name of psychological laws of human action, that might be reduced to, and so explained by reference to, whatever neurologi-

\textsuperscript{52} Ibid., p. 20.
\textsuperscript{53} Bennett and Hacker, \textit{Philosophical Foundations}, p. 356.
\textsuperscript{54} Ibid., p. 356.
\textsuperscript{55} Ibid., p. 362.
cal laws might be discovered.”56 It is not clear, however, that neuroscience in order to offer explanations of human behavior typically seeks, or that it ought to seek, laws of human behavior that then reduce to neurological laws and then to the laws of physics.

As Bennett and Hacker note, it is not clear there are or need be any such laws, especially if laws are taken to be nomological universals. However, this does not preclude mechanistic explanation. Most of biological, and neuroscientific, phenomena are explained by complex mechanisms that are not law-like in their behavior; but the elucidation of these mechanisms is nonetheless explanatory. The explanatory need for universal laws is just not there.57 Moreover, few if any mechanisms are specifiable in terms of one kind or level of ontological entities and activities, which is what everyone seems to be assuming is necessary for philosophical reduction. Mechanisms are, most always, multilevel. While isolating tannin as the cause in the wine of the bitter taste we experience when we taste that wine, means that this chemical phenolic compound is a cause of our experience, it is does not mean that our experience is nothing but the chemical reactions in our tongue to this compound. The experience involves our discrimination of the bitter taste of the wine, and this involves numerous chemical reactions, but none or even all of these is the experience. The systemic properties of the experience of tasting the wine are not at the same level as the various parts of the mechanism that explains our ability to discern the distinctive taste of this wine. But then neither are overt behaviors such as facial expressions of disgust or ambiance provided by the friends with whom one tastes.

Neuroscientists explain human or animal behavior by describing neuronal and systemic properties relevant to the mechanisms by which the behavior is produced during a given experiential episode. Bennett and Hacker seem to recognize this, although they treat mechanistic explanation as the articulation of conditions of possibility:

Neuroscientific explanations can typically explain how it is possible for creatures with such-and-such a brain to do the kinds of things they do. They can explain what neural connections must obtain and what neural activities must take place in order for it to be possible for the animal to possess and exercise the powers it naturally possesses.58

56 Ibid., p. 362.
58 Bennett and Hacker, Philosophical Foundations, p. 364.
This modal mixing of “must” and “possible” mistakes the nature of the explanations. Neuronal mechanisms are not intended to state general conditions of possibility. Discovery of mechanisms moves from how possibly, to how plausibly, and, hopefully, to how actually the behavior is produced. The movement is driven by the evidence we find that the mechanism is the right one. This is not ontological or “is really only” reductive explanation. It is reductive in a sense that we are explaining a macro-phenomenon by appeal to systems and subsystems that are more micro. Generally, this is the way that scientists, as opposed to philosophers, use the term “reduction.” This does not preclude other explanations at other levels.

There is another important aspect of claims about reduction that needs to be addressed. Bennett and Hacker hold that our behavior is explained by our everyday descriptions: “We constantly ask for explanations of why someone did or is doing something and we are typically given answers that satisfy us.”\(^{59}\)

For example, we ask why someone raised her hand and are typically satisfied by the answer that she was hailing a taxi. Bennett and Hacker state that the “typical explanations we require take the form of citing the agent’s reason for doing what he did, or citing his motive, and these are neither causal nor nomological explanations of human behavior.”\(^{60}\) Such explanations clearly cannot be reduced to neurobiological explanations. But Bennett and Hacker do not focus on the main reason for this. Reason giving is a social phenomenon that exists among communities of people, and most neurobiological and cognitive explanations confine themselves to individuals and parts of individuals. It does not follow, however, that neuroscience cannot explain any behavior, or even that what neuroscience may tell us is not relevant to social behaviors. It does follow that explanations solely about individuals cannot explain the social presuppositions necessary for making social behavior and action intelligible.

5. Concluding Thoughts

Let us close this essay by looking briefly at some relations between explanations by mechanisms and ontic commitments. We do not believe that any science (let alone all the sciences) is in the business of providing ultimate metaphysics. However, science does provide us, at any given time, very good hypotheses for some of the “things” in the world, and these are the entities and activities that scientists use when providing explanations by mechanisms. Machamer, Darden and Craver wrote:

\(^{59}\) Ibid., p. 363.
\(^{60}\) Ibid.
Nested hierarchical descriptions of mechanisms typically _bottom out_ in lowest level mechanisms. These are the components that are accepted as relatively fundamental or taken to be unproblematic for the purposes of a given scientist, research group, or field. Bottoming out is relative […] The explanation comes to an end, and description of lower level mechanisms would be irrelevant to their interests […] But remember, what is considered the bottom out level may change […] At different historical moments, in different fields, different mechanisms, entities, and activities have been discovered and accepted. The set of types of entities and activities so far discovered likely is not complete. Further developments in science will lead to the discovery of additional ones.61

Now while we believe in mechanisms, we do not believe there are only some sort of fundamental “physical” mechanisms. We may describe the American legal system in terms of the mechanisms by which it functions, and so invoke social entities such as judges, legal precedent, jury selection, and sentencing practices. We may explain how knowledge works, which involves social norms for using the knowledge.62 We may attempt to discover and articulate cognitive mechanisms for planning that use entities and activities that do not refer to the brain at all. The point here is that our ontic commitments are dictated to some degree by theories we have about the world and by the sciences we choose to use to answer our questions. Physics is not the only science; in fact, physics is not even one unified science. Moreover, about many things we have no theory or science, or, if you will, only a folk theory that serves many of our purposes quite adequately.

Bennett and Hacker ask, “is it really intelligible to suppose that the conduct of individual human beings in the circumstances of their lives will always be rendered clearer by neuroscience?”63 Of course not – no more than would our ability to use a refrigerator be greatly enhanced by understanding the physics involved in explaining how refrigerators work. But our ability to fix them may be. Neuroscience does not aim to replace day-to-day explanations of behavior, anymore than quantum mechanics aims to replace our everyday explanations of falling bodies. Expert knowledge is not needed or called for in most of our everyday interactions.

Nonetheless, we all also know, in a day-to-day sense, that changes in our brains affect our behavior. We know what brain damage can do, what drugs can do, even what hunger can do to our behavior. We do not avoid doctors because the physiological details of their explanations are more technical.
than our day-to-day explanations. We go to them exactly because of this. One explanation need not reduce another for it to be more useful in some situations for some purposes, nor do explanations need the status of laws to be scientific. Further, no explanation need be the best explanation in every situation it could be applied to. Nonetheless, it is silly to go to the opposite extreme and deny the possibility or value of neuroscientific explanations.

The value of neuroscientific explanations can be recognized and accepted without rigid ontological commitment to the entities that frame them; likewise, this value is not diminished by recognizing the possibility of important and useful explanations in other terms. Explanations do not invoke ultimate ontological entities with a corresponding apriori metaphysical status. Explanations are historical phenomena that are context, theory and purpose relative. Explanations need to be seen as our attempts to understand the world, to predict its course, and to intervene in the furtherance of our interests. Some explanations are more useful than others in a given context. Usefulness does not require that an explanation uses a *is really only* description. Reductionism in the strict and philosophical sense has no place in science or in philosophy.

Peter Machamer
University of Pittsburgh
pkmach@pitt.edu

Justin Sytsma
University of Pittsburgh
justin@JustinSytsma.com