Behavior and Philosophy, *34*, 71-87 (2006).

OF WHAT VALUE IS PHILOSOPHY TO SCIENCE? A REVIEW OF MAX R. BENNETT AND P. M. S. HACKER'S *PHILOSOPHICAL FOUNDATIONS OF NEUROSCIENCE* (MALDEN, MA: BLACKWELL, 2003)

José E. Burgos
University of Guadalajara – CEIC

John W. Donahoe
University of Massachusetts, Amherst

ABSTRACT: The book *Philosophical Foundations of Neuroscience* (2003) is an engaging criticism of cognitive neuroscience from the perspective of a Wittgensteinian philosophy of ordinary language. The authors' main claim is that assertions like "the brain sees" and "the left hemisphere thinks" are integral to cognitive neuroscience but that they are meaningless because they commit the mereological fallacy—ascribing to parts of humans, properties that make sense to predicate only of whole humans. The authors claim that this fallacy is at the heart of Cartesian dualism, implying that current cognitive neuroscientists are Cartesian dualists. Against this claim, we argue that the fallacy cannot be committed within Cartesian dualism either, for this doctrine does not allow for an intelligible way of asserting that a soul is part of a human being. Also, the authors' Aristotelian essentialistic outlook is at odds with their Wittgensteinian stance, and we were unconvinced by their case against explanatory reductionism. Finally, although their Wittgensteinian stance is congenial with radical behaviorism, their separation between philosophy and science seems less so because it is based on a view of philosophy as a priori. The authors' emphasis on the a priori, however, does not necessarily commit them to rationalism if it is restricted to conceptual or analytical (as opposed to synthetic) truths.
Key words: cognitive neuroscience, ordinary language, Wittgenstein, mereological fallacy, Cartesian dualism, essentialism, reductionism

Philosophical Foundations of Neuroscience (2003) is a close philosophical scrutiny of neuroscience. At 480-odd pages it provides engaging reading for those interested in a truly critical appraisal of a largely unchallenged field (but see Uttal,

AUTHORS' NOTE: Work on this paper by the first author was partly funded by Grant No. 42153H from the Mexican National Council for Science and Technology (CONACYT). We thank Armando Machado for inviting us to prepare this review and for useful comments to a previous draft. Please address all correspondence to José Burgos, Francisco de Quevedo 180, Col. Arcos de Vallarta, Guadalajara, Jalisco 41130, MEXICO; Email: jburgos@cucba.udg.mx, or John Donahoe, Program in Behavioral Neuroscience, Department of Psychology, University of Massachusetts, Amherst, MA 01002, USA; Email: jdonahoe@psych.umass.edu.

2001). The extraordinary progress of this field suggests that all is fine and well. This progress has given its practitioners, some of Nobel Prize fame, a high sense of self-confidence often expressed as a celebration of its independence from philosophy. Alas, celebration degenerated into mocking derision. In this book, however, philosophers strike back—and with a vengeance—convincingly showing that neuroscience is not as healthy as it seems. One of its most appealing areas, *cognitive* neuroscience, is seriously ill. The etiology of the disease is neither empirical nor theoretical, but *logical*. This diagnosis is made from a *Wittgensteinian* philosophy of ordinary language, and it applies not only to cognitive neuroscience but also to important segments of philosophy itself.

The book is the result of an unprecedented collaborative effort by Max R. Bennett, a noted neuroscientist, and P. M. S. Hacker, a leading expert on Wittgenstein. Their prose is disarmingly candid and direct, their analysis lucid, challenging, and sharp. The book is detailed, extensively documented, well written and organized, and quite friendly (one-sentence summaries of the key ideas have been conveniently inserted at the beginning of virtually every paragraph). Love it or hate it, the arguments are not to be taken lightly. The book should be read carefully by professionals and students of neuroscience, psychology, and philosophy. We thus highly recommend it as an exemplary exercise in special philosophy of science that can serve as a guide for other disciplines, in particular psychology. Our recommendation, of course, is not driven by unreserved approval (see later) but the importance of the topic and the scholarly manner in which Bennett and Hacker ("the authors" henceforth) have treated it.

After an introductory précis of the book, the contents are divided into four parts and two appendices. Part I begins with an historical survey (Chapter 1) revolving around the contrast between the Aristotelian and the Cartesian views of the soul. Chapter 2 focuses on the work of Sherrington and his disciples (Adrian, Eccles, and Penfield), whom the authors regard as Cartesian dualists. In Chapter 3, the authors sketch their main criticism. Current cognitive neuroscientists are not substance dualists but repeatedly commit Descartes' mistake, what the authors call the "mereological fallacy." The term refers to *mereology* (from the Greek "*meros*," meaning "part" or "portion"), the branch of ontology that deals with part–whole relations (for the definitive technical treatise on mereology see Simons, 1987). This chapter also includes rebuttals of some objections to this criticism and an outline of Wittgenstein's private-language argument, which the authors use to propound a view of how the meanings of ordinary psychological terms are learned.

Parts II and III provide a detailed justification of their criticism through "connective analysis" (p. 378), a method of delineation of ordinary ("common or garden") psychological concepts and their interconnections via analyses of the use of ordinary psychological terms. This exercise targets the writings of prominent cognitive neuroscientists (Crick, Edelman, Kandel, Blakemore, Damasio, Gazzaniga, Squire, Young, LeDoux, Libet, Bennett himself, and others) on sensation and perception (Chapter 4), the cognitive powers of knowledge and memory (Chapter 5), the cogitative powers of thought and imagination (Chapter 6), emotion (Chapter 7), volition and voluntary movement (Chapter 8), and

consciousness (Chapters 9 through 12). In the latter four chapters, the authors also examine critically the views of philosophers of mind such as McGinn, Chalmers, Dennett, and Searle on consciousness. Part IV waxes more philosophical, highlighting the problem of reductionism (Chapter 13) and the relations between philosophy and neuroscience (Chapter 14). The two appendices are also philosophical and are dedicated to criticisms of the methodological proposals of Daniel Dennett and John Searle.

The Authors' Message

The mereological fallacy is discussed in detail in Chapter 3 but introduced in Chapter 1, where it is defined as follows: "ascribing to a part of a creature attributes which logically can be ascribed only to the creature as a whole" (p. 29). It is a violation of what the authors call the "mereological principle in neuroscience": "psychological predicates which apply only to human beings (or other animals) as wholes cannot intelligibly be applied to their parts, such as the brain." The authors continue thus:

> Human beings, but not their brains, can be said to be thoughtful or thoughtless; animals, but not their brains, let alone the hemispheres of their brains, can be said to see, hear, smell and taste things; people, but not their brains, can be said to make decisions or to be indecisive. (p. 73)

In their historical account, the fallacy is a descendant of the ventricular doctrine of Nemesius (ca. 400 A.D.), according to which *all* mental functions are *localized* in the ventricles. Nemesius thus departed from Aristotle's view of the soul or psyche as the "unexercised dispositional powers" or "essential, defining functions" (p. 14) of a *whole* living being. The summit of this departure is Cartesian dualism, where a soul (mind, self) is essentially a *thinking* immaterial *substance*, a body is an essentially material (spatially extended) substance, and the two are the causally interacting *parts* of a human being. Thus Descartes, like Nemesius, ascribed psychological attributes to the soul as a part of a human being.

Cartesian dualism, supplemented with Locke's account of qualities, became the main philosophical influence on cognitive-neuroscientific research for the next three centuries to the present time. Although current cognitive neuroscientists have largely abandoned substance dualism, they keep ascribing psychological attributes to parts of human beings, typically their brains. Such attributions, however, are *logically flawed* or *fallacious*. Here, "fallacious" refers to a *logically invalid argument* whose premises do not *entail* its conclusion. It is a logical error in that the argument is taken *as if* it were—when it actually is not—logically valid.

Standard analyses of fallacies in informal logic revolve around prototypical inference patterns of faulty ordinary reasoning. Two patterns concern us here: *division* and *equivocation*. The fallacy of division erroneously prescribes that what is true of something is true of its parts (e.g., "A clock gives the hour; hence a clock's wheels also give the hour"). The fallacy of equivocation results from using

certain terms in different senses throughout an argument (e.g., "The end of a thing is its perfection; death is the end of life; hence, death is the perfection of life").

The mereological fallacy can be seen as a sort of compound of these two fallacies, restricted to ordinary psychological terms and concepts. In cognitive neuroscience, the fallacy is paradigmatically manifested in expressions where brains and brain hemispheres are asserted to perceive, believe, know, reason, imagine, remember, feel, and be aware or conscious. The problem with these expressions is not that they are false but that they are *meaningless*. They assert nothing, so they are not assertions. They are nonsensical gibberish. The mereological fallacy, then, is not a factual or theoretical error that can be corrected through more experimentation or better theorizing. It is a kind of "confusion" or "incoherence," words the authors use frequently to refer to unintelligible uses of ordinary psychological terms.

For example, consider the terms "sight" and "belief," with all their grammatical variants. As epitomized in Aristotelian hylomorphism, they are ordinarily used to refer to whole creatures. One ordinarily says "*I* see a red light," not "my brain sees a red light," and "*you* believe it is raining," not "your brain believes it is raining," and so on. The two senses are conceptually connected (e.g., "I believe I saw a red light"). One might want to use "sight" more precisely, to refer to a part of a creature, like its striate cortex. In the interest of clarity, the new sense should be introduced through an *explicit definition*. For instance, one should declare at the outset that *for the purposes of the analysis*, "sight" will be defined as "a temporary activation of one or more neurons in striate cortex correlated with a temporary presence of an electromagnetic radiation of a certain wavelength." *Under this definition*, expressions like "My brain saw a red light" become meaningful.

The authors do not prohibit this kind of redefinitional move per se. They have no problem with technical or quasi-technical redefinitions of ordinary psychological terms. Such redefinitions are a common scientific practice that the authors do not dispute:

> There is nothing unusual, let alone amiss, in scientists introducing a new way of talking under the pressure of a new theory. If this is confusing to the benighted readers, the confusion can easily be resolved. Of course, brains do not literally think, believe, infer, interpret or hypothesize, they think*, believe*, infer*, interpret* or hypothesize*. They do not have or construct symbolic representations, but symbolic representations*. (p. 74)

> Our criticisms of the mereological fallacy in neuroscience do not preclude neuroscientists from using the verbs 'to think', 'to believe', 'to perceive', 'to remember' in new ways according to conditions other than the received conditions of their use, as long as they can explain what these new uses mean. They can, if they so wish, redefine 'thinking', 'believing', 'perceiving', 'remembering', and *give* a meaning to the phrases 'My brain thought that it was better to keep silent', 'Your brain believes that it is Tuesday tomorrow', 'His brain perceived that she was smiling', or 'Her brain remembered to go home.' (p. 384)

In our example, then, it is not that my brain saw a red light, but that my brain saw* a red light, where "saw*" is to be linked to the above redefinition of "sight." What would be the point of such redefinition? The same as that of many other (if not most) definitions: *abbreviation*. By and large, definitions are abbreviation devices that seek economy of expression. Why use "sight*"? This question is more difficult. One answer is to concede that the string "sight" could indeed be confusing, so it would be better to coin a *new* term, making the awkward asterisk unnecessary. Another answer is closer to, but does not quite raise, the authors' concern: sight* could be *hypothesized* as a neural correlate of a certain form of ordinary use of "sight."

The authors have no qualms with these answers either, although redefinitions are not enough:

> New formation rules would have to be stipulated, the conditions for the correct application for these innovative phrases would need to be specified, and the logical consequences of their application would have to be spelled out. (p. 384)

The result of this kind of task, however, would be a system of concepts quite different from the ordinary psychological ones that motivated the analysis in the first place.

The authors' criticism is that cognitive neuroscientists neither have done, nor seem to want to do, the additional work. Rather, "they are trying to discover the neural basis for *thinking*, *believing*, *perceiving* and *remembering*—not for *something* else" (p. 384). This task, of course, is perfectly legitimate. What is dubious is to try to accomplish it by construing brains and brain hemispheres as thinking, believing, perceiving, and remembering. Such constructions do not allow us to understand human behavior any better than construing clock cogs as giving the time allows us to understand clock behavior. They only give us the *illusion* of understanding. They are no better (in fact, they are worse) than *conceptually clear descriptions* of experimental findings.

In our example, the authors would criticize inferences of assertions about the brain believing from assertions about the brain seeing* (e.g., "your brain saw* a red light and *hence believed* it was real"). Such inferences are *violations of the semantic limits* (in the authors' words, "transgressions of the bounds of sense") imposed by "sight*" relative to those imposed by the conceptual connections between "sight" and "belief." The meaning of "sight*" is thus *mixed up* with that of "sight." It is this kind of semantic muddle that the authors regard as incoherent and confused, and it is what they insist cognitive neuroscientists have been doing ever since Nemesius.

In another example from the book, Sperry (1974; cited by the authors, p. 389) took research on split-brain patients to demonstrate that the right hemisphere is "a conscious system in its own right, perceiving, thinking, remembering, reasoning, willing, and emoting, all at a characteristically human level." This interpretation is a paradigmatic example of the mereological fallacy in cognitive neuroscience: to assert about a human brain hemisphere what, *under the meanings of the terms used*, makes sense only to assert about whole humans. The problem, again, is that

such expressions are meaningless, for the terms "conscious," "perceiving," "thinking," "remembering," "reasoning," "willing," and "emoting" have not been technically redefined, let alone their new conceptual interrelationships specified. Consequently, the *only* meanings they can have when used are the ordinary ones. However, these meanings apply only to whole human beings. Hence, the expressions assert nothing and they do not contribute to our understanding of this research at all. They only add confusion.

Instead of Sperry's account, the authors propose the following:

> What has been discovered by experiments on split-brain patients is a very strange *dissociation of functions that are normally intimately associated* and a consequent *confabulation-generating confusion*, which are manifest primarily (but not exclusively) under experimental conditions when the visual stimulus is controlled by the experimenter. (p. 391)

> This functional dissociation and associated confabulation is explained by reference to the fact that the light stimulus from the snow scene affected the right hemisphere, the severance of which from the left hemisphere deprived the patient of the ability to describe or be visually aware of what was presented to him on the left of his visual field, although, remarkably, he was, by pointing, able to associate correctly what was there (viz., the snow scene) with a shovel. Nevertheless, he did not know *why* he made that association (not being aware of the snow scene being presented to him), and confabulated a tale to explain why he had done so (a confabulation comparable to those produced by subjects to explain their post-hypnotically suggested behavior). This, in turn, is crudely explicable by reference to the fact that the visual stimulation of the right hemisphere is disconnected from the left hemisphere, so that the patient is deprived of his normal cognitive capacity to be visually aware of what is presented to him and to recognize and describe familiar objects that are thus presented. It does not however deprive him of the ability to associate what was visually presented to him on the screen with an appropriate object (viz. a shovel)—but without knowing why he is doing so. (p. 392)

It is an admittedly crude explanation, largely indistinguishable from a description of the phenomenon, except for the fact that electromagnetic radiations within a certain wavelength range affect the right hemisphere, and that the right hemisphere is functionally connected with the left hemisphere through the corpus callosum. If a more refined explanation is wanted, ascribing perception, thought, memory, reason, will, and emotion to the right hemisphere will not do. On the contrary, it will only worsen matters, for it will create conceptual muddle.

More refined explanations would appeal at least to the functioning of the visual cortex of each hemisphere, how they are functionally related, and how their normal relations are disrupted by commissurotomy. This explanation can be made as refined as wished by appealing to the structure and functioning of the different microcircuits that constitute each cortex, and so on, down to the cellular and molecular levels. This possibility, however, does not mean that the split-brain phenomenon, let alone the normal functioning of a human being, is *reducible* to

the structure and functioning of a nervous system. The authors explicitly reject ontological, explanatory, and eliminative reductionism, although more for philosophical implausibility than fallaciousness (see later).

The authors argue that the muddle is not poetic license, metaphoric license, or science popularization, which are perfectly legitimate if pursued wisely. Nor is it the result of struggling with the poverty of ordinary psychological vocabulary. The muddle is interwoven in the theoretical and experimental fabric of cognitive neuroscience. It cannot be corrected within science through experimentation or more accurate theorizing, for it is neither an empirical nor a theoretical problem. It is a purely *conceptual* problem that can only be resolved by clarifying the logic of the terms being used, and this clarification is the bailiwick of philosophy. Of course, it cannot invalidate the wealth of experimental data amassed in cognitive neuroscience, nor can it inspire new research. It will only make research presuppositions, explanations, and interpretations more meaningful.

If we have grasped the authors' point, we can hardly disagree with it. The dictum that stems from it is a sensible one: *use your terms carefully*. Using a term carefully signifies not only defining it clearly and following its definition, but also specifying and paying close attention to its *conceptual connections* with other terms. Definitions are cheap, so following them is unproblematic. Again, *the authors' concern is not with definitions in and by themselves*. Of course, definitions can be unclear and, to this extent, hinder scientific research, but this is not the kind of unclearness the authors' have in their sights. To them, being *conceptually* unclear is not merely a matter of providing unclear definitions; it is also a matter of *not abiding by* the conceptual connections among the terms used. Not doing so in cognitive neuroscience has led to meaningless research presuppositions, justifications, and interpretations. The painstaking analyses in Parts II and III show convincingly that the mereological fallacy runs rampant in current scientific and philosophical research on the mind–body nexus.

Our Concerns

Our basic agreement notwithstanding, we have a few concerns. First, it is unclear whether the mereological fallacy can actually be committed in Cartesian dualism. Moreover, it is not even clear that this doctrine needs mereological talk of souls being parts of humans, or humans being composites of souls and bodies. Second, there is an odd tension between the authors' Wittgensteinian stance, which is characteristically anti-essentialistic, and their embracement of Aristotelian hylomorphism, which is characteristically essentialistic. Third, we were unconvinced by the authors' case against explanatory reductionism. Let us elaborate each concern in turn.

Cartesian Dualism and the Mereological Fallacy

The authors' charge of "crypto-Cartesianism" against current cognitive neuroscientists, perhaps the most striking aspect of the book (see pp. 111-114, 233-

235), requires further examination. The charge rests on the claim that the mereological fallacy is at the core of the logic of Cartesian dualism. This claim, in turn, relies on the assumption that Descartes took a human being as a "composite entity" (p. 26) whose parts are a body and a mind. The authors then argue that Descartes committed the fallacy in that he ascribed psychological attributes to a soul as a *part* of a human being, when logically he should have ascribed them to the human being as a whole. The fallacy crept into cognitive neuroscience through the influence of Descartes on Sherrington and his disciples, who also ascribed psychological attributes to the mind. Substance dualism was eventually abandoned in favor of materialism, but parts of human beings, typically brains and brain hemispheres, came to replace the mind and remained as the bearers of psychological attributes.

Our concern here is whether the mereological fallacy *can* be committed within Cartesian dualism. Our rationale is as follows. In order to commit the fallacy, one must be able to assert intelligibly that a soul is a part of a human. However, it is not clear that this condition is met in Cartesian dualism. Therefore, it is equally unclear that the fallacy can be committed in Cartesian dualism.

Consider spatial parts and temporal parts as the primary candidates for parthood. Evidently, a Cartesian soul *qua* spatially unextended substance cannot be a spatial part of anything. Can it be a temporal part? This question is more complicated but equally answerable in the negative. A temporal part is a phase, that is, something that exists only in a certain moment in time. Parts of events are the prototypical examples of temporal parts. A baseball game has innings, an opera has acts, a symphony has movements, and so on. Events thus exist incompletely in any moment in time during their occurrence. A baseball game in its inning four exists incompletely during that inning, for some of its parts (innings one, two, and three) lie in the past and others (innings five, six, etc.) lie in the future.

A basic intuition about parts is that they are smaller than the wholes of which they are parts. An arm is smaller than a body, an atom smaller than a molecule. This intuition is behind Common Notion 5 of Euclidean geometry: The whole is greater than the parts. It is also honored in the first disjunct of a mereological principle: x is a part of y if and only if x is a proper part of y or x is equal to y (Simons, 1987, p. 26). The proper-parthood relation in this principle is analogous to (in fact, based on) the arithmetic relation of less-than and thus denotes a strict partial ordering (i.e., irreflexivity, asymmetry, and transitivity). The second disjunct of the principle is less intuitive, so we shall ignore it under the assumption that the authors refer to proper parts when they speak of parts.

The intuition applies equally to temporal parthood. A temporal part of an event is a shorter event. An inning is shorter than a baseball game; an act is shorter than an opera. Obviously, a soul cannot be intelligibly said to be shorter than a human in Cartesian dualism, for in this doctrine souls are immortal, bodies are mortal, and humans are interactions between souls and bodies. Souls, then, cannot be intelligibly said to be proper temporal parts of humans. Nor can a soul be intelligibly said to be an improper part of a human either, because according to the

above mereological principle, improper parthood amounts to identity, and in Cartesian dualism a human is not identical to a soul.

Descartes' talk of a soul as a part of a human, then, is incoherent with his own doctrine. If the Cartesian dualist insists in asserting that a soul is a part of a human, or that a human is a composite of a soul and a body, he would have to redefine the notion of parthood. There is no textual evidence that Descartes or any other Cartesian dualist has attempted to do this, and it is most unclear whether it can be done without violating our most basic intuitions about parthood—but perhaps the Cartesian dualist need not go down this slippery slope. Cartesian dualism would not seem to depend critically on mereological talk of souls being parts of humans or humans being composites (or combinations, fusions, mixtures, or unions) of souls and bodies. Such talk would seem to be entirely dispensable in Cartesian dualism. The core of this doctrine is that substances are sharply divided into material and immaterial, and the latter interact with some of the former for some time. Talk of parts and wholes is nowhere to be found in this core.

But then, how could a human be conceived in Cartesian dualism? Not as a composite or a union of, but an *interaction* between, an immaterial substance and a material substance. This answer neither presupposes nor entails any mereological relation. Descartes' mereological talk of souls as parts of humans and humans as composites or unions of souls and bodies can thus be safely dismissed as careless and ontologically inconsequential. Such talk represents no significant aspect of Cartesian dualism.

These considerations, of course, do not solve all the problems with Cartesian dualism. In particular, the problem of how an immaterial substance can interact causally with a material substance remains—but we are not trying to rescue Cartesian dualism here. We are only arguing that it cannot be the logical root of the mereological fallacy in cognitive neuroscience, for within Cartesian dualism a soul cannot be intelligibly asserted to be a part of a human being. Nor does the core of Cartesian dualism require such an assertion. Hence, not only can the fallacy not be committed but also the mereological talk that the authors take as evidence for the fallacy is entirely dispensable in Cartesian dualism.

To be sure, the fallacy can be committed in current cognitive neuroscience, for a brain and a brain hemisphere can intelligibly be asserted to be parts of a human being. However, contrary to the authors' claim, Cartesian dualism cannot be blamed for the fallacy any more than Aristotelian hylomorphism. At most, the fallacy can be attributed to Sherrington's unreflective acceptance of Descartes' careless writing about souls being parts of humans. This acceptance, in turn, was uncritically carried over by later cognitive neuroscientists after abandoning substance dualism and replacing the soul with the brain. But the authors are not logically entitled to aver that current cognitive neuroscience "*propounds a form of crypto-Cartesianism*" (p. 111), is "*like Cartesianism*" (pp. 111, 112), or "*retain[s] the logical structure of Cartesian psychology*" (p. 113). The fallacy in current cognitive neuroscience is no more Cartesian than it is Aristotelian.

These considerations do not leave current cognitive neuroscience in a better position. On the contrary, the fallacy becomes a result of *unconditional intellectual*

consent, which is worse than misunderstanding or ignorance. Such consent is mystifying, for it goes against the critical spirit that supposedly characterizes science and philosophy. The mystification, however, results from the myth of the faultless genius, a myth that confuses brightness with perfection. A moral of the book in this respect is that even geniuses can and do make blatant mistakes. Winning the Nobel Prize, the archetype of the scientific genius, does not mean that the winner is right in everything he or she says, writes, or does. The Nobel Prize is recognition for an outstanding achievement in science, not a certificate of intellectual flawlessness.

In its most perverse form, the myth takes being an outstanding scientist as *sufficient* for being an equally outstanding philosopher of the science practiced. To be sure, a deep understanding of a science is *necessary* for philosophizing properly about it, but it is not sufficient. The book is a forceful example of the misery of this form of the myth. Crick, Edelman, and Kandel, the Nobel Prize recipients repeatedly named throughout the book, are undoubtedly outstanding scientists. However, this does not necessarily make them outstanding philosophers of the science they practice, nor does it immunize them against blatant logical mistakes.

Aristotelian Essentialism versus Wittgensteinian Anti-Essentialism

Our second concern arises from an odd tension between the authors' commitment to Wittgenstein's ordinary language philosophy (as expounded in his *Philosophical Investigations* [1953]) and their essentialistic proclivities. The latter are apparent in their endorsement of Aristotelian hylomorphism. Central to it is the distinction between form and matter, which is made in terms of the distinction between essential and accidental properties (see p. 13). The two distinctions are the basis for the view that the soul "consists of the *essential*, defining functions of a living thing with organs" (p. 14, our italics). Thus it is difficult not to read as essentialistic the assertions that "[p]sychological predicates are predicates that apply essentially to the whole living animal, not to its parts" (p. 72) and "[the ability to act for reasons and be aware of them] is essentially dependent upon language" (p. 314).

However, Aristotle conceived of essences (*to ti ên einai*, "the what it was to be," or *to ti esti*, "the what it is") as *universals* (*katholou*), which are supposed to be invariable across, or common to, multiple exemplifications.[1] Essences thus imply strict commonality whose linguistic expression requires terms to have univocal, fixed meanings. Certain languages, such as formal logic, allow for such expression. Rule-governed uses of ordinary psychological terms, in contrast, are too changeable to admit such meanings. Wittgenstein's talk of "language games" and "family resemblances" instead of "definition" was his way of conveying the futility of searching for essences in ordinary language. Although he wrote

[1] Admittedly, there is controversy over whether Aristotle conceived of essences as universals, so our present concern is grounded only on the *possibility* that he did.

"Essence is expressed by grammar" (*Philosophical Investigations*, §371), he also wrote:

> Consider for example the proceedings that we call 'games'. I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all?—Don't say: "There must be something common, or they would not be called 'games'"—but look and see whether there is anything common to all.—For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. (§66)

This depiction starkly contrasts with the authors' talk of conceptual commonality and sameness in ordinary language:

> Concepts are abstractions from the use of words. The concept of a cat is what is *common* to the use of 'cat', 'chat', 'Katze', etc. (p. 65, our italics)

> The words 'cat', 'chat', and 'Katze' are symbols in three different languages, all of which express one and the *same* concept. (p. 345, our italics)

This talk is puzzling in view of Wittgenstein's notions of language games and family resemblance, which, again, were motivated by his conviction that meaning in ordinary language, as given by rule-following everyday-life usage, is too changeable to admit an account in terms of commonality and sameness. The authors also make reference to the logicians' sense of "qualitative identity" (p. 96n), apparently without disowning it. This reference too is puzzling when compared to Wittgenstein's turn from formal logic in the *Tractatus* (1961) to ordinary language in the *Philosophical Investigations*. We are not claiming that Wittgenstein rejected the existence of essences or even universals. The issue is *linguistic*, not ontological. Universals may well exist (although demonstrating their existence is no trivial matter); however, for better or worse, *expressing them linguistically* requires a highly formalized language that extremely simplifies and, to this extent, departs considerably from ordinary language, the focus of the authors' analysis.

Explanatory Reductionism

Thirdly, we were unimpressed by the authors' rejection of explanatory reductionism (pp. 355-366). This rejection is largely independent of the authors' diagnosis of conceptual confusion in cognitive neuroscience. One can thus coherently agree with the diagnosis and still disagree with the rejection. Avoiding the mereological fallacy does *not* commit oneself to explanatory anti-reductionism, nor does explanatory anti-reductionism entail the fallacy.

The rejection in question has two aspects. On the one hand, the authors claim that human action *cannot* be explained in terms of neural laws because there are no psychological laws of human action:

> . . .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 neurological laws might be discovered. For, as far as explaining human action is concerned, it is clear enough that although there are many different kinds of explanation of why people act as they do, or why a certain person acted as they did, they are not *nomological* explanations (i.e., they are not explanations that refer to a natural *law* of human behaviour).
>
> There are, to be sure, explanations of a person's action that explain it by identifying it as an instance of a general pattern. So, we may explain why A does *V* by reference to the fact that it is a habit, or that A has a tendency to *V* in such moments as these, or that it is a custom in A's community to *V* in such circumstances and A is a conventional sort of person, or that A is in such-and-such a predicament and people with A's kind of personality traits tend to *V* in such circumstances. But these explanations do not specify anything that could possibly be deemed as strict *law*; nor do they explain the behaviour by deducing it from a law and a set of initial conditions. Instead, they identify it as an instance of one or another kind of rough regularity of the person's behaviour, which may admit of many exceptions.[2]

It is unclear exactly what the authors mean when they deny that "there is anything that can be dignified by the name of *psychological laws*" (p. 362). Two possibilities present themselves: psychological laws do not exist, or they do but remain to be discovered. The former is a strong ontological tenet that requires far more explication than is found in the book. The latter refers to a temporary historical condition that may or may not be obtained in the future. Additionally, the authors restrict their argument to *deterministic* ("strict") laws when philosophers of science largely admit *probabilistic* or *statistical* laws. If psychological laws are statistical (not too big an "if"), they admit exceptions. Hence, any of the alternative explanations the authors mention are good candidates for probabilistic laws.[3]

But no matter—for, on the other hand, even if reasoned human action were explanatorily reducible to neural laws, the resulting explanations would be inferior to those that appeal to the behaving person's *reasons*:

> We call on Jack only to find him out. We ask where he is, and are told he has gone to town. We want to know why, and are told that it is his wife's birthday,

[2] The authors refer here to the logico-positivistic, nomologico-deductive, covering-law model of theoretical explanation. In this model, a theoretical explanation is a deductive argument whose premises are one or more neural laws, and whose conclusion is a psychological law. As is well known, bridge principles or correspondence rules are required for this model to work. The authors also deny the existence of such principles, but as an argument against *ontological* reductionism. The denial, however, also applies to explanatory reductionism under the covering-law model. In any case, the denial is conspicuously similar to Davidson's (1970) anomalous monism.

[3] Additionally, probabilistic laws force us to abandon the covering-law model (see footnote 2). The authors' rejection of explanatory reductionism thus becomes inapplicable, insofar as it is restricted to that model.

> that he booked tickets for *Tosca* weeks ago, and that he has taken her to her favourite opera. Would a neuroscientific story *deepen* our understanding of the situation and events? In what way does it need deepening? Does anything remain puzzling once the mundane explanation has been given? (p. 364)

We answer the first and third questions with a resounding "yes." The authors' negative answer arbitrarily stops the explanation of Jack's behavior at social practices and conventions. But surely many (us included) are further puzzled by the practices and conventions *themselves*. The second question can thus be answered in terms of explanations of how those practices and conventions are acquired and maintained, how they are instantiated in specific individuals, what the origins of the similarities and differences observed among them are, and so on. Answers to these questions will certainly deepen our understanding of reasoned action, and cognitive neuroscience (*sans* the mereological fallacy) may have much to contribute to them.

Of course, much depends on what the authors mean by "deepen" and "understanding." Here the authors' Wittgensteinian approach can be turned toward them. One ordinary use of "deepen" is "to extend well inward from an outer surface" (Merriam-Webster Online Dictionary). By referring to certain inward anatomical macro- and microstructures and their functioning (brain hemispheres, areas, nuclei, neurons, etc.), whatever explanations cognitive neuroscience might provide will deepen our understanding insofar as they *expand* our knowledge beyond publicly observable behavior and into cerebral processes. So, the sense in which cognitive neuroscience will deepen our understanding of human behavior is perfectly consistent with that usage of "deepen."

What about "understanding"? Ordinary uses include the following: "to grasp the meaning of, to grasp the reasonableness of, to have thorough or technical acquaintance with or expertness in the practice of, to be thoroughly familiar with the character and propensities of, to accept as a fact or truth or regard as plausible without utter certainty, to interpret in one of a number of possible ways, to supply in thought as though expressed, to have the power of comprehension, to achieve a grasp of the nature, significance, or explanation of something**,** to believe or infer something to be the case, to show a sympathetic or tolerant attitude toward something" (Merriam-Webster Online Dictionary). The assumption that cognitive neuroscience will deepen our understanding of human behavior is compatible with most, if not all, of these uses of "understanding."

The authors further argue:

> It is perfectly intelligible that our knowledge of the gross observable reactions of water with various chemicals should be deepened by an understanding of the atomic and subatomic constitution of water (and other chemicals)—which will explain things that we can observe, but do not understand, about the behaviour of water. But 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? (p. 364)

Again, much depends on what the authors mean by "clearer" here. Ordinarily, to render clearer is to free from opaqueness, ambiguity, or indistinctness, to make transparent or unclouded. Neuroscientific explanations render human behavior more transparent insofar as they refer to what lies beyond what is publicly observable in ordinary situations. They will also make it less ambiguous or indistinct, thanks to their technical character.

Moreover, there is a logical connection between knowledge and intelligibility. Something is intelligible very much in that we possess some knowledge about it. Reduction of our observations of the reactions of water to atomic and subatomic laws is intelligible in the sense that we know these laws (and those about such reactions, as well as the necessary bridge principles). Without this knowledge, such reduction would obviously be unintelligible, for inconceivable. A similar consideration applies to the authors' argued unintelligibility of the assumption that human behavior will be rendered clearer by neuroscience. This unintelligibility may well be due to our present ignorance about human behavior and its neural substrates.

Granted, using water effectively in everyday life (for quenching one's thirst, bathing, boiling food, making ice, dissolving, etc.) does not *require* knowledge of quantum mechanics—our intuitive knowledge of the reactions of water suffices for that—but it would be grotesque to regard quantum-mechanical explanations of the reactions of water as somehow inferior to intuitive ones just because the latter suffice for everyday life. There are non-ordinary (scientific and technological) uses of water that require quantum mechanics. They admittedly are far removed from everyday life and common sense, but this does not make quantum-mechanical explanations inferior to intuitive ones. At worst, they are inferior *relative* to ordinary uses, but who is to say that such uses are more legitimate or important than non-ordinary ones?

Similarly, the fact that we do not need cognitive-neuroscientific laws to deal with reasoned human action in everyday life does not imply that they are inferior to intuitive explanations. It surely seems implausible that they could be improved by cognitive neuroscience, but then again, as likely as not, this is because we do not know any better. When one deals with reasoned human action in everyday life, one can rightly ask: "Who needs cognitive neuroscience to understand this?" Indeed, we need not bother with cognitive neuroscience for *that*. However, this does not *logically* exclude the possibility of future circumstances where knowledge of such laws will be required. Nor is this possibility logically guaranteed, of course, and herein lies the predicament. Our present ignorance prevents us from making any certain predictions in this respect, *one way or the other*. It is not even certain that we will be able to discover the relevant psychological laws, or that they exist. *The only way to find out is to try*, with all the concomitant risks (waste of time, money, and energy), but this is the way science works. Science inevitably involves a great deal of trial and error (and success as well).

Concluding Remarks

What should radical behaviorists, a likely audience of this journal, make of the book? There surely are important agreements between radical behaviorism and the authors' later-Wittgensteinian stance. The ten similarities mentioned by Day (1969) are found throughout the book: anti-logical positivism, anti-reductionism, anti-Cartesianism, anti-mentalism, descriptivism, private events as significant, impossibility of a purely private language, the behavioral nature of language, opposition to reference theories of language, and meaning as usage. These similarities, however, need to be complemented with two potential differences.

First, in contrast to radical behaviorists, the authors seem to admit neural causation of overt behavior:

> The term 'representation' here signifies merely causal connectedness. That is innocuous enough [referring to a quotation by Blakemore in which he refers to the relation of "the activity of the nerves to events in the outside world or in the animal's body"]. (p. 79)
>
> The correlation between [cells'] firing and features is. . .a causal one. (p. 80)
>
> Parts of the brain. . .are causally implicated in cognition, recognition and action. (p. 142)
>
> The capacity to remember various kinds of things is causally dependent on different brain areas and on synaptic modifications in these areas. (p. 159)
>
> Neural groups. . .are causally implicated in the exercise of the relevant capacities. (p. 393)

Second, there is the authors' sharp disanalogy between philosophy and science. In particular, they claim that philosophy (specifically, the philosophy of ordinary language, the kind of philosophy they practice) is independent of cognitive neuroscience, which "operates across the boundaries between . . .neurophysiology and psychology" (p. 2). This claim appeals to the a priori, non-empirical character of the philosophy of ordinary language versus the empirical character of cognitive neuroscience. On this basis, the authors claim that philosophy does not, cannot, and *should not* suggest new experimental research in cognitive neuroscience, nor the latter solve philosophical problems. Thinking otherwise reveals a deep misunderstanding of the nature of philosophy and science. Of what value, then, is philosophy to science? Succinctly put, the authors' answer is this: "What philosophy can contribute to science is conceptual clarification." By that they mean "[Pointing] out when the bounds of sense are transgressed" (p. 405). More elaborately:

> The suggestion that epistemology should be grounded in neuroscience can be proposed only by someone with an infirm grasp of what epistemology is. It is, after all, not an empirical enquiry into how, as a matter of fact, human beings can and do acquire whatever knowledge they have—that is learning theory,

> which is a branch of psychology. Rather, epistemology is an a priori enquiry into the web of epistemic concepts that is formed by the connections, compatibilities and incompatibilities between the concept of knowledge, belief, conviction, suspicion, supposition, conjecture, doubt, certainty, memory, evidence and self-evidence, truth and falsehood, probability, reasons and reasoning, etc. The relevant connections are *logical* or *conceptual*—and neuroscientific investigations can shed no light upon the *normative* connections of logic (construing 'logic' broadly). Epistemology is also concerned with the logical character of *justifications* of knowledge claims, of confirmation and disconfirmation, of the differences between deductive and inductive support, of what counts as evident and what stands in need of evidence, and so forth. This too is not an empirical investigation. It could not possibly be furthered by the discovery of facts about the brain. (p. 406)

Presumably, the same goes for the discovery of facts about behavior. The radical behaviorist's disagreement with all this seems clear. Apriorism smacks too much of rationalism, innatism, and (more modernly) nativism, all of which diminish the role of experience in the acquisition of knowledge (especially language). As is well known, Skinner's (1957) operant-conditioning interpretation of language acquisition strongly opposes the psychological nativism of Chomsky (e.g., 1957), Fodor (e.g., 1975), and Pinker (e.g., 1994).

However, it should be clarified that the authors' apriorism does not commit them to nativism. So, this potential difference may well be only apparent. In the authors' view, philosophy is a priori in that its assertions, statements, or propositions are *conceptually true* (pp. 3, 318). But what is a conceptual truth? To answer this question another distinction must be taken into account. The a priori-a posteriori distinction is only half the story of the philosophy–science nexus. The other half is the *analytic-synthetic* distinction, a veritable philosophical can of worms that the authors open and close very quickly (p. 438n). This distinction is important because in philosophy "conceptual truth" is a standard alternative expression for "analytic truth." In this use, analytic truth is a matter of *linguistic convention*. A statement or proposition is analytic if and only if it is true *entirely* in virtue of the meanings of the terms that constitute it (e.g., "all bachelors are unmarried men," "all vixens are female foxes," "all horses are animals," etc.). A synthetic statement, in contrast, is one whose truth is determined only partly by the meanings of their constituting terms, and partly by the way the world is.

To embrace the a priori leads to nativism, rationalism, and innatism only if linked to synthetic knowledge (as, for instance, Kant did). Talk of analytic a priori knowledge is perfectly acceptable even to the staunchest empiricist (neither the British empiricists nor the logical positivists had any problem whatsoever with it). Behavior scientists who take experience as the main or only source of knowledge may thus embrace the a priori without incurring any intellectual liability. Quite the contrary, they could benefit greatly from the kind of analysis presented in the book, insofar as conceptual clarity is as essential to science as experimental rigor. It is unclear whether or not radical behaviorists would be convinced by this line of argument. They might as well agree with Quine (another philosopher often

considered to have Skinnerian proclivities) in rejecting the analytic–synthetic distinction altogether. In this case, a real disagreement with the authors would ensue.

Finally, this is a *very* tendentious book in its general outlook of philosophy. Its title is thus too encompassing and, because of that, deceiving. A more precise title would have been *Conceptual Foundations of Cognitive Neuroscience: A Wittgensteinian Approach*, or something to that effect. But of course, such a title would not have been attractive to many philosophers. We ignore whether the authors' title expresses a reductionism of philosophy to Wittgensteinian philosophy of ordinary language. If it does, we advise the philosophically naive reader not to uncritically accept it at face value. For better or worse, philosophy is much richer than Wittgensteinian philosophy of ordinary language. The authors' analysis represents only a fraction of the possible benefits of philosophy to science.

References

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Malden, MA: Blackwell.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Davidson, D. (1970). Mental events. In: L. Foster & J. W. Swanson (Eds), *Experience and theory* (pp. 79-101). Amherst, MA: University of Massachusetts Press. Reprinted in D. M Rosenthal (Ed), *The nature of mind* (pp. 247-256). New York: Oxford University Press.

Day, W. (1969). On certain similarities between the *Philosophical Investigations* of Ludwig Wittgenstein and the operationism of Skinner. *Journal of the Experimental Analysis of Behavior*, *12*, 489-506.

Fodor, J. (1975). *The Language of thought*. Cambridge, MA: Harvard University Press.

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: HarperCollins.

Simons, P. (1987). *Parts: A study in ontology*. New York: Oxford University Press.

Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in* the brain. Cambridge, MA: MIT Press.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

Wittgenstein, L. (1961). *Tractatus logico-philosophicus*. New York: Routledge. (Original work published 1921).