TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 3 (2011) 399–424

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2011.01142.x

# Explaining How the Mind Works: On the Relation Between Cognitive Science and Philosophy

Jonathan Trigg,[a] Michael Kalish[b]

[a]*Program in Philosophy, University of Louisianna*
[b]*Institute of Cognitive Science, University of Louisianna*

Received 2 November 2009; received in revised form 25 October 2010; accepted 12 January 2011

**Abstract**

In this paper, we argue that under certain prevalent interpretations of the nature and aims of cognitive science, theories of cognition generate a forced choice between a conception of cognition which depends on the possibility of a private language, and a conception of cognition which depends on mereological confusions. We argue, further, that this should not pose a fundamental problem for cognitive scientists since a plausible interpretation of the nature and aims of cognitive science is available that does not generate this forced choice. The crucial difference between these interpretations is that on the one hand the aim of theories of cognition is to tell us what thinking (etc.) is, and on the other it is to tell us what is causally necessary if an intelligent creature is to be able to think. Our argument draws heavily on a Wittgensteinian conception of philosophy in which no philosophical theory can explain what thinking, perceiving, remembering, etc. are, either. The positive, strictly therapeutic, purpose of a philosophy *of* cognitive science should be to show that, since the traditional problems which constitute the philosophy of mind are chimerical, there is nothing for philosophical theorizing *in* cognitive science to achieve.

*Keywords:* Wittgenstein; Hacker; Cognition; Mental representation; Thinking; Personhood; Mereological fallacy; Private language

> To say that the soul [the *psuche*] is angry is as if one were to say that the soul weaves or builds...
>
> —Aristotle, De Anima 408[b]12–15 (350 BC)

> Without very specific neural activities one could not think…but it is I who think, not my brain.
>
> —Bennett and Hacker (2003, p. 180)

Correspondence should be sent to Jonathan Trigg, Program in Philosophy, University of Louisiana at Lafayette, LA 70504. E-mail: jon.trigg@louisiana.edu

## 1. Introduction

This paper is an exercise in conceptual geography. The aim is, as it were, to identify the territory that a cognitive science can legitimately claim as its own, rather than actually to explore any of it. According to the map we will recommend, the true homeland of cognitive science is strategically placed and replete with all manner of natural advantages, but it does not take in so much of the globe as it is often supposed to.

Both the style and the content of the project we begin here is influenced by the extraordinary cartographical achievements of Peter Hacker and Max Bennett. At one crucial point in this essay, we borrow from their findings in *Philosophical Foundations of Neuroscience* (*PFN*), and throughout we employ their basic techniques.[1] Our case for relocating cognitive science and reining in some of its expansionist tendencies is, like their recent case against the more aggressive territorial ambitions of neuroscience, essentially Wittgensteinian. While in the course of our discussion we rehearse some of the arguments they deploy so forcefully in their treatment of neuroscience, we do so mainly to remind rather than to inform. Readers who are not already persuaded that the serious attribution of psychological predicates to the brain fails to advance our understanding and entangles us in basic logical difficulties,[2] or who doubt that language is essentially public, will find no new arguments for these claims here. What is new in this paper is not the geopolitical principles we bring to bear, but the part of the globe we want to map. Our interest is in the logical propriety and explanatory value of claims prominently featuring such terms as ''cognition,'' ''computation,'' ''information-bearing state,'' and ''mental representation'' rather than terms like ''neuron,'' ''afferent fiber,'' ''neocerebellum,'' etc. Such claims do not obviously involve mereological confusions and are not addressed directly by Bennett and Hacker in *PFN*.

The most obvious difficulty in the way of an enterprise like this one is that there is no particular anxiety either inside or outside the cognitive science community about the legitimacy of its current territorial claims. Far from it. Outside the academy the idea that we are now learning how to answer age-old questions about ourselves through applications of the experimental method is all the rage, and inside it, celebrity has done nothing to encourage critical self-examination, and a great deal to encourage complacency and a certain imperialistic expansionism (e.g., Thagard, 2010).[3]

On our view, contemporary cognitive science is expansionist not primarily because it involves an aggressive attitude toward other academic states (though, especially in the case of philosophy, that is perfectly true), but because it involves a dismissive attitude toward the conception of ourselves that is manifest in everyday discourse.[4] This aggressiveness takes the form of a commitment to the perhaps innocent looking idea that in trying to explain ''how the mind works'' (as the saying goes) we are trying to explain what the mind *is*. This involves an aggressive attitude toward our everyday conception of ourselves because it depends on a failure to appreciate that this everyday conception already constitutes the framework for wide-ranging and detailed knowledge of what the mind is and what our various cognitive, cogitative, and conative capacities are.[5]

It is fundamental to the argument we present in this paper that we reject, just as emphatically, the idea that some other discipline—like philosophy—is in a position to provide us with this sort of knowledge. Philosophy too has a history characterized by exactly the same imperialist expansionism we attack here. For traditionally construed, the point of a philosophical theory of mind is precisely to tell us what thinking is and what minds are, and in pursuing this project philosophers have adopted (and still do adopt) a similarly dismissive or neglectful attitude toward the radically nontheoretical conception of ourselves that structures everyday discourse.

The obvious objection to this simple point is that philosophers and cognitive scientists do not reject or ignore ordinary conceptions of ourselves in producing their accounts of mind, thought, and will; they just go beyond them. The point about our ordinary self-conception is that it is primitive and vague; and the point about philosophers and scientists is that they want sophisticated and precise conceptions of what we are. But there can be no question that in taking their theories to be theories about what we and our various capacities *are*, cognitive scientists, like traditional philosophers, are not extending or improving our ordinary (prereflective) understanding of ourselves, but rejecting it wholesale.

Consider the following line of thought. If persons are essentially things that possess the capacity to think, and if thinking *is* information processing, and if brains are things that process information, then brains have the capacity to think, and persons might be brains.[6]

Now it should be easy to see that the view that persons might be brains is quite incompatible with our ordinary notion of personhood and so could not be a mere modification or extension of that concept. It is incompatible with that notion for the simple reason that the notion of a brain is not the notion of something essentially capable of certain sorts of ethically and epistemically significant interaction with other things of its type, but the notion of a person is. It is an essential property of persons as we ordinarily conceive them that they have *both* the capacity to think *and* the capacity to interact with other persons in ethically and epistemically significant ways, for example, by making statements and promises, or by getting married and breaking the law.[7]

The question, then, comes down to this: Might there be persons in principle incapable of morally and epistemically significant action? Our ordinary conception of ourselves (on which, roughly, we are a certain sort of living, intelligent animal) rules out this possibility because it is the concept of something essentially capable of interacting with the world and with other things of its type in ways that raise questions about truth, rationality, and moral value. A notion of personhood on which I might be a brain does not rule out this possibility, because the notion of a brain is the notion of a certain sort of bodily organ, and the things which bodily organs do, or that go on in them, even if we rightly construe some of them as consisting in the processing of information, are things that need not (perhaps cannot) involve the sorts of interaction with the world and with other brains that raise questions about truth, rationality, and moral value.

Now even though it is both true and important that our ordinary conception of ourselves is flatly incompatible with conceptions of what we are that seem to many to emerge from theories of information processing (or ''mental representation'' or ''cognition''), our strategy here is not to stress this point and urge that, since the ordinary conception does not rule

out moral and epistemic responsibility, and the technical conception does, we should stick with the friendly old concepts that we already have. Rather, we argue here that interpretations of the nature and aims of cognitive science on which its theories of cognition could require revisionary conceptions of personhood that are quite incompatible with the ordinary conception of the same, are confused, and that there is an alternative interpretation that is neither confused nor in any conflict with this ordinary conception. If that is right, then it clearly matters what this alternative interpretation is and how it changes our understanding of cognitive science. That our argument does not directly contribute to cognitive scientific research does not impugn its significance—not even for those whose primary interest is in cognitive scientific research.

### *1.1. Quine and Wittgenstein*

The key to overcoming essentially imperialist conceptions of cognitive science and to establishing a distinctive and secure position for the subject in an intellectual republic is the repudiation of the dominant Quinean conception of the relation between science and philosophy and the adoption of the almost entirely neglected Wittgensteinian conception of the same. Perhaps the most important difference between Quine (see, e.g., Quine, 1951) and Wittgenstein's (see, e.g., Wittgenstein, 1958) account of this relation is that Wittgenstein was much more negative about philosophy than was Quine—this, of course, is the very opposite of what is commonly supposed to be the case (see Hacker, 1996).

Where Quine thinks of philosophy as continuous with science, in particular with cognitive psychology, and allows for the possibility of a kind of science that is mixed up somehow with a kind of philosophy, Wittgenstein insists that there can be no such thing as a scientific philosophy or a philosophical science. This is not because he thinks that there is something wrong with science but because he thinks that, except in the most attenuated sense, there is no such thing as philosophy.

For Wittgenstein traditional philosophy is nothing but a series of interconnected conceptual confusions generated by a failure to appreciate the structural subtleties of ordinary language use; it is a disease of the intellect which can perhaps be managed, but for which there is no outright cure. As the disease runs its course grammatical platitudes (''What I say about my own experiences does not typically rest on observation of what I do, but what I say about other peoples' experiences does rest on observation of what they do.'') are transmuted into metaphysical theses (''I have immediate access to the contents of my own mind but I have only indirect access to the contents of the minds of others.''), and metaphysical theses give rise to apparently deep but chimerical problems (''I know I have a mind but how do I know that others have minds?'' or ''I know what I mean by what I say, but how does anyone else know what I mean by what I say?'').

Wittgenstein's later work has a strictly therapeutic purpose. It consists in a series of carefully chosen reminders of ordinary ways of talking and thinking which, by bringing the language in which they are formulated back from a spurious metaphysical to a respectable everyday use, are meant to make the traditional problems of philosophy ''completely disappear.'' So philosophy as practiced by the later Wittgenstein is entirely parasitic on the

conceptual confusions it aims to identify. It takes the form of a kind of standing challenge to the traditional philosopher to formulate a so-called philosophical problem that does not vanish into air as the appropriate reminders of ordinary usage are skilfully assembled. Until the traditional philosopher manages to produce such a problem, no distinctively philosophical challenge to our ordinary prereflective conception of ourselves and of the world will have been set, and we will have failed to establish that there is something distinctive for philosophical theorizing to accomplish. So it is not that Wittgenstein appeals to ordinary ways of speaking to solve philosophical problems (as even Russell accused him of doing[8]); he appeals to ordinary ways of speaking to show that nothing deserving to be called a philosophical problem arises in the first place. The danger of calling Wittgenstein's therapeutic gestures in, for example, *Philosophical Investigations*, ''philosophy'' is that we will forget this, and accuse Wittgenstein of footling with mere words in response to grand problems about the deep structure of reality.

So Wittgenstein's view, unlike Quine's, excludes in principle the possibility that the problems of philosophy might finally be solved by a new philosophical science or a scientific philosophy. And this is not because the problems of philosophy are somehow too deep to be tackled by the application of an experimental method, but because the problems of philosophy are illusory. Following Wittgenstein, then, the reason that neither cognitive science (a philosophical science) nor ''neuro-philosophy'' (a scientific philosophy) can solve the traditional problems of philosophy is the very same reason that traditional philosophical theorizing cannot solve them, namely, that there is no such thing as a solution to a problem that does not exist.

### *1.2. Talk of ''the mind''*

For present purposes, the most important application of these methodological (or geopolitical) remarks has to do with the question ''What is the mind?'' On our view the apparent depth and intractability of this question should be handled not by the construction of theories designed to explain ''consciousness'' or ''the self'' in neurological or computational terms, but by ''grammatical'' reminders about the language in which the question is formulated. As Bennett and Hacker put it:

> Talk of the mind, one might say, is merely a convenient *facon de parler*, a way of speaking about certain human faculties and their exercise. Of course that does not mean that people do not have minds of their own, which would be true only if they were pathologically indecisive. Nor does it mean that people are mindless, which would be true only if they were stupid or thoughtless. For a creature to have a mind is for it to have a distinctive range of capacities of intellect and will, in particular the conceptual powers of a language-user that make self-awareness and self-reflection possible. (*PFN*, p. 105)[9]

When we organize our thinking and research around apparently profound questions like ''What is the mind?'' and ''How does the mind work?'' we all too easily get bogged down in mereological confusions. Uncritical talk about the mind leads to talk about the

various things that minds do or the various things that happen in them (thinking, perceiving, remembering, intending, etc.). This talk, in turn, modulates quite naturally into talk about what happens in brains and what brains do. Thus, ''Tom is thinking'' becomes ''A mental event of a certain kind is taking place,'' which becomes ''A neurological event of a certain kind is taking place.'' Such ways of speaking can appear to have explanatory value because they generate a discourse about perception, memory, understanding, sensation, and the rest, which involves no essential reference to persons who perceive, remember, understand, and feel. But here the linguistic gesture that generates the appearance of an explanation generates at the same time a series of apparently deep ''metaphysical'' problems about the relations between persons, minds, and brains on the one hand and between mental and neurological events on the other. It is these pseudo-problems that seem to call for a philosophical theory of the mind. On contemplating a series of mental states, events, and processes one seems suddenly to be confronted by a family of urgent questions: ''Is a person identical to their mind?'' ''Is the mind identical to the brain?'' ''Is a person identical with a brain?'' ''Are mental events identical with neurological events (or caused by them)?'' etc. The appropriate response to such questions is not earnest theorizing, but scepticism about the grammatical prestidigitations that produced them.

We propose that cognitive science is not going to tell us what the mind is, not because the mind is metaphysically too elusive an entity to give itself up to scientific inquiry, but because it is no kind of entity at all. The everyday locutions which include the noun ''mind'' can all readily be paraphrased into ones that include only words for psychological attributes predicable of human beings (*PFN*, p. 104). To say that someone has a mind is not like saying that he/she has a dog or a car, a head or a brain; it is like saying that he/she has eyesight or the ability to think for himself/herself. At the conceptual foundation of cognitive science is the notion, not of a mind but of a person, or an intelligent creature, together with the notions of the various natural capacities and powers, possession and exercise of which makes persons, or intelligent creatures, what they are. Our primary aim in this essay is to argue for this thesis.

The argument we present turns on the distinction between constitutive features and causally necessary enabling conditions; that is, between what for example, thinking *is*, and what is causally required if it is to be possible for a creature to think. After elucidating this distinction we argue that failure to mark and observe it impales the cognitive scientist on the horns of a dilemma. On one horn, whose spike is the private language argument, thinking (remembering, perceiving, imagining, and the rest) will be construed as a computational operation *persons* perform on mental representations; on the other, whose spike is the mereological fallacy, thinking (remembering, etc.) will be construed as a certain sort of computational operation *brains* perform on mental representations. The view we ultimately recommend—which preserves an important explanatory role for the idea of cognition (or computation or mental representation or information processing) that is at the heart of anything deserving to be called cognitive science—is that the occurrence of certain sorts of events and processes in the brain (events and processes typically characterized as ''processing'' or ''computation'' or ''mental representation'') is a causally necessary enabling condition of a person's exercise of his/her natural capacity for thinking and remembering.

On our view, such events and processes deserve to be characterized as *cognitive events* and processes because they causally enable human beings to do such cognitively significant things as think, remember, and perceive.

It is vital to be clear before we present this argument that it follows immediately from our conclusions about the conceptual location of cognitive science that it is not going to tell us what a person *is*, or what, for example, remembering, thinking, imagining, or willing *are*. In other words, it is not going to discover the *constitutive features* of persons and their various affective, cognitive, and cogitative capacities. This time, this is not because the notion of a person or an intelligent creature is not the notion of any kind of entity, or that the idea of thinking is not the idea of something that intelligent creatures can do, but, roughly, because you cannot acquire what you already possess. That we already have a conception of what a person is and what thinking, remembering, imagining, and perceiving are is manifest, not in our grasp of any theory which could compete with theories propounded in science or philosophy, but in our competence in using a shared language of mentality in the ordinary circumstances of life. That we are competent users of this language is a presupposition of the cognitive sciences (along with so much else that we do). For were we not already competent users of mental terms like ''belief,'' ''sensation,'' ''perception,'' ''thought,'' and ''desire'' we would not be able to say what cognitive science and cognitive neuroscience are meant to be about, and so could have no conception of what the point of these forms of inquiry is.

Of course, as has already been pointed out, it does not follow that it falls to philosophy to tell us what persons are and what, for example, thinking is. For the reason just given, neither Wittgenstein nor Hacker can tell us these things—and unlike many Quinean philosophers of cognitive science and neuroscience, they do not presume to. Their aim is only to remind us of the tacit knowledge we already have. What Wittgenstein's ''connective analyses'' are meant to do is show that no theorist has managed to formulate a problem about thought, will, perception, or the rest which establishes the need for a philosophical theory of thought, will, or perception. The aim is to show that no deep flaws in our ordinary concepts of these natural and familiar phenomena have been revealed; no deep mystery at the heart of things, which ingenious theorizing must dispel, has been uncovered. Crucially, they do not argue that our ordinary concepts of persons, of thought, memory, imagination, consciousness, and the rest are justified in some ultimate sense; they argue that no philosopher or scientist has shown them to be unjustified in some ultimate sense. Theorizing of various kinds has indeed generated the appearance of mystery and depth, of profound inadequacy in our everyday concepts—and it is a central task of Wittgensteinian therapy to reveal the grammatical mechanics of this illusion—but it is illusions and not metaphysical insights that are at issue. The therapeutic return to the everyday is required because it is only by drawing a veil over it that we have become convinced that we need what philosophical theorizing offers. The appeal to our ordinary conceptual competences is not itself a theory of mind or language but a way of reminding us that there is no living need which such a theory could satisfy.

Despite the fact that our primary aim is to insist that once we learn to position the cognitive sciences in the region that is their proper home they have a distinctive and critical role to play in world affairs, we fear that our recommendations will meet with considerable

hostility. Once imperialistic ambitions take hold they are hard to forego (as the history of philosophy amply demonstrates). The idea of replacing with technicalities the very terms in which our ordinary discourse about ourselves is conducted intoxicates the theorist by promising a dramatic vindication of her esoteric pursuits. Such intoxication is no friend to serious inquiry even if it does a great deal to glamorize it. If, for example, Dennett is right to say that the reason everyone is so fascinated by cognitive science is that ''it so manifestly promises or threatens to introduce alien substitutes for the everyday terms in which we conduct our moral lives'' (Dennett, 2009, p. 236), then our conclusions will be unpopular for the very reasons we take them to be important.

## 2. Technical and nontechnical uses of psychological terms

Here then are two theses, which we present to focus the discussion; the first (the Cognitive Science Thesis, CST) is about cognitive science and the second (The Cognition Thesis, CT) is about cognition. Together they are meant to capture widely held assumptions about the purpose and content of cognitive scientific theories. The paper will proceed as a discussion of CT, and we will return to CST at the end.

CST: The basic aim of cognitive science is to explain how the mind works.
CT: Cognition is a computational operation performed on mental representations.[10]

The first thing to say about CT is that its meaning is hard to determine, because ''cognition'' and ''representation'' each have both a technical and a nontechnical sense. Before we are clear about whether in critical formulations like this, neither, one, or both of these terms is being used in a technical sense, it will be quite impossible to evaluate them. A great deal of the conceptual lack of clarity which attaches to the foundations of cognitive science is attributable to the fact that the ordinary sense of ''cognition'' and ''representation'' is more obscure than the ordinary sense of terms like ''see'' and ''think.'' This makes it easy for theorists to forget that these crucial terms even have a nontechnical sense, and it allows room for confusion about when they are being used in new technical ways and when in old ordinary ones. The first thing to do is to say roughly what the nontechnical senses of these terms are so that the distance between technical and nontechnical uses of ''cognition'' and ''representation'' can come clearly into view. Once these distinctions have been made we will be in a position to ask what role a technical and novel notion of cognition and representation can legitimately play in cognitive science.

In a nontechnical sense, perceiving, remembering, believing, and judging are examples of cognition; imagining, idle thinking, wondering, and intending are not. In this sense, cognition is what happens when someone takes in how things are or becomes cognizant of some thing, event, or fact. Conversely, if something is cognized it becomes known somehow to someone. To say of someone that his cognitive faculties were impaired by drunkenness on some occasion would be to say that his capacity to attend to, take in, or get to know his surroundings was impaired by drunkenness on that occasion.

In a technical sense, cognition may be said to be a certain sort of computational or algorithmic process, taking place in, or performed by the brain, on certain structures (information bearing states or representations) in, or realized in, the brain.[11]

In a nontechnical sense, Paul may be said to represent a certain gun as having gone off in Peter's hand when he asserts, ''It just went off in his hand.'' In this sense, if Paul tells Peter how things are by making a particular statement, he may be said to represent the world to him as being a certain way; and when Paul makes a judgement, but does not express this judgement by making a statement, he may be said to represent the world to himself as being a certain way. In this sense of ''representation'' the judgements and statements people make (but not the questions they ask or the commands they issue) are representations. Paul's representing the world is something he does, an action he performs, which, as an action, may properly be characterized as justified or unjustified, as reasonable or unreasonable. On the other hand, the representations he produces in performing this sort of action are the sorts of thing that can be characterized as true or false.

In a technical sense of ''represent,'' a certain pattern of neurons or neuronal firings may be said to represent various states of the environment because they causally co-vary with those states of the environment. In this sense representing the world is something that might happen in the brain, or that the brain might do. These sorts of representational events or processes could not be justified or unjustified since a brain cannot have or lack a reason (cannot be reasonable or unreasonable in doing something), and representations conceived as things (states or events) in brains could not be true or false, since a trace of something—like a footprint or the rings in a tree—is not something true or false about what it is a trace of (see, e.g., Travis, 2004).

There is obviously a great deal more to be said both about the technical and the nontechnical uses of these two pivotal terms, but enough has been said already to allow us to raise the relevant questions about how CT should be interpreted. What is important is that we do not mistake technical for nontechnical usages and vice versa. It is repeatedly emphasized in *PFN* that technical usage of a given term can diverge from nontechnical usage without any conceptual harm being done. Equally, ordinary terms can be used in novel ways in technical domains without violating the bounds of sense. The conceptual difficulties arise when an ordinary term used in a technical domain is carelessly reinvested with its ordinary, nontechnical sense, or when an ordinary term used in a nontechnical domain is carelessly invested with a technical sense. Then, for example, inferences are drawn from a technical usage of ''map'' that would only be licensed by a nontechnical usage, or from a nontechnical use of ''knowledge'' that would only be licensed by a technical one. In a technical sense of ''map,'' for example, the brain may be said to contain maps, and to have various, perhaps computational, dealings with them; but neither persons nor animals can be said to learn anything about their environments by consulting, reading, or interpreting maps of the kind that can be contained in their brain.[12] Again, in a nontechnical sense persons may be said to know about various facts or events and to use this knowledge in negotiating their environment; but in this sense neither brains nor systems of neurons can be said to know anything, nor can knowledge of the sort people have be stored in brains.

In neuroscience, as Bennett and Hacker demonstrate, this criss-crossing of technical and nontechnical usage happens with startling regularity, and various sorts of non-sense, the mereological fallacy chief among them, result (*PFN*, pp. 74–77). However, it is not so clear that this sort of thing happens regularly in cognitive science, since it is not at all clear that ascribing to ''cognitive systems'' or to ''information processing devices'' (rather than, e.g., neurons) the capacity to represent or carry information (rather than, e.g., to perceive or remember) involves the confusion of technical with nontechnical senses of key terms.

The view we defend here is that as long as we carefully mark and observe on the one hand the distinction between talk about intelligent creatures and about the possession and exercise of their various capacities (for, e.g., thought, memory, learning, imagining, perceiving), and on the other, talk about brains and their parts and the various computational processes and events which go on in them, conceptual confusion can be avoided. If, however, we *identify* intelligent creatures with information processing devices or cognitive systems, and their thinking and perceiving with information processing or mental representation, we will mire ourselves quite hopelessly in the mud of conceptual illusion.

## 3. Psychological phenomena and their causally necessary enabling conditions

Consider a fire that consists in the burning down of a certain building. The presence of oxygen in the atmosphere is not a constituent of the fire, as the burning of certain timbers might be, and neither is it its cause (which might be a certain short circuit, for example). The fire could neither start nor could it continue without oxygen in the atmosphere, so we might say that the presence of oxygen is a causally necessary enabling condition of the fire.[13]

Now consider Paul's opening a door. For the purpose of argument, suppose that this action consists in his reaching out, taking hold of the doorknob, holding onto the doorknob, turning the doorknob, and pulling the door toward him. There is reason to say that there are not five actions here but one—Paul's opening the door—but nevertheless, in opening the door, he does all these things, or he opens the door by doing all these things. When he opens the door, all sorts of events happen in his arm and hand including various muscle contractions; but contracting his muscles (etc.) is not something Paul does—he does not, for example, turn the doorknob by contracting various muscles in his hand and wrist. Here, we could say that his reaching out, his taking hold of the doorknob, etc. are partially constitutive of his action of opening the door, and that the muscle contractions in his arm are causally necessary enabling conditions of this action. Note, again, that these muscle contractions do not cause the action Paul performs (nor any of the things he does in performing this action), but if they did not happen, Paul would be quite unable to open the door, at least by doing the things we have described him as doing in this case.

Now when Paul judges that Jocasta is Oedipus's mother we could say that Paul's conceptual capacity to have thoughts about Jocasta, about Oedipus, and about mothers are constituents of his making of this judgement. He exercises those conceptual capacities, we might say, in making that judgement, or he makes that judgement by exercising those capacities in a certain way (see e.g., Evans, 1982). What is it that the exercise of these capacities jointly

constitutes—that is, what are judgings? Judgings are a little like supposings, doubtings, surmisings, considerings, and entertainings; they are stances that people adopt toward how things are or might be: We say that someone is judging that *p* rather than, supposing that *p*, for example, when she is unreservedly committing herself to its being the case that *p*, rather than committing herself to its being the case that *p* for the sake of argument. Which sort of thing it is appropriate to do when is a matter of sometimes grave importance: It can be reasonable to consider whether Jocasta is Oedipus's mother when it is not reasonable to suppose, judge, or doubt that Jocasta is Oedipus's mother, for example.

When Paul makes a judgement, certain cognitive processes or events take place in his brain which themselves consist in algorithmic computational operations performed by the brain on certain representations.[14] If these events did not take place when he made the relevant judgement, he would not have been able to make it; but they are neither the cause of his making this judgement (what caused Paul's judgement was his reason for making it, namely, shall we say, that the oracle at Delphi told him that Jocasta is Oedipus's mother) nor are they, as the exercise of various conceptual capacities are, constitutive of his making that judgement. Rather, they are causally necessary enabling conditions of his making that judgement.

We propose that this distinction between what is constitutive of a given psychological episode, process, or state, and its causally necessary enabling conditions, is essential in cognitive science. Without it cognitive scientists will be confronted with a forced choice between equally unappealing readings of CT (''Cognition is a computational operation performed on mental representations''). On one such interpretation CT entails CT1; on the other, CT2 (see below). The main business of our argument here is to establish that both CT1 and CT2 are incoherent, and thus that to be forced to choose between them is to confront a dilemma. The only way of escaping this dilemma while holding onto a significant and explanatorily valuable notion of cognition is to appeal to the distinction we have just elucidated between constitutive features and enabling conditions. If that is right, we have found a powerful argument for the claim that to explain how the mind works is not to explain what, for example, thinking and perceiving are but to discover (or at least to frame and test hypotheses about) what their causally necessary enabling conditions are. It is important to appreciate that we do not claim that CT1 and CT2 represent clearly formulated positions adopted by competing theorists. Rather, they represent conceptions of cognition that, though incompatible, are all too frequently mixed together in the work of a single theorist.

CT1: For a person to think (remember, imagine, perceive, recognize, etc.) is for a person to perform certain computational operations on mental representations.

CT2: For a person to think is for a brain or mind (or mind-brain) to perform certain computational operations on mental representations.

Problems with CT1 turn on two conceptual issues: The distinction between nonconscious events and processes on the one hand and unconscious and conscious acts and activities on the other; and on the possibility of a private language. Problems with CT2 turn on the mereological fallacy and the identification of persons with brains. We will address CT1 and CT2 in turn.

## 4. For a person to think is not for a person to perform computations on mental representations

### *4.1. Conscious and unconscious activities versus nonconscious processes*

Computations on mental representations cannot be conceived as things someone does unconsciously, since they are conceived as things that could not be done consciously or intentionally. The idea that the ordinary and perfectly familiar business of learning or recognizing, for example, consists in someone's unconsciously performing computational or algorithmic operations on certain structures (whether formal or neurological) is vulnerable to the objection that if such operations could be performed unconsciously on some occasion they could be performed consciously on another.

This point bears further elaboration. Some of the things we do voluntarily we do not do intentionally, which is to say that some of the things we do voluntarily we do habitually (e.g., making various gestures in talking).[15] But whatever counts as being done unintentionally on one occasion must be the sort of thing that could count as being done intentionally (and even deliberately) on another. I may be said to grind my teeth unintentionally when eating, but I cannot be said to secrete salivary amylase unintentionally when eating. So when Paul opens the door, he will count as, for example, turning the doorknob even if he does so without being conscious of doing so, but he will not count as contracting the *flexor carpi radialis* in his forearm. That is because turning the doorknob is, and contracting his *flexor carpi radialis* is not, the sort of thing he could do consciously if his attention was focused differently.[16] So there are two things here of which he may, on occasion, be quite unaware—his turning the doorknob (an unconscious act) and the contraction of his *flexor carpi radialis* (a nonconscious process)—but only one can be characterized as something he does unconsciously. Cognitive science cannot coherently conceive of cognitive processes performed on mental representations both as nonconscious processes in minds or brains and as activities people engage in unconsciously. So such cognitive processes—computational operations performed on neurological structures—must either be conceived as activities that people can and sometimes do engage in consciously and intentionally (which is, to say the least, profoundly implausible—is it only mathematicians and computer scientists who are able to think, learn, and remember?) or as nonconscious processes happening in their brains which are causally necessary enabling conditions for the ordinary and familiar things people do. Cognitive scientists might produce hypotheses about or models of such nonconscious processes, and neuroscientists might discover their neurological realizations. This sort of view, which is both plausible and coherent, is wholly incompatible with the idea that for a person to think (remember, learn, etc.) is for him/her to perform computational or algorithmic operations on any sort of neurological structure or mental representation.

### *4.2. Mental representation and private language*

Because Wittgenstein's private language argument has to do with the role ostensive definition plays in assigning meanings to names, and in particular with the idea that a

kind of private pointing at mental samples plays a role in naming experiences and sensations, the connection between this argument and the notion of mental representation in contemporary cognitive science is far from clear. For one thing, CT1 identifies mental representations with thoughts, and, on any remotely plausible view, thoughts are not names. Consequently, the idea that a public or private pointing could play a part in thinking (or in having thoughts) seems not to arise. For another, Wittgenstein's argument is about *language* and the present proposal is about *thinking*. If, as may be thought obvious, thought underlies intelligent discourse, we may suppose that even if discourse turned out to be an essentially public phenomenon this would not show that thinking is. It is natural to insist that thinking, in sharp contrast to talking, is precisely private, mental, and inner, and furthermore, to suppose that it is not merely an incidental inner accompaniment to talk, but the very thing that makes talk more than just noise—the thing that gives words their meaning. This picture makes it appear that an adequate explanation of language must, in the end, take the form of an explanation of thought and thinking, and this in turn encourages the idea that a cognitive scientific theory of mental representation can at last provide this sort of account (after centuries of nonscientific fumbling around in the dark.) Of course this line of thought does not merely suggest that a putative science of thought or cognition should be unaffected by Wittgenstein's private language argument. The idea that inner mental processes not only accompany the production of perceptible signs but give meaning to them is profoundly at odds with Wittgenstein's argument. Indeed, Wittgenstein's later work, taken as a whole, identifies this conception of the relation between thought and language as one of the leadings symptoms of the philosophical disease.[17] The centerpiece of that work—the private language argument—attacks this conception head-on, and, for example, the analogy between languages and games provides a calming alternative picture. This picture suggests that to think about meaning is to think about the interlocking public practices in which a word figures, not the private, metaphysically mysterious (and phenomenologically elusive) acts of mind, which allegedly underpins this use.

Now the basic problem with this classical picture of the relation between thought and language, however its details are worked out, is this: If, when Paul says ''pigs swim,'' there are two types of operation he is performing (one mental, inner and private, one physical, outer and public) and two kinds of item he is operating on (one mental, inner and private, one physical, outer and public) we will have to explain what the relation is between these operations and these items. That is, we will have to explain how, on the one hand, the combining of mental representations stands to the combining of words; and on the other, how the combined mental representations stand to the combined words. Even without appeal to Wittgenstein's devastating argument we should be able to see that the prospects of providing such an explanation look dim. If we conceive mental representations as components of a nonverbal language, we will have to explain how nonverbal thoughts are to be translated into verbal utterances; if we conceive mental representations as mental images or pictures, we will have to explain what it is to translate images into words. But, of course, there can be no correct or incorrect translations of what is not verbal into what is. While it is possible to describe, for example, an image, images cannot be translated, and furthermore—a point that

many of the great modern philosophers failed to grasp—no image determines what counts as correctly describing it.[18]

If these weighty considerations do not convince, we can turn to Wittgenstein's celebrated (misunderstood and neglected) argument which shows quite categorically that when, for example, Paul says ''pigs swim'' he does not perform two types of operation, one on mental representations and one on verbal representations, but only one. It shows this by showing that operating on representations is an essentially normative or rule-governed affair (the type of thing that can be done incorrectly or correctly) and that operations performed on items, events, or states available only to the operator could not be normative or rule-governed. What follows is an argument to the conclusion that thinking cannot be an operation on mental representations that exploits Wittgenstein's argument that language cannot be private.

### *4.3. The mental representation argument*

1. To think—in the relevant sense—is not just to think of something but to think (suppose, doubt, surmise) that something is so. If there's thinking in this sense, there's something that is thought (supposed, doubted, surmised), and what is thought is the sort of thing that can be true or false.
2. To think (suppose, doubt, surmise) is to perform an operation of some kind on representations, such that the product of this operation is the sort of thing that can be true or false. In other words, thinking is an operation on representations that produces representations. The representations that are produced must be the sorts of things that can be true or false; the representations that are combined need not be.
3. Not just any combination of representations produces something that could be true or false—some ways of combining representations are correct and others incorrect. There are many ways of combining representations that produce nonsense, and what is nonsensical cannot raise the question of truth (e.g., ''Flipper house willingly green irascible'').
4. If to think is to perform a kind of operation (by 2) and if it is possible to perform this operation incorrectly (by 3), then the combination of representations must be a rule-governed activity. This is the first significant conclusion of the argument.
5. Rules for the combination of representations which can be recognized as representations, and as the particular representations they are, only by one person, would have to be private rules: There cannot be public rules prescribing how private items, events, or states should be manipulated.
6. Representations that are *mental* must be representations that can be recognized as representations and as the particular representations they are only by one person—they must be private items, events, or states.
7. There can be no private rules.
8. (by 5, 6 and 7) There can be no rules governing the combination of *mental* representations—so the combination of mental representations cannot be a rule-governed activity.
9. But (by 4), combining representations so as to produce representations capable of truth and falsity *must* be a rule-governed activity.

10. Therefore, thinking, which (by 2) consists in the performance of some sort of operation on representations, cannot be an operation performed on *mental* representations. The argument shows that the idea that thinking is a rule-governed mental process is incoherent.

The first four premises unpack the relevant idea of thinking by connecting it to the idea of a thought, and the idea of a thought to the idea of something that can be true or false. The proposal is that it is *thinking* (not something that causally enables thinking) that *consists* in operations the *thinker* performs on mental representations. On any acceptable construal of this proposal what these operations produce must be thoughts, which is to say, must be the sorts of things capable of truth or falsity. It follows immediately (4) that the operation in question must be the sort of operation that can be done incorrectly, since not just any combination of representations will yield a thought.[19]

Premise 5—an appeal to the private language argument—says that there can be no public criterion of correctness for an operation performed on private items, events, or states. If the rule is, ''Perform operation *p* when item *y* appears, event *x* occurs, or state *y* is actualized,'' and if *p*, *x,* and *y* are items, events, and states knowable only to the performer of *p*, then performances of *p* cannot be publically checked or assessed for correctness.[20]

Premise 6 says that *mental* representations must be items, events, or states knowable only to one person, and so private in the relevant sense. Written sentences, for example, could not be mental representations precisely because they are knowable as the representations they are by more than one person.

Premise 7 is the crucial appeal to the private language argument. Private rules are impossible, Wittgenstein reminds us (see Wittgenstein, 1958, especially sections 258, 265, 268), because there can be no difference between it seeming to someone at a given time that he/she is following a private rule, *p*, and his/her following *p* at that time. Say Paul's putative rule, *p*, is—whenever an item relevantly similar to *this* one (pointing inwardly to a relevant sample) comes before my mind, I will perform operation *r* on it (or whenever I am in this state—pointing inwardly to an appropriate state—perform operation *r* on its content). In either of these sorts of case, there could be no difference between its seeming to Paul at *t* that the relevant item was before his mind, or that he was in the relevant state, and that item really being before his mind, or his really being in that state. In that case, Paul cannot have invented a rule *p* governing his performance of operation *r*.

Premises 8 and 9 together with the conclusion follow straightforwardly from here. The conclusion states that thinking cannot be an operation on mental representations; it implies that, if it is an operation on representations of any kind, it must be an operation on public, which is to say, verbal or linguistic representations. So we cannot appeal to a wholly different kind of private inner representing in order to explain the familiar outer representing we all so obviously go in for. If to think is to represent, it must be to operate on the very representations that make up a public language, and so, in some sense, to participate in an essentially public practice. So CT1 is incoherent—requiring both that mental representations be private, because they can be known as what they are only to one person, and public, because combining them must be a rule-governed activity.

The only objection open to a defender of CT1 seems to be to reject Premise 6 and deny that mental representations are private in the sense of being knowable only to one person. Many cognitive scientists and philosophers of mind might even be eager to insist that since the mental representations in question are items in brains, states brains are in, or events and processes which happen in brains (or are realized in such items, events, processes, or states) they are precisely not knowable only to one person: Anyone could, in principle, take a look at a mental representation and, for example, see that and what Paul is thinking at a given time.

This objection fails because it forgets that the relevant proposal is not that for Paul to think is for certain operations on mental representations to occur in or be performed by a certain mind or brain. That proposal (CT2) faces different objections. The relevant proposal (CT1) is that for Paul to think is for *Paul* to operate on mental representations.

CT1 requires that there be an internal conceptual connection between what Paul takes himself to be thinking, or what it seems to Paul that he is thinking, at a given time, and what he is thinking at that time. If thinking consists in the performance of mental operations that produce mental representations (thoughts), then talking requires that the speaker express these mental representations by producing certain perceptible verbal signs, or that he translate them into such signs (e.g., sentences). This conception of the relation between thinking and talking allows room for the possibility that a speaker might mistakenly use one perceptible sign to express or translate his thought when he should have used another, but it emphatically does not leave room for the possibility that he might mistakenly take himself to be thinking one thing when he is really thinking another (or not thinking at all): On this view, there can be a gap between what I think I am saying and what I am saying, but not between what I think I am thinking and what I am thinking. Here is an argument for that claim.

If the possibility of being wrong about what I am saying is explained by appeal to the difference between what I am saying and what I am thinking on a given occasion, the possibility of my being wrong about what I am thinking must be excluded on pain of generating an infinite regress. For were it possible for me to be wrong about what I am thinking at a given time that possibility would have to be explained, too. This would require appeal to a difference between what I am really thinking at $t$, and what I mistakenly think I am thinking at $t$. Since what I think I am thinking can differ from what I am really thinking at $t$ I can be wrong about what I am thinking at $t$—that is, I can think I am having one thought when I am really having another. But now, if I can do that, it will have to be allowed that I can also be wrong about what I think I am thinking. Now this regress is vicious since it requires a thinker to have an infinite number of numerically distinct thoughts at a given time, the contents of which are all appropriately related, if he is to know what he is thinking at that time. For Paul to be right about or know what he thinks at $t$ it is not enough for him just to think it. He has also to think—correctly—that he thinks it. But for him to think correctly that he is thinking a particular thought, it is not enough for him just to think that he is thinking it. He will also have to think—correctly—that he thinks he is thinking it—and so on. Thus, absolutely no explanation has been given of how Paul could be wrong about what he is saying at $t$, since no explanation has been given of how he could be right about what he is thinking at $t$.

So, on CT1 it is inconceivable that Paul be operating on certain mental representations and thus count as thinking, but yet be ignorant of or wrong about which thought he is

having. On this conception, Paul's knowledge of what he is thinking must be both incorrigible and evident; incorrigible because if he thinks he is thinking that *p* he is thinking that *p*, and evident because if he thinks that *p* he thinks that he thinks that *p*. This thoroughly Cartesian commitment is built into the conception of thought as an essentially inner accompaniment to speech, and of speech as an outward expression of private mental processes, and it is the only way to avoid the relevant regress.

Now the point is that this requirement is straightforwardly incompatible with the claim that mental representations are knowable for what they are by more than one person. This claim requires not just that Paul's mental representations are, say, brain states, and that Paul's brain states are publically observable, it requires that it be possible to come to know that and what Paul is thinking at a given time by observing the brain states which are in fact constitutive of his thoughts. If that were possible, then a situation would be conceivable in which Paul thinks that *p* in virtue of his brain being in a particular state, George knows that Paul thinks that *p* by observing his brain, but Paul, ignorant of the state of his brain, does not know he is thinking that *p*. Any event, process, item, or state which any number of people can, in principle, know for what it is, must be the sort of event, process, item, or state which any particular person may happen not to know for what it is.

It follows, not just that a thinker might sometimes count as thinking, for example, that *p* despite wrongly taking themselves to be thinking that q, or not thinking at all, but also that a thinker might *always* be wrong about whether and what they were thinking (and a third party might always be right). These possibilities are flatly incompatible with the Cartesian notion of thinking and thought to which the defender of CT1 appeals as soon as they distinguish between linguistic representation and mental representation, and identify thinking with mental representing.

So CT1 cannot be defended by rejection of Premise 6. If mental representations are constitutive of thoughts, as on CT1, then they must be private; if mental representations are private, then CT1 is refuted by the mental representation argument.

These applications of Wittgensteinian principles refute the thesis that for a person to think is for him/her to perform operations on mental representations. They do not show that cognition (the occurrence of computational operations on representations or information bearing structures in brains) is not a causally necessary condition on a person's thinking. It seems profoundly plausible that if certain very complex events did not take place in a person's brain at *t* he/she would not be able to think at *t*. Nothing in the argument just presented is incompatible with this idea.

## 5. For a person to think is not for a mind or brain to perform computations on mental representations

CT2 raises just the same problem *PFN* claims is at the heart of the conceptual confusions in neuroscience, namely, the mereological fallacy. If for a person to think is for his/her brain or mind to perform certain computational operations on mental representations, then ''Paul is thinking that the gun went off in Peter's hand'' is made true, when it is, by the occurrence

in his mind or brain of a certain computational operation on mental representations. Since it is being claimed that these computational operations *are* judgements, it is being claimed that the mind or brain, and not Paul, judges; and, since Paul is not identical either with his mind or his brain, and judging is a psychological predicate that can meaningfully be attributed only to Paul, not to some part of Paul, that is the mereological fallacy.

It may seem at this point that an important objection may be made. The mereological fallacy consists in the ascription of characteristics or capacities, etc. to a part of something that can meaningfully be applied only to the whole thing. While it makes sense to say of a motorcar, for example, that it has good acceleration and low fuel consumption, it does not make sense to say of its carburettor either that it has good acceleration or low fuel consumption. That such ascriptions are senseless is evident, but, the objection runs, it is far from evident that the case in point involves them. If Paul *is* his brain, then for Paul to make a judgement just is for a certain brain to make a judgement (which, cognitive science tells us, is for the relevant brain to perform a certain computational operation on mental representations). Nothing is being ascribed to a part of Paul which can meaningfully be ascribed only to Paul here, since, on the current proposal, the brain in question is not a part of Paul, it is Paul.[21]

The apparent availability of this response to our initial objection does a great deal to obscure the significance of the mereological fallacy in cognitive science. It is good to remind ourselves first of all just how profoundly at odds with our ordinary notion of a person the current proposal is. (Call it PB—''a person is a brain''). Our ordinary notion of a person allows us to say such things as that a person has a brain but not that a brain has a brain. We think of persons as easily visible—not as the sorts of things it takes advanced technology to get a look at—as commonly weighing a good deal more than three pounds, and as being the sorts of things that can be religious, politically engaged, funny, sunburned, athletic, selfish—and a host of other things that according to the conventions that regulate ordinary talk, brains cannot be. So there is no question whatever that PB does not accord with our ordinary ways of speaking about persons.[22]

It is at this point that philosophical theses about the status of our ordinary ways of speaking about ourselves have played a decisive role in cognitive science. For the dramatic incompatibility of PB with the logic of our ordinary talk about ourselves has been explained away by appeal to the idea that this logic is nothing but an artefact of an inadequate ''folk-psychology.'' It is only because we uncritically accept the basic principles of our folk-psychology that it seems odd (and perhaps even shocking) to say that a person is identical to a brain. A scientifically respectable psychological theory would show that there is nothing implausible or shocking about PB, and indeed, that we ought to reject the basic conceptual commitments of our folk-theory of mind.

The first thing to appreciate is that the motivation for this rejection of our ordinary conception of mentality and for acceptance of PB is not and could not be empirical. PB is a purely conceptual or a priori proposal in the spirit of traditional philosophy (reminiscent of the phenomenalist's identification of physical objects with actual and possible perceptual experiences). It has to be a purely a priori or conceptual proposal because we have no idea at all what should be treated as evidence for it. If we have no idea what it would be to discover

that PB is true (no idea, that is, how things would have to be for PB to be true) we cannot design an experiment that would test for its truth.

The next thing to appreciate is that PB is more closely connected to mereological confusions than might at first be obvious. What it is for something to count as a person is for him or her to have and exercise a certain range of capacities, including, for example, the capacity to perceive, think, reflect and deliberate. Possession and exercise of such capacities is necessarily the sort of thing that can show up or become manifest in the behavior of the creature who has and exercises them.[23] This rules out, for example, that what makes a person a person is something conceptually quite unconnected to behavior, something, that is, like privileged introspective access to an inner mental world. The thought here is that to enjoy privileged access to a private world of mental objects is to enjoy something that need never show up in behavior of any kind, but to enjoy, for example, perceptual access to an objective environment or the capacities for reflection, intellect, and will which are criterial for personhood, is to enjoy capacities that must sometimes show up in, for example, discriminatory and linguistic behavior of various kinds.

Two things immediately follow from these claims. First, that because we have no more idea of what would count as a brain manifesting in its behavior the capacity to perceive than we have of what would count as a brain manifesting in its behavior the capacity to look, squint, peer, peek, or stare, etc., we can make no sense of the proposal that a person is a brain. (The non-sense involved here is mereological non-sense). Second, that to insist in the face of the obvious absence of such behavioral criteria for ascription of psychological predicates to brains that such ascriptions do indeed make sense, is implicitly to appeal to a most anti-scientific, even mystical, conception of mentality as something like the presence to an inner eye of a private world of mental objects. It is only this sort of radically Cartesian conception of mentality that could sustain the idea that a person is a brain despite the fact that brains do not, and could never, do the sorts of things that make logical room for talk of persons.

The only response open to the defender of PB at this point is to insist that there is, after all, a behavioral basis for ascriptions of psychological predicates to brains, that is, that PB is a genuine empirical discovery rather than an unmotivated and logically spurious proposal about how we ought to speak. In fact, it seems that this thought is more appealing to many cognitive scientists and neuro-philosophers than might be imagined. Certain of the things neurons and sets of neurons do seem to many theorists to resemble things that persons do. Long-term potentiation (LTP), for example, can easily seem to be a kind of synaptic analog of learning or remembering, because the more certain synapses are activated the more sensitive they become to what activates them. Such cases are often said (e.g., Churchland, 2005; Dennett, 2007) to fire the imaginations of researchers: Suddenly the idea that the brain teems with events and processes which are strikingly like the familiar things people do (interpreting, coding, hypothesising, inferring, reasoning, learning, remembering, ordering, obeying, etc.) seems to open up a new, even a revolutionary, explanatory space.

As before there is a sense in which this sort of talk is innocent and a sense in which it is not. To speak, for example, of certain forms of synaptic plasticity as cases of learning or remembering might be logically innocent; it might simply be a catchy way of characterizing neural phenomena conceived as causally necessary conditions of a person's exercise of one

or another of his/her ordinary cognitive, affective, or cogitative capacities. But if what is spoken of in this way (because it is spoken of in this way) is conceived as providing an evidential basis for PB, it involves serious confusion. For in that case the idea is not that there is some feature of a neurological phenomenon which is a little bit like some feature of a familiar thing, like learning, that people do (an idea which already requires the imagination to work very hard indeed), the idea is rather that the familiar thing that people do really *consists* in the neurological phenomenon in question. Only that barely intelligible claim gives the defenders of PB what they need. For what is needed now is the idea that it makes sense to identify a person with a brain because it is really brains that do the familiar things (the learning, remembering, perceiving, and the rest) that provide the behavioral basis for talk of persons.

What must above all be remembered, if we are to avoid logically pernicious flights of fancy of this kind, is that the capacities whose possession and exercise are at the conceptual heart of our notion of a person are linguistic. As Bennett and Hacker remind us so forcefully in *PFN*, the celebrated capacities for reflection and self-awareness that are criterial for personhood are not quasi-perceptual ones. The capacity to reflect on one's motives, actions, and feelings is not the capacity to take a kind of look with a kind of inner eye into a kind of private world. Rather, it consists in a mastery of the many uses of the first-person pronoun (*PFN*, pp. 346–351). So if the claim that persons are brains were based on observations of how brains actually behave, the behavior in question would have prominently to feature brains talking, and in particular, brains exhibiting the capacity to refer to themselves by using the first-person pronoun! The readiness to treat phenomena like LTP as evidence that neurons learn and remember, and to treat neuronal learning as providing a respectable evidentiary basis for PB, reflects, perhaps more than anything else, a powerful failure to appreciate the role of language in human life. If we remembered that persons are above all language-using animals, we would not be so easily confused by extravagant metaphysical theses like PB.

So we cannot defend CT2—the thesis that for a person to think is for a computational operation to occur in an appropriate brain—by appeal to PB because PB either requires appeal to a profoundly anti-naturalistic Cartesian conception of mentality (on which their need be no behavioral criteria for the ascription of psychological predicates), or to the outlandish claim that the behavioral criteria which support the identification of persons with brains are provided by such phenomena as LTP.

## 6. Escaping the dilemma

Our discussion of CT1 and CT2 shows that neither is a coherent interpretation of CT. We can conclude that the events and processes going on in and around the bodies of intelligent creatures which it is the business of the cognitive sciences to discover, are causally necessary enabling conditions of such familiar phenomena as thinking, remembering, learning, perceiving, and imagining. So ''cognition,'' as far as cognitive science is concerned, is a technical term denoting the still largely obscure biological processes that causally enable creatures like ourselves to do the familiar things we do. In modelling these processes, there may well be room for meaningful talk about computation and mental representation, but it

is only metaphysical excess which leads theorists to suppose that such talk is talk about what thinking, learning, remembering, and the rest (really) are.

If we formulate CST and CT in a way which makes commitment to the distinction between enabling conditions and constitutive features of thinking, imagining, remembering, etc. explicit, they will look something like this:

CST: The basic aim of cognitive science (to explain how the mind works) is to explain what must happen in the brain (body/environment) of an intelligent creature if it is to exercise its natural capacities for thought, perception, memory, and will, etc.

CT: Cognition is a causally necessary enabling condition of an intelligent creature's exercise of its natural capacities for thought, perception, memory, and will. It consists in certain possibly computational operations on states or structures in (or around) the brain that, in virtue of their causal roles, count as representations.

We expect that these formulations will meet with considerable resistance both from cognitive scientists and philosophers of mind despite the fact that they constitute the only way out of the dilemma we have presented here. We will end by suggesting that, like the resistance put up to Bennett and Hacker's formulations of the aims and claims of neuroscience in *PFN*, this reaction is rooted in a particularly earnest commitment to an overblown philosophical theory about the ultimate nature of the universe.

## 7. Conclusion: Naturalizing the mind

One of the most negative consequences of allowing the concept of the mind to organize our thinking about mentality is that it is all too easy to suppose that talk of minds can be ''naturalized'' simply by translating it into talk of brains. It cannot. The problematic issue here is not whether something metaphysically special (''consciousness,'' ''qualia,'' ''the self,'' etc.) is left out of such a translation, but that an intelligent creature cannot be identified either with a mind or a brain. Failing to appreciate the importance of this logical point does a lot to encourage a conception of thought (perception, memory, belief, etc.) as a private inner process involving representations that occur in, and are perhaps carried out by, a mind or brain. This conception fascinated and bewitched philosophers throughout the seventeenth century (Kant was really the first to come close to breaking free of it by thinking of concepts not as items in minds but as constituents of capacities to make judgements).[24] Locke, for example, conceived thinking as a kind of inner, mental analog of combustion or bonding, consisting in a combinatorial operation performed by the mind on various sorts of ideas (simple, complex, abstract). Essentially the same conception, reanimated by the idea of a computer and of computational processes, is alive and well in far too much contemporary cognitive science and philosophy of mind (Hacker, 2006).

It is important to appreciate that as well as being logically flawed any such view is profoundly antithetical to a sensible naturalism about the mental. As long as we conceive of the cognitively significant things that intelligent creatures do (thinking, remembering,

perceiving) as ''processes'' involving ''representations'' (or ''ideas'') happening in minds or brains we will be entangled in a bad old metaphysical picture that opposes an inner psychological world to an outer physical one. Our ordinary psychological concepts do not in fact commit us to such a picture (though it is all too easy to forget this)—and it is one of the great merits of *PFN* that it reminds us of this in detail again and again. Our ordinary concepts of thinking, perceiving, judging, believing, remembering, reasoning, etc. are concepts of cognitive capacities possessed and exercised by intelligent creatures (notably persons), not of algorithmic operations or processes which go on in them or between them and their surroundings. ''The mind'' is not, as Descartes thought, a name for a metaphysically controversial entity, but rather, as Aristotle thought, a non-agential and idiomatic way of picking out a set of capacities of intellect, imagination, and will possessed and exercised by certain sorts of animal. So explaining how ''the mind'' works is not a matter of producing something like a physics of representation, but of identifying and characterizing (whether in the formal mode of the cognitive sciences or the material mode of neuroscience) what has to go on in and around a person or animal if it is to be possible for them to possess and exercise these marvellous natural capacities.

To seek to remove all reference to persons and animals from the things we say in cognitive science about thought, perception, and the rest, is to cut ourselves off from the level-headed naturalism we all crave, and land ourselves with an unnaturalizable conception of the mind or brain (or mind-brain) as a kind of inner theater for one, featuring thoughts and perceptions and the rest. What makes this disastrous and exhausted conception seem compulsory is, above all, a commitment to an overblown metaphysical materialism. In the grip of this philosophical theory anything that is hard to think of as composed of elementary particles or stuffs (such as plays, governments, legal systems, wars, words, assertions, smiles) will start to seem metaphysically subversive. But, we suggest, following *PFN*, there is nothing metaphysically suspicious about the ordinary idea of an intelligent creature as something that possesses and exercises a range of cognitively significant natural capacities (for thought, perception, and will, etc.), despite the fact that neither the possession nor the exercise of a capacity, ability, or skill is composed of anything (let alone of particles or stuffs). Rather, the metaphysical subversion is perpetrated by the idea of a mind or brain thinking, or of a person performing computational operations on neurologically realized representations. We have argued here that it is possible to resist this subversion and establish an internally secure and friendly state without giving up the basic aims and claims either of neuroscience or of cognitive science.

## Notes

1. For negative reactions to *PFN* see, for example, Brook (2009); Burgos and Donahoe (2006); Churchland (2005); Dennett (2007); Hodgson (2005); Janzen (2008); Kotchoubey (2005); and Searle (2007). For positive reactions see, for example, Cockburn (2005); Kohler (2003); Pitici (2005); Robinson (2004); Schaal (2005); and Smith (2005). For simplicity, we have divided reviews into two camps. This should not be taken to imply that there are not those, who while broadly sympathetic to the

thrust of Bennett and Hacker's treatment, worry about some of its details. Of course there are: for example, Smith (2005).

2. The most important of these is the mereological fallacy. To commit this fallacy is to attribute to a part of something characteristics or capacities that can meaningfully be attributed only to the thing taken as a whole. *PFN* turns on a discussion of this fallacy (see, e.g., pp. 68–107).
3. As just one example, take this at the beginning of Thagard (2010): ''...I will use evidence from psychology and neuroscience to show how love, work, and play make life meaningful, for most people, whether or not they are religious'' (Thagard, 2010, p. 2). One is tempted to insist, remembering Louis Armstrong's famous dictum, that if you do not already know how love makes life meaningful, neither psychology nor neuroscience (nor religion for that matter) is going to tell you.
4. There can be little doubt that there is much excitement among philosophers, cognitive scientists, and neuroscientists about the idea that the traditional problems of philosophy can finally be solved if not by neuroscience then in light of what we learn from it. Here is an example of a typical way of putting the point: ''...certain basic questions about human cognition, questions that have been studied in many cases for millennia, will be answered only by a philosophically sophisticated grasp of what contemporary neuroscience is teaching us about how the human brain processes information'' (Brook & Akins, 2005, p. 1).
5. For a nice example of how easy it is to forget how much one already knows about the cognitive capacities of intelligent creatures like us, see Hacker's reminders about the extraordinary complexities built into our ordinary concept of thinking (*PFN*, pp. 175–180).
6. Actually, many of the relevant capacities are capacities to acquire capacities—in many cases the important question is not whether something possess the capacity to speak, etc. but whether it possess the capacity to acquire the capacity to speak, etc. This refinement of the relevant claims is not relevant here since brains can neither possess the capacity to speak (etc.) nor the capacity to acquire that capacity.
7. That there is no simple answer to a question like ''just how many of the ordinary things that human beings in fact do does something have to be capable of doing to count as a person?'' does not impugn this claim about personhood. That our ordinary notion is vague does not entail that it is inadequate. Of course there could be persons who in fact were not also, say, citizens or communicators; but this is irrelevant. The question is whether there could be persons that in principle could not be citizens or communicators.
8. For example, when he says, ''To discuss endlessly what silly people mean when they say silly things may be amusing but can hardly be important'' (Russell, 1953, p. 305).
9. See also Hacker (1993) especially pp. 64–67.
10. While it is tempting to try to cite a series of works in which cognitive scientists commit themselves to this sort of view, the list would be far too long and it would invite pointless controversy. The idea that certain basically combinatorial operations on

information-bearing states play an essential role in human and animal intelligence surely has the status of a platitude in cognitive science. Precisely how the notion of a combinatorial operation on information bearing states should be articulated (e.g., whether it must be conceived as a form of computation), and exactly what it means to characterize an internal state as information-bearing or as a representation is, of course, much more controversial.

11. Or the technical sense could refer to a noncomputational process (van Gelder, 1995) or to a process occurring in the brain's vicinity (Clark, 2001).
12. For example, Churchland (2005) speaks of the information (technical causal sense) encoded (technical sense) in maps (technical sense) both intelligibly as having ''useful effects on downstream populations of neurons'' and unintelligibly as what ''constitutes any creatures perception'' (pp. 471–472). That Churchland makes mereological mistakes in the very paragraphs in which he (testily) tries to respond to Hacker's mereological objections shows how poorly understood these objections are.
13. See, for example, Ducasse (1926), or Hart and Honore (1985).
14. And, indeed, many processes and events take place in his brain, throughout his body, and in various other relevant locations. Which of these are the concern of cognitive science is open to parochial wrangling.
15. We do not wear out the leather on the soles of our shoes unintentionally because we do it habitually but because it is a foreseeable but unintended consequence of what we do. A fuller discussion of these distinctions is presented in *PFN*, section 8.1.
16. Of course Paul can cause his *flexor carpi radialis* to contract by doing something else, just as he could cause his heart beat to rise by jumping up and down, but in neither case does flexing his muscle or increasing his heartbeat count as an action he performs. It may be conceivable too that he can train himself somehow to decrease his heartbeat or contract various muscles, but these would be very strange tricks which we should not expect to fit in with ordinary usage. *PFN* discusses these sorts of apparently problematic constructions in Chapter 8.
17. See Hacker (1993, pp. 161–182) for a concise presentation both of Wittgenstein's diagnosis and his therapeutic treatment of this condition.
18. See, for example, G. McCulloch (1989) *The Game of the Name*, pp. 152–163, for an exemplary discussion of the pitfalls of an ''ideational'' theory of meaning.
19. The notion of a representation used here is just the notion of whatever can be combined so as to produce something capable of truth or falsity (so e.g., both words and sentences would count as representations).
20. See Wittgenstein (1958), for example, section 258.
21. A person is sometimes identified with a brain by being identified with a mind that is identified with a brain. This makes no essential difference to the conceptual situation addressed in the text. It is an important point that a cognitivist identification of persons with brains generates essentially the same spread of logical and metaphysical difficulties as a Cartesian identification of persons with minds. Both positions require rejection of the very ordinary idea that persons are a certain sort of intelligent creature.

22. It may be objected here that there is no simple incompatibility between the identity theory and our ordinary notion of ourselves because this theory is underdetermined by our ordinary ways of speaking about ourselves. The claim that there is such an incompatibility, it might be said, is no more plausible than the claim that water is $H_20$ is fundamentally at odds with our ordinary notions of water. The ordinary notion of water, it may be said, neither allows for nor excludes the possibility that water is $H_20$. But it is important to appreciate, first, that PB is not a form of the identity theory as traditionally conceived. That view identifies, for example, sensations with events in brains; it does not identify the things that we ordinarily take to have sensations with brains. PB is the claim that a person is really a brain. Now whereas the identification of sensations with brain processes *may* plausibly be said to be neither ruled out nor in by our ordinary notion of sensation, the same cannot plausibly be said about the identification of person with brains. For example, as we pointed out in the introduction, persons as we ordinarily conceive them are essentially the sorts of things that can both think and speak, and brains are not. So this objection misfires. It should be noted that the discussion that follows in the text does not simply serve to remind us of this incompatibility, however (for such reminders see Hacker, 2007). Rather, we argue that PB is an untestable metaphysical thesis that generates a conception of personhood that is antithetical to naturalism. For a much fuller statement of this argument, see Trigg (2010).
23. This is not a commitment to behaviorism: The claim is that to exercise the capacity, for example, to perceive is to exercise something which necessarily can be manifest in behavior this is altogether different from the claim that it is to exercise something which consists in behavior or the disposition toward it. Just as important it is not a commitment to any philosophical theory about personhood but rather a reminder about how we ordinarily use the term ‘‘person.’’ See, for example, Hacker’s discussion of this concept (in Hacker, 2007, pp. 285–316). ‘‘Persons are essentially things that can behave in various ways’’ is what Wittgenstein would have called a ‘‘grammatical proposition’’; it is not a controversial metaphysical thesis that goes beyond or seeks to extend ordinary usage, but it captures a rule for the (ordinary) use of the relevant expression.
24. As Kant says in the *Critique of Pure Reason* (A68, B93) (Kant, 1929, p. 105), ‘‘The only use the understanding can make of [these] concepts is to judge by means of them.’’ Here, we have the notion that concepts ‘‘rest on functions’’ rather than on peculiar kinds of item, that is, that they are abstractions from capacities to make judgements, not components of complex mental things, but we also seem to have the logically problematic idea that it is ‘‘the understanding’’ that judges, not persons.

## Acknowledgment

This work was supported by grant 0544705 from the National Science Foundation to Kalish.

## References


Bennett, M., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford, England: Blackwell.

Brook, A. (2009). Introduction: Philosophy in and philosophy of cognitive science. *Topics in Cognitive Science*, *1*, 216–230.

Brook, A., & Akins, K. (Eds.) (2005). *Cognition and the brain*. New York: Cambridge University Press.

Burgos, J., & Donahoe, J. (2006). Of what value is philosophy to science? A review of Max R. Bennett and P. M. S. Hacker's *Philosophical Foundations of Neuroscience*. *Behavior and Philosophy*, *34*, 1–87.

Churchland, P. M. (2005). Cleansing science. *Inquiry*, *48*, 464–477.

Clark, A. (2001). Reasons, robots and the extended mind. *Mind and Language*, *16*, 121–145.

Cockburn, D. (2005). Review of M. R. Bennett and P. M. S. Hacker, *Philosophical Foundations of Neuroscience*. *Philosophical Investigations*, *28*, 193–196.

Dennett, D. (2007). Philosophy as naive anthropology: Comment on Bennett and Hacker. In M. Bennett, D. C. Dennett, P. M. S. Hacker, & J. R. Searle (Eds.), *Neuroscience and philosophy: Brain, mind, and language* (pp. 73–95). New York: Columbia University Press.

Dennett, D. (2009). The part of cognitive science that is philosophy. *Topics in Cognitive Science*, *1*, 231–236.

Ducasse, C. J. (1926). On the nature and observability of the causal relation. *Journal of Philosophy*, *23*, 57–68.

Evans, G. (1982). *The varieties of reference*. Oxford, England: Oxford University Press.

van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, *92*, 345–381.

Hacker, P. M. S. (1993). *Wittgenstein meaning and mind*. Oxford, England: Blackwell.

Hacker, P. M. S. (1996). *Wittgenstein's place in twentieth-century analytic philosophy*. Oxford, England: Blackwell Press.

Hacker, P. M. S. (2006). Passing by the naturalistic turn: On Quine's cul-de-sac. *Philosophy*, *81*, 231–253.

Hacker, P. M. S. (2007). *Human nature: The categorial framework*. Oxford, England: Blackwell.

Hart, H., & Honore, A. (1985). *Causation in the law*. 2nd ed. Oxford, England: Clarendon Press.

Hodgson, D. (2005). Goodbye to qualia and all that? *Journal of Consciousness Studies*, *12*, 84–88.

Janzen, G. (2008). Bennett and Hacker on neural materialism. *Acta Analytica*, *23*, 273–286.

Kant, I. (1929). *Critique of pure reason*, trans. N. K. Smith. London: MacMillan.

Kohler, A. (2003). Book review: Wittgenstein meets neuroscience. *Human Nature Review*, *3*, 459–460.

Kotchoubey, B. (2005). Neuroscience through the looking glass. *Journal of Psychophysiology*, *19*, 232–237.

McCulloch, G. (1989). *The game of the name*. Oxford, England: Oxford University Press.

Pitici, F. (2005). Review of philosophical foundations of neuroscience. *Philosophical Psychology*, *18*, 277–281.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, *60*, 20–43.

Robinson, D. N. (2004). Review Philosophical Foundations of Neuroscience. *Philosophy*, *307*, 141–146.

Russell, B. (1953). The cult of 'common usage'. *The British Journal for Philosophy of Science*, *12*, 303–307.

Schaal, D. W. (2005). Naming our concerns about neuroscience: A review of Bennett and Hacker's Philosophical Foundations of Neuroscience. *Journal of the Experimental Analysis of Behavior*, *84*, 683–692.

Searle, J. (2007). Putting consciousness back in the brain: Reply to Bennett & Hacker, *Philosophical Foundations of Neuroscience*. In M. Bennett, D. C. Dennett, P. M. S. Hacker, & J. R. Searle (Eds.), *Neuroscience and philosophy: Brain, mind, and language* (pp. 97–125). New York: Columbia University Press.

Smith, J. (2005). Review of philosophical foundations of neuroscience. *Mind*, *114*, 391–394.

Thagard, P. (2010). *The brain and the meaning of life*. Princeton: Princeton University Press.

Travis, C. (2004). The silence of the senses. *Mind*, *113*, 57–59.

Trigg, J. D. (2010). The philosophy of ordinary language is a naturalistic philosophy. *Essays in Philosophy*, *11*(2), Article 6, 197–215.

Wittgenstein, L. (1958). *Philosophical investigations*, trans. G.E.M. Anscombe. New York: MacMillan.