The Philosophy of Ordinary Language Is a Naturalistic Philosophy

Jonathan Trigg

Published online: 16 July 2010
© Jonathan Trigg 2010

Abstract

It is argued that the only response to the mereological objections of the ordinary language philosopher available to the scientistic philosopher of mind requires the adoption of the view that ordinary psychological talk is theoretical and falsified by the findings of brain science. The availability of this sort of response produces a kind of stalemate between these opposed views and viewpoints: the claim that attribution of psychological predicates to parts of organisms is nonsense is met with the claim that it is only nonsensical if our ordinary ways of talking are—naively—taken to be sacrosanct. The aim of the paper is to show that the ordinary language philosopher has a reply here that the scientistic philosopher is not in a position to ignore. Namely, that the only way to resist mereological objections is to adopt conceptions of personhood that are inimical to naturalistic accounts of mentality.

Only of a human being and what resembles (behaves like) a living human being can one say: it is has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious. (Wittgenstein 1953)

The main reason everybody…is fascinated with and troubled by, work in 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. Will we still have free will? Will we still be conscious, thinking agents who might be held responsible? Does suffering really exist? (Dennett 2009)

Contemporary philosophy of mind has two salient features. First, there is a widespread commitment to the idea that the philosophical study of mind can no longer be continued in isolation from the scientific study of it; and second, there is an almost universal
commitment to the idea that the basic goal of the philosophy of mind is to produce an account that is robustly naturalistic (in the sense of being consistent with current science). Typically, it is assumed that the second of these commitments depends on the first; that is to say, that only if the philosophy of mind becomes appropriately scientific will a properly naturalistic conception of mind be forthcoming.\(^1\) In this paper I argue that this assumption is precisely false, and that the first of these commitments serves to undermine the second. On the view I defend here, a naturalistic account of mentality will consist in a purely scientific account of the causally necessary enabling conditions for the possession and exercise of cognitive, cogitative, conative and affective capacities by intelligent creatures (especially persons). This sort of account will require extensive co-operation between neuroscience and the cognitive sciences (the precise form of which will, no doubt, be controversial), and it will consist in the framing and testing of hypotheses about what has to go on in (and perhaps around) the bodies, and in particular the brains, of intelligent creatures if they are to possess and exercise their natural capacities of intellect, perception and will.

On this account quasi-philosophical theories that generalize from the latest results of neuroscience or cognitive science, or that analyze or refine the technical concepts invoked in these sciences have no part to play in a successful naturalism. The philosopher’s contribution to a naturalistic account of mind is, on this view, purely negative or therapeutic (though not, for that reason, insignificant). It is to show, by a careful assembling of reminders of how we ordinarily talk and think, that the traditional problems in the philosophy of mind reflect the theorist’s misuse of ordinary language, not the inadequacy of our ordinary psychological concepts, or the mysteriousness of mental reality. This contribution promises to prevent conceptual confusions from distorting our understanding of the aims and results of experimental work in brain science and in the cognitive sciences.\(^2\) The notion of a person, and of a person’s various capacities for thought and will, together with the distinction between what is constitutive of possession and exercise of these capacities and what is causally necessary for a person to possess and exercise them – the notions that I take to provide the essential conceptual framework for a successful naturalism – are perfectly ordinary or non-theoretical notions, our possession of which is manifest in the things we say and do in the ordinary circumstances of life.

I will concentrate in what follows on a certain issue about the notion of a person or intelligent creature that Bennett and Hacker have brought to prominence in their recent work on the conceptual foundations of neuroscience.\(^3\) The argument I will present anticipates and responds to a certain sort of objection to the main argument of that work that is likely to occur to a scientifically minded philosopher who believes that there is no sharp divide between the philosophy and the science of mind. If successful, it should show that philosophers committed to naturalizing the mind have reason to question the currently very widespread Quinean conviction that philosophy is continuous with science (scientism).
But this sort of argument is important not only because of its relevance to our conception of the philosophy of mind, cognitive science and brain science. Because it concerns the legitimacy of our ordinary notion of a person it bears on ethical issues too. It is widely assumed that the principal obstacle to the development of a thoroughly naturalistic account of mind is our, perhaps understandable, reluctance to give up well established ethical and political modes of thinking about ourselves. The argument against scientism that I present here is meant to begin to show that this assumption is precisely false. A leading consequence of the view I defend here, is that it is only if we maintain our commitment to these ethical and political conceptions – in particular to the ordinary notion of a person on which they turn – that we stand a chance of developing a convincing and stable naturalism.

The idea that the philosophy of mind should somehow incorporate the latest findings of the brain sciences, and perhaps also of the cognitive sciences (psychology, linguistics, computer science, artificial intelligence, robotics etc.) is very closely connected to the idea that these sciences, whether jointly or severally, should be expected to deliver solutions to traditional philosophical problems about the mind. Both the idea that philosophy of mind should be scientific and the idea that the sciences of mind should be philosophical, require a questionable Quinean holism on which there is no sharp distinction between conceptual and empirical truths. Although Wittgenstein rejected traditional conceptions of conceptual truths as descriptions of the necessary structure of reality (treating an important class of such truths as grammatical propositions that capture rules for the use of the relevant expressions), he would have had no sympathy for Quinean holism. Perhaps the feature of Quine’s philosophy that clashes most dramatically with Wittgenstein’s is that Quine treats language as a kind of theory, replete with a variety of ‘ontological commitments’ that are confirmed or disconfirmed not singly but collectively, where Wittgenstein treats language as (roughly) the medium of a shared ‘form of life’, and so as something which serves predominantly practical rather than theoretical purposes.4

In contemporary philosophy of mind this Quinean conception of language provides the background for the view that our ordinary talk about psychological phenomena like belief and desire constitutes or embodies a rustic folk-theory of mind one day to be replaced by an urbane neuro-theory that will make no reference to spurious mental states and events.5 This conception of our ordinary ways of talking about ourselves can take stronger or weaker forms according to whether our everyday psychological notions are treated as requiring modification or replacement by technical notions drawn from brain science.

If this sort of account of ordinary ways of talking about ourselves is accepted, it can serve to insulate philosophers of mind and neuroscientists from all forms of criticism that bring ordinary psychological concepts to bear on their work. If it is objected, for example, that neither brains nor their parts can see, believe, interpret, order, obey etc. it
will be replied that they can be said to do all of these things, as long as ‘see’, ‘believe’ and the rest are given new technical meanings derived either from brain science or cognitive science. On the other hand, acceptance of this account of ordinary psychological language might make it very difficult for philosophers of mind and neuroscientists to say what it is that their work is about, since doing that will, at least as things stand, involve the use (and not just the mention) of psychological concepts that, by their own lights, have no legitimate applications. This constitutes a rather frustrating stalemate between those committed to a descriptive, non-revisionary philosophical methodology (and a sharp distinction between the grammatico-conceptual investigations of philosophy, and the experimental work of science), and those committed to an explanatory, revisionary philosophical methodology (and to the continuity of philosophy with science and science with philosophy).

Issues about the status of ordinary ways of speaking and thinking about ourselves come to a head in questions about the notion of a person. This notion plays a fundamental role in ordinary psychological talk since it is persons, not minds, brains or brain-parts, that we speak of as having beliefs and desires, as perceiving, imagining, thinking, remembering and the rest. These obvious facts about ordinary linguistic practices may be thought to provide the philosopher of ordinary language with a very simple and direct argument against a large battery of views to be found in contemporary philosophy of mind, cognitive science and neuroscience that involve the (explicit or implicit) ascription of psychological predicates to things other than persons as we ordinarily conceive them. All such views, it may be said, are senseless (not false) because they involve the attribution of a range of characteristics to a type of thing that could neither possess nor lack them. Just as numbers can neither be smooth nor rough, and battles can neither be colored nor colorless, so brains or neurons can neither be thoughtful nor thoughtless.

In fact, the philosopher of ordinary language may specify the variety of nonsense involved in the attribution of psychological predicates to brains more precisely than this. She may say that to speak of the brain (or of neurons or systems of neurons) as having thoughts and beliefs etc. is to attribute predicates to a part of a thing that can meaningfully be attributed only to the thing as a whole. It is, for example, the person – the living human being – that sees, understands, remembers etc. not the brain or the brain-part. Brains can no more be sighted or blind than they can be skeptical or dogmatic. To attribute such capacities to brains, she may say, is to run afoul of the distinctive logical principles that constrain our thought about wholes, parts and their relations; it is to commit a ‘mereological’ fallacy.

Armed with this powerful logical principle derived directly from salient facts about ordinary usage the philosopher of ordinary language may suppose that she can refute all views that involve ascriptions of psychological predicates to things other than persons.
But, given the Quinean atmosphere of contemporary philosophy of mind, the dialectical situation is nothing like that simple.\(^9\)

If we treat ordinary psychological talk as theoretical (and so reject ordinary language philosophy in favor of a Quinean holism) we allow room for the idea that it is just wrong to attribute psychological predicates to persons in the way that we ordinarily do. Many contemporary theorists (both inside and outside philosophy) believe that the sciences of mind currently strongly suggest, and will one day show, that brains or brain-parts are the real subjects of psychological states and events. On this sort of view either our ordinary talk of persons should be dispensed with altogether, or it must be translated into talk about brains or brain-parts. In this way, an appeal to the deep structure of reality may be thought to trump appeals to ordinary linguistic usage. If persons really are brains (or if persons really are minds, and minds really are brains) then to attribute psychological predicates to brains is certainly not to commit a mereological fallacy.

It seems to me that it is primarily because many contemporary philosophers of mind and cognitive science intent on advancing the naturalistic cause are so heavily invested in this sort of position that they are largely unimpressed by the groundbreaking work of Bennett and Hacker. Be that as it may, it seems clear that only a strong metaphysical view of this kind is capable of rebutting the relevant mereological objections.\(^10\)

Ultimately there are two problems with this sort of response to the charge of mereological confusion. First, since our ethical and political thinking is, as a matter of fact, rooted in the ordinary or pre-reflective notion of a person as a self-conscious and at least partly rational creature (the sort of thing, it is worth remembering, that, unlike brains and brain parts, might have a brain) and since, if persons are really brains, the idea that persons are self-conscious and at least partly rational creatures must be wrong, these allegedly urbane neuro-theories of mind are (or will be) flatly inconsistent with our ordinary ethical and political thinking.\(^11\) Whereas partly rational animals are the sorts of things that might have rights and obligations, that could act kindly or cruelly, and be exploited or respected, brains, brain parts or 'cognitive systems', are not. Of course, on its own, this objection has little force in the present context. The idea that our ordinary ethical and political thought is fundamentally wrong, and stands to be corrected by future science, is not only acceptable to many contemporary philosophers of mind but, remarkably enough, attractive to them.\(^12\)

The second problem is more important as far as the shape of the present argument is concerned. I will argue here that the view that persons are brains entails either Cartesianism, or a revisionism about persons so violent that it amounts to eliminativism. What, it seems to me, is so significant about these positions, is that they are at least unfriendly to, if not downright inconsistent with, a robust naturalism about mind. I will rely here largely on the intrinsic plausibility of the claim that neither a
strong Cartesianism about personhood, nor the view that there are no persons, is well suited to contribute to a naturalistic account of mentality. Since an important part of my argument is that an appealing alternative to this pair of views – one that is conspicuously friendly to a sensible naturalism – is to be found in the notion of personhood we already have (the notion that is implicit in our everyday thought and talk about ourselves), the lack of a conclusive argument against them should not weaken my case too much.

Careful inspection of the ordinary practices of ascribing psychological predicates to various types of animal reveals that the criteria for such ascriptions are behavioral: it is because of the things animals do, and the things people say and do, that we ascribe psychological predicates to them in the way that we do. It would be quite senseless to attribute such predicates to things that were in principle incapable of exhibiting the relevant sorts of, for example, linguistic and discriminatory (in general, intelligent) behavior. It is not because we believe that trees and tables lack minds or brains, nor because we believe that they do not have an inner life, nor yet that we believe that certain types of psychological or neural process do not go on inside them – that we do not (seriously) ascribe psychological predicates to them; it is because we believe that they do not and could not behave in the right sorts of ways.

These points bear directly on our ordinary notion of personhood. Persons are essentially such that psychological predicates can be applied to them; since the criteria for ascription of such predicates is behavioral, the criteria for personhood are behavioral too. To be a person (to satisfy the criteria for personhood) is to have a distinctive range of capacities of intellect and will that are essentially such as can be manifest in or expressed by behavior. Now the crucial point is just that it is inconceivable (not merely unlikely) that brains or systems of neurons (etc.) could exhibit the appropriate behavior. For example it makes no more sense to suppose a brain capable of exhibiting the sort of behavior that is criterial for self-consciousness – namely, proper use of the first person pronoun – than it does to suppose that the liver or lungs could exhibit this sort of linguistic competence. The most prominent of the ordinary criteria for personhood is possession and exercise of a series of predominantly linguistic capacities that jointly allow for participation in a shared ‘form of life’. Possession of such capacities is essentially something that can be manifest in behavior of various sorts (behavior that is conspicuously interactive, at least partly rational, and prominently includes appropriate use of the first person pronoun). Nothing that could not in principle exhibit this sort of behavior in the more or less familiar circumstances of life could count as a person. If that is right, and if it is inconceivable that brains could behave, or learn to behave, in these ways, and so count as participating in a shared linguistic life, then the view that persons are brains requires rejection of these ordinary criteria for personhood; in particular, it requires the denial that the criteria for personhood are essentially behavioral.
It is worth noting straight away that the view that the criteria for personhood are essentially behavioral – call it CPB – does not entail behaviorism. To say that to be self-conscious, for example, is to have the capacity to use the first person pronoun is not to identify self-consciousness either with certain forms of behavior or with certain sorts of behavioral disposition. For one thing, to possess a capacity or a power is not to be in a dispositional state of a certain sort. (For a motorcar to have the capacity to travel at a hundred miles an hour is not for it to be in a certain sort of state, for example). Rather, the idea is that whatever makes a person a person must be the sort of thing that can be manifest in behavior – it is not that it must be the sort of thing that consists in behavior or in a behavioral disposition.

For another, whilst the grounds for the ascription of psychological predicates to other people are behavioral there are typically no grounds, behavioral or otherwise, for their ascription to oneself. Such ascriptions do not depend on observations of one’s own behavior or introspective observation of one’s inner states, and they typically do not serve to describe or report but to express or make manifest. To say what one thinks, or to cry out in pain, are typically ways of expressing one’s thoughts and feelings rather than ways of describing them. When one does sincerely express one’s thoughts or feelings by, for example, making statements or crying, the relevant thoughts and feelings are neither mental entities somehow hidden behind the behavior, nor are they physical events or gestures that constitute the behavior – rather they are what is manifest in the behavior. The idea that thinking, willing and feeling are essentially the sorts of thing that can be manifest in or expressed by the behavior of intelligent creatures is a crucial component of our ordinary notion of these phenomena, and it plays a fundamental role in making our ordinary conception of ourselves a conception that is both consistent with and conducive to a sensible naturalism about mentality.

Now the consequences for naturalism of a denial of CPB must be made clear. That denial can take one or other of two forms (typically these forms are not clearly articulated, so confused and confusing combinations of them are widespread). On one, the view that the criteria for personhood are non-behavioral comes out as the view that they are mental. Adherents of this view will place particular emphasis on the idea that if something is a person it must enjoy direct or otherwise privileged access to its current mental states and events, and perhaps also on the idea that if something is a person there will be ‘something it is like to be’ it. This view is strongly Cartesian because, on it, the criteria for personhood turn on the existence of phenomena knowable directly only through introspection; it is because there are a series of mental events, states and processes to which Sam enjoys direct access, and perhaps also because there is ‘something it is like’ for Sam to enjoy such access, that he satisfies the criteria for personhood. The existence of such states, events and process, and his enjoyment of ‘consciousness’, is immediately evident to Sam in introspection, but not directly observable by others. This mentalistic conception of personhood is consistent with the identification of persons with brains because it requires rejection of CPB and it can be
combined with the view that for there to be such a series of mental events and states, and for Sam to enjoy consciousness and privileged access to his current mental states, is for certain processes to go on in, or be performed by, a certain brain.

On a non-mentalistic denial of CPB the view that the criteria for personhood are not behavioral comes out as the claim that they consist in the occurrence of certain sorts of processes in brains (and perhaps in other non-organic sorts of processing device as well). The claim here is that the occurrence of certain processes in brains constitutes the logically adequate evidential basis for talk of persons (in our case at least). On this view, metaphysically legitimate talk of persons would be talk based on observation of the behavior of the brain: it is because certain kinds of events and processes go on in brains (or because the brain performs certain sorts of activities), that it is appropriate to talk about persons. The occurrence of these processes is publically observable (at least in principle) in a way that consciousness and the introspectible contents of a mind are not; so this non-mentalistic denial of CPB is not strongly Cartesian. Nevertheless, because both mentalistic and non-mentalistic views hold that the readily observable intelligent behavior of living human beings – their smiles and frowns, assertions and denials, demands and compliances, thanks and admonishments – has nothing to do with personhood, they might both count as Cartesian in at least a weak sense. A view will be Cartesian in this weak sense if it holds that what matters both logically and metaphysically about persons is something that – in one way or another – lies behind the readily observable intelligent behavior of ordinary human beings. On a mentalistic and so strongly Cartesian position, what satisfies the criteria for personhood lies behind observable behavior as Descartes thought the mental lay behind the physical; on a non-mentalistic and so weakly Cartesian position, it lies behind observable behavior as neural processes in brains lie behind bodily movements and the emission of sounds.

What shall we say about these two ways of denying CPB? It seems to me that because it invests so heavily in the notion of a private inner world made up of events, states and processes to which a single ‘thinker’, ‘mind’ or ‘subject’, enjoys a special kind of access, the mentalistic denial of CPB is straightforwardly incompatible with a satisfying naturalism. Possession and exercise of introspective capacities (often conceived as quasi-perceptual capacities), or enjoyment of what is typically called ‘consciousness’, need have absolutely no behavioral manifestations. So something utterly incapable of observable behavior of any kind – something like an unembodied spirit – could possess consciousness, and enjoy introspective knowledge of its internal states. Of course, so could something like a brain. Brains are unlike disembodied spirits in being capable of being observed, but like them in being incapable of the interactive and prominently linguistic behavior that is criterial for personhood as we ordinarily conceive it. That this conception of personhood is consistent with the thesis that unembodied spirits could be persons hints at its incompatibility with naturalism; that it makes the identification of persons turn on the exercise, not of ordinary
observational capacities but of introspective ones, surely settles the question. That means there could be persons in principle knowable as such only by themselves.

Whilst many philosophers are still optimistic about providing a naturalistic account of these conspicuously first-personal or phenomenological notions, it seems to me that this optimism is misplaced. As soon as we allow the notion of private mental space to play an essential role in our thinking about persons we put a naturalistic account of personhood out of reach. The idea that room has to be found for such notions if the naturalist is to do justice to subjectivity (or ‘the first person’ or even ‘consciousness’), reflects nothing more than a residual commitment to revisionary Cartesian conceptions of subjectivity. It is in interaction with others and in practical action (work, play) that subjectivity manifests itself, not in (allegedly) private inner space. The better naturalizing strategy is to abandon the Cartesian assumptions that make essentially private and ineffable phenomena criterial for personhood; and we can do that simply by articulating – perhaps with the help of ordinary language philosophy – the non-theoretical notion of personhood we already have.

That notion of personhood is precisely not the notion of something essentially private, but of something essentially capable of being manifest in various kinds of public and prominently linguistic interaction. It thus involves no commitment either to a fundamental epistemic opposition between what is knowable only introspectively and what is knowable by ordinary observation, or to a fundamental metaphysical opposition between mental and physical realms. On this ordinary notion, thinking, for example, is criterial for personhood, but it is not conceived as an essentially private phenomenon that goes on in a non-spatial mental medium, but as something essentially capable of being expressed in various sorts of behavior – most obviously in, for example, the making of assertions, the asking of questions, the issuing of warnings and reprimands, the making of jokes and the telling of stories (etc.). Treatments of personhood that focus more or less exclusively on the nature of mental states and events (offering theories of ‘intentionality’ for example), or on what Kant called ‘inner sense’, or on what is nowadays often called ‘consciousness’, allow the familiar and complex forms of interaction between persons in the ordinary circumstances of life to drop out of their account of personhood. These interactions presuppose the body as the vehicle of expressive behavior, a shared language as the medium of communication, and ultimately, a shared culture and history. That they do so, should make the claim that these interactions are criterial for personhood congenial to the philosopher intent on naturalizing ‘the mind’.

Non-mentalistic denials of CPB have different failings. These sorts of views are motivated not only by a distaste for the Cartesian flavor of mentalistic denials of CPB but also by the commitment to make philosophy scientific. In particular what is wanted is something like an empirical basis for the claim that persons are brains. On the face of it, this is not an easy thing to come by; for we simply have no idea what sort of things a
brain would have to do in order to provide us with reasons to speak of it as a person. As already noted, it is hard to imagine a brain addressing us verbally by asking us a question, or by using the first person-pronoun in telling us what it thought. It is equally hard to imagine being led to say of a brain that it is in perceptual touch with its surroundings – for the behavior that provides the criteria for ascribing perceptual experiences to persons and animals is largely discriminatory behavior manifest in successful negotiations of various environmental obstacles that it is impossible to imagine a brain exhibiting. If we have no idea of what should be allowed to count as evidence that a brain is thinking, remembering or perceiving, that is to say, of what it would be like for a brain’s thinking, remembering or perceiving to be manifest in its observable doings, we have no idea of what it might be to discover that brains thought, remembered or perceived, or of how to design an experiment that would settle the question of whether they did these things. That makes it look very much like the identification of persons with brains must be a metaphysical thesis in the old-style, reminiscent, for example, of the phenomenalist’s identification of material objects with actual and possible experiences, or the Humean identification of persons (or ‘selves’) with bundles of experiences.

Perhaps it is mindfulness of this sort of difficulty that leads so many contemporary philosophers and cognitive scientists to insist that many of the things that we have recently discovered going on in brains are actually strikingly like many of the familiar things that people do. According to Dennett, for example, we have recently discovered that the brain is conscious, that it gathers information and makes inferences from it, that it makes simplifying assumptions, uses information to come to conclusions and interprets the information it receives. He confidently upbraids (even lampoons!) Hacker for a kind of intellectual conservatism that refuses to countenance the ascription of familiar psychological predicates to brains and brain-parts. In his reply to Hacker’s presentation at the American Philosophical Association Eastern Division, on 28th December 2005 in New York he put it like this: ‘...it is an empirical fact, and a surprising fact, that our brains, or more particularly parts of our brains, engage in processes that are strikingly like guessing, deciding, believing, jumping to conclusions and so forth. It is enough like these personal level behaviors to warrant stretching ordinary usage to cover it’. It is certainly true that many theorists in brain science and cognitive science are struck by such alleged similarities. Many of the things that neurons and sets of neurons do, seem to such theorists to be ‘strikingly like’ the familiar things that people do. Long term potentiation (LTP), for example, has struck many neuroscientists and cognitive scientists as a kind of synaptic analogue of learning or remembering. Neuroscientists tell us that the more often certain synapses are activated, the more sensitive they become to what activates them, which is to say, the more readily they are activated by those activators. Discoveries like this may seem to justify stretching ordinary usage so as to accommodate talk of such synapses learning or remembering.
There is a way to construe this sort of talk on which it is perfectly innocent. If it amounts to no more than a catchy method of identifying certain of the causally necessary neural conditions for an intelligent creature to exercise its natural capacities for learning or memory then it might be commended for its imaginativeness and any charge of incoherence against it dismissed. But construed in this way it cannot provide the empirical basis for the serious ascription of psychological predicates to brains, and so the behavioral criteria for the identification of persons with brains. For if it is to provide an evidential basis for the identification of persons with brains the thesis that requires support is not just that some neural phenomena are strikingly like some of the familiar things that people do; it is that these neural phenomena and not the familiar things that people do constitute the real criteria for personhood. On this view it is not the familiar adaptive behavior of human beings but the exotic and recently discovered behavior of certain synapses and synaptic networks that provide the true metaphysical basis for talk of persons. Only if we came to that conclusion could we have reason to say that the real subject of learning is not the human organism but certain systems of synapses or the brains that contain them.

The idea that neural phenomena like LTP are enough like the familiar adaptive behavior of human beings to count as the real criteria for personhood, is not commendable imaginativeness, and resistance to it is not condemnable intellectual conservativeness. It is rather a pathological flight of fancy that has potentially devastating ethical and political consequences. Of course it is a flight of fancy that promises to spice up the painstaking experimental labor of science with the apparently lofty thrills of revisionary metaphysics – but that, it seems to me, is precisely a reason to be wary of it.

The scientific revisionist about persons might give up this implausible attempt to justify the identification of persons with brains by appeal to a resemblance between recently discovered neural processes and the ordinary intelligent behavior of human beings. They might say instead that it is simply the intrinsic features of these neural goings on that makes them fit to count as the true criteria for personhood. I think that, as a matter of fact, the enthusiasm in contemporary philosophy of mind and cognitive science for revisionary conceptions of persons of the kind I have been discussing comes most directly from typically computational theories of what tends to be called ‘cognition’ or ‘mental representation’. It is perhaps because so many philosophers and cognitive scientists are persuaded that thinking (which on any plausible view must at least enter into the criteria for personhood) is mental representing, and that mental representing is ultimately a certain sort of neural event or process, that they are so warmly disposed to revisionist accounts of personhood.28

There is indeed a close conceptual connection between personhood and the capacity to think.29 Plausibly, anything that is a person, is a thing that can think; so if thinking is a process that happens in or is performed by a brain, brains might count as things that
think, and so might satisfy the criteria for personhood. If that is right, then a theory of mental representation could be construed as a theory that explains what it is for certain of the most basic criteria for personhood to be satisfied. That sort of theory of mental representation might claim both to be based on experimental work in brain science and to reveal what our talk of persons has always really been about. With this sort of theory in mind it might be maintained that it is because (by dint of difficult experimental work) we can observe computational operations on information-bearing states going on in brains – because we can (or at least will be able to) observe the relevant brain ‘behavior’ – that we have good reason to speak of persons. This is the sort of view that will be taken to combine philosophy with experimental science: it claims to be a theory of personhood based not on mere analysis of the concept of a person, but on detailed observation of the brain.30

The identification of thinking with combinatorial operations on information bearing states in brains (or computers etc.) faces difficulties that parallel those faced by the identification of persons with brains. The basic problem with this identification of thinking with mental representing is that our ordinary notion of thinking is the notion of something that can be, and typically is, manifest in discourse and other forms of interaction between persons. Our notion of a neural process, in apparently sharp contrast, is a notion of something that might be manifest in certain sorts of events (some detectible only by the use of the latest scanning technology) involving certain sorts of brain-part. It is hard then even to understand the proposal that what our ordinary notion of thinking is really a notion of, is a certain type of process realized in the neural architecture of the brain. Even if we can rightly characterize certain neural processes as combinatorial operations on information bearing states (and if we grant that this is enough to distinguish them from renal processes for example) it is not clear why possession of that property qualifies them to replace the familiar intelligent behavior of human beings as what provides the true evidential basis for talk of persons. Indeed, so dramatic is the gulf between the intelligent behavior of human beings and the allegedly combinatorial or computational behavior of brains, that it is hard to see the difference between the claim that there are persons only because brains perform certain computational functions, and the claim that there are really no persons. If we thought that it had turned out that the occurrence of certain types of neural processes are criterial for personhood, we would surely be right to feel that all talk of persons should be given up as a bad job. Eliminativism about persons is not, I think, an attractive or sustainable position (it is much harder to understand and defend than an eliminativism about ‘consciousness’ or ‘qualia’ etc.); if naturalism depends on it, it is a bad thing for naturalism.

Eliminativist or not, it should be noted too that this conception of thinking is bedeviled by Cartesian elements. If, for example we follow Fodor and say that the combinatorial operations constitutive of thinking are operations that are carried out in a ‘language of thought’, we will conceive these operations as standing in merely contingent relations
to possible translations into a natural language and thus to the making of statements and
the asking of questions etc.". So something in principle incapable of exhibiting any of
the familiar forms of linguistic behavior that we ordinarily take to be criterial for
intelligence, could nevertheless count as a thing that thinks. By this standard,
unembodied spirits, brains-in-vats and desktop computers could all count as thinking
things simply because the appropriate combinatorial operations go on in, or are
performed by, them. These things will be thinkable, of course, whether we follow
Fodor or not; for as long as thinking is conceived as a combinatorial operation on
information bearing states or structures it is conceived as something that need stand in
no relation at all to observable behavior in which it is manifest.

Such a conception of thinking is not only consistent with the possibility of unembodied
thinkers, it also brings a familiar array of skeptical issues into play, for on this
conception of it, thinking need have no bearing on objective reality. What matters, on
this notion of thinking, is not how the relevant processes happen to come about, but just
that they take place. That makes them logically and metaphysically independent both of
the existence of thinkers (embodied or otherwise) and of a public world of ordinary
things and events.

There is a way to avoid these treacherous and tedious difficulties whilst maintaining
commitment to the significance of theories of mental representation (and so to the
significance of a great deal of the most prominent work in the philosophy of mind of the
last forty years). It requires that we give up both the identification of thinking with
mental representation (or ‘cognition’, ‘computation’ or ‘information-processing’) and
the revisionary scientism on which ordinary thought and talk about ourselves is
theoretical, and the true criteria for personhood will emerge from experimental work in
brain and cognitive science. If we can learn to take our ordinary notion of personhood
as we find it, we will make conceptual room for the view that a theory of mental
representation is a theory about what has to happen in the bodies, and especially the
brains, of intelligent creatures if it is to be possible for them to possess and exercise
their natural capacities of intellect and will. On this conception of such theories they are
not about what thinking (etc.) is; they are about what certain of its causal preconditions
are.

But learning to take our ordinary conception of ourselves as we find it is not an easy
thing to do. It requires that we learn to resist the constant and powerful temptation to
philosophize about ourselves – that we put up a more or less effective struggle against
the ‘bewitchment of our intelligence by means of language’. This struggle is ordinary
language philosophy. I have argued here that if we want to naturalise mentality we
should take up this struggle and give up Quinean scientism. Because Quinean scientism
is largely motivated by strong naturalizing ambitions and ordinary language philosophy
is not, this is ironic. That scientifically minded philosophers who tend to be suspicious
of ‘ordinary language’ have reason to favor Wittgenstein’s neglected therapeutic and
non-revisionary conception of philosophy over Quine’s dominant scientistic and revisionary conception of it, will doubtless strike many theorists as unlikely. They should remember that Wittgenstein’s concern to capture the detail of ordinary thought led him to a skepticism about traditional philosophical theorizing much more sweeping than that to which Quine’s scientism led him.

The really important point to emerge from this critique of scientism is that our ordinary conception of ourselves as partly rational creatures could not come into conflict with a successful science of mind. The line of thought contained here strongly suggests that in taking interpretations of mind and brain science seriously on which they seem to force us to consider such questions as ‘Am I really a responsible agent?’ and, the remarkable, ‘Does suffering really exist?’33 we do not only ignore and so offend against our humanity, we put a naturalistic conception of ourselves quite beyond reach. For it is only if such familiar phenomena as shame and suffering are treated as *mere* behavior and, as such, are opposed to exotic interior processes (mental or neural) taken to constitute the essence of mentality, that such questions can acquire the appearance of seriousness. This opposition between what can be readily observed but isn’t essential to personhood, and what is essential to personhood but can’t be readily observed, is irredeemably Cartesian, and it cannot be tamed simply by being given a neural gloss. On the broadly Wittgensteinian view I recommend here the ordinary notion of a person, far from standing in the way of a naturalistic science of mentality, is actually its indispensable conduit. With this notion doing its job at the center of our conception of mind, we can conceive perceiving, desiring, remembering, believing, judging and the rest, not as events, states and processes taking place in a mysterious mental or neural world, but as an intelligent creature’s natural capacities (capacities that are essentially such as can be exercised by it in various kinds of action and interaction). The scientific project of framing and testing hypotheses about the causal conditions of the possession and exercise by intelligent creatures of these capacities, just is the project of producing a naturalistic account of mentality. This sort of project – unburdened of the impossible task of naturalizing private inner mental space, or explaining what it is for a brain to think – is as dependent on the ordinary notion of a person as is the ethical and political thinking with which it is so often taken to conflict. If that is right, then ordinary language philosophy not only promises to remove the conceptual obstacles that have, for so long, frustrated attempts to naturalize mentality, it promises to do that by reminding us of an ordinary conception of ourselves that is at least not obviously inconsistent with the very idea that we (and our actions) are apt for genuinely moral kinds of evaluation. That I, and others like me, are the sorts of things that can rightly be held responsible for what we do and that suffer when we are harmed, is not only a necessary condition of a shared ethical life and a decent politics, but also of a stable and satisfying science of mind.
References


---


2 The leading example of this sort of contribution is, and is likely to remain, M.R. Bennett, and P.M.S. Hacker. *Philosophical Foundations of Neuroscience*. (Oxford: Blackwell, 2003).

3 Bennett and Hacker 2003.

4 This means that Wittgenstein is decidedly closer to Heidegger than is Quine, despite the latter’s pragmatist leanings. The crucial issue here is that, on Wittgenstein’s view a language is not any sort of theory and so cannot be confirmed or disconfirmed. Language makes the practices of confirmation and disconfirmation possible, but it cannot itself fit or fail to fit the facts. These themes in Wittgenstein could also be related to central themes in Heidegger’s philosophy.


6 It is important to note that there is some conceptual space between the notion of a person and the notion of a living human being; and also between that notion and the notion of an intelligent creature. Arguably, something could be a living human being without being a person; and it is uncontroversial that many non-human animals are intelligent creatures. It is also the case that something could be a person without being a living human being (super-sophisticated robots or members of alien species for example). I am rather casual about these distinctions in the text, but that is because the crucial distinction in play is that between a brain, or brain-part, or mind, or mind/brain on the one hand, and the thing – whatever it is – that has a brain etc. on the other. There is no doubt that this distinction is quite independent of the relevant controversies. Living human beings that possess an appropriate range of capacities count as persons; but capacities can be possessed without being exercised, and certain of the relevant capacities are second order capacities to acquire capacities (for example the capacity that human infants possess to acquire linguistic capacities.) That means, for example, that infants and those suffering various sorts of disability or disorder will nevertheless count as persons. Many thorny issues lurk in these conceptual woods; but they do not have to be settled here.

7 In an important sense the whole of Bennett and Hacker 2003 is devoted to reminding us of this fact.

8 Discussion of this type of fallacy can be found in Aristotle: ‘To say that the soul is angry is as if one were to say that the soul weaves or builds...’ *De Anima* 408b 12-15 (350 BC); the same issues are taken


10 See Bennett and Hacker 2003 pp.74-81 for a discussion of other types of objection.

11 There is no scope here for a utilitarian objection to this Kantian sounding formulation since brains are no more plausibly things that can suffer than they are things that can reason and reflect.


14 This is a rendering of Wittgenstein’s dictum that only of a living human being or something relevantly like one can we say that it thinks etc. etc. See Wittgenstein 1953, § 109.

15 It is notable that when, as we often do, we fantasize about situations in which we treat trees or tables as persons, we imagine them as having face-like features, and as talking and behaving in other familiar ways. Absent the person-like behavior and the fantasy lapses.

16 I simply help myself to this weighty claim here – it is, I believe, a claim that plays a leading role in Wittgenstein’s later philosophy. For a wonderfully condensed discussion of its significance see Hacker 2002.

17 For a defense of this claim, see, for example, Bennett and Hacker 2003, pp.334-351.


21 See, for example, Bennett and Hacker 2003, p.82, for elucidation of the distinction between inductive and criterial reasoning.
Hacker claims in Bennett and Hacker 2003 that it was characteristic of Cartesianism to attribute psychological characteristics primarily to the mind and only derivatively to the human being; views that identify persons with brains do something similar and so count as Cartesian in this sense.

It is worth noting that the basic naturalising strategy here is functionalism; the aim is to define beliefs, desires and even sensations (for example) in terms of their causal profiles and without appeal to their intrinsic qualitative features. It is notable that this functionalist definition of mental phenomena is consistent with the possibility that unembodied minds could have beliefs and desires etc. Obviously on the ordinary conception of beliefs and desires that I defend here, on which beliefs and desires are defined in terms of the kinds of observable behavior in which they may be manifest, no such Cartesian possibility is kept open.

Of course ways of thinking about perception are still popular – indeed dominant - on which to perceive is to stand in some sort of relation to an internal mental representation – and these ways of thinking are striking symptoms of the Cartesian disease. (See, for example, A. Pautz ‘Intentionalism and Perceptual Presence’ Philosophical Perspectives, 21, pp. 495-541) The present point about such views of course is that they make something that need have no behavioural manifestation – standing in a relation to a private entity or state (or to some proposition that is realised by such an entity or in such a state) – criterial for perceiving.

Dennett’s remark is 39 minutes and 40 seconds into the recording of the proceedings available on this page.


For a fuller discussion of this view see ‘Thought, Language and Mental Representation’ forthcoming in Proceedings of the 32nd Annual Conference of the Cognitive Science Society. The claim that mental representing is ultimately a neural process may involve the idea that mental representing is something that is realised in various sorts of brain process.

It is a connection Cartesians are wont to overplay however – taking it to be a sufficient and not just a necessary condition on being a person.

This type of view is Cartesian in yet another sense because it tends to make the occurrence of essentially ‘internal’ processes seem sufficient for personhood. To be a person is to have a range of capacities that are essentially such as can be manifest in various forms of action – prominently involving practical interactions with other persons. Hacker rightly stresses this point in his discussion of Strawson’s concept of a person (Hacker 2002); it also seems to emerge strongly from Heidegger’s discussion of what he calls ‘being-in-the-world’. It is worth noting that an immediate implication of this conception of personhood is that persons are essentially things capable of standing in morally significant relations to one another.


33 See D. Dennett 2009.