The personal/subpersonal distinction

Zoe Drayson

To appear in Philosophy Compass

Abstract

Daniel Dennett's distinction between personal and subpersonal explanations was fundamental in establishing the philosophical foundations of cognitive science. Since it was first introduced in 1969, the personal/subpersonal distinction has been adapted to fit different approaches to the mind. In one example of this, the 'Pittsburgh school' of philosophers attempted to map Dennett's distinction onto their own distinction between the 'space of reasons' and the 'space of causes'. A second example can be found in much contemporary philosophy of psychology, where Dennett's distinction has been presumed to be equivalent to Stephen Stich's distinction between doxastic and subdoxastic states. Both these interpretations of the personal/subpersonal distinctions, and also Dennett's own philosophical views of the mind, go beyond the personal/subpersonal distinction itself. They each involve supplementing the distinction between personal and subpersonal explanations with metaphysical claims about the relationship between the two kinds of explanation and the entities they posit.

Introduction

The personal/subpersonal distinction was first explicitly drawn by Daniel Dennett in his book Content and Consciousness in 1969, when it was put forward as a distinction between two ways of explaining human behaviour. Dennett's distinction is interesting in its own right, but it also been re-used in different forms in other philosophical projects, each of which approaches the distinction, and in particular the relationship between personal and subpersonal explanations, in a different way.

This survey article begins by introducing Dennett's original distinction, and situating it in the psychological developments of the time. The very notion of an explanation that could be both subpersonal and genuinely psychological had been viewed with suspicion, and the
personal/subpersonal distinction was an essential part of Dennett’s defence of subpersonal psychology, and with it the associated methodology of cognitive science.

The next section considers how philosophers associated with the ‘Pittsburgh school’ attempted to map Dennett’s personal/subpersonal distinction onto their own distinction between what they called ‘the space of reasons’ and ‘the space of causes’. They likened Dennett’s separation of a personal level of explanation to their own insistence on the autonomy of our everyday psychological explanations. There are several points of tension, however, between Dennett’s distinction and the way in which the Pittsburgh scholars used it to emphasise the normativity, rationality, and autonomy of the mind.

While the Pittsburgh school focused on the difference between the personal and subpersonal levels of explanation, psychological functionalists wanted to reduce the personal level of psychological explanation to the subpersonal level. They claimed that beliefs, desires, and suchlike (the posits of personal psychology) corresponded to functionally-identified, often computational, states of psychological subsystems (the posits of subpersonal psychology). In doing so, they found that some posits of subpersonal psychology had no corresponding posit in personal psychology: these are states that Stephen Stich called ‘subdoxastic’. Over the subsequent years, however, Stich’s distinction between doxastic and subdoxastic states is increasingly referred to as a distinction between personal and subpersonal states.

In the final section of the article, I focus on the relationship between the personal/subpersonal distinction and the mind-body problem, and ask what part the distinction can play in helping us understand the metaphysics of mind. Drawing on the previous sections, I show that the issue is a complicated one, not least because of Dennett’s own changing approach to philosophy of mind.

**Dennett and psychological explanation**

Dennett introduced the personal/subpersonal distinction in 1969 as a distinction between "personal and subpersonal levels of explanation" (93). Dennett argued that in addition to
“the explanatory level of people and their sensations and activities”, there is a second kind of psychological explanation, which focuses on "the sub-personal level of brains and events in the nervous system" (93). Dennett’s distinction was a controversial one, because he claimed that both the personal and subpersonal styles of explanation were psychological explanations, in the sense that they account for people’s behaviour in terms of their mental states. Traditionally, mental states like beliefs and desires were only attributed to whole persons rather than to their proper parts. But in the 1960s, psychologists were ascribing mental states to parts of persons and yielding results. Using the method of functional analysis, psychologists such as Deutsch and Attneave broke down each psychological capacity of a person (e.g. the capacity to perceive depth, or understand a language) into separate sub-capacities (e.g. discriminating between inputs, or evaluating information), and attributed each sub-capacity to a subsystem. Like the overall capacity, each sub-capacity is specified in psychological terms, and so each subsystem is like a ‘subperson’ who discriminates, evaluates, calculates, remembers, or suchlike; hence the term ‘subpersonal psychology’. Dennett’s aim was to defend and validate these increasingly popular accounts of human behaviour that attributed mental states below the level of the whole person.

Critics of subpersonal psychological explanation drew on the philosophical ideas of Wittgenstein and Ryle to accuse subpersonal psychology of committing two fallacies. First, they claimed that subpersonal psychology commits the ‘mereological fallacy’ by applying predicates true of the whole to its proper parts. Just as we would be wrong to move from describing water as wet to describing its component atoms (hydrogen and oxygen) as wet, so we would be wrong in attributing thoughts to the component parts of persons, according to this view. The second challenge levelled at subpersonal psychology is that it commits the ‘homunculus fallacy’: it attempts to explain intelligent agency by positing further intelligent agencies. These intelligent agencies (or ‘homunculi’) would then need to be explained by further intelligent agencies, supposedly leading to a regress.

Dennett defended subpersonal psychology from both these accusations. First, he suggested that the method of functional analysis in psychology works precisely because each component ‘subperson’ behaves sufficiently like the person to make the explanation an appropriate one. Where the parts have similar properties to the whole, it is not fallacious to describe them using the same terms: one does not thereby commit the mereological fallacy.
Secondly, he pointed out that homuncular regress only occurs if each intelligent agent is being explained by positing another intelligent agent. If, on the other hand, each intelligent agent is being explained a team of progressively less intelligent agents, then there is no need for regress: at each level of decomposition, the mental states predicated of subsystems will become simpler and simpler until there is no need to ascribe mental states at all. (For further details of both fallacies and how subpersonal psychology is supposed to avoid them, see Drayson 2012.)

Dennett’s aim in introducing the personal/subpersonal distinction was first and foremost to highlight that explanations of behaviour could be both subpersonal and psychological, and thus to demonstrate that there are two types of psychological explanation: personal and subpersonal. The abstract functional explanations of subpersonal psychology could act as a bridge between personal-level psychological explanation and lower-level neural explanation, allowing for the birth of cognitive science. Cognitive science proposed computational models to account for human cognitive capacities, where this involved positing computational processes operating over internal representational states. This talk of computational states ‘representing’ states of affairs seemed worrying to some, because it was normal to assume that only people can genuinely represent the world as being a certain way. Stephen Stich admits that back in the 1960s he “and many other philosophers were skeptical about the idea of invoking internal representations in psychological theories” (2011, xix). But Dennett’s demonstration of how subpersonal psychology could avoid the mereological and homunculus fallacies was found by Stich and others to be “particularly compelling” (xix). The terminology of the personal/subpersonal distinction also allows philosophers of mind to avoid confusion in this regard: they can specify whether their representation-talk is part of a personal or a subpersonal psychological explanation.

It’s important to notice that the personal/subpersonal distinction, as drawn by Dennett, is a distinction between two types of psychological explanation or theory. In order to make claims about the entities to which the explanations refer (if indeed they refer at all), the notion of personal and subpersonal explanatory accounts needs to be supplemented with further claims. Similarly, the distinction itself is silent as to the relation between the two kinds of psychological explanation. The following sections will explore different ways of using the personal/subpersonal distinction to make more substantial claims.
Rationality, autonomy, and normativity

The terminology of Dennett’s personal/subpersonal distinction was quickly picked up by members of the so-called ‘Pittsburgh school’, a group of philosophers influenced by Wilfrid Sellars. Sellars had suggested that certain domains are normatively constrained: they are governed by prescriptive principles rather than descriptive causal laws. He coined the phrase ‘space of reasons’ to cover the former, and distinguished it from the latter ‘space of causes’. Sellars had originally created this distinction as a way of dealing with the normativity of epistemic facts, but John McDowell proposes that our everyday psychological concepts also belong in the space of reasons. McDowell, like Donald Davidson and others, thinks that psychological explanation is governed by normative constraints of rationality:

“[T]he concepts of the propositional attitudes have their proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be.” (389)

McDowell and followers assume that Dennett’s insistence on distinguishing a ‘personal level’ of psychological explanation is motivated by similar concerns to their own emphasis on a “special sort” of explanation: one which shows how mental states can act as reasons for action in a way that makes the action intelligible in light of the beliefs and desires of the agent. Jennifer Hornsby, for example, credits Dennett with the insight that “the point of insisting on a personal level of explanation can be grasped in a context in which persons are seen as (among other things) rational agents” (11). Having identified Dennett’s personal-level explanations with their own explanations in the space of reasons, the Pittsburgh school then map Dennett’s subpersonal explanations onto their own idea of explanations in the space of causes (or the ‘realm of law’ as McDowell characterizes it). There are several important differences, however, between Dennett’s personal/subpersonal distinction and the version adopted by the Pittsburgh school:

- *There is a difference in the kind of explanations involved.* Recall that Dennett’s aim had been to show that the *same* sorts of explanation (ascribing psychological predicates) that we apply to persons in everyday psychology can be usefully applied to the components of persons in scientific psychology. But for the Pittsburgh school,
the personal/subpersonal distinction becomes a distinction between two very different sorts of explanations: normative explanations constrained by rational intelligibility, on one hand, and descriptive explanations constrained by natural laws, on the other.

- **The two approaches differ over whether the personal/subpersonal distinction is a whole/part distinction.** The point of the distinction, for Dennett, is that it distinguishes whether the strategy of psychological explanation is being applied to the whole person or to their parts. Followers of the Pittsburgh school explicitly deny that the personal/subpersonal distinction is a whole/part distinction: Hornsby thinks “it would be a mistake to suppose that the difference between subpersonal and personal levels could be fully caught in part/whole terms” (9) and Matthew Elton cautions against viewing the subpersonal level as “explanations couched in terms of parts of persons” (26).

- **The approaches differ on the scope of personal-level explanations.** The Pittsburgh school restricts their notion of personal-level explanation to reason-giving explanation involving propositional attitudes (beliefs, desires, etc.). Dennett’s personal level does not seem to be so restricted: his original (1969) example of personal-level explanation focuses on pain, for example.

- **Dennett takes subpersonal explanation to be genuinely psychological.** Dennett introduced the term ‘subpersonal’ to qualify a type of psychological explanation, i.e. an explanation ascribing mental state terms to cognitive subsystems. For the Pittsburgh school, however, mental state terms can only correctly be applied in personal-level explanation, and so they use the term ‘subpersonal’ to refer to non-psychological (e.g. physiological) explanations.

All of these differences are a result of the distinction that the Pittsburgh school makes between the ‘space of reasons’ and the ‘space of causes’. The normativity of the space of reasons, according to this approach, guarantees the explanatory autonomy of reason-giving explanations. Dennett’s distinction between personal and subpersonal explanations, on the other hand, is neutral with regard to whether psychological explanation is essentially normative, and whether normativity can be accommodated by descriptive explanations. For Dennett, the precise relation between personal and subpersonal explanations is (initially at least) left unspecified. I’ll say more about this in the final section.
**Subpersonal explanation and subdoxastic states**

Dennett’s original version of the personal/subpersonal distinction between types of psychological explanation is neutral with regard to the relationship between the personal and subpersonal explanations. Subpersonal psychology offers theories of how a person’s psychological capacities might arise: it need not claim that the functional components described by subpersonal explanations correspond to the particular propositional attitude states (e.g. beliefs, desires) that we ascribe in everyday personal explanation. In order to make such claims, one needs to posit an appropriate relation between the states picked out by personal psychology and the states picked out by subpersonal psychology. Philosophers such as William Lycan and Jerry Fodor do just this: for each everyday psychological state we attribute to persons, they propose there is a functionally-characterized proper part of a person posited by subpersonal psychology. Lycan’s ‘homuncular functionalism’, for example, identifies our everyday mental states like belief with “the property of having such-and-such an institutionally characterized state of affairs obtaining in one (or more) of one’s appropriate homuncural departments or subagencies” (41). Since many subpersonal psychological theories are computational theories, these functionally-characterized proper parts of persons are often understood as computational states. Fodor’s computational theory of mind, for example, suggests that “having a particular propositional attitude is being in some computational relation to an internal representation” (1975, 198).

Once this claim about the relationship between personal and subpersonal psychological explanations is in place, something interesting happens. While each psychological state posited by personal explanation corresponds to a functional state posited by a subpersonal theory, the reverse does not hold: there will be functional states posited by subpersonal psychology that do not correspond to anything posited by personal psychology. To see this, consider how functional analysis works in subpersonal psychology, and in particular how it must ‘bottom-out’ in order to avoid the homunculus fallacy. Each intelligent system is analysed into less intelligent subsystems (or subpersons), which are further analysed into even less intelligent sub-subsystems. In the higher-level subsystems, it seems reasonable that we might find functional states that correspond to our everyday psychological states. In
the lower-level subsystems, however, it seems unlikely that any of the functional states will correspond to any of the posits of personal psychology.

The two most often-cited examples of these lower-level subsystems are those posited to account for language-learning and visual processing:

- **Language learning.** Chomsky famously claimed that the ability of children to learn language can't be explained purely on the basis of the external stimuli they receive. Psychological theories of children’s linguistic competence need to posit the existence of internal grammar states: stored information that allows children to become competent speakers despite having insufficient input from their environments.

- **Early visual processing.** How do our sparse retinal data lead to a rich and detailed conscious percept? Vision scientists like David Marr propose that we have computational processes converting information about reflectance properties and light intensity in information about surfaces and edges, for example.

The states posited by these subpersonal psychological theories don’t correspond to anything that we find in personal explanation: we don’t seem to have beliefs about the complex mathematical equations that convert luminosity values into edges; and we can’t experience the contents of our stored grammatical rules, or use the information to draw inferences, for example. So even if we identify some of the functional states posited by subpersonal explanations with states posited by personal explanations, it looks like there will remain other functional states that don’t correspond to anything posited by personal explanation.

This observation was first made by Stephen Stich, who coined the term ‘subdoxastic’ for the lower-level functional states posited by subpersonal psychology, distinguishing them from higher-level functional states that correspond to doxastic states like belief. Drawing on examples including Chomsky’s internal grammar and Marr’s early-visual states, he pointed out that our subdoxastic states are isolated from our doxastic states and our mental lives in general: we can’t use the information they carry in our reasoning or speech, and we have no conscious access to them.
When Stich draws his distinction between doxastic and subdoxastic states, he (like Lycan and Fodor above) is assuming that beliefs and other posits of personal psychology correspond to functionally-characterized proper parts of persons posited by subpersonal psychology. This entails that doxastic states like belief are not only the posits of personal psychology, but also the posits of subpersonal psychology. In other words, he thinks that both doxastic and subdoxastic states appear in subpersonal theories, where these are defined by Dennett as theories which “proceed by analyzing a person into an organization of subsystems” (1978, 154). Our everyday talk of beliefs and such (in personal explanation) refers to parts of functionally-characterized subsystems described by subpersonal theories, according to Stich:

“If we think in terms of a cognitive simulation model, the view I am urging is that beliefs form a consciously accessible, inferentially integrated cognitive subsystem. Subdoxastic states occur in a variety of separate, special purpose cognitive subsystems.” (1978, 508-509)

Notice, therefore, that Stich’s distinction between doxastic and subdoxastic states is not a distinction between those states posited by personal explanations and those states posited by subpersonal explanation. It would be more accurate to describe it as a distinction between those states posited by subpersonal explanation that are also referred to by personal explanation, and those states that appear only in subpersonal explanation. It is unfortunate, therefore, that when Stich’s distinction is used in the current literature, it is usually without his terminology: beliefs and other doxastic states are now referred to more commonly as ‘personal states’, while subdoxastic states are now termed ‘subpersonal states’.

As a result of this terminological change, the term ‘subpersonal state’ has become ambiguous. Some people use ‘subpersonal state’ to refer to any state posited by subpersonal psychology, including those states that correspond to everyday mental states like belief. Other people conflate the terms ‘subpersonal’ and ‘subdoxastic’, and so use the term ‘subpersonal state’ to refer exclusively to those states (like Chomsky’s grammar and Marr’s early-visual states) posited by subpersonal psychology that aren’t also the posits of personal-level psychology.
To see this ambiguity at work, consider two descriptions of a Fodor-style computational theory of cognition. (For further evidence of the ambiguity, see Drayson 2012.) Fodor describes his theory as follows:

“At the very top are states which may well correspond to propositional attitudes that common sense is prepared to acknowledge [...] But at the bottom and middle levels there are bound to be lots of symbol processing operations that correspond to nothing that people – as opposed to their nervous systems – ever do. These are the operations of what Dennett has called “sub-personal” computational systems” (1987, 24)

Fodor’s theory is a subpersonal theory: an explanation which breaks the cognitive system into computational subsystems. While propositional attitudes are the posits of personal explanation, the computational states to which they “may well correspond” are the posits of subpersonal explanation. But Fodor restricts the term ‘subpersonal’ to those computational states that don’t correspond to propositional attitudes. This makes it clear that although he cites Dennett, Fodor is actually using the term ‘subpersonal’ in the way that Stich uses the term ‘subdoxastic’. Now consider Martin Davies’ description of Fodor’s theory:

“We can begin from the assumption that personal-level events of conscious thought are underpinned by occurrences of physical configurations belonging to types that figure in the science of information-processing psychology. These physical configurations can be assigned the contents of the thoughts that they underpin. So we assume that, if a person consciously or occurrently thinks that p, then there is a state that has the representational content that p and is of a type that can figure in subpersonal-level psychological structures and processes.” (370)

When Davies talks of “subpersonal-level psychological structures”, he is clearly not using ‘subpersonal’ as equivalent to Stich’s term ‘subdoxastic’. Davies explicitly says that these subpersonal structures correspond to “personal-level events of conscious thought”, i.e. the propositional attitudes of everyday personal psychology. Davies says nothing in this passage about subdoxastic states: he is using the term ‘subpersonal’ to refer to the states posited by computational psychology in general, not just the subset of subdoxastic states.

These two different ways of understanding and using the term ‘subpersonal state’ are unfortunate, but would be relatively harmless if each was adopted consistently. This is not the case, however: it is common to find someone switching from one interpretation to the
other. When José Bermúdez, for example, describes Fodor’s computational theory of cognition as “a hypothesis about subpersonal cognitive architecture” (159), he is using the term ‘subpersonal’ as Davies does above. In the very next sentence, however, Bermúdez claims that “subpersonal states are inferentially insulated from the conscious processes of cognitive evaluation and self-criticism [...] see, for example, Stich 1978” (159). Here, he is clearly using the term ‘subpersonal’ as Fodor does above, to refer to Stich’s subdoxastic states. Bermúdez puts these two claims together and concludes that Fodor’s theory cannot account for our cognitive evaluation capacities. But his argument is invalid: the first step is only true on one interpretation of ‘subpersonal’, and the second step is only true on the other interpretation of ‘subpersonal’. These are the sorts of problems that arise due to the ambiguity of the term ‘subpersonal’.

Notice that on both these interpretations of the term ‘subpersonal state’, subpersonal states are psychological states: they are the posits (or a subset thereof) of subpersonal psychological explanations. The Pittsburgh school, on the other hand, takes subpersonal explanations to be non-psychological explanations. On their view, subpersonal states are physiological or anatomical states. There can be further confusion, therefore, when either of the two senses of ‘subpersonal’ above is conflated with this third interpretation. Further confusions can arise from different understandings of the term ‘personal’: although everyone seems to agree that personal-level explanations are psychological explanations, only the Pittsburgh school restricts this to reason-giving explanations. In cognitive science, the ‘personal level’ is usually taken to include not just propositional attitudes like beliefs and desires, but also emotions, perceptions, and sensations. There is, of course, nothing preventing us from mixing-and-matching these different notions of ‘personal’ and ‘subpersonal’ to form new versions of the distinction, but it’s not clear what the payoff would be. If we simply want to distinguish between parts and wholes (without any reference to psychology, for example) or between conscious and non-conscious states, we can do this without using the terminology of the personal/subpersonal distinction. Lastly, because the terms ‘personal’ and ‘subpersonal’ are predicated of both explanations and states, it’s important to use precise terminology: talk of ‘the subpersonal’ or the ‘subpersonal level’ tout court should be avoided.
The personal/subpersonal distinction and the mind-body problem

The mind-body problem is a problem about the ontological nature of mental states: what mental states are. The personal/subpersonal distinction, in its original formulation at least, does not seem to shed much light on the matter. The distinction between personal and subpersonal psychology is a distinction between two kinds of explanations or theories: semantic entities, the ontological commitments of which will vary, depending on the views of reference and theoretical entities with which one supplements them.

As we saw in the previous section, some philosophers have taken the methodological strategy of functional analysis from subpersonal psychology and turned it into a functionalist metaphysics. But it's important to notice that the methodology of functional decomposition is just as compatible with the type-identity theory of mental states: a type-identity theorist might claim that subpersonal psychology is a useful heuristic for eventually locating the (non-functionally defined) neurological states to which the terms of personal psychology refer. And even if one does adopt the functionalist view that the posits of personal psychology correspond to the posits of subpersonal psychology, the metaphysical details are yet to be settled. Lycan, for example, proposes to type-identify each posit of personal psychology with a functionally-defined subpersonal component, while Fodor remains neutral as to whether the relation between token psychological states and token computational states is one of identity or supervenience. In his later work, he suggests that it had been a mistake to confuse the methodology of subpersonal psychology with the functionalist program in the metaphysics of mind.

Using the relationship between personal and subpersonal psychological explanations to address the mind-body problem has become an increasingly popular tactic in philosophy of mind. Instead of focusing on the ontological issues, this approach looks instead at how our theories about the mind relate to one another. John Bickle suggests that just as in science, ontological conclusions can be gained from focusing on the relation between theories, looking at our theories of the mind “allows us to reformulate the traditional mind-body problem as first and foremost a question about intertheoretic relationships, and only secondarily as an ontological question” (1). Questions about ontology become questions about theory reduction: Does our everyday theory of psychology reduce to the theories of
subpersonal psychology, as the functionalist hopes? Or do personal psychological theories reduce directly to neurological theories without the need for subpersonal psychology at all, as the type-identity theorist supposes?

Treating the mind-body problem first and foremost as a problem of theory-reduction, however, only works if we take the personal level of psychological explanation to be a proto-scientific theory in the first place. The Pittsburgh school denies this assumption in the first place: for them, the personal level of explanation is not scientific explanation governed by causal laws, but rational explanation governed by norms of reasoning. This, they claim, is what guarantees the autonomy of personal level explanations. But this stance of explanatory autonomy is compatible with different approaches to the mind-body problem, because explanatory autonomy is not necessarily ontological autonomy. Donald Davidson, for example, claims that personal explanation is autonomous but that the states it posits are token-identical with states posited by scientific explanations. John McDowell and Jennifer Hornsby, on the other hand, deny token-identity and couple their claims about the explanatory autonomy of the personal level with a claim of ontological autonomy.

Dennett himself has complicated views on the matter. When he initially draws the personal/subpersonal distinction in *Content and Consciousness*, he seems drawn towards the idea that personal psychological explanations are not proto-scientific theories. He suggests towards the end of that book that mental state terms are non-referential and that personal-level explanation is not causal: this is what leads the Pittsburgh school to adopt his distinction in the first place. In his later book *The Intentional Stance* he suggests that personal explanations are like instrumentalist scientific theories, and thus that the states they posit (beliefs, desires, etc.) are useful fictions. In the more recent paper, ‘Real Patterns’, however, Dennett’s view is somewhat more realist: he proposes that terms like ‘belief’ and ‘desire’ refer to complex behavioural dispositions.

Dennett’s complex and changing metaphysical views of psychological explanation can often obscure our understanding of the distinction between personal and subpersonal explanations. But the distinction itself can be considered completely separately from the development of his ideas about the nature of the mind, and Dennett himself advocates “a stance of ontological neutrality” (1969, 90) when introducing the distinction. Distinguishing
between personal and subpersonal psychological explanations can be a useful undertaking in itself, prior to any theorising about metaphysical commitments. And when we do delve further into the metaphysics of mind, the personal/subpersonal distinction remains a helpful framework for getting clear on the commitments of different philosophical approaches.

**Works cited**


