Draft, October 2012
Forthcoming in *Philosophical Perspectives* 26 (2012)

The uses and abuses of the personal/subpersonal distinction

Zoe Drayson

University of Stirling

1. Introduction

It is a commonplace assumption throughout contemporary philosophy of mind that there is a distinction to be made between *personal* and *subpersonal*. What it distinguishes, however, is a matter of confusion: one finds the terms 'personal' and 'subpersonal' predicated of states, facts, explanations, events, and levels, to name a few. Opinions on the grounds of the distinction are just as wide-ranging. As a result, the personal/subpersonal distinction has prompted confusion; philosophers confess to "not grasping exactly how this distinction is to be drawn" (Rey 2001, 105), describe it as a "somewhat obscure distinction" (Machery 2009, 25), or complain that it "isn't very often made clear" (Boghossian 2008, 133). This befuddlement has not prevented the personal/subpersonal distinction being adopted beyond contemporary philosophy of mind: one finds it in metaethics, legal theory, psychiatry, and economics,[1] and being used to reinterpret the work of past thinkers including Descartes, Kant, and Nietzsche.[2] The aim of this paper is to clarify what the personal/subpersonal distinction is and is not, and to caution against the common confusions that surround it.

[1] Ferrero (2009) questions whether the principle constitutive of agency operate at the personal or subpersonal level; Moore (2010) includes the notion of 'subpersonal intentions' in his theory of criminal law; Hughes (2011) distinguishes understanding the dementia patient as a person from understanding the subpersonal causes of the dementia; Ross (2007) discusses various attempts to model an economics of the subpersonal.

[2] See, for example, Sorell's *Descartes Reinvented* (2005), Brook's *Kant and the Mind* (1994), and Janaway's *Willing and Nothingness* (1998).

In this paper, I claim that the personal/subpersonal distinction is first and foremost a distinction between two kinds of psychological theory or explanation: it is only in this form that we can understand why the distinction was first introduced, and how it continues to earn its keep. I go on to examine the different ontological commitments that might lead us from the primary distinction between personal and subpersonal *explanations* to a derivative distinction between personal and subpersonal *states*. I argue that on one of the most common metaphysical interpretations of the explanatory distinction, talk of a distinction between personal and subpersonal states simply makes no sense. When people insist on applying the personal/subpersonal terminology to psychological states, I allow that they are often making a genuine distinction, but one that it is best understood in terms of Stich's (1978) distinction between doxastic and subdoxastic states. I end the paper by considering some other common misinterpretations of the personal/subpersonal distinction, such as those involving consciousness, normativity, or autonomy.

2. Personal and subpersonal explanation

2.1 Horizontal and vertical explanation

The personal/subpersonal distinction is best understood as a distinction between two types of psychological explanation. The distinction is an instance of a more general distinction, not particular to psychology, between so-called 'vertical' and 'horizontal' approaches to explanation.

The practise of separating vertical and horizontal explanation can be found in our everyday explanations as well as in many domains of scientific explanation.[3]

Horizontal explanations are singular and dated: the *explanandum* is a particular event, and the explanans cites temporally antecedent events, usually of a causal nature. Calling these 'horizontal' explanations reflects the standard practice whereby "we usually represent diachronic causal relations on a horizontal line, from past (left) to present (right)" (Kim 2005, 36). When we explain why a window broke by citing a sequence of events involving the throwing of a stone, for example, we are giving a horizontal explanation.

Vertical explanations focus on accounting for the features of an event rather than its occurrence. Instead of explaining why this particular window broke when it did, a vertical explanation might focus on why the glass shatters in this particular way: what is it about the glass itself that could account for this phenomenon? The *explanandum* of a vertical explanation is often thought of in terms of a thing's capacities or dispositions, and the *explanans* tends to cite the thing's parts or components, e.g. the molecular structure of the glass. 'Vertical' explanation is so-called "to reflect the usual practice of picturing micro-macro levels in a vertical array" (Kim 2005, 36).

In summary: horizontal explanations attempt to account for an event's occurrence by citing a sequence of preceding events, while vertical explanations attempt to account for a thing's features by citing its componential features. Notice that the distinction, as I have introduced it, is between explanations considered as semantically-evaluable. In other words, it is a distinction

[3] Further discussion on the notion of horizontal and vertical explanations can be found in Hoffman 1997, Bermudez 2005, Gaukroger 2010, and Dear 2012.

between two types of *account*, two *approaches* one can take, or two kinds of explanatory *story* or *theory*. This leaves it open whether the explanations are true or not, and how horizontal and vertical explanations relate to each other. The distinction between horizontal and vertical explanations is therefore a more metaphysically-neutral version of such distinctions as Salmon's (1984) distinction between etiological and constitutive explanations and Schaffer's (forthcoming) distinction between causal and grounding explanations. Like the neutral horizontal/vertical framework, the personal/subpersonal distinction in psychology is primarily between the explanatory accounts themselves rather than between their truthmakers. Similarly, the practice of distinguishing between personal and subpersonal explanations does not rely on any claims about one providing the grounds of the other, or any sort of competition between the two.

2.2 Psychological explanation

What happens when we apply the vertical/horizontal distinction to psychological explanation? It is clear to see that our everyday "folk" psychological explanations are horizontal: when we explain a person's behaviour, we cite the sequence of mental events that preceded the behaviour, primarily in terms of propositional attitudes such as the person's beliefs and desires. Horizontal psychological explanation of this sort is our default method of accounting for the actions of other people, and is often termed 'personal' explanation.

The idea of vertical psychological explanation has traditionally been considered problematic, because it is not clear how an explanation could be both vertical and psychological. To give a vertical account of a person's psychological features would require citing features of the person's components, but the obvious way to think about such components is in physiological

terms: brain regions, cells, or neurotransmitters. While these sorts of physiological components might account for a person's physiological features like movements or reflexes, it's hard to see how they could account for a person's *psychological* features: their ability to learn a language, for example, or their capacity for mental arithmetic. Accounting for psychological features in terms of the features of physiological components seems to leave an explanatory gap. One might take this merely as evidence that we don't currently know how to give a satisfying physiological account of human psychology, but some philosophers have made a stronger claim: following in the tradition of Wittgenstein and Ryle, it is suggested that to use non-psychological concepts to account for psychological concepts is to change the subject or make a category mistake.

Does this mean there can't be vertical psychological explanations? The alternative approach would be to ascribe *psychological* instead of physiological predicates to a person's components*:* this would seem to count as genuinely psychological vertical explanation. Here, however, vertical psychological explanation faces what is known as the *mereological fallacy*: ascribing to a part of something a predicate that can only correctly be ascribed to the whole thing. In this case, the fallacy would be to ascribe a psychological predicate to a person's component part, where that psychological predicates can on be ascribed correctly to the whole person. Worries about the mereological fallacy can be found in the work of Wittgenstein and followers:

“It comes to this: Only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious.” (Wittgenstein 1953, para.281)

"It makes no sense to ascribe psychological predicates (or their negations) to the brain […] Psychological predicates are predicates that apply essentially to the whole living animal, not to its parts." (Hacker and Bennett 2003, 72)

By the 1960s, however, psychologists were often to be found ascribing psychological predicates to parts of persons. The justification for such ascriptions was based on the observation that parts of persons - parts of their brains - behaved in way sufficiently similar to whole persons that there was no fallacy involved. In the above quotation, notice that even Wittgenstein allowed that psychological predicates can be ascribed to what *resembles* or *behaves like* a person. When researchers in psychology, robotics, and computer science began to notice that parts of intelligent systems often seem to function in ways similar to the intelligent systems themselves, it looked like this similarity was "sufficient to warrant an adjusted use of psychological vocabulary to characterize that behavior" (Dennett 2007, 78).

2.3 Functional analysis and subpersons

Ascribing psychological predicates to parts of persons was the result of psychologists adopting the approach of *functional analysis*. Functional analysis is an approach to understanding complex systems, both biological and non-biological, in which we attempt to explain how a system works by understanding the functional contributions of its components. It involves singling out the functional phenomenon that we want to comprehend, and analysing it into a number of simpler functions. Each of these subfunctions can, if needed, be decomposed into further sub-subfunctions. The functional analysis ends when we can explain the lowest level of functions by appealing to natural laws such as mechanical or biological principles.[4] Many

[4] See Cummins (1983) for a detailed treatment of functional analysis.

complex systems can be understood in this way, such as the fuel-injection system of a car or the digestive system of a cow.

In psychology, the adoption of functional analysis began with psychologists Deutsch (1960) and Attneave (1961). In opposition to the behaviourist trend of the time, Deutsch realized that functional analysis could provide a way of thinking about psychological states as internal states, without descending to the level of neuroscience:

"An event is explained by being deduced as the property of a structure, system or mechanism and not as an instance of events in its own class. […] The precise properties of the parts do not matter; it is only their general relationships to each other which give the machine as a whole its behavioural properties." (Deutsch 1960, 1)

The early of work of philosopher Fodor (1965, 1968) was also influential in emphasising the view that psychological theories work by providing descriptions of psychological *functions*.[5]

So how does functional analysis in psychology work? It explains complex psychological capacities, such as depth perception or language acquisition, by breaking them down into simpler subcapacities that combine to produce the complex phenomena. The following quotations illustrate how functional analysis works when applied to the science of psychology:

"a large part of the psychologist's job is to explain how the complex behavioral capacities of organisms are acquired and how they are exercised. Both goals are greatly facilitated by analysis of the capacities in question, for the acquisition of the analyzed capacity resolves itself into acquisition of the analyzing capacities and the requisite organization, and the problem of performance resolves itself into the problem of how the analyzing capacities are exercised." (Cummins 1975, 761)

[5] Piccinini (2004) contains an extended discussion of the birth of functional analysis in psychological theorizing.

"the psychologist will first explain the behaviour and behavioural capacities of the whole person in terms of the joint behaviour and capacities of the person's immediately subpersonal departments, and if deeper and more detailed explanation is desired, the psychologist will explain the behaviour of the departments in terms of the joint behaviour and capacities of their own components, and so on down as far as anyone might care to go." (Lycan 1988, 5)

Functional analysis in psychology involves a particular kind of decomposition: the decomposition of the person into *subpersons* to whom we ascribe the sorts of psychological predicates that can explain the personal-level capacities. The ascription of subcapacities to subpersons is thought to avoid committing the mereological fallacy because the component parts of persons appear to function like persons themselves:

"It is an empirical fact, and a surprising one, that our brains–more particularly, parts of our brains–engage in processes that are strikingly like guessing, deciding, believing, jumping to conclusions, etc. And it is enough like these personal level behaviors to warrant stretching ordinary usage to cover it." (Dennett 2007, 86)

2.4 Subpersons and homunculi

Functional analysis in psychology purports to offer vertical psychological explanation that is genuinely psychological, while avoiding the mereological fallacy. All this talk of subpersons or subagents, however, should worry us: it looks like we are trying to explain intelligent beings by positing internal intelligent beings. Worries about such 'homunculi' had earlier been problematic for introspectionist psychology, and one of the motivations behind behaviourist psychology. As a result, behaviourist psychologists like Skinner (1964) were very suspicious of any attempt to reintroduce psychological predicates into vertical explanations.

“Skinner sees – or almost sees – that there is a special way that questions can be begged in psychology, and this way is akin to introducing a homunculus. Since psychology’s task is to account for the intelligence or rationality of men and animals, it cannot fulfil its task if anywhere along the line it presupposes intelligence or rationality.” (Dennett 1978, 58)

The homunculus fallacy is closely related to Ryle's (1949) worry about regress: if our explanations of intelligent mental states or activities require positing further intellectual mental states or activities, then the explanation will lead to an infinite regress. If we have to posit an internal learner or thinker or decider in order to explain a person's capacity to learn, think, or decide, then we have merely postponed the problem rather than solved it.

(Notice that the homunculus fallacy differs from the mereological fallacy. The latter concerns the correct use of our concepts, while the former concerns our explanatory practices. Even if we establish a case where ascribing psychological predicates to parts of persons, the question still remains regarding the explanatory work that can be done this way.)

Psychologist Attneave (1961) was the first to point out that as long as homunculi were used the right way, there was nothing either ghostly or regressive about them. Homunculi are only problematic if we posit an internal agent's psychological capacity to explain that same psychological capacity of the person. If, on the other hand, we posit *several* less-intelligent agents with a *range of capacities* to account for a more-intelligent agent's capacity, it looks like less of a problem.

“It was Attneave’s insight that homunculi can after all be useful posits, so long as their appointed functions do not simply parrot the intelligent capacities being explained. For a subjects’s intelligent performance can be explained as being the joint product of several

constituent performances, individually less demanding, by subagencies of the subject acting in concert. We account for the subject's intelligent activity, not by idly positing a single homunculus within that subject whose job it simply is to perform that activity, but by reference to a collaborative team of homunculi, whose members are individually more specialized and less talented." (Lycan 1991, 259)

One of the important points to remember about functional analysis is that it bottoms out: at some point the functions are subsumed under basic laws. In the case of psychological functional analysis, this means that when the person's capacities are understood in terms of the subcapacities of various subpersons, each of these is further analysed into the subsubcapacities of various subsubpersons, getting progressively less intelligent at each level of decomposition. Eventually, we reach a level where we don't need to ascribe psychological predicates to the components at all: we can understand them in terms of mechanical or biological laws.

"The AI researcher *starts* with an intentionally characterized problem (e.g., how can I get a computer to *understand* questions of English?), breaks in down into sub-problems that are also intentionally characterized (e.g., how do I get the computer to *recognize* questions, *distinguish* subjects from predicates, *ignore* irrelevant parsings?) and then breaks these problems down still further until finally he reaches problem or task descriptions that are obviously mechanistic." (Dennett 1978, 80)

Any worries about homuncular regress are taken care of by this idea that the intelligent components become progressively less intelligent until they can be explained without the need for an psychological predicates at all. This leaves us with a notion of vertical explanation which is genuinely psychological, in virtue of decomposing the person into subpersons, without invoking a regress. This is what is meant by subpersonal explanation.

2.5 The personal/subpersonal distinction

We are now in a position to see what subpersonal explanations are, why there are important to the science of psychology, and why it matters that we distinguish them from personal explanations.

The point of the personal/subpersonal distinction is to emphasise that there is a type of psychological explanation which is not folk-psychological: it is not horizontal, and it does not consist in ascribing psychological predicates to whole persons. In addition to our everyday 'personal' psychological explanations, there are 'subpersonal' psychological explanations that use functional analysis to understand the person in terms of their components, where these components function sufficiently like persons to be thought of as subpersons.

“Sub-personal theories proceed by analyzing a person into an organization of subsystems [...] and attempting to explain the behaviour of the whole person as the outcome of the interaction of these subsystems.” (Dennett 1978, 154)

The practice of offering such psychological theories began with Deutsch (1960) and Attneave (1961), but the term 'subpersonal' was first coined by Dennett (1969) in his distinction between "personal and subpersonal *levels of explanation*" (Dennett 1969, 93). Dennett pointed out that while horizontal explanations in psychology focus on “the explanatory level of people and their sensations and activities", there was a second kind of psychological explanation, vertical explanation, which focused on "the sub-personal level of brains and events in the nervous system” (Dennett 1969, 93).

Subpersonal explanation offered an alternative to personal explanation that was still genuinely psychological, which in turn allowed the birth of cognitive science: explaining intelligence with

intelligence became a reputable approach once it was clear how the mereological and homunculus fallacies are avoided. And it was in this sense that the personal/subpersonal distinction can be understood as being “instrumental in the development and flourishing of the cognitive sciences” (Kriegel 2012, 77). Subpersonal psychology allows us to give vertical explanations of psychological capacities without requiring detailed knowledge of neural factors, thereby providing "precisely the enabling move that lets us see how on earth to get whole wonderful persons out of brute mechanical parts" (Dennett 2007, 89). The introduction of a distinctively subpersonal level of psychological explanation provides an additional way to understand people, supplementing our traditional folk-psychological explanations and the theories of neural circuitry.

The personal/subpersonal distinction not only allows us to distinguish between vertical and horizontal psychological explanations in general, but also enables us to clarify individual instances of psychological predicate ascription. Since both horizontal and vertical explanations in psychology use the same terminology of mental states and intentionally specified capacities, it’s important that we can distinguish the ascription of a particular psychological predicate to a part of person from the ascription of the same psychological predicate to a whole person. The personal/subpersonal distinction is what allows us to avoid confusion.

As should now be apparent the personal/subpersonal distinction is first and foremost a distinction between kinds of psychological *explanation* or *theory*. When Dennett (1969) first introduces the distinction, he also talks of personal and subpersonal “stories” (78), “accounts” (92), and “points of view” (69) - all of which emphasise that the distinction is between semantically-evaluable entities. Dennett's distinction has become widely used throughout

philosophy of mind and beyond, but it is no longer confined to explanation. In some of the more recent literature, the personal/subpersonal distinction is applied to entities such as events (Hurley 1998), processes (Rowlands 2006), states (Machery 2009), contents (McGinn 1988), mechanisms (Metzinger 2003), and facts (Hornsby 2001). In the following section, I focus on the concept of personal and subpersonal *states* to explore how the original distinction might be developed in such a way as to account for the contemporary usage.

3. Personal and subpersonal states

3.1 What is a personal state?

It is not obvious that we should expect a distinction between explanatory accounts to correspond to a distinction between other entities. First, notice that we might be anti-realist about explanation in general, and we might hold that an account is an explanation in virtue of its internal structure rather than in virtue of its relation to external states of affairs. And in order to be an explanation, recall that the account doesn't have to be true: 'how-possibly' explanation still count as explanations on this view. One might think that explanations can be instrumental, and that their terms can fail to refer. But even if we focus our attention on referential explanations, there's no one clear answer to what sort of states they posit. In the case of *personal* explanation, for example, notice that terms like 'belief' and 'desire' don't necessarily pick out internal states of the person. Following Steward, for example, we might think that “the 'ontological commitments' of folk psychology are usually best understood as commitments to abstract entities, not unobservable concrete ones” (Steward 1997, 242). On this view, to be in a

mental state is to instantiate a certain relation to a proposition, rather than to possess an internal state: “Subjects are in mental states, not vice versa” (Williamson 2009).[6]

On an alternative construal of folk psychology, one might think that we instantiate a relation to a proposition in virtue of having an internal representation: an internal state that bears the content of the proposition. This representational theory of mind involves a three-place relation rather than a two-place relation: the relation between people and the content of their thoughts is mediated by internal representations. But even on this view, we are not committed to any particular view of the nature of mental representations.

Merely to indulge in the practise of personal psychological explanation, therefore, is not to be committed to any particular notion of a 'personal state'. What about subpersonal explanations: do they posit a certain kind of ‘subpersonal state’?

3.2 What is a subpersonal state?

Subpersonal explanations, just like personal explanations, need not be referential. Some philosophers and scientists treat subpersonal psychological explanations as *heuristics* to help identify the underlying physiological explanations of behaviour, while denying the existence of subpersonal psychological states. Bickle (2003) for example, holds that subpersonal psychological explanations merely serve as methodological tools to help us locate the "real neurobiological explanations" (Bickle 2003, 110) that don't cite mental states:

[6] See also Thau (2007) on the distinction between internal and instantial states.

"They tell us where in the brain to look [...] But that is all they do, and all they can do. When they've exhausted this descriptive and methodological function they fall away, much like Wittgenstein's ladder" (Bickle 2003, 130)

Bechtel and McCauley (1999) hold a similar view called 'heuristic identity theory'.

If we reject such instrumentalist views of subpersonal explanations and claim that subpersonal explanations are more than mere heuristics, we can ask what sorts of states these explanations posit. As we saw, the standard method of subpersonal psychological explanation takes the form of functional analysis, which posits functional components of the persons that combine to account for the psychological capacities of the person. But notice that a commitment to subpersonal components is not a commitment to any particular relationship between those subpersonal components and the states (whatever they might be) posited by personal explanation.

Thus far, I have said nothing about computational states. The practice of functional analysis is logically independent of computation: the traditional view of functional analysis proceeds by specifying component types and their functions "without specifying the precise state types and state transitions that must occur within the analyzed system" (Piccinini 2004, 818). In practice, however, the project of functional analysis in psychology is often supplemented with computational theory. Computational theory can help us to understand how basic low-level capacities can result in more complex capacities in the way that functional analysis describes: it "supplies us with extremely powerful techniques for constructing diverse analyses of very sophisticated tasks into very unsophisticated tasks" (Cummins 1975, 764). In addition to psychological capacities and activities, computation adds the postulation of discrete internal

psychological states. In order to explain the capacities of the subpersonal components, the computational states themselves have psychological predicates ascribed to them. In other words, computational states are understood as representational states.[7] And notice that thinking about subpersonal psychological explanations as computational is not to be committed to any particular view of the relation between personal and subpersonal psychological theories.

All of this serves to highlight that there is no definitive notion of a 'subpersonal state' that comes out of the distinction between personal and subpersonal explanations. The distinction between personal and subpersonal psychological explanations does not licence any particular commitment to personal or subpersonal states, without further metaphysical claims.

3.3 Personal/subpersonal relations

The existence of personal and subpersonal psychological explanations does not directly result in a commitment to personal and subpersonal psychological states: for that, one needs to supplement the claim about our explanatory accounts with some metaphysical claims. One way to do this is to combine the subpersonal explanatory strategy of functional analysis with personal explanation, and claim that the mental states posited by personal explanations correspond to the functional components posited by subpersonal explanation. An example of this approach is Lycan's 'homuncular functionalism':

"I propose to type-identify a mental state with the property of having such-and-such an institutionally characterized state of affairs obtaining in one (or more) of one's appropriate homunctional departments or subagencies." (Lycan 1987, 41)

[7] Notice that this still leaves us with more than one way of cashing out the notion of a computational state. One might think, like Cummins (1989) that the attribution of mentalistic terms to computational states is just a matter of interpretation. Alternatively, one might have a semantic view of computation, according to which computational states are essentially representational: see Fodor (1975), for example.

Lycan makes it clear that his metaphysical position of homuncular functionalism does not follow merely from the practise of giving subpersonal psychological explanations. He describes his position as "a metaphysics inspired by an epistemology" (Lycan 1991, 259), where the epistemological aspect is the explanatory approach of functional decomposition used in psychology.

This metaphysical position can be adapted to take into account the computational approach to subpersonal psychology. In this case, the beliefs and desires posited by personal explanation correspond to the computational states posited by subpersonal theories. In Fodor's words, "having a particular propositional attitude is being in some computational relation to an internal representation" (Fodor 1975, 198).[8] On this view, the computational state carries the content that we ascribe to the propositional attitude: we ascribe the same content to the mental state posited in personal explanation and to the computational state posited in subpersonal explanation.[9]

"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

[8] Fodor remains neutral on the precise relation (e.g. identity, supervenience) between tokens of propositional attitudes and tokens of computational states. In his later work he claims that he is "by no means convinced that such issues have much substance" (Fodor 2008, 6).

[9] It is possible to use computational theory to give subpersonal psychological explanations without first adopting the programme of functional analysis: a capacity can be analysed into a list of instructions for a sequence of operations without having first to be analysed into subcapacities of functional components (see Piccinini 2004). But most approaches to computational psychology, including Fodor's own, begin by analysing the person's psychological capacities into functional subcapacities. (Fodor's modularity view is a clear example of functional decomposition.) When Fodor identifies propositional attitudes with relations to computational states, those computational states are understood as the states of a *subsystem*. Even thought that subsystem is labelled as the 'central system' or the 'general reasoner', it is still a functional *component* of the overall system.

is of a type that can figure in *subpersonal-level psychological structures and processes*." (Davies 2005, 370, my italics)

Notice that when we identify the posits of personal and subpersonal explanations, whether or not computational theory is involved, we lose any notion of a distinction between personal and subpersonal *states*: the terms of personal and subpersonal explanations refer to the same entities. So not only does the original personal/subpersonal distinction fail to licence *any clear distinction* between personal and subpersonal states, personal and subpersonal states become *indistinguishable* when common metaphysical claims are combined with the explanatory form of the personal/subpersonal distinction. This should make us very wary of some philosophers' tendencies to switch from talking of personal and subpersonal explanations to talking of the 'corresponding' personal and subpersonal states.

And although Lycan is clear on the difference between adopting the explanatory practice of psychology and using it to provide a metaphysics of mental states, not everyone keeps the distinction in mind. Fodor later acknowledged that some his own work may have contributed to what he calls "the widespread failure to distinguish the computational program in psychology from the functionalist program in metaphysics" (Fodor 2000, 105). Partly as a result of this failure, the distinction between personal and subpersonal psychological explanations has been entangled with talk of personal and subpersonal states. What's particularly troubling is that even those people who (like Lycan and Fodor) identify propositional attitudes with functional or computational states persist in drawing a distinction between what they call 'personal' and 'subpersonal' states – despite having a metaphysical position that prevents any such distinction. How should we understand this?

3.4 Doxastic and subdoxastic states

Once the personal/subpersonal distinction is adapted into a metaphysical claim about the relation between the two types of explanation, as in the Fodor and Lycan examples above, something interesting happens. Each propositional attitude posited by personal explanation is identified with a functional or computational state posited by a subpersonal theory, but not every functional or computational state is identified with a propositional attitude. This is a result of how functional analysis works, and in particular the way it avoids committing the homunculus fallacy. Each intelligent capacity is analysed into less intelligent subcapacities, which are further analysed into even less intelligent sub-subcapacities. While the higher-level capacities might be identified with folk-psychological mental states, the lower-level capacities might play important roles in psychological explanation despite not corresponding to any of our standard 'mental states'. The two most often-cited examples of these lower-level capacities involve language-learning and visual processing. In the first of these, 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. The second example concerns the early stages of visual processing: how does our sparse retinal data lead to a rich and detailed conscious percept? Vision psychologists propose that we have computational processes converting information about reflectance properties and light intensity in information about surfaces and edges, for example. Such capacities as these don't correspond to anything that we find in personal explanation: we don't have *beliefs* about the complex mathematical equations that convert luminosity values into edges; 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 functions or computations posited by subpersonal explanations with states posited by personal explanations, there will remain other functions or computations that don't correspond to anything posited by personal explanation.

This observation was first made by Stich (1978), who labelled these lower-level states 'subdoxastic', in contrast to doxastic states like belief. He noticed that our subdoxastic states are isolated, in the sense that we can't use the information they carry in our reasoning or speech, and we have no conscious access to them. Stich concluded that "[s]ubdoxastic states occur in a variety of separate, special purpose cognitive subsystems" (Stich 1978, 508). Our doxastic states, on the other hand, "form a consciously accessible, inferentially integrated cognitive subsystem" (Stich 1978, 508). This talk of cognitive subsystems makes it clear that Stich is drawing a distinction between two kinds of functional component that appear in subpersonal psychological explanations. His distinction is between those components that map onto everyday mental states, and those components that don't.

Stich's distinction does not appear to be used frequently in the current literature, at least if we look for the terminological indications. But the distinction itself, between those functional or computational states that correspond to our folk psychological states and those that don't, is still alive and well. But more often than not, the distinction is labelled as between 'personal' and 'subpersonal' states rather than 'doxastic' and 'subdoxastic' states.

Fodor (1983), for example, endorsed Stich's distinction in its original form. But a few years later, we find him describing the computational components of his metaphysical view 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" (Fodor 1987, 24)

The distinction that Fodor is drawing here is not Dennett's distinction between personal and subpersonal explanations: the computational hierarchy he describes is found in subpersonal rather than personal explanations. He is instead drawing a distinction between two types of computational component, on the basis of whether or not they correspond to propositional attitudes. This is quite clearly a reference to Stich's distinction rather than Dennett's distinction.

A more recent example of the confusion between the two distinctions can be found in Kriegel's (2012) discussion of the 'two visual systems' hypothesis. This is the suggestion that there are two computational pathways in the brain which process visual input in different ways: the dorsal stream leads to conscious perception, while the ventral stream guides fine-grained hand movements in the absence of conscious control. Kriegel makes the claim that "online, on-the-fly visually guided action turns out to be determined by dorsal stream (sub-personal) representations, not ventral stream (potentially personal) ones" (Kriegel 2012, 84). But the computational theories he is discussing are subpersonal psychological theories, which attempt to account for our psychological capacities in terms of subcapacities of our psychological components. Kriegel cannot be referring to the distinction between personal and subpersonal psychological explanations, because only the subpersonal explanations are relevant here. The distinction he seems to be drawing is between those computational states that correspond to the states posited in personal explanation (conscious perception, in this case) and those

computational states that don't correspond in this way. If this interpretation is correct, then the distinction Kriegel is in fact using is Stich's distinction between doxastic and subdoxastic states.

Similar ways of using the terms 'personal' and 'subpersonal' can be found throughout philosophy of mind. The term 'subpersonal' is almost always used to refer to the states of the early visual system, or to the grammatical information in the language system, and almost never to refer to functional or computational states that correspond to our everyday mental states. This suggests that 'subpersonal' is being used instead of 'subdoxastic' to make Stich's (1978) distinction. Notice that in the following quotation, Burge cannot be using 'subpersonal level' to refer to the subpersonal level of explanation:

"I take the subpersonal level to be a level that is not only not conscious, but is not accessible to introspective or reflective consciousness and must be gotten at only theoretically. This is true of the basic grammatical structures underlying our linguistic competence and the information-processing structures underlying our perceptual experience." (Burge 2003, 384)

The subpersonal level of explanation can posit conscious states, accessible to introspection. The most charitable reading of Burge would interpret him as meaning 'subdoxastic' by 'subpersonal'.

4. What the personal/subpersonal distinction isn't

In the first part of this paper, I showed what the personal/subpersonal distinction **is**: a distinction between two kinds of psychological explanation, one horizontal and the other vertical. I also showed why this is an important distinction: the very introduction of subpersonal psychology, via the method of functional analysis, allows us to form psychological explanations

that are genuinely vertical, i.e. distinct from our folk psychological horizontal explanations. This is turn allows us to bridge the explanatory gap between folk psychology and neural circuitry.

The remainder of the paper has been concerned with the abuses, rather than the uses, of the personal/subpersonal distinction. I showed that merely having a distinction between personal and subpersonal psychological explanations does not involve a commitment to any particular kind of psychological state: the explanatory distinction is consistent with a number of ways of thinking about the sorts of states referred to. Talk of personal and subpersonal psychological *states* only makes sense, therefore, within an established framework of metaphysical commitments. But on one of the most common metaphysical frameworks, personal and subpersonal explanations pick out the same set of psychological states: on this view, there is no distinction between personal and subpersonal states, only between the explanations that posit them. I suggest, therefore, using personal/subpersonal distinction as if it coincided with or licenced a distinction between two types of psychological states constitutes an abuse of the original distinction. Furthermore, I argued that in many instances where people take themselves to be implementing the personal/subpersonal distinction, they are in fact using Stich's distinction between doxastic and subdoxastic states. This abuse of the personal/subpersonal terminology has resulted in much confusion in philosophy of mind and beyond.

Any way of using the personal/subpersonal distinction takes it to be a distinction between ontological categories is misleading. The personal/subpersonal distinction should not be understood as a distinction between the mental and physical, for example: the whole point of subpersonal explanations is that they involve the ascription of mental states, just as personal explanations do. Similarly, it is wrong to portray the personal/subpersonal distinction as

equivalent to the distinction between intentional and mechanistic explanations. Even where subpersonal explanations posit mechanisms, they are intentionally described.

One common misinterpretation of the personal/subpersonal distinction involves taking it to distinguish what is conscious from what is unconscious. This is a misinterpretation whether we're talking about the cognitive unconscious or a Freudian notion of the unconscious. If we're interested in the cognitive unconscious, then Stich's distinction would be more suitable, for reasons already discussed. And if we're interested in the Freudian unconscious, then it's not clear that subpersonal explanations come into play at all: psychoanalytic theories tend to involve personal explanations in terms of beliefs, desires, and so on. These may be unconscious mental states, but they arguably account for behaviour by giving a horizontal explanation rather than a vertical explanation.[10]

One of the most pervasive misinterpretations of the personal/subpersonal distinction uses the distinction to support the Sellarsian idea that there are two distinct realms, the space of reasons and the space of causes. Proponents of this view, including Hurley (1998) and Hornsby (2001), take personal explanations to be essentially normative and reason-giving, and subpersonal explanations to be non-normative and mechanistic or causal. On this view, the point of the distinction is to emphasise the autonomy (irreducibility) of propositional attitude explanations that results from the normative constraints on them. This sort of position requires further argument, and cannot be derived from the personal/subpersonal distinction alone.

[10] Garvey (2008) also suggests the explanations in cognitive science are subpersonal whereas explanations in psychoanalysis are personal.

Acknowledgements

Versions of this paper have been presented at the University of Bristol, the University of Edinburgh, the Australian National University, and Macquarie University, and it was also presented at 'The Personal and Subpersonal' conference at the Institute of Philosophy in London in May 2012. Many thanks to everyone involved. Particular thanks are due to Kathleen Akins, David Chalmers, Andy Clark, John Collins, Tim Crane, Daniel Dennett, Anthony Everett, Patrick Greenough, Frank Jackson, Peter Menzies, Michelle Montague, Daniel Nolan, Matthew Nudds, Barry Smith, Mark Sprevak, Daniel Stoljar, and Tillman Vierkant.

References

Attneave F. *(*1961) In defense of homunculi*.* In *Sensory Communication*, Rosenblith WA, ed., pp. 777-782. New York, NJ: MIT Press and John Wiley.

Bechtel, W. P. & McCauley, R. N. (1999). Heuristic identity theory (or back to the future): The mind-body problem against the background of research strategies in cognitive neuroscience. In Martin Hahn & S. C. Stoness (eds.), *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum.

Bermúdez, J. L. (2005) *Philosophy of Psychology: A Contemporary Introduction*. Routledge.

Bickle, J. (2003) Philosophy and Neuroscience: A Ruthlessly Reductive Account. Kluwer Academic Publishers.

Boghossian, P. A. (2008) *Content and Justification: Philosophical Papers*. Oxford University Press.

Brook, A. (1994) *Kant and the Mind*. Cambridge University Press.

Burge, T. (2003). Concepts, conceptions, reflective understanding: Reply to Peacocke. In Martin Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Mit Press.

Cummins, R.C. (1975). Functional Analysis. *Journal of Philosophy* 72 (November):741-64.

Cummins, R. C. (1983). *The Nature of Psychological Explanation*. MIT Press.

Cummins, R. C. (1989). *Meaning and Mental Representation*. MIT Press.

Davies, M. (2005) Cognitive science. In F. Jackson & M. Smith (eds.), *The Oxford Handbook of Contemporary Philosophy*. Oxford University Press.

Dear, P. (2012). Horizontal explanations in the Enlightenment. *Studies in History and Philosophy of Science Part A* 43 (1):221-223

Dennett, D. C. (1969) *Content and Consciousness*. Routledge and Kegan Paul.

Dennett, D. C. (1978) *Brainstorms*. MIT Press.

Dennett, D.C. (2007) Philosophy as naïve anthropology: comment on Bennett and Hacker. In Bennett, M., Dennett, D. C., Hacker, P. M. S. & Searle, J. R. (eds.) *Neuroscience and Philosophy: Brain, Mind, and Language*. Columbia University Press.

Deutsch, J. A. (1960). *The structural basis of behavior*. Cambridge: Cambridge University Press.

Ferrero, L. (2009) Constitutivism and the inescapability of agency, *Oxford Studies in Metaethics* 4: 303–333.

Fodor, J. A. (1965). Explanations in psychology. In M. Black (Ed.), Philosophy in America. London: Routledge and Kegan Paul.

Fodor, J. A. (1968). *Psychological explanation*. New York: Random House.

Fodor, J. A. (1975) *The Language of Thought*. Harvard University Press.

Fodor, J. A. (1983) *The Modularity of Mind*. MIT Press.

Fodor, J. A. (1987) *Psychosemantics: The problem of meaning in the philosophy of mind*. MIT Press.

Fodor, Jerry A. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. MIT Press.

Fodor, J.A. (2008) *LOT2: The Language of Thought Revisited*. OUP.

Garvey, B. (2008) Quasi-beliefs and crazy beliefs: Subdoxastic states and the 'special characteristics' of the unconscious. In *Origins and Ends of the Psyche: Philosophical Essays on Psychoanalysis*, ed. by C. Kerslake and R. Brassier, Leuven University Press.

Gaukroger, S. (2010). *The Collapse of Mechanism and the Rise of Sensibility: Science and the Shaping of Modernity, 1680-1760*. OUP Oxford.

Hacker, P. M. S. & Bennett, M. R. (2003). *Philosophical Foundations of Neuroscience*. Malden MA: Blackwell Publishing.

Hoffmann, R. (1997), *The Same and Not the Same*, New York: Columbia University Press.

Hornsby, Jennifer (2001). *Simple Mindedness: In Defense of Naive Naturalism in the Philosophy of Mind*. Harvard University Press.

Hughes, J. C. (2011) *Thinking Through Dementia*. Oxford University Press.

Hurley, S.L. (1998) *Consciousness in Action.* Harvard University Press.

Janaway, C. (1998) *Willing and Nothingness: Schopenhauer as Nietzsche's Educator*. Oxford University Press.

Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton University Press.

Kriegel, U. (2012). Personal-level representation. *Protosociology*. 28:77-114.

Lycan, William G. (1987). *Consciousness*. MIT Press.

Lycan, W. G. (1988). *Judgement and Justification*. Cambridge University Press.

Lycan, W. G. (1991). Homuncular functionalism meets PDP. In William Ramsey, Stephen P. Stich & D. Rumelhart (eds.), *Philosophy and Connectionist Theory*. Lawrence Erlbaum.

Machery, E. (2009) *Doing Without Concepts*. Oxford University Press.

McGinn, Colin (1988). Consciousness and content. *Proceedings of the British Academy* 74:219-39.

Metzinger, Thomas (2003). *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.

Moore, M. A. (2010) *Placing Blame: a theory of the criminal law*. Oxford University Press.

Piccinini, G. (2004). Functionalism, Computationalism, and Mental States. *Studies in the History and Philosophy of Science* 35 (4):811-833.

Rey, G. (2001) Physicalism and psychology: a plea for a substantive philosophy of mind. In C. Gillett & B. M. Loewer (eds.), *Physicalism and its Discontents*. Cambridge University Press.

Ross, D. (2007) The economics of the sub-personal: two research programs, in B. Montero and M.D. White (eds.), *Economics and the Mind*, 41-57. Taylor & Francis.

Rowlands, Mark (2006). *Body Language: Representation in Action*. Cambridge MA: Bradford Book/MIT Press.

Ryle, Gilbert (1949). The Concept of Mind. Hutchinson and Co.

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Schaffer, J. (forthcoming) Grounding, transitivity, and contrastivity. In Correia and Schnieder (Eds.), *Grounding and Explanation*. CUP.

Skinner, B. F. (1964) Behaviorism at fifty. In T. W. Wann (Editor), Behaviorism and phenomenology: Contrasting bases for modern psychology. Chicago: University of Chicago Press.

Sorell, T. (2005*) Descartes Reinvented*. Cambridge University Press.

Steward, H. (1997) *The Ontology of Mind: Events, Processes, and States*. Oxford University Press.

Stich, S. P. (1978) Beliefs and subdoxastic states. *Philosophy of Science* 45: 499-518.

Thau, M. (2007) Response to Jackson. *Philosophical Studies* 132 (3): 607 – 623.

Williamson, T. (2009) Replies to Critics. In D. Pritchard and P. Greenough (eds.) *Williamson on Knowledge*. 279-384. Oxford University Press.

Wittgenstein, L. (1953*) Philosophical Investigations*. G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.