

## **Wittgenstein and Davidson on First-Person Authority and the Univocity of Mental Terms**

William Child

‘The question can be raised: Is a state that I recognize on the basis of someone’s utterances really the same as the state he does not recognize this way?’ (Wittgenstein 1992, 8-9)

‘If the mental states of others are known only through their behavioural and other outward manifestations, while this is not true of our own mental states, why should we think our own mental states are anything like those of others?’ (Davidson 1991, 207)

### **1. Introduction**

There are striking parallels between Davidson’s conception of the mental and Wittgenstein’s. At the most general level, Davidson’s pronouncement that ‘the mental is not an ontological but a conceptual category’ (Davidson 1987, 114) echoes Wittgenstein’s statement that “‘mental’ . . . is not a metaphysical, but a logical, epithet’ (Wittgenstein 1992, 63). More specifically, Davidson’s idea that mental concepts belong to a *sui generis* system of description and explanation, which is governed by norms distinct from those that govern the physical scheme of description and explanation (see Davidson 1970), echoes Wittgenstein’s idea that the language-game of applying mental terms to ourselves and others is ‘autonomous’ (Wittgenstein 1992, 40), and that the ‘the inner differs from the outer in its *logic*’ (Wittgenstein 1992, 62). And Davidson’s insistence that ‘there are no rules’ for correctly interpreting a person’s words and thoughts, ‘no rules in any strict sense, as opposed to rough maxims and methodological generalizations’ (Davidson 1986, 107), echoes Wittgenstein’s idea that there are no ‘exact rules of evidence’ for ascribing inner states (Wittgenstein 1992, 94): that such rules as there are ‘do not form a system . . . Unlike calculating rules’ (Wittgenstein 1953, part II xi §355 [p. 227])<sup>1</sup>. There are many others points of contact between Davidson’s views and Wittgenstein’s.

Despite the deep similarities, however, there is little direct discussion of Wittgenstein in Davidson’s writings.<sup>2</sup> But one point at which Davidson does explicitly engage with Wittgenstein comes in his writings on first-person authority. And in this case, his tone is critical; he presses an objection to a view about our knowledge of our own and others’ minds that he associates with Wittgenstein and Strawson. The Wittgenstein/Strawson view, as Davidson presents it,

tr[ies] to relieve worries about ‘our knowledge of other minds’ by remarking that it is an essential aspect of our use of certain mental predicates that we apply them to others on the basis of behavioural evidence but to ourselves without benefit of such aid (Davidson 1987, 16).

Davidson agrees that this *is* an essential feature of mental predicates. But, he complains, if that is all we say about the relation between first-person and third-person uses of mental terms, we invite scepticism about whether mental terms have the same meaning in their first-person and third-person uses; and we invite scepticism about whether the mental properties we ascribe to others on the basis of their behaviour are the same as the mental properties we ascribe to ourselves without evidence.<sup>3</sup> For ‘why should we think that a predicate that is

sometimes applied on the basis of observation, and sometimes not, is unambiguous’? (Davidson 1984, 8.) An adequate account of mental terms and concepts, Davidson insists, must answer that question. But the Wittgensteinian approach does not even attempt to address it. It ‘correctly describe[s] the asymmetry between first and other person ascriptions of mental predicates [but it does] nothing to explain it (Davidson 1984, 8).<sup>4</sup>

In §2, I explore Davidson’s own answer to the explanatory challenge he poses. I argue that Davidson’s account explains less than he seems to suggest; he must, in the end, accept that it is simply a basic, irreducible fact that a person can, without observation, reliably self-ascribe the mental properties that others ascribe to her on the basis of her behaviour. Then I turn to Wittgenstein’s position. In §3 I consider and reject the suggestion that Wittgensteinian considerations show Davidson’s challenge to be in some way illegitimate or misconceived. In §4, I argue that Wittgenstein’s discussion of mental concepts offers a plausible answer to Davidson’s challenge.

## **2. Davidson on Self-Ascription and Other-Ascription**

Davidson asks what reason there is to think that mental terms have the same meanings, and pick out the same properties, in their first-person and third-person uses. That question seems a legitimate one – though, as we shall see below, some might question its legitimacy on supposedly Wittgensteinian grounds. Consider a different kind of case. We make claims about birds on the basis of different sorts of perceptual evidence: I can know that there is a cuckoo nearby by hearing its distinctive call; but also, by seeing it fly by. How would we respond to the question, what reason there is to think that the term ‘cuckoo’ has the same meaning, and picks out the same thing, when applied on the basis of auditory experience and on the basis of visual experience? We would certainly not say that it is just a primitive part of the meaning of the word ‘cuckoo’ that it can be applied on the basis both of auditory and of visual experience. Rather, we would appeal to our conception of what a cuckoo is – a physical object of a certain kind – to show how it is that one and the same thing can be known about in these different ways. Davidson’s general strategy in answering his own question about the mental case is broadly the same. His idea is to set out a conception of thought – more specifically, a conception of the *content* of thought – that will show how it is that one and the same property (the property of *having the content, that p*) can be known about in two quite different ways.

We can distinguish two elements in Davidson’s explanatory challenge: two questions he thinks an account of mental concepts or properties must address.

1. Are the mental properties that I ascribe to others on the basis of their behaviour the same as the properties that I self-ascribe without reference to my behaviour?
2. Given that they are the same properties:
  - a) how is it possible for me to know on the basis of behavioural evidence that others have those properties; and
  - b) how is it possible for me to know without evidence that I have those same properties?

Davidson’s answer to his own challenge turns on the idea that, in the simplest and most basic cases, the meanings of my words – and so, according to Davidson, the contents of my thoughts – are determined by the way I apply those words to things and kinds in my environment. As Davidson puts it, ‘whatever she regularly . . . appl[ies] them to gives her words the meaning they have and her thoughts the contents they have’ (Davidson 1987, 37). On this view, using the word ‘cat’ to mean *cat*, for instance, is a matter of being generally disposed to apply the word to cats (and to withhold it from things that are not cats).<sup>5</sup> How

does that conception of meaning and content help in addressing Davidson’s explanatory challenge?<sup>6</sup>

At first sight, Davidson’s broadly dispositional conception of meaning might seem to give an immediate and satisfying answer to question 1. Using a word with a given meaning is a matter of being disposed to apply it in a certain way: I ascribe that dispositional property to others when I say that they use the word ‘cat’ to mean *cat*; I ascribe the same property to myself when I say that I use the word ‘cat’ to mean *cat*. There seems no temptation (as there may be a temptation in the case of properties like pain) to think that there are two different properties in play: a dispositional property that we ascribe to others, and a non-dispositional property that we ascribe to ourselves.

But on reflection, we can see that the mere appeal to dispositional properties does not by itself answer question 1. As Davidson himself points out, a creature can interact with its environment, and with other creatures, in complex and regular ways without having thoughts about an objective world and without acting for reasons. Such a creature could produce sounds in regular ways in response to objects and events in its environment without thereby using a language: its sounds could carry information (in a familiar, causal sense) without having meanings; in producing those sounds, it would not be *saying that* things were some way. (See for instance Davidson 1997, 128; Davidson 2001b, 11-12). So we can distinguish two different kinds of disposition that a creature might have if it is disposed to produce the sound ‘cat’ in response to cats in its environment. It may, like me, have a genuinely linguistic disposition: the disposition to use the word ‘cat’ to say that something is a cat. Alternatively, it may have something more basic: the mere disposition to produce the sound ‘cat’ in the presence of cats. So even if we agree that the property I ascribe in the first-person case and the property I ascribe in the third-person case are both dispositional properties, there is room to doubt whether the *same* dispositional property is involved in each case. We are left with the same question as before: what reason is there to think that the dispositional properties I ascribe to others on the basis of their behaviour are the same as the dispositional properties I self-ascribe without observation?

Davidson thinks that we do know that our fellow human beings are thinking, reasoning, language-using subjects, and not mere information-processors that react in regular ways to the environment and each other. So we do know that the dispositional properties we ascribe to others are the genuinely mental and linguistic properties that we ascribe to ourselves. And he says something about how we know it, with his story about triangulation: about the complex pattern of interactions between two or more people, communicating about a shared environment, which he takes to be a necessary condition for thought.<sup>7</sup> Davidson emphasizes, however, that the difference between a full-blown subject of thought and a mere information-processor is basic and unanalysable; it cannot be explained in terms of anything else. A genuine subject grasps the distinction between how she takes the world to be and how it is: she has the idea of error; the concept of objective truth. That is the difference between a subject of thought and a mere information-processor. But we cannot explain that difference in other terms. In Davidson’s picture, then, this most fundamental feature of the mental – what makes something a genuinely mental state at all – must be accepted as basic and unanalysable. We might wonder why this kind of unanalysability or inexplicability is more acceptable than the inexplicability that Davidson objects to in the Strawson/Wittgenstein account. I will return to that point shortly.

What about question 2? Davidson’s view of meaning makes it easy to see how I can know what someone else means by observing their behaviour. Using a word with a given meaning is a matter of being disposed to apply it in a particular way; and I can tell how someone is disposed to apply a word by observing how she does in fact apply it. That answers the first half of the question: 2(a).<sup>8</sup> But how does Davidson answer the second half

of the question: 2(b)? How do I know what my own words mean, without reference to my linguistic behaviour?

There is a debate in the literature about how exactly we should understand Davidson’s account of our knowledge of the meanings of our own words. It is clear that, if the meanings of a speaker’s words are determined by the way that he himself applies those words, he cannot generally misapply his own words. As Davidson puts it: ‘a person cannot generally misuse his own words, because it is that use which gives his words their meaning’ (Davidson, 1993, 250). (Analogously: a person cannot generally act out of character, because it is the way she generally acts that determines what her character is.) But how does that relate to the idea that a speaker generally knows what his words mean? How do we get from the claim that a speaker cannot generally misapply his own words to the claim that he generally knows what those words mean? (In terms of our analogy: How do we get from the claim that a person cannot generally act out of character to the claim that a person generally knows what her character is?)

Some commentators take Davidson to hold the deflationary view that knowing the meaning of a word just consists in the ability to use it correctly.<sup>9</sup> On this deflationary view, there is no gap between the claim that a speaker cannot generally misapply her own words and the claim that she generally knows what her words mean; for all there is to knowing what one’s words mean is being able to apply them correctly. If we read Davidson in this way, it will follow directly from his account of what determines meaning that speakers generally know what their words mean – and what their thoughts are about.

Others interpret Davidson as taking a less minimalist view of knowledge of the meanings of one’s words.<sup>10</sup> On this less minimalist interpretation, knowing what my word ‘cat’ means is not merely a matter of having a practical ability to use it correctly. It is a piece of higher-level, propositional knowledge: the knowledge that my word ‘cat’ refers to *cats*. But, on Davidson’s view of meaning, this is a kind of higher-level knowledge that speakers cannot fail to have – at least in the simple, basic cases on which we are focusing. The reason is this. For Davidson, using a word with a given meaning is, in the simplest, most basic cases, a matter of being disposed to apply it to things of a particular kind in one’s immediate environment. So a speaker who means something by a word must in general be able to give a correct ostensive statement of its meaning: ‘*That’s a cat*’, for instance. Explaining the meaning of a word in that way draws directly on the very disposition that using the word ‘cat’ to mean *cat* consists in. So it is no mystery that I can know the meaning of my word, and the content of the associated thought, without observing my linguistic behaviour.<sup>11</sup>

However, even supposing that Davidson offers a completely satisfying explanation of our knowledge of the contents of our own and other people’s thoughts, he does not give anything like a general explanation of our knowledge of mental states. For his account only deals with our knowledge of the *contents* of thoughts. He says nothing at all about how I know what *attitudes* I and others have towards those contents: about how I know that I or someone else *believes* that *p*, or *intends* to  $\Phi$ , and so on. But could we develop the basic intuitions behind Davidson’s account of our knowledge of the contents of attitudes to give a more general account of knowledge of those attitudes themselves?

The starting point would be the idea that mental properties are essentially dispositional properties. Davidson certainly accepts that idea. He says explicitly, for instance, that a pro-attitude is ‘a disposition to act under specified conditions in specific ways’ (Davidson 1987, 108); and he takes a parallel view of belief, intention, and the rest. Now as before, if mental properties are dispositional properties of this sort, it is easy to see how we can in principle know other people’s mental properties on the basis of their behaviour. But how easy is it, in the general case, to understand how a dispositional property can be self-ascribed without evidence? We have many dispositional properties that we are

not in a position to know that we have without observation. Being strong, for example, or being allergic to pollen, or being friendly, are dispositional properties. But I cannot know that I have any of those properties without observing how I behave or react in the relevant circumstances. What explains why it is different with mental properties? How is it possible to know without observing my behaviour that I have ‘a disposition to act under specified conditions in specific ways’; how do I know without observation that I believe that *p*, or intend to  $\Phi$ ?

However we interpret Davidson’s explanation of our knowledge of the meanings of our words, it is hard to see how that explanation could be adapted to answer the more general question. On the minimalist interpretation of Davidson’s view, knowing what one’s words mean just consists in being disposed to apply them correctly. But we cannot explain a subject’s knowledge of her own attitudes in a parallel way. Believing that *p* may be a disposition to act in certain ways. But knowing that I believe that *p* is not merely a matter of having that very disposition; it involves more than simply believing that *p*. On the less minimalist interpretation, Davidson’s explanation of our knowledge of the meanings of our words turns on the fact that judging that one’s word ‘cat’ means *cat*, for instance, is a direct manifestation of the disposition that makes it true that one’s word has that meaning: the disposition to apply the word ‘cat’ to cats. But again, that is not a plausible model for our knowledge of what we believe: for judging that I believe that *p* is not, in the same way, a direct manifestation of the disposition that makes it true that I believe that *p*.

Where does that leave Davidson’s response to his own challenge? As far as I can see, Davidson’s account ultimately takes it as a basic, unanalysable fact about such properties as belief, desire, and intention – which he conceives as dispositions to act in certain ways – that subjects can reliably self-ascribe those dispositional properties without observing their own behaviour. That propositional attitudes have that feature is part of the difference between genuine propositional attitudes and the more basic dispositional properties possessed by simpler creatures. But, by Davidson’s own insistence, there is no way of explaining that difference in other terms. Davidson complains that the Wittgenstein-Strawson view correctly describes the asymmetry between self-ascriptions and other-ascriptions, but does not explain it. But, if I am right, the same is true of his own account.

### **3. Davidson’s Challenge and Wittgenstein’s Conception of Philosophy**

How would Wittgenstein respond to Davidson’s demand for an explanation of how mental terms can have the same meanings, and ascribe the same properties, in their first-person and third-person uses? It might be thought that Wittgenstein would either reject Davidson’s challenge altogether or else claim that the issue it raises is trivial and entirely linguistic. In the current section, I explore and reject those two reactions; from Wittgenstein’s point of view, I argue, the question Davidson poses is legitimate and substantive.

#### **3.i Is the demand for an explanation of sameness of meaning legitimate?**

On one view, Wittgenstein would simply reject Davidson’s question; the demand for an explanation of the univocality of mental terms in first-person and third-person ascriptions, he would say, is illegitimate. I shall consider two possible arguments for ascribing that position to Wittgenstein.

A first argument appeals to Wittgenstein’s general insistence that it is not the job of philosophy to explain anything: ‘Philosophy just puts everything before us, and neither explains nor deduces anything’ (Wittgenstein 1953, §126); ‘There must not be anything hypothetical in our considerations. All *explanation* must disappear, and description alone

must take its place’ (Wittgenstein 1953, §109).<sup>12</sup> And Wittgenstein explicitly applies that general point to the case of mental concepts in particular: for example, ‘I look at this language-game [of applying mental terms to ourselves and others] as autonomous. I merely want to describe it, or look at it, not justify it’ (Wittgenstein 1992, 40). So from Wittgenstein’s point of view, it will be said, it is simply a mistake for philosophy to address explanatory questions like Davidson’s.

How should we respond to that argument? It is quite true that Wittgenstein rejects the demand for explanation in philosophy – in the sense of ‘explanation’ expressed by the German word *Erklärung*: that is to say, hypothetical or causal explanation.<sup>13</sup> But Wittgenstein does not reject the idea of explanation in philosophy altogether. After all, he is clear that philosophy can yield understanding: *Verstehen*. A key aim for philosophy, as he sees it, is to produce ‘surveyable representations’ of our practices. And ‘a surveyable representation’, he says, ‘produces precisely that kind of understanding which consists in “seeing connections”’ (Wittgenstein 1953, §122). But producing understanding is, precisely, a kind of explaining. There is, then, no general Wittgensteinian prohibition on explanation in philosophy. There is only a prohibition on a particular kind of explanation: causal or hypothetical explanation. And when Davidson asks for an explanation of the sameness of meaning of mental terms in their first-person and third-person uses, he is obviously not asking for a *causal* explanation. He is asking for an account of our use of mental terms – a surveyable representation, we could say – that will allow us to understand how those terms can be univocal, given the very different bases on which they are applied.<sup>14</sup> There is nothing essentially unWittgensteinian in Davidson’s question; nor is there any presumption about the form of an adequate answer that offends against any Wittgensteinian principle.

A second possible argument is that Wittgenstein would reject Davidson’s demand for an explanation of the univocality of mental terms on the grounds that it misconstrues the dialectical position. We could put the point like this. ‘There is in Wittgenstein’s view no genuine question about the univocality of mental terms. On the contrary, it is perfectly clear in ordinary life that the term “pain”, say, has the same meaning in its first-person and third-person uses. There is nothing intrinsically puzzling about that, and no need for a philosophical account to explain it. If we are sometimes tempted to think that the word “pain” must mean something different in its first-person and third-person applications, or that the claim that it has the same meaning stands in need of a substantive philosophical justification, those are not ideas that we ordinarily take seriously; nor are they the kind of data from which philosophical reflection must begin. On the contrary, they are artefacts of a distinctively philosophical picture of the mental: the picture of the mental as an inner realm that lies concealed behind a person’s behaviour; a realm with which only the subject herself is acquainted and which others can only know indirectly, if at all. We are easily gripped by that picture when we stand back from the practice of applying mental terms to ourselves and others and try to achieve a reflective understanding of the practice. And when we conceive of the mental in terms of this “inner-outer picture”, questions about the univocality of mental terms can seem genuinely pressing. But in Wittgenstein’s view, we should not accept the inner-outer picture and then look for a substantive answer to the question, how it is possible for our mental terms to be univocal. The task for philosophy is, rather, to diagnose the sources of the inner-outer picture and to expose the picture as an illusion. Once that is done, the suggestion that the word “pain” is ambiguous will be seen to have no plausibility, and there will be no need for any positive explanation to justify the common-sense conviction that “pain” and other mental terms are univocal.’

That line of thought plainly reflects an important strand in Wittgenstein’s philosophy. And it is clearly true that Wittgenstein does not think that the univocality of mental terms is genuinely in doubt. But that does not show that, from Wittgenstein’s point of view,

Davidson’s question is misplaced or illegitimate. One strand in Wittgenstein’s philosophy is the aim of achieving a reflective understanding of a practice that we have mastered, but of which we do not command a clear view. In the current case, it is one thing to say that we have a practice of ascribing mental states to others on the basis of their behaviour and to ourselves without evidence. It is another thing to have a reflective understanding of the practice: an understanding that allows us to see why we are entitled to treat the self-ascription and other-ascription of mental terms as elements of a single practice, rather than thinking of our terms as ambiguous. That is a recognizably Wittgensteinian project. And it is reasonable to describe the project as that of giving an account of our practice that makes it intelligible that the property we ascribe to ourselves is the same as the property we ascribe to others. It is wrong, then, to think that Wittgenstein would simply reject the request for an account that shows why mental terms are not ambiguous in their first-person and third-person uses. That, at any rate, is what I shall aim to show.

There is a further and related point. Wittgenstein regards the inner-outer picture as a bad picture of the mental and of our use of mental terms. If the bad picture were right, we can suppose, then ‘pain’ and other mental terms really would be ambiguous in their first-person and third-person uses. But if there is a bad picture of the mental, on which the word ‘pain’ really would be ambiguous, there must be a good view – an accurate account of our practice – that will allow us to see why the term ‘pain’ is not in fact ambiguous. It cannot be that there is literally nothing to say about how we actually use the term and why that use does not make it ambiguous. All that Davidson is asking for when he presses his explanatory question is an account of our mental terms that does that. And, I shall argue, when we look at the detail of Wittgenstein’s treatment of mental terms we can see that he does offer a positive account of just that kind.

### **3.ii Is the answer to Davidson’s challenge trivial?**

A second response to Davidson’s explanatory challenge holds that, from Wittgenstein’s perspective, the answer to Davidson’s question is trivial. On this view, whether or not the property of pain that we ascribe to ourselves is the same as the property we ascribe to others is determined by how we choose to count properties. And that is a merely verbal matter; nothing substantial turns on the choice. As things are, we regard the property we self-ascribe when we use the word ‘pain’ of ourselves, and the property we ascribe to others in third-person uses of the word ‘pain’, as the same property. But we could just as well have chosen to regard it as a different property. And if we had done, we would not be getting things wrong: failing to recognize an objective fact about the identity of the properties ascribed in first-person and third-person uses of the word. For there is no absolute or objective standard of sameness of property: whether or not the same property is involved in the two kinds of case is entirely a matter of what we choose to count as sameness of property.

Someone might cite passages like the following in favour of ascribing that view to Wittgenstein:

If I were to reserve the word ‘pain’ solely for what I had previously called ‘my pain’, and others ‘LW’s pain’, I’d do other people no injustice, so long as a notation were provided in which the loss of the word ‘pain’ in other contexts were somehow made good. Other people would still be pitied, treated by doctors, and so on. It would, of course, be *no* objection to this way of talking to say ‘But look here, other people have just the same as you!’ (Wittgenstein 1953, §403)

On the present interpretation, the point Wittgenstein is making is this. If we use the same word, ‘pain’, in each case, then we will say that the property that other people have is the same as the property that I have. If we reserve the word ‘pain’ for the first-person case, and use a different word in the third-person case, then we will say that other people have a

different property from me. But nothing important turns on which we say. Provided everything else stays the same, the difference between the two positions is purely verbal.

And consider this passage, where Wittgenstein explicitly raises Davidson’s question: The question can be raised: Is a state that I recognize on the basis of someone’s utterances really the same as the state he does not recognize this way? And the answer is a decision (Wittgenstein 1992, 8-9; see Wittgenstein 1982 §428 for a similar remark).

Wittgenstein says it is ‘a decision’ whether to regard the states we self-ascribe and the states we ascribe to others as the same. His point, on the present view, is that there is nothing essential to the practice of using mental terms of ourselves and others that makes it true either that the states are the same or that they are different. We can decide to call them ‘the same’; or we can decide to call them ‘different’. But that is just a matter of how we choose to represent things; it is not a substantial question in its own right.

There is something right and something wrong in that account of Wittgenstein’s approach to questions about sameness of meanings and properties. It is true that, in Wittgenstein’s view, standards of sameness of property are not given by nature but are determined by human practices of description and classification. And it is true that, in his view, a system of classification that counted the property of pain that we self-ascribe without evidence as different from the property we ascribe to others on the basis of their behaviour would not be wrong; it would simply be a different system from ours. But it is wrong to think that Wittgenstein regards these questions about sameness of meaning, or sameness of property, as merely verbal and entirely trivial. And it is wrong to think that he takes such questions to have no substantive answers. That, at least, is what I shall argue in §4.

#### **4. A Wittgensteinian Response to Davidson’s Challenge**

There is a general question, when we should say that a word has a single meaning in its various uses, and when we should say that it has two (or more) different meanings.

Wittgenstein writes:

What does it mean to say that the ‘is’ in ‘The rose is red’ has a different meaning from the ‘is’ in ‘Two times two is four’? If it is answered that different rules are valid for these two words, the retort is that we have only *one* word here. – And if I attend only to the grammatical rules, these do allow the use of the word ‘is’ in both kinds of context. – But the rule which shows that the word ‘is’ has two different meanings in these sentences is the one allowing us to replace the word ‘is’ in the second sentence by the sign of equality, and forbidding this substitution in the first sentence (Wittgenstein 1953, §558).

He returns to the same question a few sections later:

Now isn’t it remarkable that I say that the word ‘is’ is used with two different meanings (as copula and as sign of equality), and wouldn’t want to say that its meaning is its use; its use, namely, as copula and sign of equality?

One would like to say that these two kinds of use don’t yield a single meaning; the union under one head, effected by the same word, is an inessential coincidence. (Wittgenstein 1953 §561).<sup>15</sup>

He discusses an analogy:

Let’s say that the meaning of a piece is its role in the game. – Now let it be decided by lot, before a game of chess begins, which of the players gets white. For this, one player holds a king in each closed hand, while the other chooses one of the two hands, trusting to luck. Will it be counted as part of the role of the king in chess that it is used to draw lots in this way?

...

But, after all, the game is supposed to be determined by the rules! So, if a rule of the game prescribes that the kings are to be used for drawing lots before a game of chess, then that is an essential part of the game. What objection might one make to this? That one does not see the point of this prescription. Perhaps as one likewise wouldn't see the point of a rule by which each piece had to be turned round three times before one moved it. If we found this rule in a board-game, we'd be surprised and would speculate about the purpose of the rule. ('Was this prescription meant to prevent one from moving without due consideration?')

If I understand the character of the game aright, I might say, then this isn't an essential part of it (Wittgenstein 1953 §§563, 567-8).

We can spell out some of the lessons of that discussion.

In the case of chess, Wittgenstein's reflections suggest the following view. The use of the kings to draw lots is not essential to the game, even if it is included in the rules. And the reason it is not essential is that it does not matter to the point of the game; no use is made in the game of the fact that the kings are used to draw lots; it would make no significant difference to the game if we used something else to draw lots. So the role of the king in the game of chess, and its role in drawing lots, are distinct roles. Of course we can yoke the two roles together, by stipulating in the rules of the game that the king is to be used in drawing lots. But for all that, they remain distinct roles; they do not form a unity.

Wittgenstein takes a parallel view about questions of sameness and difference of meaning. Why should we say that the word 'is' has two different uses, and two different meanings, rather than saying that it has a single use, 'as copula and sign of equality'? Because the two meanings are explained in distinct and unconnected ways; there is no point in using the same word in both cases; no use is made of the fact that it is the same word.<sup>16</sup> For those reasons, the two uses do not form a unity. Now Wittgenstein suggests that it is (or can be) a 'decision' whether a given case is a case of ambiguity or, rather, a case of a word's having one meaning and two kinds of use.<sup>17</sup> But the lesson of the quoted discussion from Wittgenstein 1953 §§558 ff (and of the related discussion in Wittgenstein 1982 §§272 ff) is that in answering the question, 'One meaning or two?', we are not making an *arbitrary* decision. On the contrary, what we say is answerable to independent facts about the practice of using the word. Those facts do not absolutely *require* us to answer the question in one way or the other. But if it is to keep contact with the ordinary notion of meaning, the way we count meanings must answer to the kinds of fact about our practice of using the word that we have just mentioned.

In the light of all of that, what should we say about the use of the word 'pain' in first-person and third-person ascriptions? Wittgenstein says that it is 'a decision' whether to say that the properties we ascribe to ourselves without reference to our behaviour are the same as the properties we ascribe to others on the basis of their behaviour. His point is that it would not be wrong to use different words in first-person and third-person ascriptions of what we currently regard as the same mental properties; nor would it be wrong to regard the properties ascribed as different properties. Nonetheless, it is clear that, in Wittgenstein's view, the overall use of the word 'pain', encompassing both first-person and third-person applications, has an internal unity. Its use in first-person ascriptions and third-person ascriptions are not distinct, self-standing practices, arbitrarily yoked together under the single label 'pain'. Rather, they are parts of an integrated whole. Wittgenstein puts the point like this:

One has to look at the concepts 'to be in pain' and 'to simulate pain' in the *third and first* person. Or: the infinitive covers all persons and tenses. Only the whole is the instrument, the concept (Wittgenstein 1992, 37).

Similarly:

One may have the thought: ‘How remarkable that the *single* meaning of the word “to feel” (and of the other psychological verbs) is compounded of so heterogeneous components, the meanings of the *first* and of the *third* person.’

But what can be more different than the profile and the front view of a face; and yet the concepts of our language are so formed, that the one appears merely as a variation of the other. And of course it is easy to give a ground in facts of nature for this structure of concepts. (Heterogeneous things; arrow-head and arrow-shaft.) (Wittgenstein 1980a §45)

Similarly, Wittgenstein suggests, though the first-person and third-person uses of mental terms are in themselves very different, they belong to a single overall use with a genuine, natural unity.

But it is one thing to claim that the first-person and third-person uses of ‘pain’ form a single meaning, with an internal unity. It is another thing to justify that claim: to explain in detail why we *should* say, with Wittgenstein, that only the whole use of the word ‘pain’, in the first and third person, is an ‘instrument’. The basic reason, Wittgenstein thinks, is that the two uses are mutually interdependent; they cannot be given distinct explanations. We can bring out the interdependence in both directions.

In the first place, the first-person use of mental terms is not independent of their third-person, behaviour-based application: their application to other people on the basis of their circumstances and behaviour. In teaching the child to replace her natural, pre-linguistic expressions of pain with the linguistic expression ‘I’m in pain’, we teach her to make first-person uses of the word that are not grounded in observations of her behaviour. But, at the outset, we teach her to make such first-person uses only in circumstances where she already expresses her pain non-linguistically and, therefore, where we can apply the word ‘pain’ to her on the basis of her behaviour. And she only qualifies as having grasped the meaning of the word if her first-person applications are consistent with the way we apply the word to her on the basis of her circumstances and behaviour. Later on, we accept her self-ascriptions of pain as decisive in cases where we have no independent basis for making a third-person ascription of pain. But even then, her overall use of the term must not part company from the third-person criteria. If she regularly applied the word ‘pain’ to herself in a way that conflicted with the external evidence, we would start to doubt whether she really did understand the word after all. By contrast, if the word ‘pain’ were ambiguous, there would be no reason why a person’s applications of the word ‘pain’ to herself would need to harmonize with the way that others applied the word ‘pain’ to her on the basis of her behaviour; for whether or not the first-person term ‘pain’ was true of a person would have nothing to do with whether or not the third-person term was true of her.

The first-person use of mental terms is not, then, a distinct, self-standing practice, independent of their third-person use. Similarly, Wittgenstein thinks, the third-person use of mental terms is not independent of the first-person use of those terms by the subjects of our third-person ascriptions. A word that was applied to others on the basis of their behaviour, but which subjects could not learn to apply to themselves without reference to their behaviour, he suggests, would not be a word for a kind of mental phenomenon at all. He writes:

One might distinguish between two chimpanzees with respect to the way in which they work, and say of the one that he is thinking and of the other that he is not.

But here of course we wouldn’t have the complete employment of ‘think’. The word would have reference to a mode of behaviour. Not until it finds its particular use in the first person does it acquire the meaning of mental activity (Wittgenstein 1980b §§229-30).

In other passages, too, Wittgenstein imagines counterparts of our ordinary mental terms that relate only to behaviour.

Look at it purely behaviouristically: Someone says, ‘Man thinks, wishes, is happy, angry, etc.’. Imagine that these words were only about certain forms of behaviour on certain occasions. One could suppose that whoever talks about human beings in this way had first observed these kinds of behaviour in other beings and was now saying that these phenomena could also be observed in human beings. That would be like our saying this of a species of animals (Wittgenstein 1980b §33).

One could imagine a concept of fear, for instance, that had application only to beasts, and therefore pertained only to behaviour (Wittgenstein 1980b §333).

The cases are not developed. But Wittgenstein seems to trace the fact that these terms pick out behaviour, or behavioural dispositions, rather than genuinely mental phenomena to the fact that all there is to their use is their application to others on the basis of their behaviour; they have no relation to any first-person use by the creatures to which the terms are applied. The lesson of that is that the third-person use of a term only counts as the use of a *mental* term if it is connected in appropriate ways with related first-person uses.<sup>18</sup> As before, the first-person and third-person uses are not distinct, and self-standing: they are parts of a unified whole. That is why Wittgenstein says that it is the ‘whole use’ of mental terms, encompassing both first-person and third-person uses, that is ‘the instrument’.

#### **4.i An objection to the Wittgensteinian view; and a reply**

We can anticipate the following objection to the Wittgensteinian view we have just described. ‘Wittgenstein suggests that the third-person use of mental terms is not independent of their first-person use: a term that we apply to a creature on the basis of its behaviour only picks out a mental phenomenon if the creature itself can make first-person uses of that term without reference to its behaviour. That may be true for concepts like *belief* and *intention*; it is quite plausible to think – with Wittgenstein, Davidson, McDowell, and others – that such concepts only properly apply to creatures that can think of themselves as accepting and rejecting things for reasons.<sup>19</sup> But it seems completely implausible to hold that it is a condition for the truth of the third-person ascription “That creature is in pain” that the creature should be capable of self-ascribing pain. It may be true that it is not until the word “thinking” finds its particular use in the first person that it acquires the meaning of mental activity. But it is surely not true that it is not until the word “pain” finds its particular use in the first person that it acquires the meaning of sensation.’

How should a defender of Wittgenstein respond to this objection? One possibility would be to bite the bullet, insist that the dependence of the third-person use of mental terms on their first-person use is a perfectly general feature of such terms, and accept that creatures that cannot self-ascribe pain cannot experience pain: or at least, that they cannot experience pain in exactly the same sense that we do.<sup>20</sup> It is possible that Wittgenstein would respond in that way. Consider the question: ‘Is the sensation we ascribe to a fox when we say that the fox feels pain the same as the sensation we ascribe to a person when we say that she feels pain?’ Wittgenstein would certainly allow that we could legitimately count the fox’s sensation as different from ours; it would not be wrong to classify sensations like that. However, the current proposal is not simply that that way of classifying sensations is one possible form of classification: the idea is, rather, that it is implicit in the ordinary notion of sensation. And it is far from clear that Wittgenstein would endorse that stronger view.

A different response – which seems closer to Wittgenstein – would involve distinguishing between the conditions of applicability of different mental concepts. On this view, concepts like *thought*, *belief*, and *intention* only fully apply to creatures that are capable of self-ascribing those concepts, as Wittgenstein insists. But for concepts like *pain* and *visual*

*experience*, there is no similar condition. There is a precedent in Wittgenstein’s discussion of experience for distinctions of this kind between the conditions of application of different mental concepts. He says there that there are some experiences that a person can only have if she has mastered a particular technique: for instance, a person looking at a triangle can only have the experience of seeing ‘now *this* as apex, *that* as base – now *this* as apex, *that* as base’ if she is ‘capable of making certain applications of the figure with facility’ (Wittgenstein 1953, part II xi §222 [p. 208]). He continues:

But how odd for this to be the logical condition of someone’s having such-and-such an *experience*! After all, you don’t say that one ‘has toothache’ only if one is capable of doing such-and-such. – From this it follows that we cannot be dealing with the same concept of experience here. It is a different concept, even though related (Wittgenstein 1953, part II xi §223 [p. 208]).

It seems plausible that Wittgenstein would draw a similar distinction in the present case: between concepts like thought, wishing, happiness, anger, and fear, which in his view apply only to creatures that are capable of self-ascribing them, and concepts like pain, for which there is no such requirement.

#### **4.ii Conclusion**

What should we conclude about Wittgenstein’s attitude to the question, whether mental terms mean the same in their first-person and third-person uses, and whether the mental properties we ascribe to ourselves without evidence are the same as the properties we ascribe to others on the basis of their behaviour?

In the first place, from Wittgenstein’s point of view these questions about the univocality of mental terms and the sameness of mental properties are perfectly legitimate. And they are to be answered by achieving a reflective understanding of our practice of ascribing mental properties to ourselves and others – rather than by reference to any supposedly more basic metaphysical facts about the sameness or difference of the properties our terms pick out.

The point of saying that we make a ‘decision’ when we answer these questions is that no particular answer is forced on us by the facts: it is possible to count properties in different ways; and no way of counting properties is simply wrong. Notwithstanding that point, however, the question ‘same property or different?’ does not in this case call for any active decision; it is so overwhelmingly natural to count one’s own pain and other people’s pain, say, as instances of the same property that it would be quixotic to say anything else.

Relatedly, the naturalness of counting mental properties in the way we do is not a merely psychological fact about us. And the ‘decision’ to count properties that way is not arbitrary. It has a deep rationale in the underlying character of our practice of applying mental terms to ourselves and others: specifically, in the interdependence of the first-person and third-person uses of those terms.

There remains a question in this area that Wittgenstein’s philosophy does not attempt to answer: how is it possible for a person reliably to self-ascribe without evidence properties that have the dispositional element that mental properties do; what explains that ability? Wittgenstein’s response to that question would be this. When we train people in the use of mental terms, they do in fact learn to make mental self-ascriptions, without reference to their behaviour, that cohere with the ascriptions that others can make on the basis of that behaviour. That we can learn to use mental terms in that way is a deep, contingent fact about us. But it is a fact that philosophy, at least, cannot explain. That, I have argued, is a further parallel between Wittgenstein’s views and Davidson’s.<sup>21</sup>

## Bibliography

- Child, W. 2013 ‘First-Person Authority’, in E. LePore and K. Ludwig eds., *A Companion to Donald Davidson*, Oxford: Wiley Blackwell, 533-49.
- Davidson, D. 1970 ‘Mental Events’, in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 1980.
- Davidson, D. 1984 ‘First Person Authority’, in Davidson 2001a.
- Davidson, D. 1986 ‘A Nice Derangement of Epitaphs’, in Davidson 2005.
- Davidson, D. 1987 ‘Problems in the Explanation of Action’, in D. Davidson, *Problems of Rationality*, Oxford: Oxford University Press, 2004.
- Davidson, D. 1988 ‘The Myth of the Subjective’, in Davidson 2001a.
- Davidson, D. 1990 ‘Epistemology Externalized’, in Davidson 2001a.
- Davidson, D. 1991 ‘Three Varieties of Knowledge’, in Davidson 2001a.
- Davidson, D. 1992 ‘The Second Person’, in Davidson 2001a.
- Davidson, D. 1993 ‘Reply to Bernhard Thöle’, in Stoecker 1993, 248-50.
- Davidson, D. 1994 ‘The Social Aspect of Language’, in Davidson 2005.
- Davidson, D. 1997 ‘The Emergence of Thought’, in Davidson 2001a.
- Davidson, D. 1999 ‘Reply to Anita Avramides’, in U. Žegleń ed., *Donald Davidson: Truth, Meaning and Knowledge*, London: Routledge, 154-6.
- Davidson, D. 2001a *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.
- Davidson, D. 2001b ‘Externalisms’, in P. Kotatko, P. Pagin, and G. Segal eds., *Interpreting Davidson*, Stanford: CSLI Publications, 1-16.
- Davidson, D. 2005 *Truth, Language, and History*. Oxford: Oxford University Press.
- Higginbotham, J. 1998 ‘On Knowing One’s Own Language’, in Wright, Smith, and Macdonald 1998, 429-41.
- Jacobsen, R. 2009 ‘Davidson and First-Person Authority: Parataxis and Self-Expression.’ *Pacific Philosophical Quarterly*, 90, 251-66.
- LePore, E. and Ludwig, K. 2005 *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford: Oxford University Press.
- McDowell, J. 1994 *Mind and World*. Cambridge, Mass.: Harvard University Press.
- Smith, B. 1998 ‘On Knowing One’s Own Language’, in Wright, Smith, and Macdonald 1998, 391-428.
- Stoecker, R. ed. 1993 *Reflecting Davidson: Donald Davidson Responding to an International Forum of Philosophers*. Berlin: de Gruyter.
- Thöle, B. 1993 ‘The Explanation of First Person Authority’, in Stoecker ed. 1993, 213-47.
- Wittgenstein, L. 1953 *Philosophical Investigations*, 4th edition, eds. P. M. S. Hacker and J. Schulte, trans. G. E. M. Anscombe, P. M. S. Hacker and J. Schulte, Oxford: Blackwell, 2009.
- Wittgenstein, L. 1980a *Remarks on the Philosophy of Psychology* Volume I, eds. G. E. M. Anscombe and G. H. von Wright, trans. G. E. M. Anscombe, Oxford: Blackwell.
- Wittgenstein, L. 1980b *Remarks on the Philosophy of Psychology* Volume II, eds. G. H. von Wright and H. Nyman, trans. C. Luckhardt and M. Aue, Oxford: Blackwell.
- Wittgenstein, L. 1982 *Last Writings on the Philosophy of Psychology* Volume I, eds. G. H. von Wright and H. Nyman, trans. C. Luckhardt and M. Aue, Oxford: Blackwell.
- Wittgenstein, L. 1992 *Last Writings on the Philosophy of Psychology* Volume II, eds. G. H. von Wright and H. Nyman, trans. C. Luckhardt and M. Aue, Oxford: Blackwell.
- Wright, C., Smith, B., and Macdonald, C., eds. 1998 *Knowing Our Own Minds*. Oxford: Oxford University Press.

## NOTES

<sup>1</sup> References to passages in *Philosophical Investigations* Part II are given in two forms: by the § numbers introduced in the 4th edition (in which Part II appears as ‘Philosophy of Psychology: A Fragment’), and by the page numbers of the 1st and 2nd editions, which are cited in square brackets.

<sup>2</sup> A notable exception is Davidson’s explicit endorsement of what he takes to be Wittgenstein’s reason for holding that thought and language are essentially social: that ‘unless a language is shared, there is no way to distinguish between using the language correctly and using it incorrectly; only communication with another can supply an objective check’ (Davidson 1991, 209-10). For other references to Wittgenstein in connection with that point, see e.g. Davidson 1992, 116; 1994, 124; 1997, 129; and 2001b *passim*.

<sup>3</sup> Davidson’s discussion moves fairly freely between talk of mental terms and concepts, on the one hand, and talk of mental states or properties, on the other. My discussion follows his in that respect.

<sup>4</sup> See Davidson 1987, 16 and Davidson 1991, 207 for other statements of the same objection. Davidson is characteristically cautious about explicitly attributing ‘the Wittgensteinian style of answer’ (Davidson 1987, 17) to Wittgenstein himself; he writes of ‘Wittgenstein’s insight (if it is Wittgenstein’s)’ (Davidson 1987, 16). Note also that Davidson 1999 acknowledges that there is more to Strawson’s view than Davidson’s original critique allowed.

<sup>5</sup> Davidson recognizes, of course, that this is an oversimplification. It is possible for someone to use the word ‘cat’ to mean *cat* without being able to distinguish cats from non-cats. Such a person will have no general disposition to apply the word to the cats in her environment and withhold it from the non-cats (see Davidson 1988, 48-9). But language and thought, Davidson insists, depend on there being a core of cases for which something close to the simple dispositional view is true. These are the cases ‘that anchor language to the world’ (Davidson 1988, 45; for other statements of the same idea, see Davidson 1987, 29; 1990, 197). And it is these cases that are the focus of his explanation of the univocality of, and asymmetry between, first-person and third-person ascriptions of content.

<sup>6</sup> For a full discussion of Davidson’s account of first-person authority, see Child 2013.

<sup>7</sup> See e.g. Davidson 1990, 202-3; 1991, 212-13; 1992, 117-21; 1997, 128-9; 2001b *passim*.

<sup>8</sup> I bracket any difficulties relating to the point just discussed in the text, about how we know that others are genuinely using language, rather than simply producing sounds in regular and consistent ways.

<sup>9</sup> See for example Jacobsen 2009, 255; LePore and Ludwig 2005, 355, 364-5; Smith 1998, 417; Thöle 1993, 243-4.

<sup>10</sup> See Higginbotham 1998, 432-6; Child 2013, 540.

<sup>11</sup> I have argued elsewhere that this account of our knowledge of the meanings of our own words does not in fact adequately explain the character of that knowledge (see Child 2013, 541-2). I will not repeat that criticism here.

<sup>12</sup> For other well-known statements of opposition to explanation in philosophy, see Wittgenstein 1953 §496 and §654.

<sup>13</sup> *Erklärung* is the word used in each of the passages quoted or cited above: Wittgenstein 1953 §§109, 126, 496 and 654.

<sup>14</sup> In his response to a paper by Anita Avramides about his views on this topic, Davidson comments on the kind of explanation at issue: ‘it seems to me . . . that, in practice, at least in the present context, there is no significant difference between description and explanation’ (Davidson 1999, 155).

<sup>15</sup> See Wittgenstein 1982 §§272-291 for related considerations about sameness and difference of meaning.

<sup>16</sup> For the first of these points, see also Wittgenstein 1982 §280: ‘If you say that [a word] has two meanings then you have to use an explanation to distinguish between them. (There can be various reasons for doing that.)’

<sup>17</sup> For the notion of decision in this context, see e.g. Wittgenstein 1982 §275: ‘You visit a tribe; they have a language; in this language you hear a word (a sound) – does it have one meaning, or several? How will you find out, *how will you decide?*’ (emphasis added).

<sup>18</sup> Notice the close similarity between, on the one hand, Wittgenstein’s distinction between the purely behavioural concepts of fear, belief, desire etc. and their fully mental counterparts and, on the other hand, Davidson’s distinction between the concepts we can apply to creatures that merely react to their environment in regular ways and the fully mental concepts that apply to genuine subjects of thought.

<sup>19</sup> We should note, however, that many other philosophers take this to be too demanding a requirement.

<sup>20</sup> For a position of that kind, see McDowell 1994, chapter 6.

<sup>21</sup> Earlier versions of this paper were presented at conferences and workshops at Birkbeck College, London; Keio University, Tokyo; the American University of Beirut; and York University, Toronto. I am grateful to the organizers of those events, in particular Claudine Verheggen; and to the participants, for much helpful discussion.