

## Meaning, Use, and Supervenience

William Child

### 1. Wittgenstein’s Anti-Reductionism

What is the relation between meaning and use? Wittgenstein says that ‘the meaning of a word is its use in the language’ (PI §43).<sup>1</sup> He makes a parallel claim about the sense of a proposition: ‘the use of a proposition – that is its sense’ (BT 80). But what sort of illumination are we supposed to derive from those ideas?

Consider a particular expression: the word ‘red’, for instance. Part of Wittgenstein’s point is that the word ‘red’ means what it does because we use it in the way we do. But the significance of that point depends on how we understand the notion of use. We can distinguish between a reductionist and an anti-reductionist view. The anti-reductionist thinks of the use of a word in a wholly quietist or pleonastic way. On this view, all we can say about how use determines meaning is this: the word ‘red’ means *red* because we use it to mean *red*; the words ‘add 2 each time’ mean *add two each time* because they ‘are used by us to mean that two is to be added each time’ (Stroud 2012, p. 27); and so on. As Barry Stroud puts it, a description of the use of an expression that ‘suffices to fix its meaning’ must itself ‘employ the idea of meaning’ (2012, p. 27). That is the view that Stroud both endorses and attributes to Wittgenstein.<sup>2</sup>

For the reductionist, by contrast, the point of the idea that meaning is use is to explain linguistic meaning in more basic terms. She agrees that we use the word ‘red’ to mean *red*. But she thinks we can spell out what is involved in using the word ‘red’ to mean *red* in a way that does not employ semantic concepts: in terms, for instance, of people’s dispositions to produce and respond to sounds or symbols containing ‘red’ in specified observable circumstances. That is a view that many readers have ascribed to Wittgenstein. According to Michael Dummett, for instance, when Wittgenstein describes the use of language,

what is described is the complex of activities with which the utterances of sentences are interwoven; and . . . the description does not invoke psychological or semantic concepts, but is couched entirely in terms of what is open to outward view (Dummett 1978, p. 446).

Paul Horwich agrees: Wittgenstein’s ‘examples of the meaning-constituting uses of words’, he writes, ‘are never couched in semantic or intentional terms’ (Horwich 2012, p. 112).

My own view is that Wittgenstein is an anti-reductionist about meaning and intentional content. And I think Wittgenstein is right; facts about meaning and content cannot be constructed from or reduced to facts about use characterized in wholly non-semantic, non-intentional terms. But Wittgenstein does not adopt the most flat-footed, uncompromisingly anti-reductionist position on these matters. For, though he insists that meaning cannot be explained or accounted for in other terms, he does think that there are interesting and non-pleonastic things to say about what it takes for an expression to be used with a particular meaning, including things about the relation between a word’s meaning what it does and facts about its use, characterized in non-semantic terms. That strand in his thinking emerges in many passages. I will give two examples.

Wittgenstein writes:

Let us consider very simple rules. Let the expression be a figure, say this one:

|—|

and one follows the rule by drawing a straight sequence of such figures (perhaps as an ornament).

|--| |--| |--| |--| |--|

Under what circumstances should we say: someone gives a rule by writing down such a figure? Under what circumstances: someone is following this rule when he draws that sequence? It is difficult to describe this.

If one of a pair of chimpanzees once scratched the figure |--| in the earth and thereupon the other the series |--| |--| etc., the first would not have given a rule nor would the other be following it, whatever else went on at the same time in the minds of the two of them.

If however there were observed, e.g., the phenomenon of a kind of instruction, of showing how and of imitation, of lucky and misfiring attempts, of reward and punishment and the like; if at length the one who had been so trained put figures which he had never seen before one after another in sequence as in the first example, then we should probably say that the one chimpanzee was writing rules down, and the other following them (RFM VI §42).

That passage illustrates a point that Wittgenstein makes a few remarks earlier: ‘What, in a complicated surrounding, we call “following a rule” we should certainly not call that if it stood in isolation’ (RFM VI §33). What each chimpanzee does can be described in terms that do not presuppose that a rule is being given and followed; the first chimpanzee scratches the figure |--| in the earth and the second scratches the series |--| |--| etc. If that is just a one-off occurrence, no rule has been given or followed. When it happens in the right kind of context, we do have a case of giving and following a rule. But what is the right kind of context? What exactly are the circumstances under which we should say that a rule is being given and followed? The most uncompromisingly anti-reductionist response to that question would be simply to say this: that the circumstances under which we should say that a rule is being given and followed are those in which people (or chimpanzees) are participating in a practice of giving and following rules or, more simply, in which a rule is indeed being given and followed. But in the passage I have quoted, and others like it, Wittgenstein does not give that unhelpful answer. Instead, he tries to say something genuinely informative and non-pleonastic about what it takes for there to be a custom of giving and following such rules, and what it takes for two people (or chimpanzees) to be participants in such a practice.<sup>3</sup> We can highlight two points.

First, Wittgenstein thinks that the existence of a custom or practice of following rules requires there to be a whole pattern of rule-involving activity. Some of the activities he mentions in RFM VI §42 are specific to the situation of learning: instruction, showing how, imitation, etc. Others are more general: reward and punishment, for instance. But the most basic feature of rule-following, which is implicit in all the activities Wittgenstein mentions, is that giving or following rules involves *acknowledging* or *understanding* actions as being correct or incorrect. The idea of acknowledging an action as correct or incorrect is no more basic than the idea of following a rule itself: to the extent that it is unclear whether two chimpanzees are really giving and following rules (as opposed to merely acting in a regular way), it will be equally unclear whether they are really acknowledging their actions as being correct or incorrect; and vice versa.<sup>4</sup> As before, there is no prospect of a reductive account of rule-following. Nonetheless, it is philosophically illuminating to set out the kind of structure of holistically related activities that provides the necessary background for something to count as an instance of rule-following.

Second, it is a characteristic feature of Wittgenstein’s work to approach the task of saying something substantial about the circumstances in which an interaction counts as a case of a rule being given and followed by focusing on the procedure of someone’s *learning* to follow a rule.<sup>5</sup> In describing the process of learning, we describe a process in which there is a transition from the learner’s doing things that do not yet involve following a rule (making marks, copying the teacher, etc.) to her engaging in rule-following activity. In giving such a description we are not explaining what rule-following consists in in non-rule-involving terms. Nonetheless, we are charting one kind of link between non-rule-involving facts and facts about following rules. And that, Wittgenstein

suggests, is as close as we can get to giving an informative account in non-rule-involving terms of what rule-following consists in; ‘we can go no further’ (RFM VII §26).

Now consider a second passage, in which Wittgenstein insists that the existence of a practice of multiplying requires the existence of lower-level regularities in linguistic behaviour.

What if we said that mathematical propositions were prophecies in *this* sense: they predict what result members of a society who have learnt this technique will get in agreement with other members of the society? ‘ $25 \times 25 = 625$ ’ would thus mean that men, if we judge them to obey the rules of multiplication, will reach the result 625 when they multiply  $25 \times 25$ . -

That this is a correct prediction is beyond doubt; and also that calculating is in essence founded on such predictions. That is to say, we should not call something ‘calculating’ if we could not make such a prophecy with certainty. This really means: calculating is a technique. And what we had said pertains to the essence of a technique. (RFM III §66. For related comments, see e.g. RFM VI §23.)

We could put the lesson of that passage like this. Mathematical propositions are not prophecies, in any sense. Nonetheless, take a group of people whom we judge to have learned to multiply and to obey the rules of multiplication. Now say to them, ‘What is  $25 \times 25$ ?’ We can predict that they will, by and large, respond by saying ‘625’. That is an empirical prediction about their behaviour, non-semanticly characterized. And it is a correct prediction; people who have been through a certain training and whom we judge to have mastered the rules of multiplication do generally respond in that way in these circumstances. Wittgenstein’s point is that the existence of a practice of multiplication in a community depends upon the obtaining of patterns of non-semantic facts in virtue of which such prophecies, if someone chose to make them, would be true.<sup>6</sup> If we could not make such prophecies ‘with certainty’, the people in question would not be calculating. In this particular case, Wittgenstein is talking about the rules of multiplication and the technique of calculating. But he would make the same points about rules and meaning in general.

## **2. Supervenience**

It is natural to characterize the picture that emerges from passages like those discussed in Section 1 in terms of supervenience. Facts about meaning cannot be reduced to, or explained in terms of, non-semantic facts about use. But facts about meaning do supervene on non-semantic facts about use. That is to say, two worlds cannot differ in a semantic respect without differing in some non-semantic respect. Equivalently, if two worlds are alike in all non-semantic respects, they must be alike in all semantic respects. In the rest of this essay I shall defend this supervenience thesis against some significant criticisms. My primary goal is to defend the supervenience thesis itself. But I shall also argue that the position I defend is consistent with Wittgenstein’s views.

The thesis that semantic facts supervene on non-semantic facts about use can be understood in many different ways. I shall follow Stephen Kearns and Ofra Magidor in spelling out the target thesis in the following way.<sup>7</sup> First, the *semantic facts* that are claimed to supervene on non-semantic use facts are to be understood as inclusively as possible, as encompassing both facts about the meanings of words (‘The English word ‘plus’ refers to the *plus* function’; ‘Jane said that the rose was red’) and facts about the contents of thoughts and attitudes (‘Jane thought that Brexit was a terrible idea’, ‘Jim intended to buy his son a present’). Second, the *non-semantic facts about use* that make up the supervenience base are also to be understood as widely as possible. In particular, the facts about an expression’s use are to include the following. (1) Facts about the community’s use of the expression, and not simply about any particular individual’s use. (2) Facts about the physical environment in which the expression is used. That accommodates semantic externalism’s insight that the meanings of many expressions are determined in part by which natural kinds and which individual things are present in speakers’ environments.<sup>8</sup> (3) Facts, if there are such facts, about the relative naturalness of the properties and individuals to which expressions refer. As Kearns and Magidor note, a common picture in contemporary philosophy is this. Speakers’ linguistic dispositions, taken by themselves, are compatible with numerous different assignments of meaning.

Nonetheless, the English word ‘plus’ refers to the *plus* function rather than to any quus-like alternative, the word ‘green’ refers to the property of being green rather than to any grue-like alternative, and so on. And what makes ‘plus’ mean what it does is in part that the *plus* function is objectively more natural than any such alternative, and hence more eligible to be picked out by our word. Similarly for the word ‘green’ and the property of being green, and so on.<sup>9</sup>

There is room for debate as to how far this conception of the non-semantic facts on which semantic facts supervene is consistent with Wittgenstein’s views. I cannot resolve that question here. But I will make two observations. First (concerning point (2) above), some readers of Wittgenstein think that semantic externalism in the style of Kripke and Putnam is incompatible with Wittgenstein’s views, because it severs the essential connection between the meanings of people’s words and their own use of those words.<sup>10</sup> I have argued elsewhere that that is a mistake; there is nothing in Wittgenstein that conflicts with the idea that we may use terms to refer to natural kinds in the way described by Kripke and Putnam.<sup>11</sup> Second (concerning point (3)), the idea that objective naturalness plays any role in determining meaning will seem to many readers to be directly opposed to Wittgenstein’s views. In my view, however, the issue is not clear cut. I have argued elsewhere that Wittgenstein gives an important role in the determination of meaning to a notion of naturalness that is not constrained by the limits of our actual classificatory capacities, and thus that there is less distance between Wittgenstein’s position and the ‘natural-properties’ view than is generally assumed.<sup>12</sup> But whether or not I am right about that, the important point for present purposes is a conditional one; if facts about the naturalness of properties and individuals *do* play a part in fixing the meanings of expressions, then the defender of semantic supervenience should count those facts as part of the non-semantic supervenience base.

The thesis I want to defend, then, is this. Semantic facts, taken to include facts about both linguistic meaning and intentional content, supervene on non-semantic facts, taken to include facts about the community’s use of words and concepts, facts about the physical environment, and facts (if such there be) about the naturalness of properties and individuals.

### **3. Supervenience and Dispositional Properties**

In ‘Wittgenstein on Following a Rule’, John McDowell presents an anti-reductionist view of meaning and rule-following in terms of a distinction between two levels at which we can describe linguistic behaviour. On the one hand there are descriptions at (or above) ‘bedrock’: descriptions in which language-use is characterized in semantic terms (‘She said that the post-box was red’, ‘He said that 1000 plus 2 equals 1002’). That, says McDowell, is the ‘deepest level at which we can sensibly contemplate the place of language in the world’ (McDowell 1984, p. 341). On the other hand, there are ‘sub-bedrock’ descriptions, in which speakers’ behaviour is characterized without reference to meaning or rules (‘She made the sound “The post-box is red”’, ‘He put “1002” after “1000”’). There is, McDowell says, ‘an intimate relation’ (1984, p. 349) between facts about rule-following and patterns of sub-bedrock facts such that ‘a certain disorderliness below “bedrock” would undermine the applicability of the notion of rule-following’ (1984, p. 349). But ‘recognizing the intimate relation must not be allowed to obscure the difference of levels’ (1984, p. 349); no account of language in sub-bedrock terms can capture facts about meaning and rules. That is the kernel of McDowell’s anti-reductionism.

McDowell briefly considers the idea that ‘statements about rule-following *supervene*, in Wittgenstein’s view, on sub-“bedrock” statements’. He comments: ‘There may be an acceptable interpretation of this; but on the most natural interpretation, it would make statements about rule-following vulnerable to future loss of mutual intelligibility’ (1984, p. 362, n. 43), in a way that would falsify the epistemology of meaning and fail to accommodate common-sense truths. So, he suggests, the ‘intimate relation’ between bedrock and sub-bedrock facts is not to be captured in terms of supervenience. I agree with McDowell that the particular version of semantic supervenience he has in mind is unacceptable. But working through his objections will point us to a

better version of the supervenience thesis, as well as preparing the ground for a response to further objections that we will meet below, in section 4.

McDowell associates the supervenience thesis with a particular picture of the epistemology of meaning. In that picture, what we detect when someone speaks are in the first instance non-semantic facts about the sounds he makes and the circumstances in which he makes them. Knowledge of what someone means by an expression depends on an inference from such non-semantic facts about his use of the expression and involves hypotheses about his future use, non-semantically characterized. So, when I claim to understand someone, ‘I bind myself to a prediction of the uses of language he will make in various possible future circumstances, with these uses characterized in sub-“bedrock” terms’ (1984, p. 349). The consequence of that picture, McDowell objects, is that any claim to know what someone else means by an expression is indefinitely vulnerable to the possibility of an unfavourable future. Below ‘bedrock’ there is nothing but contingency; so at any time in the future my interlocutor’s use of the expression in question may simply stop conforming to the pattern I expect. And that would retrospectively undermine my present claim to be able to vouch for the character of his understanding (1984, p. 348).

But that, McDowell thinks, is a mistake. It is true that mutual understanding rests on a ‘tissue of contingencies’ (1984, p. 349). And it is true that those contingencies might break down; an interlocutor who currently uses a word in a way I seem to understand might go on to use it in a bizarre and unexpected way that I did not understand. If that were to happen, however, the fact that I did not then understand her would not entail that I do not currently know what she means. Common sense distinguishes two different possibilities.

- (a) I know what she currently means by the word. Then her use of the word changes. I do not know what if anything she means by the word after the change.
- (b) I do not know what she currently means by the word, though I initially appear to understand her. Her use of the word is consistent over time. What emerges from my failure to understand her later uses of the word is that I did not understand what she meant by it earlier.

The thesis that semantic facts supervene on sub-bedrock facts, McDowell suggests, represents every case in which my interlocutor’s use of an expression stops ‘conforming to the pattern I expect’ as a case of type (b); it implies that a current claim to know what someone means will always be undermined if her subsequent use of the expression diverges from the pattern I expected. But that overlooks the possibility of cases of type (a).

McDowell is right that we must allow for the existence of both kinds of case: (a) and (b). But the defender of supervenience can do that perfectly well. In criticizing supervenience, what McDowell has in mind is the thesis that what a person means by an expression at a given time supervenes on non-semantic facts about her actual applications of the expression at that and subsequent times. But it is natural to offer this different formulation: what a person means by an expression at a given time supervenes on non-semantic facts about her use of the expression *at that time*, including facts about how she is then disposed to use it. We can explore this latter formulation in connection with a different example of Wittgenstein’s.

Wittgenstein writes:

Let us imagine the following example: A writes down series of numbers; B watches him and tries to find a rule for the number series. If he succeeds, he exclaims: ‘Now I can go on!’ (PI §151)

He considers a particular case: ‘A has written down the numbers 1, 5, 11, 19, 29; at this point B says he knows how to go on’ (PI §151). The discussion continues:

Suppose B says he knows how to go on – but when he wants to go on, he hesitates and can’t do it. Are we then to say that it was wrong of him to say he could go on; or rather, that he was able to go on then, only now is not? – Clearly, we shall say different things in different cases. (PI §181. See PI §323 and BB 115-16 for further discussion.)

It seems clear that the truth or falsity of the claim ‘Now I can go on!’ does not supervene on facts about what actually happens when the speaker does in fact try to go on. If it did, we would lose the common-sense distinction between someone (call her Able) who was able to go on when she said could but cannot now and the person (call her Unable) who was never able to go on but only thought she could. But that does not mean we should abandon the idea that semantic facts supervene on non-semantic facts. It just means that we need to specify the right non-semantic supervenience base. And the obvious thought is that the non-semantic facts on which the truth or falsity of the claim ‘Now I can go on!’ supervene include dispositional facts. At the time when she spoke, Able was disposed, *ceteris paribus*, to put ‘41, 55, 71, 89, 109, . . .’ after ‘1, 5, 11, 19, 29 . . .’. Unable was not.

There are well-known difficulties in any attempt to spell out dispositional conditions for the truth or assertability of individual claims about meaning and rule-following, taken one by one.<sup>13</sup> For one thing, what someone who can develop a particular series would do if she were to attempt to develop it depends not just on her knowledge of the series but also on her desire to develop it correctly, the strength of any competing desires, her belief about what stage of the series she has reached, and so on. For another thing, even if she tries to develop the series correctly, she may make a mistake. And so on. So there is no question of pairing individual semantic facts about a person with facts about individual dispositions, characterized non-semantically. If it is to be plausible, the supervenience thesis must be that the semantic facts about a person, as a whole, supervene on non-semantic facts about her, including dispositional facts, as a whole.

It is tempting to think that differences in non-semantic dispositions must in turn supervene on something more basic: on differences in underlying physical states. After all, if two people differ in some dispositional respect, there will be possible circumstances in which one person would do something that the other would not do. We naturally think that there must be some causal explanation of any such difference in behaviour and that the explanation must ultimately come down to some difference in underlying brain states.<sup>14</sup> So, it is tempting to argue, semantic facts ultimately supervene on occurrent physical facts. But the defender of semantic supervenience need not accept this argument. And if dispositional properties can be fundamental features of things, she should certainly not accept it.

Wittgenstein, for one, explicitly accepts that a person’s dispositional properties need not supervene on her occurrent brain states; dispositional properties may be fundamental.

It is . . . perfectly possible that certain psychological phenomena *cannot* be investigated physiologically, because physiologically nothing corresponds to them.

I saw this man years ago: now I have seen him again, I recognize him, I remember his name. And why does there have to be a cause of this remembering in my nervous system? Why must something or other, whatever it may be, be stored up there *in any form*? Why *must* a trace have been left behind? Why should there not be a psychological regularity to which *no* physiological regularity corresponds? If this upsets our concepts of causality then it is high time that they were upset.

. . .

Why should there not be a natural law connecting a starting and a finishing state of a system, but not covering the intermediary state? (Only one must not think of *causal efficacy*.) (Z §§609-10, 613)

Applied to our case, the possibility that Wittgenstein is envisaging is this. Able and Unable both say ‘Now I can go on’ but cannot continue the series when they try to do so. What Able said was true; she really could go on when she said she could. What Unable said was false; she thought she could go on but she couldn’t. There is a dispositional difference between them: at the time when they spoke, Able was disposed *ceteris paribus* to continue the series correctly; Unable was not. But there is no further physical difference underlying and explaining this dispositional difference. It is just a brute fact that Able’s exposure to the initial steps in the series left her with the disposition to

develop the series in that way, while Unable’s exposure to the same thing left her with no such disposition.

The point I want to stress here is that the possibility we have just described is entirely consistent with the idea that semantic properties supervene on non-semantic properties. There is (in our extended sense) a semantic difference between Able and Unable: Able understands the series and knows how to continue it; Unable does not. That semantic difference is underpinned by a non-semantic difference; there is a difference between what Able is disposed to do and say, characterized non-semantically, and what Unable is disposed to do and say, similarly characterized. The dispositional difference is fundamental; it does not supervene on any further, more basic, physical difference between Able and Unable. But to make that point is not to reject the thesis that semantic facts supervene on non-semantic facts. It is just to acknowledge that the non-semantic facts on which semantic facts supervene may be, or include, irreducibly dispositional facts.

With that in mind, we can return to the distinction McDowell draws between two cases in which someone’s future use of an expression confounds our expectations: case (a) where we now understand our interlocutor but no longer understand her in future; and case (b) where we never understand what she means by the expression, though we initially seemed to understand. As I said above, I agree that an adequate account of meaning must make room for that distinction. And I agree that the specific version of supervenience that McDowell has in mind does not make room for it. But a different version can accommodate the distinction perfectly well, treating it as we treated the distinction between Able and Unable. There is a semantic difference between the two cases; the speaker is using the expression with one meaning in case (a) and a different meaning in case (b). That semantic difference does not supervene on any difference in the non-semantic facts about her actual past or future applications of the expression; those are the same in both cases. But there is a non-semantic difference between the two cases: a difference at the level of the two speakers’ non-semantic dispositions. In case (a) she is currently disposed, *ceteris paribus*, to use the expression in one way, non-semantically characterized. In case (b) she is currently disposed to use it in a different way. As before, we naturally expect the dispositional difference between the two cases to supervene on a difference at the level of underlying physical states. As before, however, there is no reason why things must work like that; the difference in non-semantic dispositions could be a brute difference. But even if the dispositional difference is a brute difference, that does not threaten the thesis of semantic supervenience. For there will still be a non-semantic difference between the two cases, (a) and (b); it will simply be an ineliminably dispositional difference. And, as we saw in the case of Able and Unable, that is consistent with the thesis that semantic facts supervene on non-semantic facts. We will return to the lessons of these cases in the next section.

McDowell is certainly right to reject the inferential picture of knowledge of another person’s meaning that he associates with the claim that semantic facts supervene on non-semantic, ‘sub-bedrock’, facts about use. But that is consistent with accepting the supervenience thesis itself, which is a claim about the metaphysics of meaning and rule-following, not a claim about our knowledge of meaning and rule-following. Of course, an acceptable account of the metaphysics of meaning has to be consistent with a plausible account of the epistemology of meaning. But nothing in the thesis of semantic supervenience, as I have presented it, conflicts with that requirement.

#### **4. Intentional Ghosts and Semantic Magic**

In a recent paper, Kearns and Magidor offer a series of arguments against the thesis of semantic supervenience and in favour of ‘semantic sovereignty: the thesis that semantic facts do not supervene on use facts’ (Kearns and Magidor 2012, p. 322). Their arguments fall into two general kinds. First, they offer a range of counterexamples which, they argue, show that the thesis of semantic supervenience is false. Second, they consider a family of familiar arguments, due to Quine, Putnam, and Kripke’s Wittgenstein, that ‘purport to show that use facts are insufficient to determine semantic facts’ (2012, p. 335)<sup>15</sup>. They argue for the plausibility of a ‘neglected response’ to such

arguments: that semantic facts are a fundamental and irreducible feature of reality and thus that the fact (if it is a fact) that use facts underdetermine semantic facts is no threat to the reality of semantic facts.<sup>16</sup> Though these more general considerations certainly deserve consideration, I shall focus here on Kearns and Magidor’s argument from counterexamples.<sup>17</sup>

Kearns and Magidor start with a series of counterexamples that are targeted at the claim that semantic facts supervene on *physical* facts (which, they say, is the ‘commonly accepted picture’ (2012, p. 324)), rather than at the more general claim that semantic facts supervene on non-semantic facts about *use*. These initial counterexamples involve pairs of worlds,  $w_1$  and  $w_2$ , such that  $w_2$  is a physical duplicate of  $w_1$  but also contains non-physical subjects (‘ghosts’), or non-physical properties, that are absent from  $w_1$ . Kearns and Magidor argue that these non-physical differences between  $w_1$  and  $w_2$  make room for words in  $w_2$  to have different semantic properties from their counterparts in  $w_1$ . For instance, a word that has one meaning in  $w_1$  may have a different meaning in  $w_2$  because of the way that ghosts use the word in  $w_2$ , even though all the physical facts about its use by physical language-users are the same in both worlds (2012, pp. 327, 328-9). Or a word that refers to a physical thing or property in  $w_1$  may refer to a non-physical thing or property in  $w_2$  (2012, pp. 329-30). So physically duplicate worlds can differ in their semantic properties; semantic facts do not supervene on physical facts.

Examples of this kind may challenge the thesis that semantic facts supervene on physical facts. But, as Kearns and Magidor acknowledge, they do not by themselves threaten the thesis that semantic facts supervene on use facts (2012, pp. 339-40). For in these cases the differences between the semantic facts in  $w_1$  and in  $w_2$  are explained by differences in the use of words between  $w_1$  and  $w_2$ ; it is just that, in the presence of non-physical language-users and non-physical properties, there can be differences in use between two worlds without those worlds differing in any physical respect. For present purposes, then, we must concentrate on those of Kearns and Magidor’s counterexamples that specifically target the thesis that semantic facts supervene on use facts.

The first of these more specific counterexamples appeals to the possible existence of ‘purely semantic (and in particular, purely intentional) entities’: ‘purely intentional ghosts’, in their terminology. Purely intentional ghosts, they say, ‘have various mental states (beliefs, desires, etc.) but lack any (interesting) non-semantic properties. Such ghosts seem readily conceivable. Indeed, they seem to be what are normally called Cartesian minds’ (2012, p. 340). Suppose that purely intentional ghosts are indeed possible. And suppose that  $w_2$  is a non-semantic duplicate of  $w_1$  but contains in addition a number of purely intentional ghosts, which are absent from  $w_1$ . That creates an immediate challenge to the thesis of semantic supervenience. For, though  $w_2$  does not differ from  $w_1$  in any non-semantic respect, it will differ semantically from  $w_1$ : most obviously, by containing the thoughts and attitudes possessed by these purely intentional ghosts; but also, because the existence of the ghosts and their thoughts will affect the reference and truth-value of some of the thoughts and utterances of ordinary subjects.<sup>18</sup> How should a defender of semantic supervenience respond to this kind of example?

In the first place, when Kearns and Magidor offer counterexamples that appeal to the possible existence of purely intentional ghosts, they are not really offering an *argument* for the falsity of semantic supervenience; they are in effect simply asserting that the supervenience claim is false. For the supposition that there can be subjects that have intentional (and thus, semantic) properties without having any non-semantic properties *just is* the supposition that semantic properties do not supervene on non-semantic properties. If there is an argument here, it is one that proceeds from the conceivability of purely intentional ghosts to their possibility.<sup>19</sup> Such an argument invites one of the standard responses to arguments of this kind; either purely intentional ghosts are not really conceivable or else they are conceivable, but their conceivability does not entail their possible existence.

However, even if Kearns and Magidor have given us no positive reason to think that purely intentional ghosts are possible, they might say that the defender of supervenience, for her part, has given us no reason to think that purely intentional ghosts are not possible. The onus, they might

argue, is on the defender of supervenience to give us some positive reason to think, first, that every possessor of semantic properties must have some non-semantic properties and, second, that its semantic properties must supervene on its non-semantic properties. That is a legitimate challenge. What positive reason is there to think that the thesis of semantic supervenience is true?

There is a rich history of philosophical argument against the intelligibility of Cartesian subjects, or purely intentional ghosts. There are arguments, originating with Kant and Strawson, that contend that we can only make sense of the existence of individual subjects if they are substantial entities that possess non-intentional as well as intentional properties.<sup>20</sup> And there are arguments, originating with Wittgenstein, that contend that we cannot make sense of the existence of mental phenomena in the purely first-person way that is all that remains if we try to think away the embodiment and behavioural expression of the mental.<sup>21</sup> I think those arguments make a compelling case against the conceivability of Cartesian subjects. But not every philosopher finds them convincing. Can we appeal to arguments of a different kind to support the thesis of semantic supervenience?

One obvious suggestion is that we could argue for semantic supervenience on causal grounds, in a way that mirrors a standard argument for the supervenience of mental properties on physical properties. In outline, the argument would go like this. (i) Semantic phenomena play a causal role in producing and explaining other phenomena. For instance, my turning right is causally explained by your saying ‘Turn right’. And the fact that ‘Turn right’ means what it does plays a role in that explanation; if ‘Turn right’ had meant *turn left*, your saying what you did would not have caused the effect it did. (ii) Every phenomenon has a complete non-semantic cause. (iii) The effects of semantic causes are not overdetermined by their semantic and non-semantic causes. So (iv) Semantic phenomena must supervene on non-semantic phenomena. Despite the popularity of that form of argument, however, it seems dialectically ineffective to appeal to that line of thought in the current debate with Kearns and Magidor.

In the first place, if Kearns and Magidor are right that there could be purely intentional ghosts, that will immediately undercut the causal argument for supervenience. For, in a world containing purely intentional ghosts, it would not be true that every phenomenon had a complete non-semantic cause; in particular, the actions of purely intentional ghosts would not have complete non-semantic causes. In the second place, even without appealing to the alleged possibility of purely intentional ghosts, the opponent of supervenience may simply deny that every phenomenon must have a complete non-semantic cause. It is perfectly possible, she may argue, for two things to have different effects, in virtue of having different semantic properties, without differing in any non-semantic respect. So, she will say, we should reject premise (ii) of the causal argument; it is not true that every phenomenon must have a complete non-semantic cause. And that undermines the causal argument for supervenience. It is no coincidence that Kearns and Magidor’s second series of counterexamples to semantic supervenience take precisely this form (2012, pp. 341-2).

Consider two worlds,  $w_1$  and  $w_2$ , that are exactly alike in every non-semantic respect. They are also alike in every semantic respect, with one exception. In both worlds, the word ‘cat’ refers to cats. In both worlds, the word ‘cat’ has never in fact been uttered by anyone to refer to cats.<sup>22</sup> But  $w_1$  and  $w_2$  differ in the following respect. In  $w_1$ , if someone were to utter the word ‘cat’, thereby referring to cats, their doing so would have no special effect. But in  $w_2$ , if someone uttered ‘cat’ to refer to cats, their doing so would have the effect of magically turning their interlocutor into a unicorn. Since  $w_1$  and  $w_2$  are alike in all non-semantic respects, there would be no natural, non-semantic causal explanation of that effect; that is the point of describing the effect as magical. And, Kearns and Magidor stress, the magical effect really would be causally explained by the semantic properties of the action of uttering the word ‘cat’ to refer to cats; merely making the sound ‘cat’, rather than uttering the word ‘cat’ to refer to cats, would have no such effect. (I will return to this point shortly.) Kearns and Magidor then argue as follows. The difference between  $w_1$  and  $w_2$  is a semantic difference. But  $w_1$  and  $w_2$  are non-semantic duplicates. So worlds can differ semantically without differing non-semantically; semantic facts do not supervene on non-semantic facts.<sup>23</sup>

Is this argument convincing? I agree with Kearns and Magidor that there is a semantic difference between  $w_1$  and  $w_2$  in their example. The semantic action of uttering the word ‘cat’ to refer to cats would have different causal consequences in  $w_1$  and  $w_2$ ; and that is itself a semantic difference between the two worlds. But, I shall argue, when we reflect on the causal differences between  $w_1$  and  $w_2$ , it is clear that, *pace* Kearns and Magidor, there is also a non-semantic difference between the two worlds. So the case is not an effective counterexample to the thesis of semantic supervenience.

The basic structure of Kearns and Magidor’s case is this. The semantic difference between  $w_1$  and  $w_2$  is a difference in dispositional properties: a difference in what would happen if someone were to utter ‘cat’ to refer to cats. And the difference in dispositional properties, or causal powers, is fundamental: it does not supervene on any underlying difference in occurrent non-semantic properties;  $w_1$  and  $w_2$  are alike with respect to all such non-semantic properties. Now that, of course, is exactly the structure that we saw in the previous section in the case of Able and Unable. There is a semantic difference between Able and Unable. The semantic difference involves a difference in dispositional properties: at the point when they both said ‘Now I can go on’, Able understood the series and would *ceteris paribus* have continued it correctly had she tried to do so; Unable did not understand the series and would not have continued it correctly. And, we supposed, the dispositional difference does not supervene on any underlying difference in occurrent non-semantic properties; Able and Unable are alike with respect to all such properties. Nonetheless, we said, the case does not threaten the supervenience thesis. For there is a non-semantic difference between Able and Unable; they have different non-semantic dispositional properties. In particular, Able is disposed, *ceteris paribus*, to put ‘41, 55, 71, 89, 109, . . .’ after ‘1, 5, 11, 19, 29 . . .’; Unable is not.<sup>24</sup>

We can treat the case of words whose utterance would have magic effects in  $w_2$  but not  $w_1$  in essentially the same way. In particular, even though  $w_1$  and  $w_2$  are alike with respect to all occurrent non-semantic properties, they differ with respect to their non-semantic dispositional properties.

In order for me to utter the word ‘cat’ to refer to cats, two conditions must be satisfied. (i) I must make the sound ‘cat’; and (ii) I must do so in a context in which making the sound ‘cat’ counts as uttering the word ‘cat’ to refer to cats. With that in mind, consider the two worlds,  $w_1$  and  $w_2$ , in the ‘cat’ example. In both worlds, the context is such that, if I were to make the sound ‘cat’, I would count as uttering the word ‘cat’ to refer to cats. Given the context, therefore, my making the sound ‘cat’ would have different effects in  $w_1$  and in  $w_2$ . But then there is, after all, a non-semantic difference between  $w_1$  and  $w_2$ : holding context fixed, my making the sound ‘cat’ in  $w_2$  would turn someone into a unicorn; doing the same thing in  $w_1$  would have no such effect. That is a difference at the level of dispositional properties, non-semantically characterized. There is, *ex hypothesi*, no further non-semantic difference between  $w_1$  and  $w_2$  that *explains why* making the sound ‘cat’, in this context, would have such different effects in the two worlds. It is just a brute fact that it would. It remains the case, however, that there is a non-semantic difference between the two worlds:  $w_1$  and  $w_2$  are not non-semantic duplicates. So the case of words with magical properties is not an effective counterexample to the thesis of semantic supervenience.

It might be objected that this argument overlooks Kearns and Magidor’s specification that what causes my interlocutor to turn into a unicorn in  $w_2$  is not my making the sound ‘cat’ but my uttering the word ‘cat’ to refer to cats. I have argued that there is a non-semantic difference between  $w_1$  and  $w_2$  at the level of dispositional properties, on the grounds that making the sound ‘cat’ in  $w_2$  would cause someone to turn into a unicorn but doing the same thing in  $w_1$  would have no such effect. But, the objector will protest, that is not true. Making the sound ‘cat’ would have no special effect in either world; what does the causal work in  $w_2$  is uttering the word ‘cat’ to refer to cats. So, contrary to what I have claimed,  $w_1$  and  $w_2$  really are alike with respect to all non-semantic properties, both occurrent and dispositional.

But the idea that what turns my interlocutor into a unicorn in  $w_2$  is not my making the sound ‘cat’ but my uttering the word ‘cat’ to refer to cats needs to be handled with care. The context we are considering is one in which, in both  $w_1$  and  $w_2$ , my making the sound ‘cat’ counts as uttering the word ‘cat’ to refer to cats. In that context, my action of making the sound ‘cat’ *just is* an action of uttering the word ‘cat’ to refer to cats. So it cannot be true that what causes the magical effect in  $w_2$  is my uttering the word ‘cat’ to refer to cats *rather than* my making the sound ‘cat’. The point the objector is reaching for can be put in terms of counterfactuals: if I were in a different context, in which my making the sound ‘cat’ did not count as my uttering the word ‘cat’ to refer to cats, then making the sound ‘cat’ would have no magical effect. That is true. But it is irrelevant to the current argument. The crucial question is whether there is a relevant non-semantic difference between  $w_1$  and  $w_2$  in the case in which making the sound ‘cat’ *would* count as uttering the word ‘cat’ to refer to cats, and *would* therefore produce different effects in  $w_1$  and  $w_2$ . And, as I have argued, there clearly is such a difference: given the context, there is a difference in what would happen in  $w_1$  and  $w_2$  if I were to make the sound ‘cat’. So the semantic difference between  $w_1$  and  $w_2$  is underpinned by a non-semantic difference between them. The case does not challenge the thesis of semantic supervenience.

## **5. Conclusion**

We have rejected two arguments against the thesis that semantic facts supervene on non-semantic facts about use. McDowell, I argued, was right to reject the particular version of the supervenience thesis that he had in mind. But a better version, which includes non-semantic dispositions in the supervenience base, is not vulnerable to McDowell’s criticism. Kearns and Magidor offered two kinds of counterexample to semantic supervenience. The first depended on the possible existence of purely intentional subjects. But to suppose that purely intentional ghosts are possible, I said, is in effect simply to suppose that semantic supervenience is false; it does not constitute an argument against the supervenience thesis. And, I suggested, there are good reasons for thinking that purely intentional ghosts are not possible, though I have not defended the arguments here. Kearns and Magidor’s second kind of counterexample involved the possibility of semantic phenomena having magical effects: effects that cannot be causally explained in non-semantic terms. I agreed that that kind of case is possible. But, I argued, that possibility is consistent with semantic supervenience, once we acknowledge, as we should, that the non-semantic facts on which semantic facts supervene can include brutally dispositional facts.<sup>25</sup>

## References

- Boghossian, Paul 1989. “The Rule-Following Considerations,” *Mind* 98: 507-49.
- Child, William 2010. “Wittgenstein’s Externalism,” in D. Whiting (ed.) *The Later Wittgenstein on Language*, Basingstoke: Palgrave Macmillan, 63-80.
- Child, William 2011. *Wittgenstein*. Abingdon: Routledge.
- Child, William 2017. “Sensations, Natural Properties, and the Private Language Argument,” in Kevin Cahill and Thomas Raleigh (eds.) *Wittgenstein and Naturalism: Wittgenstein’s Thought and Legacy*, Abingdon: Routledge, 75-95.
- Dummett, Michael 1978. *Truth and Other Enigmas*. London: Duckworth.
- Ginsborg, Hannah 2011. “Primitive Normativity and Skepticism about Rules,” *Journal of Philosophy* 108: 227-54.
- Glock, Hans-Johann and Preston, John 1995. “Externalism and First-Person Authority,” *The Monist* 78: 515-33.
- Horwich, Paul 2012. *Wittgenstein’s Metaphilosophy*. Oxford University Press.
- Kant, Immanuel [1781] 1998. *Critique of Pure Reason*, translated and edited by Paul Guyer and Allen Wood. Cambridge University Press.
- Kearns, Stephen and Magidor, Ofra 2012. “Semantic Sovereignty,” *Philosophy and Phenomenological Research* 85: 322-50.
- Kripke, Saul 1980. *Naming and Necessity*. Oxford: Blackwell.
- Kripke, Saul 1982. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
- Lewis, David 1983. “New Work for a Theory of Universals,” *Australasian Journal of Philosophy* 6: 343-77
- Lewis, David 1984. “Putnam’s Paradox,” *Australasian Journal of Philosophy* 62: 221-36.
- McDowell, John 1984. “Wittgenstein on Following a Rule,” *Synthese* 58: 325-63.
- McGinn, Colin 1984. *Wittgenstein on Meaning*. Oxford: Blackwell.
- Putnam, Hilary 1975. “The Meaning of ‘Meaning’,” *Minnesota Studies in the Philosophy of Science* 7: 131-193.
- Putnam, Hilary 1978. “Realism and Reason,” *Proceedings of the American Philosophical Association* 50: 483-98.
- Quine, W. V. O. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.
- Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.
- Strawson, P. F. 1966. *The Bounds of Sense: An Essay on Kant’s Critique of Pure Reason*. London: Methuen.
- Stroud, Barry 2000. *Meaning, Understanding, and Practice: Philosophical Essays*. Oxford University Press.
- Stroud, Barry 2012. “Meaning and Understanding,” in J. Ellis and D. Guevara (eds.) *Wittgenstein and the Philosophy of Mind*. Oxford University Press, pp. 19-36.
- Williams, Bernard 1978. *Descartes: The Project of Pure Enquiry*. Harmondsworth: Penguin.

---

### Notes

<sup>1</sup> Wittgenstein restricts this ‘explanation’ to ‘a *large* class of cases of the employment of the word “meaning”’. For present purposes we can focus on that class of cases, leaving aside cases to which Wittgenstein’s explanation is not intended to apply.

<sup>2</sup> For other statements of Stroud’s anti-reductionism about meaning, see Stroud 2000, pp. ix, 91-2, 130. For similarly anti-reductionist readings of Wittgenstein’s view of meaning and use, see McGinn 1984, McDowell 1984, Child 2011, pp. 95-104. Boghossian 1989 advocates an anti-reductionist view of meaning as the best response to Kripke’s Wittgenstein; he does not take a stand on Wittgenstein’s own position.

<sup>3</sup> Wittgenstein writes, in a related context: ‘Here there is nothing more difficult than to avoid pleonasm and only to say what really describes something’ (RFM VI §21). He thinks it is difficult to say something substantial and non-pleonastic; he does not think it is impossible.

<sup>4</sup> Here I disagree with Hannah Ginsborg, who argues in recent work that there is a primitive way of taking a performance to be appropriate in its context, which is independent of any prior grasp of meaning or rules (see Ginsborg 2011). She appeals to this ‘consciousness of . . . primitive appropriateness’ (2011, p. 248) to offer a ‘partly reductionist’ explanation of facts about meaning and rule-following in terms of ‘facts that are in a sense more primitive’, though not purely naturalistic (2011, p. 230). I plan to discuss this interesting proposal elsewhere.

<sup>5</sup> Compare RFM VII §26: ‘what the correct following of a rule consists in cannot be described *more closely* than by describing the *learning* of “proceeding according to the rule.”’

<sup>6</sup> In other passages, Wittgenstein stresses a distinction between genuine predictions (e.g. ‘if you follow the rules of multiplication as best you can, you will get 625’) and propositions that look like predictions but are really pleonasm – ways of stating what the rule in question requires (e.g. ‘if you follow the rules of multiplication, you will get 625’). ‘It is not a prediction’, he writes, ‘if the concept of *following* the rule is so determined, that the result is the criterion for whether the rule was followed’ (RFM VI §15). Someone might object that the ‘predictions’ Wittgenstein describes in RFM III §66 are not supposed to be genuine predictions at all: that they fall on the ‘pleonasm’ side of his distinction. In the context of the quoted passage, however, it is clear that Wittgenstein does mean to be describing genuine predictions: ‘In a technique of *calculating*, prophecies must be possible’ (RFM III §67).

<sup>7</sup> See Kearns and Magidor 2012, pp. 323-4.

<sup>8</sup> See Kripke 1980, Putnam 1975.

<sup>9</sup> For this picture, see Lewis 1983 and 1984.

<sup>10</sup> For an objection of this form, see Glock and Preston 1995.

<sup>11</sup> See Child 2010, pp. 65-9.

<sup>12</sup> See Child 2017.

<sup>13</sup> See Kripke 1982, pp. 22-37.

<sup>14</sup> We can for these purposes safely leave aside differences in behaviour that depend entirely on differences in the external environment, such as the difference between the behaviour I exhibit in drinking a glass of water and the behaviour my Twin Earth doppelgänger exhibits in drinking a glass of XYZ.

<sup>15</sup> See Quine 1960, Putnam 1978, Kripke 1982.

<sup>16</sup> Kearns and Magidor cite Boghossian’s proposal that ‘the so-called “Kripkenstein puzzle” ought to be solved by accepting a non-reductive view of semantic facts’ as a ‘rare exception’ to the general neglect of anti-reductionism as a plausible option in this debate (2012, p. 336, n. 30). As noted above, anti-reductionist responses to Kripke’s Wittgenstein are also offered by McDowell, McGinn, and Stroud, amongst others.

<sup>17</sup> For reasons of space, I also set aside one class of Kearns and Magidor’s counterexamples: those involving worlds that are non-semantic duplicates but which allegedly differ in haecceitistic semantic properties (2012, pp. 331-5, 342-4). I hope to consider that class of counterexamples elsewhere.

<sup>18</sup> In this phase of their discussion, Kearns and Magidor also offer a different counterexample, which appeals to the possibility of ordinary human thinkers having purely intentional properties: intentional properties whose possession by a person is completely independent of her possession of any non-intentional properties (2012, p. 340). The issues raised by that case are not fundamentally different from those raised by the alleged possibility of purely intentional ghosts.

<sup>19</sup> There are some indications that that is indeed how Kearns and Magidor are thinking of things. When they introduce the idea of purely intentional ghosts they point out that such beings ‘seem readily conceivable’ (2012, p. 340). And they stress in a different but related case that ‘there is nothing clearly incoherent about the scenario we have presented’ and argue that ‘our case presents a challenge to the proponent of [supervenience] precisely because the scenario we describe seems perfectly possible’ (2012, p. 342).

<sup>20</sup> For Kant’s arguments, see ‘The Paralogisms of Pure Reason’ in Kant [1781]1998, pp. 411-58. For Strawson’s arguments, see Strawson 1966, pp. 162-9.

<sup>21</sup> See PI §§243-315, 350-1. For related considerations, see Strawson 1959, chapter 3, and Williams 1978, pp. 100-1.

<sup>22</sup> How can the word ‘cat’ refer to cats without ever having been uttered by anyone? Well: it may have been written down, described, gestured at, and so on. (Compare the case of a society with a deity whose name, for religious reasons, is never uttered by anyone. It seems unproblematic that the name really is a name of the deity.)

<sup>23</sup> I have slightly adapted the case, for ease of presentation. But its essentials are taken from Kearns and Magidor.

<sup>24</sup> Someone might object that to characterize something as a case of ‘putting “41, 55 . . .” after “. . . 19, 29”’ is already to characterize it in semantic terms (in our extended sense), on the grounds that it involves characterizing it as an intentional action. Even if that is true, it remains the case that there will be *some* way of characterizing non-semantically what would have happened had Able tried to develop the series. That is all that the defender of supervenience requires.

<sup>25</sup> Earlier versions of some of this material were presented at the University of Helsinki and at the Higher School of Economics in Moscow. I am grateful to the audiences on both occasions for very helpful discussion.