WITTGENSTEIN ON PHILOSOPHY, OBJECTIVITY, AND MEANING

EDITED BY

JAMES CONANT
University of Chicago

SEBASTIAN SUNDAY
University of Oxford

CAMBRIDGE UNIVERSITY PRESS
CHAPTER 9

Demystifying Meaning in Horwich and Wittgenstein

Silver Bronzo

For more than two decades, Paul Horwich has been refining and defending a use-theory of meaning that seeks to demystify meaning by reducing it to pure regularities of use. He has presented this theory as a development of ideas from later Wittgenstein. Even though he acknowledges that his proposal goes beyond what Wittgenstein explicitly says, he doesn’t take it to go against Wittgenstein’s central commitments. Horwich’s views have changed in some respects over time, and he has occasionally distanced himself from Wittgenstein, but these differences do not affect the reductionist ambitions, the main structure, and the alleged Wittgensteinian roots of his proposal. This chapter argues, contra Horwich, that Wittgenstein’s demystification of meaning involves no reduction.

1. Horwich’s Reductionist Reading

Consider this passage from Horwich:

[Wittgenstein] was aiming to demystify the concept of meaning (and derivative intentional concepts such as belief and desire) by specifying in comparatively unproblematic terms, what meaning is . . . He wanted to explain how “life” is injected into signs that are otherwise “dead.” And his answer, to put it bluntly, is that meaning facts reduce to underlying non-intentional facts of word use . . . (including physical, behavioral, and certain psychological aspects). (Horwich 2010b, 19)

According to this reading, Wittgenstein takes the notion of meaning to be problematic. More specifically, he takes it to be mysterious, and thus in need of demystification. In order to demystify it, one must give a reductive account of what it is for a sign to be meaningful. A “sign” is here an item “individuated non-semantically,” for example in terms of its shape or acoustic properties (Horwich 2005, 28 n.4; see also 1998, 1, and 2005, 6). The “comparatively unproblematic terms” to which meaning is to be
reduced are dispositions to internally assent to sentences in conformity with certain regularities, plus feelings of satisfaction and dissatisfaction issuing in acts of self-corrections (see Horwich 2010b; and 2012, chapters 4 and 5). It is crucial that all the elements that figure in the base of reduction should be construed in a manner that does not presuppose the notion of linguistic or mental content; otherwise the account would not be truly reductive, and thus not truly demystifying. This applies, remarkably, to the notions of “sentence,” “assent,” and “self-correction.” Moreover, the base of reduction must not include normative notions. Horwich states explicitly that meaning, according to Wittgenstein, is ultimately constituted by “regularities of word-use” that are “wholly non-normative” and “non-regulative” (2010b, 24 n.10, 21).

Horwich emphasizes that the use-theory of meaning he attributes to Wittgenstein is not *behavioristic*. This is because the base of reduction includes psychological phenomena – such as the internal assent to sentences – that need not be expressed in outward behavior (see 2010b, 23 n.1). The theory is also not committed to *physicalism*: it demands that the psychological items figuring in the base of reduction be free of intentionality and normativity, but leaves open the question of whether they should in turn be reduced to physical phenomena. Horwich, for example, cashes out the notion of “internal assent to sentences” in terms of “cognitive role”: “The sentences to which S internally assents are those deployed as premises in his reasoning (i.e. computations), and are the ground for his dispositions to *overtly* assert their vocal correlates” (Horwich 2012, 113 n.5; see also 1998, 94–6; 2005, 30; and 2003a, 139). This characterization, which exploits the apparatus of cognitive science, is meant to be non-intentional and non-normative; but Horwich does not suggest that the success of his theory is conditional on the reducibility of cognitive science to physics. On the other hand, the theory does not *rule out* the prospect of a physicalist reduction of all the notions that figure in its own base of reduction – which perhaps explains why Horwich refers to such notions as “comparatively” unproblematic.

Horwich often characterizes his use-theory of meaning as “naturalistic” (see, for instance, 1998, 114; and 2005, 44, 64, 105, and 116). This makes good sense, since many forms of contemporary naturalism share the ambition of reducing meaning to a non-intentional and non-normative base. It can be puzzling, at first, to find out that Horwich firmly rejects “naturalism” and denies that Wittgenstein was committed to it (Horwich 1993, 156; 2013, 124–7; 2014, 37–40). But the puzzle is simply due to equivocation. The sort of naturalism that Horwich opposes and does not
ascribe to Wittgenstein is the view that “everything that exists is located within [the] spatiotemporal, causal domain” (2014, 38; see also 1993, 155 and 2013, 112). Against this view, Horwich holds that we should recognize “phenomena that fall outside the spatiotemporal causal order,” such as “numbers, values, universals, possibilities, and so on” (2014, 38, 40; see also 2013, 26–7); and he maintains that Wittgenstein was perfectly open to this recognition (1993, 156–7). This position is consistent with the view that all the phenomena belonging to the “spatiotemporal causal order” – which for Horwich include meaning – are reducible to non-intentional and non-normative notions. This is the form of naturalism that Horwich accepts and attributes to Wittgenstein.

There are two features of the context in which Horwich began to present his use-theory of meaning that are worth mentioning. First, it had already become widely accepted that any theory of meaning must account for its “normative” character. The idea is that the meaning of a word places constraints on how it should be used; it is a standard against which the use of the word is to be measured. Some applications of a word accords with its meaning and others fail to accord with it. To use a meaningful word is precisely to subject oneself to the constraints dictated by its meaning, and thus to expose oneself to the possibility of violating them, which amounts to committing a mistake. For example, if I want to use the word “green” in its most common English sense and I want to use it to make a serious assertion, then it will be correct for me to predicate it, say, of the grass, but not of the sky. I shall refer to this idea as the “normativity constraint.”

Secondly, many had already become convinced that the normativity constraint poses a serious challenge to dispositional accounts of meaning. This was largely due to Kripke’s influential critique of those accounts (see Kripke 1982, 22–37). An important part of this critique is that dispositional accounts, in order to be even prima facie plausible, must appeal to dispositions that operate under “ideal conditions” – which, however, cannot be specified without circularity. One way to see this is to notice that we are disposed to make mistakes in the application of our words under certain circumstances. For instance, I might be disposed to misapply the word “green” when I take certain drugs. So what I mean by “green” can be identified, at most, with how I am disposed to use the word under circumstances that exclude my taking those drugs. But – so the challenge goes – the idea that we can specify all the relevant “distorting factors,” and

---

1 For some early formulations of the normativity constraint on theories of meaning, see Wright 1980, 19; Kripke 1982, 37; and McDowell 1998, 221 and 235.
thus the correlative ideal conditions, without appealing to the very notion that we are trying to analyze is hopeless: the ideal conditions that we need are simply those in which I apply the word “green” in a manner that accords with the relevant meaning of the word (see Kripke 1982, 28–32).

Kripke argues, however, that this sort of difficulty is only a symptom of a deeper problem with dispositional theories. The fact that they cannot specify non-circularly the ideal conditions in which our dispositions are supposed to operate shows that they cannot give an account of the meaning of a word that tracks our competent judgments about which applications of the word accord or fail to accord with its meaning. But the deeper problem with dispositional theories is that they leave no room for the idea that an application of a word can accord or fail to accord with its meaning. This is because they construe the relation between the meaning of a word and its application in “descriptive” rather than “normative” terms (1982, 37). What they tell us is that, given what one “means” by a word, one will apply it in certain ways and not in others. But what they should account for, in order to satisfy the normativity constraint, is that given what one means by a certain word, one should apply it in some ways and not in others. Dispositional theories do not so much as fail to satisfy the normativity constraint, but simply choose to ignore it. And by doing so – the challenge continues – they relinquish any ambition to provide an account of meaning.²

Horwich is well aware of these challenges and has devoted great efforts to showing that the normativity constraint, when appropriately construed, can in fact be satisfied by his theory. Here we need to have in view at least the main structure of his theory. There are actually two versions of the theory: a more complex one, which Horwich ascribes to Wittgenstein, and a simpler one, which Horwich takes to be “neater and cleaner” (2005, 77 n.14; see also 2010a, 133–5; 2010b, 24 n.7; and 2012, 121 n.15). The simpler version involves three main steps. Starting “from above,” we have:³

1) An account of the “normative implications” of the truth-theoretic import of words in terms of various sorts of pragmatic and moral values. For example, given the extension of the word “green,” we ought

² In connection with the failure of dispositional theories to satisfy the normativity constraint, Kripke speaks of these theories as being “misdirected,” “off-target,” and as having an “air of irrelevance” (1982, 23, 37).

³ My reconstruction of the three main steps that Horwich’s simpler version of his theory involves is based on Horwich 2005, chapters 1–3 and 5; 2010a, chapters 6 and 7; 2010b; and 2012, chapters 4 and 5.
to apply it only to objects that fall under that extension, because truth is valuable.

2) A “deflationist” account of the truth-theoretic import of words in terms of a prior and independent notion of meaning and a number of schemas defining the truth-theoretic notions of truth, reference, and being true of. For example, the fact that “green” is true of the green things and nothing else follows by definition, and thus trivially, from the fact that “green” means the concept green. This is because the notion of being true of something is defined by a schema that has among its instances the following conditional: If “green” means the concept green, then “green” is true of x if and only if x is green.

3) An account of meaning in terms of law-like regularities of word use. The account is modeled on constitution theses advanced by the empirical sciences, such as the thesis that water consists of H₂O. The criterion of adequacy of these reductions, for Horwich, is their capacity to explain in terms of the base of reduction all the “symptoms” of the target phenomenon — for instance, the fact that water boils at a certain temperature, freezes at another temperature, and so on. But the symptoms of meaning, for Horwich, are the patterns of actual deployment of words (characterized in non-intentional and non-normative terms). On these grounds, Horwich maintains that for a word to have a certain meaning consists in its being governed by an ideal law of word use (more specifically, by an ideal law concerning the acceptance of a basic set of sentences containing the word). The use of the word, here, is “governed” by the law in the same sense in which the movements of the planets are governed by Kepler’s laws; and the law is “ideal” in the same sense in which Kepler’s laws are ideal (for example, in treating the planets as point masses). The process of sorting out “ideal laws” and “distorting factors” in the explanation of the overall use of a word is subject to the same epistemic norms that apply elsewhere in the empirical sciences (such as considerations of explanatory power, simplicity, and coherence with other accepted theories). Finally, the identification of the specific ideal laws of word use is an empirical task for linguistics, and it is to be expected that these laws will be different for different classes of words. For example, my meaning the concept green by “green” might consist in the fact that what I do with the word is governed by the ideal law “Silver accepts any sentence of the form ‘x is green’ when he has a visual experience normally produced in humans by observed green things.” (This is the
same as saying that my meaning green by “green” might consist in my being disposed to assent to those sentences under those circumstances.) By contrast, my meaning the concept truth by “true” might consist in the fact that what I do with the word is governed by the ideal law “Silver accepts any instance of the schema ‘<p> is true iff p.’”

It is worth noting that, in spite of verbal appearances, the part of the account that is meant to address the normativity constraint, as specified above, is not the first step, but the conjunction of the other two. There are of course many different kinds of reasons for using words one way or the other. The point of the normativity constraint is that the meanings of words impose characteristic demands on how to use them. So the question is not, say, why I should use “green” to make true predications as opposed to false ones; rather, it is why, assuming that I want to use “green” to make a true predication and that I want to use it in its common English sense, I should apply it to the grass but not to the sky.

Another feature of the account that is worth emphasizing is the crucial role of the second step, which severs the concept of meaning that Horwich purports to reduce from any truth-theoretic notion. This constitutes the core of Horwich’s response to Kripke’s circularity objection to dispositional theories. Horwich concedes that it is indeed hopeless to give a direct reduction of truth-theoretic content to use-regularities, but holds that the demand for such a direct reduction rests on an illegitimate, “inflationist” theory of truth. Truth-theoretic content can indeed be reduced to use-regularities, according to Horwich, but only in an indirect way via the definitional schemas mentioned in the second step of the reduction.

This was the simpler version of Horwich’s proposal. In the more complex version, the last step splits into two. So we have:

3a) An account of meaning in terms of implicit rule-following; that is, rule-following that does not involve the explicit articulation of the rule. For example, my meaning the concept green by “green” consists in my implicitly following certain rules for its use (for instance, possibly the rule “Accept any sentence of the form ‘x is green’ when you have a visual experience normally produced in humans by observed green things”).

3b) An account of implicit rule-following in terms of ideal natural laws (which operate in the absence of “distorting factors,” like Kepler’s laws) and acts of self-correction caused by feelings of satisfaction and dissatisfaction with one’s performances. For example, my implicitly following the rule mentioned in (3a) would consist in the fact that my
use of the word “green” is governed by the ideal law “Silver accepts any sentence of the form ‘x is green’ when he has a visual experience normally produced in humans by observed green things,” in conjunction with the fact that I am occasionally dissatisfied with how I am inclined to use the word and correct myself accordingly, which shows that I have a desire to conform to the ideal natural law governing my use of the word.

As one can see, Horwich’s reductionist account in either of its versions is a mechanism with many moving parts. Evaluating whether it succeeds in meeting, or debunking, the normativity constraint is not an easy task, and goes beyond the scope of this chapter. As I anticipated, I am only going to take issue with Horwich’s grounds for attributing that sort of account to Wittgenstein and for thinking that a demystification of meaning must take the form of a reduction. I will begin, in Section 2, by looking at several well-known passages from the Investigations that deal with meaning, understanding, and rule-following. In those passages, I will argue, we can see what Wittgenstein’s demystification of meaning actually looks like, and how it differs from Horwich’s reconstruction. Then, in the subsequent three sections, I am going to examine the main textual evidence that Horwich takes to support his reductionist reading of Wittgenstein.

2 Demystification in Philosophical Investigations §§185–242

A pervasive theme of the Investigations is how, when we do philosophy, some of the most familiar things in the world – such as naming something, understanding the meaning of a word, or obeying an order – can come to strike us as being mysterious: for instance, as involving “a strange connection,” an “occult” or “odd process” (§§ 38, 196), “a strange medium,” “an odd kind of being” (§196), “a pure intermediary” (§94), or “a shadow” of actual events (§194).

As we have seen, it is natural to think, in accordance with the normativity constraint, that the meaning of an expression determines which applications of the expression accord and fail to accord with it – and thus how one ought to use it, if one wants to use it in accordance with its meaning. To take one of Wittgenstein’s examples: if you understand what the order “Add 2” means, it seems to be already determined that if you were asked to add 2 to 1000, and you wanted to satisfy the order, you ought to answer 1002 (and not, say, 1004). Moreover, it seems that this is the answer you ought to give even if you never explicitly thought about that particular
application of the order. But now it can seem that understanding the
meaning of an expression, as well as meaning it in some particular way,
must be an extraordinary mental accomplishment. Here is how
Wittgenstein describes the aura of extraordinariness that may be thought
to surround the act of meaning the order “Add 2” in the usual way:

Your idea was that this meaning the order had in its own way already taken all
those steps: that in meaning it, your mind, as it were, flew ahead and took all
the steps before you physically arrived at this or that one.

So you were inclined to use such expressions as “The steps are really
already taken, even before I take them in writing or in speech or in thought.”
And it seemed as if they were in some unique way predetermined, antici-
pated – in the way that only meaning something could anticipate reality.

(PI § 188)

To use an image that Wittgenstein introduces later on in the course of the
same discussion, we are led to think of the meaning of a linguistic expres-
sion as a pair of “rails invisibly laid to infinity” (§218), that is, rails that go
through the space of all the possible applications of the expression and
determine in each case what counts as a correct and what as an incorrect
application. Understanding the meaning of the expression (or meaning it
in a particular way) is then pictured as a mental act that consists in
instantaneously going through these infinitely long rails. The actual appli-
cations that we are going to make of the expression (in writing, speech, or
even merely in thought) are going to be correct or incorrect precisely
because they have been anticipated by the initial act of understanding.
But how can our finite mind go through – and in an instant! – all the
countless possible applications of an expression? And how can it do so
without even actually thinking of them? Understanding the meaning of an
expression, we are inclined to conclude, is an impenetrable mystery.

But according to Wittgenstein there is no real mystery here. We come to
think that there is one, he believes, because we misunderstand the forms of
expressions that we use when we talk about meaning and understanding:

We do pay attention to the way we talk about these matters, we don’t
understand it, but misinterpret it. When we do philosophy, we are like
savages, primitive people, who hear the way in which civilized people talk,
put a false interpretation on it, and then draw the oddest conclusions from
this. (PI §194)

Similar remarks occur earlier in the Investigations, in the context of
a discussion of the impression that a proposition is something “remarkable,”
“extraordinary,” and “unique” (§93) – for instance, “a pure intermediary
between the propositional *sign* and the facts” (§94). This impression, Wittgenstein writes, is due to a “misunderstanding of the logic of language” (§93): “For our forms of expressions, which send us in the pursuit of chimeras, prevent us in all sorts of ways from seeing that nothing extraordinary is involved” (§94).

Now, what are the forms of expression that we savagely misunderstand when we think about meaning and understanding? A prominent example is discussed right after the passage that compares us to savages:

“But I don’t mean that what I do now (in grasping the whole use of a word) determines the future use *causally* and as a matter of experience, but that, in a *strange* way, the use itself is in some sense present.” – But of course it is, ‘in *some* sense!’ Really, the only thing wrong with what you say is the expression “in an odd way.” The rest is right; and the sentence seems odd only when one imagines it to belong to a different language-game from the one in which we actually use it. (PI §195)

We can rephrase the form of expression that causes trouble here this way: *Understanding the meaning of an expression does not determine its future use in a causal way, but in some sense already contains it.* Wittgenstein emphasizes that the form of expression is, in itself, all right: there is nothing strange and mysterious about it when it is used in its original language-game. It *becomes* strange and mysterious only when we fail to get into view its original language-game and think of it on the model of expressions that belong to quite different language-games, thereby casting on it a strange interpretation.

So far we have identified a representative example of the forms of expression that mislead us into thinking that meaning and understanding are mysterious phenomena. Now we need to distinguish the correct and incorrect interpretations of this form of expression. How should it be understood in order to be “all right”? And how exactly should we characterize its misinterpretation? The most direct answer to these questions, I suggest, comes in sections 219–21:

219 “All the steps are really already taken” means: I no longer have any choice. The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space. – But if something of this sort really were the case, how would it help me?

No; my description made sense only if it was to be understood symbolically. – I should say: *This is how it strikes me.*

When I follow the rule, I do not choose.

I follow the rule *blindly.*
220 But what is the purpose of that symbolical proposition? It was supposed to bring into prominence a difference between being causally determined and being logically determined.

221 My symbolical expression was really a mythological description of the use of a rule.

Wittgenstein speaks here of obeying a rule, but the discussion is closely connected to – and indeed a special case of – the overarching discussion of meaning and understanding. For a rule is said to be “stamped with a particular meaning”; thus a “rule,” here, is equivalent to what Wittgenstein calls elsewhere the “expression of a rule” (§198). It can be, for instance, a signpost that tells us to go in some particular direction. The passages I quoted, then, are about the idea that the meaning of a rule determines what counts as obeying and violating the rule. The troublesome forms of expressions that Wittgenstein discusses in this connection are variants of the one we encountered before: “[When I understand a rule,] all the steps are really already taken”; “The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space.” The first thing to notice is that, for Wittgenstein, there is a way of interpreting these propositions so that they make sense. They make sense, he says, only if they are understood symbolically. I take it that symbolical understanding contrasts here with literal understanding (see also Baker and Hacker 1980/2009, 197). In order to make perspicuous the symbolical character of such propositions, Wittgenstein points out, it would be helpful to frame them with a phrase such as “This is how it strikes me.” A prefix of that sort would help to make clear that what follows should not be taken at face value.

But what do the troublesome propositions mean, when they are understood symbolically? Their meaning, Wittgenstein suggests, can be paraphrased by means of more straightforward propositions such as the following: “When I follow the rule, I do not choose”; “I follow the rule blindly”; “[When I follow a rule,] I no longer have any choice” (see also §230). These are good examples of what Wittgenstein calls “grammatical remarks” (§232). Grammatical remarks are meant to remind us of the way we use our words (so, of their grammar, or logic); they are not meant to be controversial, but to capture what everybody would agree on (see §§ 89–90 and 127–8). The grammatical remarks in question concern our use of the expressions “rule” and “following a rule.” It is part of our concepts of rule and following a rule that we don’t have to arbitrarily choose at each step what counts as following the rule: to follow a rule is precisely to subject our will to constraints. I suggest that, in spite of its technical language, the
following is also presented as a grammatical remark: “[There is] a difference between being causally determined and being logically determined.” The meaning of a rule determines what course of action accords or fails to accord with the rule; but this form of determination is different from causal determination.

The symbolical propositions, when properly understood, are ways of expressing these grammatical points. They are, Wittgenstein says, “mythological description[s] of the use of a rule.” They point out aspects of the grammar of “rule,” and they do achieve their purpose, in so far as the mythology is understood symbolically (rather than misunderstood literally). The problem arises when we forget, so to speak, the symbolical operator in which the mythological descriptions are embedded, and take them literally. Only at that point the propositions appear to express something astonishing and mysterious.

To summarize, I have argued that in the cluster of sections I have focused on, Wittgenstein distinguishes three kinds of proposition. We have grammatical remarks, which are meant to remind us—in a straightforward and uncontroversial manner—of features of our concepts of meaning, rule, and understanding:

G1) “When I follow the rule, I do not choose.” (§219)
G2) “I follow the rule blindly.” (§219)
G3) “[When I follow a rule,] I no longer have any choice.” (§219)
G4) “[There is] a difference between being causally determined and being logically determined.” (§220)

Then we have symbolical propositions, which serve the same purpose but seek to achieve it through a symbolical employment of images and “mythologies”:

S1) Understanding the meaning of an expression does not determine its future use in a causal way, but in some sense already contains it.
S2) “[When I understand a rule,] all the steps are really already taken.” (§219)
S3) “The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space.” (§219)

And finally, we have propositions that express a misunderstanding of symbolical propositions. They give a literal construal of symbolical propositions, and thus take them to describe mysterious phenomena. We may call them mysterian propositions:
Demystifying Meaning in Horwich and Wittgenstein

M1) “I don’t mean that what I do now (in grasping the whole use of a word) determines the future use causally and as a matter of experience, but that, in a strange way, the use itself is in some sense present.” (§195)

M2) “[When I understand a rule,] all the steps are [in some mysterious way] really already taken.” (§219)

According to Wittgenstein, these are propositions to which we are drawn “when we do philosophy” (§194), and which are part of what philosophy (as he seeks to practice it) aims to criticize.

The first moral I want to draw from this discussion is that there is an obvious sense in which Wittgenstein is concerned to demystify meaning. When doing philosophy, he argues, we are inclined to construe meaning and understanding as mysterious phenomena. But these are misconceptions, mainly due to the fact that we misunderstand the forms of expressions that we use to talk about those phenomena. They are mystifications in the sense that they render mysterious something that, in itself, is not mysterious. Wittgenstein seeks to undo these mystifications by clarifying the way we talk about meaning and understanding. In the passages I have examined, mysterian propositions are unmasked as misunderstandings of symbolical propositions, which can in turn be paraphrased by means of grammatical remarks about our concepts of meaning and understanding. This process of demystification – and this is the second moral I want to draw – does not involve any attempt to reduce meaning and understanding to something else. What we are left with, when the process is over, are remarks such as G1–G4. These remarks aim to spell out features of our concepts of meaning and understanding by connecting them with a number of other concepts – such as an appropriate notion of choicelessness, or of determination. But there is no intention to show that what meaning and understanding (really) are can be specified in a different and supposedly more fundamental conceptual apparatus. (See also Hans-Johann Glock’s “What Is Meaning?,” Chapter 10 in this volume.)

3 The Shopkeeper and the Builders

Horwich claims that his interpretation is supported by the discussion of various simple language-games in the opening sections of the Investigations. He refers, in particular, to the shopkeeper’s language-game described in section 1, and to the builders’ language-game as described in sections 2 and 8. For Horwich, the presentation of these and other similar language-games is meant to illustrate the “meaning-constituting uses” of words,
which are “never couched in semantic or intentional terms” (2012, 112; see also 2010b, 19). But this reading can be questioned.

Consider, to begin with, the shopkeeper’s language-game. The section where it is described begins with a quotation from Augustine’s Confessions that conveys, Wittgenstein says, a certain idea about the essence of language: the idea that the meaning of a word is the object it stands for. He goes on to suggest that we come to that idea only if we think primarily of words such as “table” and “chair,” and not about words such as “five” or “if.” At that point, he tells the shopkeeper story:

Now think of the following use of language: I send someone shopping. I give him a slip of paper marked “five red apples.” He takes the slip to the shopkeeper, who opens the drawer marked “apples”; then he looks up the word “red” in a chart and finds a colour sample next to it; then he says the series of elementary number-words – I assume that he knows them by heart – up to the word “five”, and for each number-word he takes an apple of the same colour as the sample out of the drawer. – It is in this and similar ways that one operates with words. (PI §1)

The main point of the story is to show how different are the uses of words such as “apple,” “red,” and “five.” In particular, it is meant to show, as Wittgenstein goes on to point out, that the question of what the word “five” stands for does not come up: all that matters is how the word is used. Now, I agree with Horwich that the story illustrates the general idea that the meaning of a word is constituted by its use – by what we do with it. But I reject Horwich’s claim that the story is meant to show that this use can be specified in non-semantic and non-intentional terms. The meaning of the words used by the shopper and the shopkeeper are indeed constituted, at least in part, by what they do with them. But their doing is (already) informed by meaning and understanding. The shopper is fulfilling the request to go shopping. The slip marked “five red apples” is something like a shopping list: by giving it to the shopkeeper, the shopper is requesting the shopkeeper to sell her what is written on the list. And the shopkeeper goes on to fulfill that request. Making requests, fulfilling them, and reading shopping lists are activities that presuppose semantic or intentional notions. The whole episode, moreover, takes place within the framework of a practice of economic transactions; and it is far from obvious that such a practice can be intelligibly specified in non-semantic and non-intentional terms, as Horwich’s reading requires. Furthermore, the claim that a reduction in such terms is possible is completely unnecessary for the kind of claim that Wittgenstein actually appears to be making, namely,
that words have very different functions and do not have to stand for anything in order to have meaning. Horwich’s reductionist reading of the shopkeeper example is only a projection of his own theoretical agenda.

Next, here is how Wittgenstein describes the builders’ language-game:

Let us imagine a language for which the description given by Augustine is right: the language is meant to serve for communication between a builder A and an assistant B. A is building with building stones: there are blocks, pillars, slabs and beams. B has to pass him the stones and to do so in the order in which A needs them. For this purpose they make use of a language consisting of the words “block,” “pillar,” “slab,” “beam.” A calls them out; B brings the stone which he has learnt to bring at such-and-such a call. – Conceive of this as a complete primitive language. (P.I § 2)

As before, we need to look at the role that the description of this language-game is intended to play in Wittgenstein’s argument. He says he introduces it in order to show that Augustine’s description of language can be regarded, on the one hand, as expressing a conception of meaning that is “at home in a primitive idea of the way language functions” – and thus as an inadequate description of our language – but also, on the other hand, as an adequate description “of a language more primitive than ours” (§2). The builders’ language-game is meant to be such a language. Wittgenstein asks us to conceive it “as a complete primitive language.” But as several commentators have pointed out, this request is not easy to interpret (see, for example, Goldfarb 1983 and Cavell 1996). We need to pause and ask what exactly we are supposed to imagine here. There are various options.

One option is to suppose that the life of A and B is by and large like ours. But then it is mysterious why they choose to speak only when they are working on the building site, and why they choose to use only those four words. Maybe because of some kind of prohibition or vow? In that case, it is not clear why their language should be called “primitive.” Moreover, while what they do with their words may be regarded as meaning-constituting, it won’t be describable in purely non-semantic terms: as in the shopkeeper’s case, their actions will be permeated by semantic and intentional notions. A, for example, won’t be merely producing noises, but will be uttering words to make requests, which B understands and goes on to satisfy, and all this on the background of a collaborative activity of building that presupposes the capacity for intentional action, practical reasoning, and social coordination.

Alternatively, we might suppose that A and B are in fact like marionettes, or rather simple automata. At the end of the day, when the building
activity is over, they simply shut down. In that case, it is not clear that they can be said to have a “primitive language.” What they do may indeed be describable in purely non-semantic and non-intentional terms, but it won’t be meaning-constituting. A and B communicate with one another, and have a language, only in the derivative sense in which, say, my cellphone and laptop may be said to communicate with one another and use a language.

Yet another option is to suppose that A and B resemble those monkeys that, apparently, use a system of communication consisting exclusively of a small number of signals in order to alert the group about the presence of different kinds of predators. In that case, it makes good sense to say that they master a “primitive language.” There are enormous similarities, as well as striking differences, between the life of the monkeys and our own, and between the roles that the respective forms of communication play in these lives. In virtue of these similarities and differences, we are able to recognize some primitive form of meaning in their signals. Nevertheless, there appears to remain a qualitative difference between the scream of one of these monkeys and what we do when we shout “Snake!” in order to warn another about the presence of a snake, even though the monkeys’ scream is unquestionably more similar to our warning than, say, the sounds emitted by a smoke-detector. We may hold, with Horwich, that the form of meaning sensibly ascribable to the monkeys’ signals is constituted by how they use those signals. But it is plausible to maintain that the extent to which we may regard that use as meaning-constituting is directly proportional to the extent to which it is not describable in non-intentional and non-semantic terms. The proto-semantics that we can sensibly ascribe to the signals of the monkeys – or to the “calls” of Wittgenstein’s builders – goes hand in hand with the proto-semantic and proto-intentional character of the activities in which those signals are embedded.

I have distinguished three ways of understanding the builders’ language-game, arguing that none of them supports the view that the meaning-constituting use of words is describable in non-intentional and non-semantic terms. What is crucial for our present purposes is that, as we have noted about the shopkeeper example, here too there is no need of saddling Wittgenstein with that reductionist claim; good sense can be made of what Wittgenstein wants to do with the builders’ language-game without ascribing to him any such reductionist intentions. His aim, we have seen, is simply to give an example of a “primitive language.” Indeed, as I have shown in my discussion of the third way of understanding
the builders’ language-game, one can achieve this aim, while insisting that use is meaning-constituting only to the extent that it is already infused with meaning and intentionality.

4 Meaning, Truth, and Deflationism

Another piece of textual evidence that Horwich mentions in support of his interpretation is Wittgenstein’s supposed endorsement of a “deflationary view of the truth-theoretic notions” in section 136 of the Investigations (Horwich 2012, 110). After arguing that Wittgenstein wants to explain meaning in terms of use, Horwich appeals to that section in order to show that the sort of use that Wittgenstein invokes is specifiable in non-semantic terms, and hence without any reliance on truth-theoretic notions such as truth conditions, reference conditions, and satisfaction conditions. In the relevant section, Horwich maintains, Wittgenstein accepts the schema “p is true = p (where ‘p’ stands for the proposition that p)” (2012, 110). This shows, he argues, that Wittgenstein wants to explain the notion of truth in terms of a prior and independent notion of sentence meaning:

> The notion of truth . . . is explained in terms of that of [proposition] (i.e. sentence meaning), which will on pain of circularity have to be explained independently of truth. Therefore, he cannot be supposing that the notion of proposition be analyzed in terms of the notion of truth condition. (Horwich 2012, 110)

If truth is to be analyzed non-circularly in terms of sentence meaning, then the analysis of sentence meaning in terms of use cannot take for granted the notion of truth. Analogous considerations, Horwich argues, apply to the other truth-theoretic notions and to Wittgenstein’s use-theoretic account of non-sentential meaning.

Wittgenstein does indeed mention, in section 136, the schema “‘p’ is true = p,” and I agree with Horwich that Wittgenstein mentions it in an approving manner. I can also grant Horwich’s rephrasing of the schema as “The proposition that p is true = p.” But I question whether there is any evidence that Wittgenstein takes the schema to express the sort of deflationism that Horwich wishes to ascribe to him (namely, a reductive analysis of truth in terms of sentence meaning). No direction of reductive analysis can be inferred from the schema itself; it only states an equivalence. Moreover, anyone who cares to look at the relevant context in the Investigations will find that there is nothing that would support Horwich’s reading.
Wittgenstein mentions the schema in the course of a discussion of his earlier specification of the general form of the proposition. In the *Tractatus*, he said that the general form of the proposition is “This is how things are” (TLP 4.5). And now he says:

At bottom, giving “This is how things are” as the general form of propositions is the same as giving the explanation: a proposition is whatever can be true or false. For instead of “This is how things are,” I could just as well have said “Such-and-such is true.” (PI §136)

If anything, this passage suggests that the notion of a proposition is defined in terms of truth and falsity, rather than the other way around. However, Wittgenstein actually goes on, from the passage just quoted, to introduce the equivalence schema with the adversative phrase “Nun ist aber,” translated simply as “but”; “But ’p’ is true = p.” I suggest that the role of the schema, in this context, is to block the impression that the notion of a proposition can be defined in terms of a prior and independent notion of truth. However, its role is not to make the reverse claim, as Horwich would have it, that truth can be defined in terms of a prior and independent notion of proposition. Rather, Wittgenstein wants to convey the idea that the notions of truth and proposition come together (so that neither can be non-circularly defined in terms of the other).

Support for this reading comes from the way section 136 continues – Wittgenstein writes that the use of the words “true” and “false” is correctly treated “as belonging to our concept ‘proposition’” – as well as from the following later sections of the book:

224 The word “accord” and the word “rule” are related to one another; they are cousins. If I teach anyone the use of the one word, he learns the use of the other with it.

225 The use of the word “rule” and the use of the word “same” are interwoven. (As are the use of “proposition” and the use of “true.”)

The notions of *proposition* and *truth* are interwoven, just as the notions of *agreement* and *rule*, and *rule* and *same*. There is no way of learning or understanding the one without learning and understanding the other. In other words, there is no definitional or conceptual priority among them.

Thus, Horwich’s reading of section 136 of the *Investigations* is incorrect. Wittgenstein’s discussion of the relation between propositions and truth in this section provides no reason to believe that he aims to specify the meaning-constituting use of words in non-semantic terms.
5 Meaning and Dispositions

Horwich purports to find direct textual evidence for his reading in this passage:

“But I already knew, at the time when I gave the order [to continue the series +2], that he should write 1002 after 1000.” – Certainly; and you may even say you meant it then; only you shouldn’t let yourself be misled by the grammar of the words “know” and “mean.” For you don’t mean that you thought of the step from 1000 to 1002 at that time – and even if you did think of this step, still, you didn’t think of other ones. Your “I already knew at the time...” amounts to something like: “If I had then been asked what number he should write after 1000, I would have replied ‘1002.’” And that I don’t doubt. This is an assumption of much the same sort as “If he had fallen into the water then, I would have jumped in after him.” (PI §187)

Horwich comments: “Thus Wittgenstein’s view, quite clearly, is that the meaning consists in the disposition” (2012, 141). For Horwich, of course, the disposition must be specifiable in non-semantic and non-intentional terms. According to Horwich, then, what Wittgenstein is saying is that meaning the order “Continue the series +2” in the usual way is partially constituted by one’s being causally disposed to write down certain marks after hearing certain noises (for instance, to write down “1002” after hearing “What should be written after 1000, if you want to continue the series +2?”), where “writing” and “hearing” must be construed without presupposing semantic or intentional notions. As we saw in Section 1 above, to have such a disposition is equivalent to being governed by a ceteris paribus causal law, where the ceteris paribus conditions of the law (which correspond to the activation conditions of the disposition) must be specifiable without any appeal to the meaning of the order under discussion, and indeed (assuming that we are dealing with the meaning-constituting uses of primitive expressions) without appealing to any linguistic or mental content.

Once again, Horwich’s reading is far from obvious. Here is an alternative interpretation of Wittgenstein’s point. Meaning the order “Continue the series +2” in the usual manner entails having the capacity to determine that the only answer that satisfies the order, when one reaches 1000, is “1002.” More generally, meaning something by an expression – and, conversely, understanding the meaning of an expression – requires the capacity to determine which uses of the expression accord or fail to accord with its meaning. This capacity differs from Horwich’s dispositions in three interrelated respects.
First, it is the capacity to carry on with the use of the expression *in accordance with its meaning*: the meaning of the expression enters into the specification of the capacity, which cannot therefore be used to provide a reductive account of the meaning.

Second, the capacity in question is specified in *normative* terms: it can be successfully or unsuccessfully exercised, where the standards of success are fixed by the requirements of the meaning of the expression that one is using. By contrast, the purely causal dispositions which Horwich invokes do not admit of such normative distinctions. There is no such thing as *failing* to act in accordance with a causal disposition. If something does not behave in accordance with a causal disposition (say, if a planet does not move in accordance with Kepler’s laws), it simply means that the *ceteris paribus* conditions under which the disposition becomes operative are not satisfied – or, alternatively, that the thing in question does not actually have that disposition.

Third, the capacity to carry on with the use of an expression in accordance with its meaning will be successfully exercised in some conditions and not in others, but there is no reason to expect that these conditions can be specified *without circularity*: the conditions for the successful exercise of the capacity are simply the conditions under which one manages to live up to the demands of the meaning of the expression.

Taken in isolation, the passage from Wittgenstein that Horwich considers (PI §187) does not rule out his reductionist reading. But when we look at the immediate context of the passage, there are some good reasons for preferring the kind of reading that I have suggested over Horwich’s reading. As we saw in Section 2, in section 220 of the *Investigations* Wittgenstein contrasts causal and logical determination. From the perspective of Horwich’s reading, Wittgenstein’s drawing of this contrast makes little sense, because the whole point of Horwich’s reading is that Wittgenstein wants to *reduce* logical determination to causal determination. Wittgenstein discusses an analogous contrast in sections 193–4. (These sections have been plausibly interpreted by some commentators as an explicit criticism of dispositional theories of meaning; see, for example, Kripke 1982, 25 n.24.) In those sections, Wittgenstein distinguishes two concepts of machine. When we talk about a “machine,” he argues, we can be talking about a “machine *qua* symbol,” or about an “actual machine.” In the former case, we look at the machine from a functional perspective (that is, from the perspective of how the machine is *supposed* to work), and so we may derive from it the movements that *accord* with its function. In the latter case, we look at the machine from
Demystifying Meaning in Horwich and Wittgenstein

a merely causal perspective, and so we may derive from it the movements that it \textit{will in fact make} in different circumstances. Wittgenstein observes that “the movement of the machine \textit{qua} symbol is predetermined in a different way from how the movement of any given actual machine is” (§193). The whole point of this discussion of machines is to clarify through an analogy the sense in which the meaning of an expression predetermines or “already contains” all its possible applications. Thus, Wittgenstein is here anticipating the contrast drawn in section 220 between causal and logical determination. Again, it is not clear why Wittgenstein should emphasize such a distinction in the way he does, if he really wanted to claim, as Horwich maintains he did, that logical determination is \textit{reducible} to causal determination.\footnote{I have not discussed all of the passages that Horwich marshals in support of his reductionist reading. I believe that in each of the remaining cases, the text can be shown to be at least compatible with an anti-reductionist conception of meaning. Some of the relevant passages have already received interpretations that are explicitly at odds with Horwich’s reading. See, for example, Gustafsson 2014b, 1200 and Bronzo 2017, 1351–4.}

6 Concluding Remarks

I have argued that for Wittgenstein meaning becomes mysterious when we misunderstand its grammar or logic – that is, when we misconstrue the language-games in which we talk about meaning and related phenomena, mainly because we model them on quite different language-games. His attempt to demystify meaning takes therefore the form of an attempt to clarify the concept of meaning. In order to achieve this purpose, Wittgenstein investigates the connections between meaning and a host of other notions, including appropriate notions of choicelessness, determination, truth, and capacity. All these notions are interdependent. In particular, the notions that Wittgenstein takes to be part of the conditions of intelligibility of meaning have meaning as part of \textit{their own} conditions of intelligibility. We are here dealing, we might say, with a distinctive metaphysical layer of reality, which is thoroughly colored with meaning and intentionality. Meaning becomes visible, in all its ordinariness and unmysterialousness, when we attend to the place that it occupies within that layer of reality.

Horwich is well aware of extant anti-reductionist conceptions of the demystification of meaning. For example, he gives the following apt description of John McDowell’s approach to intentionality: “Since our puzzlement about meaning is merely an artifact of self-inflicted
mystification, the illumination we need will have to come from a rooting out of confusions rather than from the development of a reductive theory” (Horwich 2005, 7). Indeed, this statement would also serve as an accurate description of the approach ascribed to Wittgenstein in the present chapter. Horwich finds it unviable: he rejects as “perverse” an account of meaning that does not seek to avoid circularity (2012, 113 n.5). However, he assures us that his reason for demanding a reductionist account of meaning is not “some gut metaphysical conviction that all facts must be grounded in physical (or ‘naturalistic’) phenomena” (2005, 36). Horwich’s substantial objection to anti-reductionist accounts of meaning is that they render mysterious how meaning and understanding can cause physical events, such as actions and speech acts. The main difficulty, he thinks, is to vindicate this causal efficacy without denying the causal closure of the physical world and without positing causal overdetermination (Horwich 2010a, 103 n.4; see also 2005, 6 and 36; 2008, 472 n.4; and 2012, 160).

Horwich is right that this is a serious challenge. However, reductionist theories of meaning also face a serious challenge. In order to be successful, they must refrain from presupposing in any way the notion of meaning, while also managing to provide an account that is recognizable as an account of meaning (rather than simply changing the topic and describing a scenario in which meaning is completely absent). If the argument of the present chapter is correct, Wittgenstein is open to the prospect of meeting the challenge raised by Horwich about the causal efficacy of meaning and understanding, but regards a reductionist theory of meaning à la Horwich as a hopeless undertaking.