Book Review

Truth — Meaning — Reality, by Paul Horwich. Oxford: Oxford University Press, 2010. Pp. x + 341. H/b £55.00, P/b £21.00.

One of the most important philosophical developments in the last quarter century is the emergence of Paul Horwich's systematic account of thought and language. The account is remarkable for its plausibility, originality, and explanatory power. The volume under review contains eight essays that present this account, extending and deepening the formulations in Horwich's earlier writings, and also six essays that work out its implications for a range of important questions in metaphysics and epistemology. The chapters all derive from papers that were published between 2001 and 2010. They are all accessible, and they are full of provocative and well motivated ideas. Other readers will, I am sure, join me in feeling grateful to Horwich for writing them. Bravo!

One of Horwich's main contributions is a deflationary theory of truth that he calls *minimalism*. A theory of truth is deflationary if it denies that truth has a robust nature that can be elucidated by science, metaphysics, or normative inquiry. In consequence, a deflationary theory denies that truth can be explained in terms of the properties that were favoured by traditional theories — correspondence, coherence, and convergence of opinion. Minimalism is the deflationary theory which claims, first, that we are disposed to accept all instances of the following *equivalence schema*:

(T) The proposition that p is true just in case p

and secondly, that our use of the concept of a true proposition can be exhaustively explained in terms of this disposition. In other words, it maintains that the disposition exhaustively determines the content of the concept of propositional truth. Minimalism goes on to make additional claims. For example, it asserts that a proposition is false just in case it is not true, and that our use of the concept of a true *sentence* can be explained in terms of the fact that we are disposed to accept all instances of (ST):

(ST) If a sentence S means that p, then S is true just in case p. (p. 164)

But the key idea is the foregoing thesis about the explanatory adequacy of (T). Much of the book is devoted to the elaboration and defence of this idea. Thus, one chapter presents Horwich's favourite rationale for minimalism, which argues that it provides the best explanation of the role that the concept of truth plays in generalized and indefinite endorsements of propositions (such as *Everything Obama said at the meeting is true*); another chapter reviews the main deflationary alternatives to minimalism, and argues that they all have disabling flaws; a third responds to ten objections to minimalism, due principally to Davidson, Dummett, Field, Gupta, Richard, and Soames; a fourth maintains that minimalism is superior to Tarski's theory of truth; and a fifth criticizes the anti-minimalist doctrine that propositions owe their truth to the existence of facts.

I cannot discuss all of these lines of thought here. Instead I will describe the current status of a particularly important objection to minimalism. This objection has been discussed by a number of philosophers, but it was first raised by Anil Gupta ('A Critique of Deflationism', *Philosophical Topics*, 21, 1993, pp. 57–81.).

It seems that anyone who possesses the concept of truth and the basic logical concepts is able to see, *a priori and indeed ipso facto*, that certain generalizations about truth are correct. Thus, for example, simply in virtue of possessing the concept of truth and the concept *if*, it is possible to appreciate the correctness of (A):

(A) Every proposition that has the form *if p then p* is true

Now, on the face of it, it seems that minimalism is unable to explain this fact. Thus, according to minimalism, our grasp of the concept of truth consists in a disposition to accept a number of particular propositions — propositions of the form (T). Our mastery of the concept does not involve knowledge or acceptance of any general propositions. But it seems that our mastery of truth would have to involve some sort of generality in order for that mastery to suffice for the appreciation of generalizations like (A). Certainly it would not be possible to *derive* the generalization from premisses consisting only of particular propositions, unless the premisses were infinite in number. And how could a human mind construct an infinite proof?

Horwich has long struggled with this objection. In the present book he gives a version of his reply that is more fully developed than previous versions. Here is the key idea:

> Suppose it were the case that whenever anyone is disposed to hold, concerning each F, that it is G, then he comes, on that basis, to believe that every F is G. Our disposition to accept, for each proposition of a certain form, that it is true would then suffice to explain our acceptance of the generalization, 'Every proposition of that form is true'. (p. 44)

Of course, it can fail to happen that someone who believes every instance of a generalization also believes the generalization itself. This would occur, for example, if someone's theory of a certain domain was ω-inconsistent. But Horwich thinks that his principle holds in cases like (A), because he thinks that, in such cases, we cannot conceive of there being additional Fs that fail to be Gs (p. 44). Consider our rationale for thinking that *if torture is wrong then torture is wrong* is true. Could someone who has that rationale in mind conceive of propositions of the form *if p then p* that fail to be true? Horwich thinks not, and because of this perception, he thinks that the rationale in question is capable of carrying us to acceptance of the generalization (A).

Minimalism is an exciting development, one of the very best ideas about truth that philosophers have thus far managed to devise. Even while admiring it tremendously, however, I have doubts about the ultimate adequacy of the theory, and also doubts about some of the moves that Horwich makes in defending it.

We have vivid intuitions to the effect that truth consists in correspondence with fact. To be adequate, a theory of truth must either honour these intuitions or explain them away. Now in an earlier book, Horwich showed that he can go some way toward explaining correspondence intuitions in minimalist terms (1999a, pp. 105–8), but that effort succeeded only in explaining what it is for *atomic* propositions to correspond to reality. It seems unlikely that it can be extended to propositions with complex logical structures. Perhaps for this reason, Horwich takes a different line in the present volume, devoting an entire chapter to criticizing the idea that a true proposition owes its truth to the existence of a fact, and to attempting to explain correspondence intuitions away. Among other things, he defends the familiar Fregean idea that 'fact' just means 'true proposition', and he maintains that to explain why the proposition that Mars is red is true, it suffices to observe that Mars is red — there is no need to invoke the fact that Mars is red, or a correspondence relation linking the proposition to the fact. These discussions are interesting, but as I see it, they do not do justice to the alternative conception of facts that Russell defended, according to which facts are complexes of objects and properties. As is widely recognized, this conception of facts is of considerable importance — it is presupposed by much of what we say about causal relations. Accordingly, we are committed to it independently of any intuitions we may have about correspondence and truth. But given that there is this independent motivation for believing in Russellian facts, it is perfectly natural and appropriate to ask whether facts of this sort provide a basis for a robust account of correspondence intuitions. And there is good reason to think that the answer is 'yes'. (Given that we must accept Russellian facts in order to make sense of causal relations, the task of formulating a correspondence theory reduces to that of defining a relation of correspondence between Russellian facts and propositions. I argue elsewhere that this can be done using substitutional quantification (see *Thought and World: An Austere Portrayal of Truth, Reference, and Semantic Correspondence*, Cambridge: Cambridge University Press, 2002, chapter 3).

Further, it seems that Horwich still has trouble with Gupta's generalization objection. Clearly, there is no *logical* guarantee that someone who is disposed to accept every particular proposition of the form *a is an F and a is a G* will also be disposed to accept the generalization *every F is a G*. Accordingly, if it is true that dispositions of the first sort are accompanied by dispositions of the second sort in certain cases, this will have to be true in virtue of a *psychological law*. But if there is a law of the given sort, then there must be a causal process that is operative in such cases, a process that leads from individual dispositions involving particular propositions to dispositions to accept appropriate generalizations. Unfortunately, Horwich does not say what this process is. This is a serious omission, especially because there is reason to doubt that a process of this sort could be rational unless there was something explicitly or implicitly general about the rationale for acceptance that figures in the dispositions to accept particular propositions.

To elaborate, suppose that an agent has a disposition to accept the particular proposition (P):

(P) The proposition that if torture is wrong, then torture is wrong is a proposition of the form *if p, then p*, and that proposition is true.

Clearly, if an agent has this disposition, it will be because (i) the agent is disposed to accept the proposition that if torture is wrong, then torture is wrong, (ii) the agent is disposed to appreciate that this proposition is a proposition of the form *if p, then p*, and (iii) the agent is disposed to accept the following instance of (T):

> The proposition that (if torture is wrong, then torture is wrong) is true just in case (if torture is wrong, then torture is wrong).

But how could these highly specific dispositions involving particular propositions possibly serve as a rationale for the *generalization* that all propositions of the form *if p, then p* are true? The answer is that they cannot. To turn (i)–(iii) into a rationale for the generalization, we would have to replace its highly particularized propositions with generalizations. For example, it would be necessary to replace (iii) with the claim that the agent is disposed to accept the generalization *For all p, the proposition that p is true just in case p*. Moreover, the situation would not change if the agent was disposed to accept an infinite number of propositions of form (P), provided that the dispositions in question were grounded in local, particularized dispositions like (i)–(iii). In view of these considerations, it is clear that at the very least, Horwich owes us a more detailed version of his reply to Gupta's objection.

It may be useful to reformulate this objection. It depends on four main claims. First, if there is a psychological law of the sort that Horwich has in mind, linking sets of dispositions to accept particular propositions to dispositions to accept appropriate generalizations, then there must be a causal process that explains why the law holds. Second, Horwich owes us a description of this process. More specifically, he owes us a description of how the process takes the basic dispositions posited by his theory of truth as inputs, and transforms them into dispositions to accept generalizations about truth. Third, the description should provide a basis for classifying beliefs formed by the process as rational. This is because it is obviously rational to accept generalizations like (A). And fourth, Horwich does not provide such a description, and it seems very unlikely that one could be given. Since the basic dispositions that the theory posits are all dispositions like (i)–(iii), that is, highly specific dispositions involving particular propositions, they do not provide a sufficient basis for explaining dispositions to accept generalizations. They are not sufficiently general to do the necessary explanatory work.

It is worth emphasizing that there are other deflationary theories that have no problem with explaining acceptance of generalizations about truth. This is true of all theories that explain truth in terms of propositional quantification.

So far we have been concerned only with questions about the explanatory adequacy of Horwich's theory of truth. But there are also grounds for concern about his criticisms of other theories. Thus, for example, he dismisses deflationary theories that are based on propositional quantification by charging that they are circular, since any explanation of propositional quantification must invoke the concept of truth (p. 25). For the case of propositional quantification that is substitutional in character, this view was shown to be false some time ago (Hill 2002, pp. 17–22). As with the more familiar logical connectives, it is possible to explain substitutional quantifiers by describing their roles in inference — that is, by stating introduction and elimination rules. (Horwich briefly considers the option of explaining substitutional quantification in terms of rules of inference in another work, but he erroneously concludes that that any such approach would be unsatisfactory. See pp. 25–6 of Horwich's, *Truth*, 2nd edition, Oxford: Oxford University Press, 1998.)

Horwich also maintains that propositional quantification is foreign to our conceptual scheme. This seems questionable in view of the availability of propositions like these:

(a) When Bill claims that matters are arranged in such and such a way, then it always turns out that matters are in fact arranged in that way — no matter what being arranged in that way may involve.

(b) If the content of a thought is that matters stand thus and so, then the thought is true just in case matters really do stand thus and so. This holds for any thought whatsoever.

To be sure, as Horwich points out, there are various ways of articulating the logical structures of claims of this sort, and not all of them represent such claims as making use of propositional quantification (p. 25). While acknowledging this point, however, we should also acknowledge that propositional interpretations are very much in the running. After all, it is arguable that the introduction and elimination rules that govern the relevant quantifiers are propositional in character. (Thus, for example, starting with (a) as a premiss, it is clearly possible to infer *If Bill claims that Biden likes trains, then Biden does in fact like trains.*) Propositional interpretations cannot be dismissed with a flick of the pen. (For further discussion, see Hill 2002, pp. 24–7.)

I turn now to the other main component of Horwich's philosophical system, his elaboration of Wittgenstein's doctrine that meaning is use.

The use of any word is governed by a number of different patterns or regularities, but in each case, Horwich maintains, it is possible to find a single regularity, or a small set of regularities, that is explanatorily fundamental, in the sense that it provides a sufficient basis for explaining all of the others. Here are a couple of examples of explanatorily fundamental regularities that Horwich cites in other works:

(c) The acceptance property that governs a speaker's overall use of 'and' is the tendency to accept 'p and q' if and only if the speaker accepts both 'p' and 'q'.

(d) The acceptance property that governs a speaker's overall use of 'red' is the disposition to apply 'red' to an observed surface when and only when it is clearly red.

According to Horwich, it is possible to explain all of the uses of 'and' in terms of (c), and possible to explain all of the uses of 'red' in terms of (d). Or rather, it is possible to explain all of the uses of these words when (c) and (d) are combined with certain basic laws of psychology and certain assumptions about the learning histories and psychological constitutions of individual users of the terms (see *Meaning*, Oxford: Oxford University Press, 1998, p. 45; and *Reflections on Meaning*, Oxford: Oxford University Press, 2005, pp. 21–7.)

Horwich focuses on regularities that are explanatorily fundamental because he holds that they constitute meanings. Here is one of his formulations of this view:

> [I]n looking for the property of a word that constitutes its meaning we should be looking for something whose possession — in conjunction with other factors (such as environmental, psychological laws, and meaning-constituting properties of other words) — will explain the various conditions in which the various sentences containing the word are accepted and rejected. (p. 174)

But why should we say this? What is the motivation for maintaining that meanings are constituted by regularities of use that are explanatorily fundamental? Horwich maintains that this view is warranted by the general rule that a phenomenon is constituted by the property that explains its characteristic symptoms or effects. We are following this rule, for instance, when we say that the property *being made of water* is constituted by the property *being made of H_2O molecules*, and cite as our reason the fact that the latter property suffices to explain all of the effects of the former. Now the symptoms or effects of the meaning of a word are its characteristic uses or deployments. Accordingly, applying the rule to the case of meaning, we should say that the meaning of a word is constituted by whatever it is that explains its characteristic uses. But this will of course be the regularity governing the use of the word that is explanatorily fundamental. (p. 107, pp. 129–30)

In addition to laying out the basic structure of this theory of meaning, Horwich undertakes a number of related tasks, including responding to the Kripkenstein paradox, explaining how a regularity-based theory can accommodate the idea that the uses of words are governed by linguistic rules, constructing a multi-dimensional account of linguistic normativity, criticizing the idea that compositionality poses a fundamental obstacle to theories that identify meaning with use, and explaining how a use-based theory can answer the Frege-Geach objection to expressivist theories of the meanings of normative terms.

Among its many virtues, Horwich's explanationist theory of meaning contains one of the very few good ideas that have thus far appeared as to how Quine's critique of meaning might be answered. Fodor and Lepore have argued that any use-based account of meaning must claim that certain of the sentences containing a term are analytic, thereby making the account vulnerable to Quine's strictures against analyticity (see Jerry A. Fodor and Ernest Lepore, *Holism: A Shopper's Guide*, Oxford: Blackwell, 1992). Horwich's theory avoids this problem, for he shows how we might represent certain sentences as meaning-constituting without claiming that they are analytic. According to the theory, it suffices to show that the sentences figure in regularities that are explanatorily fundamental. A sentence can have this status without possessing the absolute immunity to revision that analyticity has been thought to entail. To have worked this out is an outstanding achievement.

Inevitably, however, there are a number of problems that remain to be addressed. I will mention one. If it is required that a set of regularities must explain *all* uses of a term in order to count as meaning-constituting, then it must explain uses that are incorrect or erroneous. But a set of regularities that explains erroneous uses will be quite complex, and will in general fail to sustain our intuitions about the meanings of particular words. To see this, consider a case in which an agent's disposition to use the word 'wolf' has been primed by an earlier conversation, and in which the agent has a visual experience as of a coyote. Suppose that this combination of factors causes the person to exclaim 'Wolf!' This is an error. The exclamation occurs because the agent's thought and speech are governed by the following regularities:

> If *A* has an experience as of a wolf, then *A*'s disposition to exclaim 'Wolf!' is activated to degree *D*.
>
> If A has an experience as of a coyote, then A's disposition to exclaim 'Wolf!' is activated to degree *D* minus *X*.
>
> *A*'s disposition to exclaim 'Wolf!' can be activated to degree *X* by priming.

> The level of activation of a linguistic disposition on a particular occasion is the sum of the activation that is due to current experience and the activation that is due to priming.
>
> If *A*'s disposition to exclaim 'Wolf!' is activated to degree *D*, A exclaims 'Wolf!'

Suppose now that we join Horwich in supposing that a set of regularities must explain all of the uses of a word in order to count as constitutive of its meaning. Then we will feel obliged to cite all of these generalizations in accounting for the meaning that our agent assigns to 'wolf'. But this seems wrong. If all of the generalizations were constitutive of the meaning of 'wolf', then it would be very hard to see why 'wolf' should be thought to mean *wolf*— as opposed, say, to *wolf or coyote* or *wolf or coyote-seen-after-a-conversation-about-wolves*. This problem is quite general in its significance, for priming can cause any word to be misused. Moreover, priming is just one of a number of sources of error. For example, any word can be triggered by a cognitive malfunction, and many words can be triggered by emotional interference with the cognitive system. (For a theory of meaning that is like Horwich's in a number of respects, but that is immune to the present objection, see my forthcoming 'Concepts, Teleology, and Rational Revision' (to appear in Albert Casullo and Joshua Thurow (eds), *The A Priori in Philosophy*, Oxford: Oxford University Press).

The problem can be restated as follows: if we need to include a generalization linking uses of 'wolf' to coyotes in the explanatory base that accounts for uses of 'wolf', then, on the assumption that 'wolf' means *wolf* rather than something more complex, there has to be a factor other than explanatory relevance that plays a role in determining whether a generalization is meaning-constituting. But Horwich appeals only to explanatory relevance and simplicity in explaining meaning. So, unless simplicity can do the necessary work, Horwich's theory of meaning is incomplete. There is a missing principle.

Can simplicity do the necessary work? More specifically, is it possible to use appropriate principles of simplicity to squeeze the uses of a term that count as data into a more or less uniform assortment, thereby eliminating such erroneous uses as exclamations of 'Wolf!' in the presence of coyotes? Alternatively, is it possible to apply principles of simplicity directly to the class of explanatory generalizations, thereby reducing them to a more or less uniform set? (On this approach, exclamations of 'Wolf!' in the presence of coyotes would be accepted as data, but it would no longer be an ambition of a theory of meaning to explain all of the data.) Both of these approaches would block the foregoing objection, by eliminating the generalization linking exclamations of 'Wolf' to coyotes from the class of meaning-constituting principles. But both of the approaches face serious problems. In the first place, to adopt either of them would require changing Horwich's theory, for as it stands, the theory explicitly claims that the meaning of a word is constituted by the minimal set of generalizations that is needed to explain *all* of the uses of the word. If our ambition is to explain *all* of the uses of 'wolf', it is clear that we need to deploy *all* of the foregoing generalizations. Secondly, more importantly, the first approach goes against the time-honoured methodological injunction to explain *all* of the data unless there are independent grounds for dismissing some of them. According to this injunction, Horwich could not be entitled to ignore exclamations of 'Wolf!' in the presence of coyotes unless he had an *independent* ground for dismissing those exclamations as erroneous. But it is hard to see how he could fund a notion of *semantic error* without *first* giving a theory of meaning. Thirdly, the second approach has essentially the same flaw as the first, for it involves a neglect of data. To neglect data is to abandon science.

Horwich clearly thinks that simplicity should play a large role in shaping a theory of meaning, but for the reasons just given I do not see how it can be of any great help in dealing with the problem of error.

Unfortunately there is not space to discuss the other contributions of this many-splendoured book. I will just note that in addition to the material that is described above, there are chapters on epistemic norms, the nature of paradox, and the tension between various forms of realism and anti-realism. (I have benefited considerably from Paul Horwich's extensive comments on an earlier version of this review.)

Department of Philosophy
Brown University
Providence, RI 02912
USA

CHRISTOPHER S. HILL

doi:10.1093/mind/fzr085