Reducing Truth through Meaning

1 In a number of books and articles, Horwich has claimed that meaning-properties are naturalistically reducible but sentential truth is not.[1] For Horwich, this difference is of great philosophical importance. He (2010: 15–6 & 103–4) claims that the literal content of the deflationist metaphor that truth is insubstantial is that truth is naturalistically irreducible. So, for Horwich, meaning is substantial but truth is not. Price (1997: §2) long ago argued that Horwich's meaning reductionism and anti-reductionist deflationism are mutually inconsistent.[2] Despite the importance of this challenge for his position, which he has most recently expounded in his 2010 book *Truth-Meaning-Reality*, Horwich has never responded to it. This oversight is perhaps forgivable as Price's argument is readily seen to go wrong in the details; however, in this paper I will show that Price's argument can be repaired.

2 In this section I will introduce meaning reductionism and anti-reductionist deflationism. Before that, though, some preliminaries. First, following Horwich I will refer to meanings using small capitals; for example, the meaning of 'dog' is DOG. Second, as I will be primarily concerned with the meanings of predicates, it will be useful to have a special name for them; unimaginatively, I will use **predicate-meanings**. Third and finally, I will operate with an abundant notion of properties according to which every predicate-meaning stands for a property. This abundant stance is contentious, but is

[1] 1995; 1997: 99–106; 1998: 27–30, 68–71, ch.4; 2005: ch.3; 2010: ch.6.

[2] Field (1992: 326–8) hints that he is sympathetic to this conclusion, and McGrath (1997: 79–85) offers a different, but to my mind weaker, argument for it.

commonplace amongst deflationists;[3] indeed it just is deflationism applied to properties. At any rate, taking this stance will not prevent us from recognising certain properties as more "elite" than others should we so wish.

Meaning reductionism is Horwich's thesis that each **meaning-property** — e.g. *x means* DOG and *x means* BLUE — reduces to a naturalistic property.[4] Of course, until some account has been given of what it is to "reduce" one property to another and of what it is for a property to be "naturalistic", this definition of 'meaning reductionism' is somewhat schematic. Much of this essay is spent considering some broad ways in which 'naturalistic reduction' might be understood.

Anti-reductionist deflationism is the conjunction of the following two theses:

(i) An adequate account of sentential truth is given by the schema

(T) $\forall x(x \text{ means } \textsc{p} \rightarrow (x \text{ is true} \leftrightarrow p))$[5]

(ii) Sentential truth is not reducible to a naturalistic property

(i) provides us with a positive account of truth; (ii) is Horwich's way of filling out the idea that truth is "insubstantial". Horwich has endorsed anti-reductionist deflationism in a number of places.[6] This may come as a surprise to some readers: it is well known that Horwich directs his deflationism primarily at the level of propositions, whereas anti-reductionist deflationism as characterised above is a thesis about sentential truth. However, Horwich

[3]For evidence that Horwich would likely be happy with abundantism, see 1998: 21, 1990: 37.

[4]Horwich subscribes to this position in 1995: 356; 1997: 106; 1998: 5; 2005: 63; 2010: 107–8.

[5]Horwich sometimes (e.g. 2010: 103–4 fn.5) writes this schema as: $\forall x(x \text{ means } \langle p \rangle \rightarrow (x \text{ is true} \leftrightarrow p))$.

[6] 1995: §1; 1998: 28–9, 70–1, ch.4; 2005: 74–5; 2010: 103–4. Occasionally Horwich weakens his position to the claim that we have no reason to expect a naturalistic reduction of truth (e.g. 2005: 75). This paper will show that even this weaker stance is incompatible with meaning reductionism: if we have a reduction of meaning then we have *every* reason to expect a reduction of truth.

makes it clear that he extends his deflationism to sentential truth; for example, he has said that according to his deflationism "sentential truth is not a deep 'substantive' property" (2010: 103) and then went on to explain that by 'substantive' he means naturalistically reducible (2010: 104). In the remainder of this paper I will mean sentential truth by the unmodified 'truth' and its cognates.

Although this paper is concerned with the incompatibility of meaning reductionism and anti-reductionist deflationism, the property of truth will be little discussed; instead, we will argue at one remove in terms of the relation *true of*. In an exact analogy with truth, anti-reductionist deflationism about *true of* can be defined as the conjunction of:

(i) An adequate account of *true of* is given by the schema

(S) $\forall x(x \text{ means } \textsc{f} \rightarrow \forall y(x \text{ is true of } y \leftrightarrow Fy))$

(ii) *True of* is not reducible to a naturalistic relation

As Horwich (1995: 360) points out, anti-reductionist deflationism about truth and *true of* are interdefinable, and so anti-reductionist deflationism about the one entails anti-reductionist deflationism about the other. Consequently, showing that meaning reductionism is incompatible with anti-reductionist deflationism about *true of* suffices to show that it is also incompatible with anti-reductionist deflationism about truth.

3 Without further ado we now turn to Price's argument that meaning reductionism and anti-reductionist deflationism are mutually inconsistent:

> If the fact that x means F reduces to a naturalistic fact about x, can't we define a naturalistic "true of" relation by
>
> (1) $P(x, y) =_{df}$ For some F, x means F and Fy?
>
> (Here F is to have the same range as it does in Horwich's discussion.) Doesn't Horwich's own view that the fact that x means F

> has an underlying naturalistic nature commit him to the conclusion that P, too, can be reduced to a naturalistic property in this way? For the fact that x means F reduces to something naturalistic in nature ensures that there is nothing irreducibly semantic on the RHS of (1) (at least when F isn't itself a term for some semantic property). (Price 1997: 114)[7]

I enthusiastically agree with the spirit of Price's argument, but it fails in the letter. According to Price, Horwich claims that each property of the form *x means F* is naturalistically reducible, but this is not quite right: Horwich's meaning reductionism is the thesis that all properties of the form *x means* F — e.g. *x means* BLUE — are naturalistically reducible. If we are to understand Price as addressing Horwich's position, then, we must read every occurrence of 'x means F' as 'x means F'. This may seem like a trivial modification, but it is not. When read in this way, (1) binds both 'F' and 'F' with a single quantifier; but 'F' takes predicate-meanings as values whereas 'F' takes properties as values, and so binding both of them with one and the same quantifier is no more legitimate than quantifying into quotation marks, that-clauses or Horwich's (1990: 19 fn.3) angle-brackets. It is for precisely this reason that Horwich was forced to use the schemata (T) and (S) rather than their universal closures.[8]

Price might hope to escape this problem by insisting that we read (1) as

(1′) $P(x, y) =_{df}$ For some F and some F, x means F and Fy and F stands

[7] I have changed every occurrence of 'non-semantic' in this quotation for 'naturalistic' to bring it in line with the terminology I have borrowed from Horwich. I have also made trivial changes in the notation to avoid clashes with that of my paper.

[8] In fairness to Price, at one time Horwich (1998: 5) did propose identifying each predicate-meaning F with the property it stands for F. If we assume this then (1) becomes less problematic. However, Horwich later retracted this thesis in his (2005: 33–4), and so Horwich's 2005 timeslice is able to resist Price's argument. And setting aside Horwich's change of mind, Price's argument as it stands at most establishes that anti-reductionist deflationism, meaning reductionism and the thesis that in general F is identical to F are together inconsistent, not the advertised conclusion that the first two theses are inconsistent with each other.

for F

But as of yet we have no reason to think that the relation F *stands for* F is reducible to something naturalistic. Indeed, as F *stands for* F seems to be (at least analogous to) a truth-theoretic relation, an anti-reductionist deflationist is likely to insist precisely that it is irreducible. Consequently, even if Price reads (1) as (1′), he still would not have established that a *true of* relation is naturalistically reducible.

The rest of this paper will be spent developing a working version of Price's argument.

4 We start by putting up front what Price hid behind parentheses. In this paper I will limit my attention to predicate-meanings which stand for naturalistic properties. As a consequence, I will only really show that meaning reductionism is incompatible with anti-reductionist deflationism about the restricted *true of* relation defined over those predicate-meanings. This limited conclusion is still of substantial importance. It would show that given meaning reductionism, *true of* is not in itself naturalistically irreducible; if it is naturalistically irreducible, it is because there happen to be other naturalistically irreducible properties causing trouble. From now on I will leave this restriction implicit.

With this matter dealt with, we can turn to the serious business of fixing Price's argument. As a first shot at fixing Price's argument we might suggest the following:[9]

> Horwich's meaning reductionism is going to require a reduction of the relation *x means* F. But the relation *x means* F is surely only going to be reducible to a relation of the form
>
> (2) $R(x, F)$

[9]Price has made precisely this suggestion to me in correspondence.

where 'F' takes naturalistic properties as values. For example, R might be *we are disposed in the right conditions to apply x to all and only Fs*, or perhaps *there is an appropriate law correlating the application of x with the property F*. I will call such a reduction of meaning **strongly relational**.

With a strongly relational reduction of meaning in hand, we can reduce *true of* to

(3) For some F, $R(x, F)$ and Fy

Unlike (1), the use of the quantifier in (3) is not (obviously) illicit:[10] by reducing *x means* F to (2) the meaning reductionist exchanged the variable 'F' for 'F', and so the quantifier in (3) is only being used to bind one type of variable.

But as tempting as this line of thought might be, it will not impress Horwich: he has repeatedly denied that the meaning reductionist needs to offer a strongly relational reduction of meaning.[11] Horwich has developed two lines of justification for this denial; I will consider them in turn over the next two sections.

5 First, Horwich has claimed that the meaning reductionist does not need to offer a reduction of the relation *x means* F at all. Horwich's meaning reductionism is the thesis that each meaning-property is naturalistically reducible, and he has argued that this thesis does not entail that the relation *x means* F is also reducible.[12] I will call this reduction of meaning **piecemeal**.

I am far from convinced that a piecemeal reduction of meaning is really a reduction of meaning at all. One obvious problem with such a reduction is

[10] It is, of course, controversial to bind predicate variables with quantifiers at all. However, given the assumption that predicates refer to properties, which Horwich (e.g. 1998a: 21) grants, such quantification is permissible.

[11] Indeed, he has denied it precisely to avoid a variant of this argument; see 1998: 65–8; 2005: 73–4.

[12] 1998: 25–7, 65–8; 2005: 73.

that it is not at all clear how it will handle a property like *there is some* F *such that x means* F. But for the sake of argument I will grant the legitimacy of a piecemeal reduction of meaning; I suppose that when we squint, the relation *x means* F looks like a kind of sum of all the meaning-properties, and so once they have all been reduced there is nothing of meaning left waiting above them.

Instead, I will show that even a piecemeal meaning reductionism is inconsistent with anti-reductionist deflationism. One obvious way of trying to argue for this conclusion is as follows:

> To begin with, we can identify *true of* with the disjunction of the relations of the form
>
> (4) x means F and Fy
>
> In other words, *true of* is identical to
>
> x means DOG and y is a dog
>
> or x means BLUE and y is blue
>
> or ...[13]
>
> Of course, this is not yet a *naturalistic* reduction of *true of*, as relations of the form (4) are obviously semantic; consequently, we have so far said nothing to worry the anti-reductionist deflationist, who is only concerned with naturalistic reductions of *true of*.
>
> However, given meaning reductionism we can convert the above semantic analysis of *true of* into a naturalistic one. Meaning re-

[13]Thanks to an anonymous reviewer for pointing out that it does not follow from (S) alone that *true of* reduces to, or is even co-extensive with, the disjunction of the relations of the form (4). (S) does not rule out the possibility that there is some x and y such that x is true of y but there is no F such that x means F. This is, however, a shortcoming of (S): that possibility should be ruled out. Horwich (2001: 158 n.21) is aware of this general problem, and his strategy, which is surely correct, is to add the principle 'For any x and y, x is true of y only if there is some F such that x means F' to his deflationist account of *true of*.

ductionism entails that each relation of the form (4) reduces to the corresponding relation of the form

(5) $U_{\text{F}}x$ and Fy

where in general 'U_{F}' refers to the naturalistic property to which the meaning-property *x means* F reduces; e.g. U_{DOG} is the naturalistic property to which *x means* DOG reduces. For example, meaning reductionism entails that the (part) semantic relation

(4′) x means DOG and y is a dog

reduces to the naturalistic relation

(5′) $U_{\text{DOG}}x$ and y is a dog

So, given meaning reductionism we can reduce *true of* to the disjunction of the relations of the form (5), i.e.

$U_{\text{DOG}}x$ and y is a dog

or $U_{\text{BLUE}}x$ and y is blue

or ...

And that is a *naturalistic* reduction.

However, few philosophers would be satisfied with this argument. There are infinitely many relations of the form (5),[14] and so the disjunction of those relations will be infinitely long. It is generally accepted that there are no legitimate infinitely disjunctive reductions. This ban on infinitely disjunctive reductions can be motivated in a number of ways: for example, Horwich[15] and many others think that a reduction ought to play an explanatory role, and it seems that an infinitely disjunctive reduction could not be appropriately

[14]That is assuming that there is a definite totality of them at all; if there are not then this reduction is in even worse shape.

[15]1995: 356; 1998: 25; 2005: 77; 2010: 106–7.

explanatory. But whatever the motivation, if we ban infinitely disjunctive reduction then the above argument must be abandoned.[16]

We will concede that the above argument goes wrong; it is important to pause on exactly *where* it does so. Let's call the relations of the form (4) ***true of*-relations**. So long as we assume meaning reductionism, I take it to be uncontroversial that we can reduce each of the *true of*-relations to relations of the form (5). The difficulty is in linking up all of these reductions to get a reduction of *true of* itself.

But let us recall, this difficulty arose only because we allowed the meaning reductionist a piecemeal reduction of meaning. And if a piecemeal reduction of meaning is reduction enough, then so is a piecemeal reduction of *true of*. In that case, there is no need to disjoin all of the reductions of the *true of*-relations to provide a unified reduction of *true of*. Instead, we reduce *true of* piecemeal by reducing each of the *true of*-relations. To go back to the metaphor guiding piecemeal reductions: when we squint, the relation *true of* looks like a kind of sum of all the *true of*-properties, and so once they have all been reduced there is nothing of *true of* left waiting above them.

So, even if we allow meaning reductionism to be piecemeal, it is incompatible with anti-reductionist deflationism.

6 Second, Horwich (2005: 74) has argued that even if the meaning reductionist does need a reduction of the relation *x means* F in addition to her reductions of the meaning-properties, there is no reason why that reduction should be strongly relational. Horwich proposes identifying each predicate-meaning, F, with the naturalistic property to which the corresponding meaning-property, *x means* F, reduces;[17] i.e. DOG is to be identified with U_{DOG}, BLUE with U_{BLUE}, and so on. Then, Horwich argues, we can reduce the relation *x means* F to:

[16] A ban on infinitely disjunctive reductions will also block the identification/reduction of *true of* to the disjunction of instances of (4).

[17] This marks a change from his 1998 (20–1) position I mentioned in fn.8.

(6) x instantiates U

where 'U' is a variable which takes as its values the naturalistic properties to which meaning-properties reduce.[18] (6), Horwich (2005: 74) tells us, is just as naturalistic as

(6′) Ux[19]

As this reduction of meaning involves introducing a variable ranging over the reductions of the meaning-properties, I will call it **higher-order**; any reduction which is not higher-order, e.g. a strongly relational reduction of meaning, is **first-order**.[20]

On the face of it, a higher-order reduction of meaning is no more compatible with anti-reductionist deflationism than a piecemeal reduction was. To repeat, given meaning reductionism we can unproblematically reduce each *true of*-relation to a relation of the form (5), and so can supply the following higher-order reduction of *true of*:

(7) There is some S such that $S(x, y)$

where 'S' is a variable which takes all relations of the form (5) as values.

But Horwich may criticise this putative reduction as follows:

> What reason do we have to think that from a naturalistic perspective the various relations of the form (5) have any (non-infinitely disjunctive) feature in common? And if they don't, then reducing *true of* to (7) is really no more acceptable than the earlier suggestion of reducing *true of* to the disjunction of the

[18] In fact, Horwich suggests reducing *x means* F to *x instantiates y*. However, I take it that Horwich is charitably interpreted as meaning (6): it obviously does not follow from the fact that Horwich instantiates some property or other, for instance the property of being a human, that Horwich means something.

[19] In moving from (6) to (6′), Horwich is making a pun. 'U' in (6) is a first-order variable, but in (6′) it is a second-order one. For the sake of argument I will ignore this slide.

[20] It is worth noting that so defined, my uses of 'first-order' and 'higher-order' is not intended to line up with their uses in 'first-order logic' and 'higher-order logic'.

> relations of form (5). Compare the type-identity theorist trying to avoid the threat which the multiple realisability of pain poses to his theory by trying to reduce *x is in pain* to *x instantiates one of the realisations of pain*.

Horwich's own higher-order reduction of meaning is, however, open to an exactly analogous challenge. In (6), 'U' is a restricted variable which ranges over the naturalistic properties to which the meaning reductionist promises that the predicate meaning-properties reduce. If Horwich's use of this restricted variable is to be any better than an infinite disjunction, he must argue that the reductive bases of these meaning-properties have a (non-infinitely disjunctive) naturalistic feature in common which we can use to group them together.

But matters get worse for Horwich. It is hard to see how the reductive bases of the meaning-properties could have some (non-infinitely disjunctive) naturalistic feature in common unless there was a strongly relational reduction of meaning. After all, there seems to be no feature in common between the properties *x means* DOG, *x means* BLUE, *x means* ELECTRON, *x means* SPIRAL GALAXY... for a reductive theory to get a handle on other than that whatever instantiates one of them has an extension.[21] But as we already saw in §4, a strongly relational reduction of meaning is straightforwardly inconsistent with anti-reductionist deflationism.

The problem facing Horwich is best put in the form of a dilemma. Suppose that a higher-order reduction of meaning is legitimate only if there is a (non-infinitely disjunctive) naturalistic feature in common between the reductive

[21] Here is a more Horwich-specific presentation of this problem. Horwich (1998: esp. ch.2–3; 2005: ch.2) claims that the reductive bases of meaning-properties are lawlike "use-properties". What Horwich must do is find a way of describing the use of each meaningful predicates which is general enough that all of the uses have a (non-infinitely disjunctive) naturalistic feature in common. It is difficult to see what common use all predicates are put to apart from being applied to the members of their extension. But if we think of use-properties in this way then they will be of the form '$R(x, F)$'. Therefore, by reducing the meaning-properties to these use-properties, Horwich will be giving a strongly relational reduction of meaning.

bases of the meaning-properties. In that case the task of providing a higher-order reduction of meaning collapses into that of providing a first-order, non-piecemeal reduction: if the reductive bases of the meaning-properties had such a feature in common, then we could surely use this feature to reduce the relation *x means* F (without recourse to an infinite disjunction). But the only obvious first-order, non-piecemeal reduction of meaning — and certainly the only one mentioned by Horwich — is a strongly relational reduction, which is easily seen to be inconsistent with anti-reductionist deflationism.

Suppose instead that the legitimacy of a higher-order reduction does not depend on there being a (non-infinitely disjunctive) naturalistic feature in common between the reductive bases of the meaning-properties. In that case, a higher-order reduction of meaning is a variation of a first-order piecemeal one. In both cases, all the work of reducing meaning is done in reducing the meaning-properties: there is no further challenge of reducing the relation *x means* F. But again, we have already seen that a piecemeal reduction of meaning is inconsistent with anti-reductionist deflationism.

Either way, then, the higher-order reduction of meaning collapses into a first-order reduction; and either way, the detour via higher-orders has not helped Horwich defend himself from the charge of inconsistency.

7 It seems, then, that *true of* is reducible in whatever sense meaning is.[22] What, then, are Horwich's options? First he could stick to his guns and try to find a novel kind of naturalistic reduction which is neither strongly relational nor piecemeal. However, it is not at all clear what such a reduction would look like, and Horwich has certainly not so far given us any hint.

[22] I suppose we might wonder whether Horwich should be interpreted as saying that truth is not reducible in the sense in which meaning is, or whether he is using two different senses of 'reducible'. But there is textual evidence that Horwich meant the two senses to be the same (e.g. 2005: 34–5); he even goes so far as to say that "we have no reason to expect *any* sort of reductive analysis of the truth-theoretic properties and relations" (2005: 75, original emphasis). But even if Horwich did mean to use different senses of 'reducible', it is still a substantial lesson to learn that there are serious problems with saying that the meaning-properties are reducible but that truth isn't reducible in the same sense.

Second, Horwich might reply by abandoning his meaning reductionism.[23] But given the central role that meaning reductionism has played in Horwich's writings since *Meaning*, this would represent a major retreat on his part.

Third, Horwich might withdraw the claim that the truth-theoretic properties are naturalistically irreducible. If this entailed abandoning his deflationism, this response would be just as unacceptable for Horwich as giving up on meaning reductionism; however, I do not think it needs to be understood in this way. Horwich introduces the claim that *true of* and the rest are naturalistically irreducible to give some sense to the metaphor that they are insubstantial (2010: 16). If he can find some alternative way of explicating this metaphor or, perhaps better yet, of abandoning the metaphor altogether, then his deflationism need not be inconsistent with meaning reductionism.

I would, though, like to caution against one way of pursuing this third option. It may be tempting for Horwich to respond to the argument of this paper by conceding that sentential truth is naturalistically reducible, but insisting that propositional truth is not.[24] Of course, this response requires us not to see propositional truth as an abstraction from sentential truth; if we did, sentential truth would be the philosophically interesting notion. Propositions (and predicate-meanings) are, then, given a life outside of the context of meaning-properties. We no longer talk about a proposition only to say that a particular sentence expresses that proposition; they are objects of interest in their own right. Consequently, a satisfactory meaning reductionism could not just offer reductions of each of the meaning-properties. It must also include a general naturalistic account of propositions/predicate-meanings. Horwich, in fact, offers such an account (e.g. 2005: 33-4). But in that case it is not clear that anti-reductionist deflationism about propositional truth will be any more consistent with meaning reductionism than anti-reductionist deflationism about sentential truth was. It seems that the arguments of this paper could be rerun, but this time replacing all talk of

[23]This is Price's (1997: 114–5) recommendation.

[24]Simon Blackburn has recommended this option to me.

reducing the property *x means* F with talk of reducing the property *x is* F.

But however Horwich responds, he faces a challenge which he can no longer ignore.

References

1 Field H. 1992. Critical notice: Paul Horwich's *Truth*. *Philosophy of Science* 59: 321–30.

2 Horwich P. 1995. Meaning, use and truth. *Mind* 104: 355–68.

3 ——. Truth and the problem of aboutness. *Philosophical Issues* 8: 95–106.

4 ——. 1998a. *Meaning*. Oxford: Oxford University Press.

5 ——. 1998b. *Truth*. 2nd edition. Oxford: Clarendon.

6 ——. 2001. A defense of minimalism. *Synthese* 126: 149–65.

7 ——. 2005. *Reflections on Meaning*. Oxford: Oxford University Press.

8 ——. 2010. *Truth-Meaning-Reality*. Oxford: Oxford University Press.

9 McGrath M. 1997. Weak deflationism. *Mind* 106: 69–98.

10 Price H. 1997. What should a deflationist about truth say about meaning? *Philosophical Issues* 8: 107–15.