

Three Norms of Assertibility, or How the Moa became Extinct

Author(s): Huw Price

Source: *Philosophical Perspectives*, Vol. 12, Language, Mind, and Ontology (1998), pp. 241-254

Published by: Ridgeview Publishing Company

Stable URL: <https://www.jstor.org/stable/2676149>

Accessed: 10-06-2020 17:21 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Ridgeview Publishing Company is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Perspectives*

**THREE NORMS OF ASSERTIBILITY,
OR HOW THE MOA BECAME EXTINCT**

Huw Price
University of Sydney

Deflationism about truth combines two claims: (i) that truth is not a substantial property; and (ii) that the key to our use of the concept of truth lies in its disquotational character—i.e., in the fact that

(DS) “p” is true if and only if p

holds for all central declarative sentences. According to deflationists, the key to an understanding of truth lies in an appreciation of the grammatical advantages of a predicate satisfying DS. As Paul Horwich puts it, “our truth predicate is merely a logical device enabling simple formulations of certain sorts of generalization.” (1996, p. 878; see also Horwich 1990)

Deflationism has been criticised by Crispin Wright, who claims in *Truth and Objectivity* (1992) that such an account of truth cannot make sense of the fact that truth is a normative notion. Wright argues that deflationism is an inherently unstable position: It follows from premises that deflationism accepts that truth is a normative property, but this is incompatible with the deflationist’s “thin”, grammatical conception of the role of the truth predicate.

Horwich (1993, 1996) has responded to this charge in two reviews of *Truth and Objectivity*. He claims that on the contrary, the disquotational property of truth makes perfectly good sense of the various considerations to which Wright appeals, so that these considerations do not require that truth be a normative property, in any sense not already entailed by the deflationary view.

My own view, in a nutshell, is that Horwich wins the battle but Wright wins the war. I think that truth is normative, in a way not explained by the deflationary theory; but that Wright has not given us a good argument for this conclusion. In this paper I want to reinforce Horwich’s objections to Wright’s argument, but then to offer an alternative argument to the same conclusion. As I’ll explain, however, this conclusion does not require that we abandon (i) above, the claim that truth is not a substantial property. It simply requires that our explanation of

the folk use of the concept of truth should not be grounded solely on (ii), but rather needs to appeal to the utility within a speech community of the distinctive kind of norm that truth provides.

1. Wright's Anti-deflationary Argument

Wright argues that deflationism is incoherent, on the grounds that it follows from DS that truth is a norm of assertoric practice, in a sense incompatible with (i). What is it to be a norm? Wright distinguishes descriptive and prescriptive versions. A descriptive norm of a practice is a feature or characteristic which, as a matter of fact, is positively correlated with the choices made by participants in the practice in question—a characteristic possessed more often by chosen moves than by possible moves as a whole. A prescriptive norm is a feature or characteristic of a move in the practice which participants would take as providing a reason for the move in question.

Wright argues that it follows from DS that truth is both a prescriptive and (with one qualification) a descriptive norm of assertoric practice. Why is it prescriptively normative? Because “any reason to think that a sentence is T may be transferred, across the biconditional, into reason to make or allow the assertoric move [the sentence in question] expresses.” (1992, p. 17) In other words, any reason to believe that *p* is true is, in virtue of DS, a reason to believe (and hence to allow the assertion) that *p*. “And ‘T’ is descriptively normative,” Wright continues, “in the sense that the practices of those for whom warranted assertibility is a descriptive norm are exactly as they would be if they consciously selected the assertoric moves which they were prepared to make or allow in the light of whether or not the sentences involved were T.” To a substantial extent, Wright says, “any actual assertoric practice will be just as it would be if T were a self-conscious goal.” (1992, p. 17)

Having thus established to his satisfaction that truth is *a* norm for assertion, Wright goes on to argue that it is a distinctive norm—in particular, that it is not the same norm as warranted assertibility. This follows, he notes, from the fact that we take it that lack of warranted assertibility need not imply lack of truth.

2. Horwich's Response to Wright, and an Alternative

As I noted, Horwich has replied to Wright's argument in two reviews of Wright's *Truth and Objectivity*. Let's begin with the earlier response:

Given that “The statement *that p* is true” always entails “*p*”, then, without assuming anything at all about what sort of entity (if any) the truth predicate stands for, we can see that believing the former provides grounds for asserting the latter. Thus it is perfectly consistent to deny that truth is a substantive property yet accept the normative principle: a statement's being true is a good reason to assert it. Moreover, none of the philosophers who have advocated “deflationism” have ever denied that truth is a

distinctive norm of assertion. On the contrary, they would say, this normative principle is precisely the sort of generalization that our deflationary truth predicate is needed to formulate. For it is a way of capturing the infinitely many instances of the schema: “If p, then one should assert that p.” (1993, p. 28)

An immediate difficulty is that Horwich seems to run together two normative principles. In the first sentence of the above passage, Horwich endorses the principle

- (1) Believing that p is true provides grounds to assert that p.

In the second sentence (and by implication in the last), he endorses the principle

- (2) The fact that p is true provides a good reason to assert that p.

These principles are both plausible, at least if we take it for granted that each is to be heavily qualified. (For example, they shouldn’t be taken to imply that there is general imperative to assert everything one believes true, let alone everything *actually* true!) But they are clearly different principles: (1) refers to what we *believe*, (2) to what is actually the case.

In Horwich’s more recent response to Wright, the ambiguity disappears—Horwich opts consistently for principle (1). For example, he notes (1996, p. 880) that the deflationary truth predicate may be used to generalise the principle that if you believe that snow is white then you have reason to say “Snow is white”. “[C]olloquially”, as he puts it, “you have reason to assert any sentence you believe to be true.” (1996, p. 880)

All the same, the ambiguity seems worth remarking. The principle (1) embodies a very weak kind of normative principle, one which relates the appropriateness of an assertion only to a speaker’s own immediate doxastic state. It leaves out of account all issues concerning the justifiability, or correctness, of the speaker’s belief that p—even such issues as the coherence of the belief that p with the speaker’s other beliefs. In one sense, the weakness of this norm counts in Horwich’s favour. If the intuitive principles to which Wright appeals involve nothing more than this weak and uncontroversial notion of normativity, then it seems clear that they pose no threat to deflationism. But perhaps Horwich has overplayed his hand. Is it really plausible that the principle that, as he himself puts it, “truth is a distinctive norm of assertion” amounts to nothing more than this? Indeed, doesn’t (2) itself suggest that truth might be a norm of assertion in some other sense?

On the face of it, however, both (1) and (2) differ from the principle Wright himself invokes. As I glossed it above, this principle is

- (3) Any reason to believe that p is true is a reason to believe (and hence to allow the assertion) that p.

It should be clear that (3) is not a paraphrase of (1). Principle (1) turns on the idea that belief itself provides a reason for making an assertion, and so leaves no room for a reading in terms of reasons *for* belief. Nor is (3) a paraphrase of (2). Unlike (2), (3) involves the notion of reason for belief on the left hand side—(2) speaks of what flows from *the fact* that *p* is true, (3) of what flows from *a reason to believe* that *p* is true. Hence Wright has grounds to claim that in so far as Horwich states his case in terms of (1) and perhaps (2), he simply hasn't engaged with Wright's argument.

All the same, I think that Wright's argument is invalid as it stands, and for the reason underlying Horwich's objections: The intuitions to which Wright appeals reveal nothing more than the formal role of the truth predicate, as embodied in DS. To show this, I want to construct a parallel argument to Wright's, using a piece of terminology whose role is purely formal by definition.

In place of

“*p*” is true

let us write

True(*p*).

Let us now define an analogous construction, as follows:

Twice(*p*) $\equiv_{\text{defn.}}$ *p* and *p*.

In other words, “Twice(*p*)” is simply to be understood as an alternative logical notation for the conjunction “*p* and *p*”.

It follows immediately that

Twice(*p*) if and only if *p*.

In other words, this construction satisfies something analogous to DS—which means, in turn, that a reason for believing that Twice(*p*) provides a reason for believing (or being prepared to assert) that *p*. “Twiceness” appears to be operating as a norm of assertion! To complete the parallel, note that Twiceness cannot simply amount to warranted assertibility. The issue as to whether it is the case that Twice(*p*) is not the same as the issue as to whether we are warranted in asserting that *p*, for just the same reason that the issue as to whether *p* (or whether “*p*” is true) is not the same issue as that as to whether *p* is warrantably assertible.

I think that this makes it clear how little can be inferred from Wright's principle (3). The principle amounts to little more than the truism that a reason to believe that *p* is a reason to believe that *p*. It rests entirely on the formal substitutions licensed by DS, and doesn't show that truth is any kind of norm of assertion, in any sense not immediately explicable by deflationism.

However, it should be emphasised that this does not show that truth is *not* a norm of assertion. It simply means that (3) does not entail that truth *is* a norm of assertion. It might be that Wright has argued for the right conclusion by the wrong means. I want to argue that this is in fact the case. By way of background, however, I want first to try to further disentangle some of the confusing threads in the debate between Horwich and Wright.

3. Two Norms Weaker than Truth

In my view, a significant source of confusion is that there are two weaker norms of assertion, in addition to any distinctive norm of truth. The first and weakest norm is that embodied in (1). It relies on the principle that it is *prima facie* appropriate to assert that *p* when and only when one believes that *p*. (*Prima facie*, because of course many other factors may come into play, in determining the appropriateness of a particular assertion in a particular context.) Let's call this the norm of *subjective assertibility*. The easiest way to see that it has very little to do with truth is to note that it is exactly analogous to norms which operate with respect to utterances which we don't take to be truth-apt. It is *prima facie* appropriate to request a cup of coffee when and only when one wants a cup of coffee, but this doesn't show that requests or desires are subject to a norm of truth. (Like many other utterances, including assertions, they are subject to a norm of *sincerity*. At least to a first approximation, subjective assertibility simply *is* the norm of sincerity, restricted to the case of assertoric utterances.)

The second kind of norm is that of *warranted* or *objective* assertibility. Roughly, "p" is objectively assertible by a speaker who not only believes that *p*, but is *justified* in doing so. (Of course, there are different kinds and degrees of justification, some of them more objective than others. For example, is justification to be assessed with reference to a speaker's actual evidence as she sees it, or by some more objective lights? Here, for definiteness, let us think of it in terms of coherence—a belief is justified if supported by a speaker's other beliefs.) This norm is not equivalent to the norm of truth either, for the reason Wright points out. We can make sense of the possibility that *p* is true, even though, through lack of evidence, we are not justified in believing that *p*.

So far, then, we have two kinds of norm. Neither is a norm of truth, in the intuitive sense—neither provides a norm that a speaker can fail to meet, even if she speaks sincerely, on the basis of a justified belief. How then does truth get into the picture in such a confusing way, even with respect to these weaker norms? Simply by virtue of the fact that the disquotational schema makes it sound as if these norms do have something intrinsic to do with truth. Given DS, for example, it is natural to express the first kind of norm like this: It is *prima facie* appropriate to assert that *p* only when one believes that *p* is true—in other words, as (1) puts it, believing that *p* is true provides a reason for asserting that *p*.

But the reference to truth does no non-grammatical work here, as the parallel with requests makes clear. What we are being told about belief is exactly analo-

gous to what we are told by the principle that wanting a cup of coffee provides a reason for requesting one. In neither case does truth come into it, except as a matter of grammatical felicity. If it is this kind of normative restriction we have in mind—i.e., subjective assertibility—it is quite misleading to describe it as the idea that truth is a distinctive norm of assertion. It is misleading in just the way that it would be misleading to say that Twiceness is a distinctive norm of assertion. True, it is possible to characterise the norm concerned in a way which uses the construction “Twice()”, but this simply reflects the logical properties of the term, and doesn’t connect it in any important way with the norm itself.

Similar remarks apply in the case of objective assertibility. In sum, then, we have two norms of assertibility—*subjective* and *objective* assertibility—in addition to any third norm of truth itself. Neither Horwich nor Wright properly identifies and distinguishes these two norms. Had Horwich done so, I think it would have seemed less plausible to him that (1) might exhaust the sense in which truth is a distinctive norm of assertion. Had Wright done so, the fact that (1) embodies a norm of some kind would surely have made it less plausible to argue that DS implies that truth is a norm as strong (or stronger) than warranted assertibility.

4. Why Wright Is Right After All—Truth as a Third Norm

In my view the real objection to the deflationary view comes from the fact that there is a third kind of norm for assertion, which—unlike the two norms distinguished above—is linked to the notion of truth in an intrinsic way. Ironically, this third norm turns out to be the one most plausibly associated with Horwich’s generalisation “If p , then one should assert that p ”, and principle (2). This principle does not say merely that *if one believes that p* , one should assert that p (“subjective assertibility”); or even that *if one has good evidence that p* , one should assert that p (“objective assertibility”). It says (as we would ordinarily put it), that *if p is true*, one should assert that p .

It might be doubted first, whether there is really any such normative principle, distinct from those of subjective and objective assertibility; and second, whether, even if so, it has anything more to do with truth than the two weaker norms. I’ll defer the second doubt for the moment (see §5 below), and concentrate on the first.

The best way to bring the third norm into focus is consider its negative form:

If not- p then it is *incorrect* to assert that p .

The crucial point is that there is a norm of assertion which a speaker may fail to meet, *even if she does meet the norms of subjective and objective assertibility*. We judge a speaker *wrong, incorrect, mistaken*, when we judge her assertion false, even if we are in no doubt that she is sincere, and in possession of the kind of evidence that would lead any reasonable person to make the same mistake.

One of the reasons why this third norm is hard to distinguish from the two weaker norms of assertibility is that when we apply it in judging a fellow speaker

right or wrong, the basis for our judgement lies in our own beliefs and evidence. It is not as if we are in a position to make the judgement from the stance of reality itself, as it were. I think this can make it seem as if application of this norm involves nothing more than re-assertion of the original claim (in the case in which we judge it correct), or assertion of the negation of the original claim (in the case in which we judge it incorrect). Construed in these terms, our response contains nothing problematic for the deflationary view, of course. Re-assertion of this sort is one of the linguistic activities disquotational truth facilitates.

But our response is not merely re-assertion, or assertion of the negation of the original claim. If it were, it would involve no commendation or criticism of the original utterance. This is hard to see, but the crucial point is that we can imagine a linguistic practice which allowed re-assertion and contrary assertion, *without this third normative dimension*. That is, we can imagine a linguistic community who use sentences to express their beliefs, but for whom disagreements have no normative significance, except in so far as it is related to the norms of subjective and objective assertibility.

What we need is the idea of a community who take an assertion—or rather the closest thing they have to what we call an assertion—to be *merely* an expression of the speaker's opinion. The relevant idea is familiar in the case of expressions of desires and preferences. It is easy to imagine a community—we are at least close to it ourselves—who have a language in which they give voice to psychological states of these kinds, not by reporting that they hold them (which depends on assertion), but directly, in linguistic forms tailored specifically for this purpose. In such a community we would expect a norm analogous to subjective assertibility: essentially, a normative requirement that speakers use these expressions sincerely. Less obviously, such a practice might also involve a norm analogous to objective assertibility. In other words, expressed preferences might be criticised on the grounds that they were not well-founded (for example, on the grounds they did not cohere with the speaker's other preferences and desires). However, in this practice there need be no place for a norm analogous to truth—no idea of an objective standard, over and above objective assertibility, which preferences properly aim to meet.

At least to a first approximation, we can imagine a community who treat expressions of beliefs in the same way. They express their beliefs—i.e., the kind of behavioural dispositions which we would characterise as beliefs—by means of a speech act we might call the *merely-opinionated assertion* (MOA, for short). These speakers—"Mo'ans", let's call them—criticise each other for insincerity (lack of subjective assertibility) and for lack of coherence, or objective assertibility. But they go no further than this. In particular, they do not treat a disagreement between two speakers as an indication that, necessarily, one speaker or other is mistaken (i.e., in violation of some norm). On the contrary, they allow that in such a case it may turn out that both speakers have spoken correctly, by the only two standards the community takes to be operable. Both may be sincere, and both, in their own terms, may have good grounds for their assertion.

This speech community could quite well make use of deflationary truth, for example as a device to facilitate agreement with an expression of opinion made by another speaker. “That’s true” would function much like “Same again”, in the context in which a group of customers is placing orders in a restaurant. Just as “Same again” serves to indicate that one has the same preference as the previous speaker, “That’s true” would serve to indicate that one holds the same opinion as the previous speaker. The crucial point is that if the only norms in play are subjective and objective assertibility, introducing disquotational truth leaves everything as it is. It doesn’t import a third norm.

The difficulty we have in holding on to the idea of such a community stems from our almost irresistible urge to see the situation in terms of our own normative standards. There really is a third norm, we are inclined to think, even if these simple creatures don’t know it. When they make incompatible assertions, at least one of them must be objectively incorrect—must have spoken *falsely*—even if by their lights they both meet the only norms they themselves recognise. But the point of the story is precisely to bring this third norm into sharp relief, and hence I am quite happy to allow challenges to the story on these grounds, which rely on the very conclusion I want to draw: *For us*, truth does operate as a norm, in a way in which its disquotational function alone does not explain.

Thus I have argued that assertion is subject to (at least) three different norms. It may be helpful to formulate these norms in the form of explicit rules. As before, it is most useful, I think, to take them in negative form—as formulations of the various ways in which an assertion may be incorrect, or wrong:

- (4) One is incorrect to assert that p if one does not believe that p.
- (5) One is incorrect to assert that p if, though one believes that p, one does not have adequate grounds for believing that p.
- (6) One is incorrect to assert that p if, in fact, it is not the case that p.

Note that the converse of (6) gives us something close to Horwich’s “If p, then one should assert that p”. (Strictly, the converse is more like “If p, then one is not incorrect to assert that p.”) However, (6) is surely closer to the relevant norm than Horwich’s own version. The fact that p does not normally impose any positive normative requirement that one assert that p. It simply cancels a normative requirement that one *not* assert that not-p.

Recall that Horwich offers this principle as a formulation of the idea that truth is a distinctive norm of assertion—which fact about truth, Horwich says, none of the philosophers who have advocated the deflationary view has ever denied. We saw that Horwich also notes that “this normative principle is precisely the sort of generalization that our deflationary truth predicate is needed to formulate.” (1993, p. 28) However, this is not to say that the deflationary view can itself account for the *normativity* involved in this particular generalisation—after all, a deflationary truth predicate allows us to express generalisations about many

matters, on most of which the deflationary view of truth provides no illumination whatsoever.

Perhaps Horwich is right that no advocate of deflationary truth has ever denied that truth is a norm of assertion in this third sense. What this would mean, I think, would be that no advocate of the view has noticed the fact that the deflationary theory cannot account for the existence of this norm. Why can't it do so? Simply because as a grammatical device, the truth predicate would have the same kind of use in an assertoric practice which lacked this third norm. (Recall the analogy with "Same again", used to make a request by "agreeing" with a previous request.)

5. What if the Third Norm Isn't Truth?

One of Horwich's themes in *Truth* (1990) is that issues which are not really the proper concern of a theory of truth may mistakenly be thought to be so. Horwich charges Wright with making this kind of mistake—with confusing the plausible idea that there are various grades of realism with the mistaken idea that there are various notions of truth (or at least significant variations from discourse to discourse in the kind of truth predicate we have in mind). It seems likely, then, that he might respond to my line of argument in the same way. He might say that although there is a third norm of assertibility, of the kind distinguished above, truth itself is not that norm (but merely facilitates our expression of the general principles underlying the norm).

To avoid prejudging the issue, let us say that the assertoric norm we are dealing with is that of correctness and incorrectness, or rightness and wrongness—in *the third degree*, in each case, when we need to mark the distinction between this norm of assertion and those of subjective and objective assertibility. Is truth the same notion as correctness in the third degree, falsity the same notion as incorrectness in the third degree? There are two ways to approach the issue. One way is to appeal to our linguistic intuitions. To my ear, these notions do seem to be pretty much interchangeable in the relevant contexts. And after all, Horwich himself says that truth is a distinctive norm of assertion. But what norm could it be, if not this one, if it isn't the norm of correctness in either of the two lesser degrees?

Perhaps more telling than this kind of appeal to intuition is the way in which the third norm is linked to notions which all sides agree are the proper concern of a theory of truth. The notion of correspondence seems to have built into it the third norm's idea of answerability to an external standard. The notion of coherence replaces this with the idea of answerability to a standard which, while internal to a community of speakers, is still external to the perspective of any individual speaker. The common idea is that truth is a dimension of correctness for assertion, or belief. To argue that the third norm is not a matter for a theory of truth would seem to be to detach the theory of truth from its traditional concerns with notions such as correspondence and coherence.

These issues require very much more discussion, of course. But what follows if Horwich is right, and the third norm of assertibility turns out not to be truth? I think the victory would be somewhat hollow. For once it is recognised that there is a third norm of assertibility, it itself becomes the focus of most of the issues which are the traditional concern of a theory of truth. What is it for a judgement to be *correct*, or *right*, in this third sense? This question is interesting and difficult in just the way that “What is it for a judgement to be *true*?” has traditionally been thought to be. So the game stays much the same, even if Horwich manages to deflate the traditional ball.

Moreover, there might be a way of playing the game, thus reformulated, which would in one sense be in keeping with the deflationists’ intuitions about truth. Deflationists say that truth is not a substantial property—that truth has no essence waiting to be discovered, as Horwich puts it. Rather, they say, the correct approach is to explain the truth predicate’s role and function in language—a task they seek to discharge in terms of DS. I want to point out that one might say the same thing about the third norm of assertibility. Here, too, there might be no hidden essence, but simply an explanatory task to be discharged—that of explaining what role the third norm plays in the lives of language users such as ourselves, and hence offering a plausible account of its genealogy. In what follows, I’ll describe an approach to this task under the assumption that the third norm *is* truth, but this isn’t essential. The real interest lies in the role of the norm, not in its label. (The following account is one I have argued for at length in Price 1988.)

6. How the MOA Became Extinct

Let’s return to the MOA, or merely-opinionated assertion. Recall that Mo’ans use linguistic utterances to express their beliefs (as well as other psychological states, such as preferences and desires). Where they differ from us is in the fact that they do not take a disagreement between two speakers in this belief-expressing linguistic dimension to indicate that one or other speaker must be at fault. They recognise the possibility of fault consisting in failure to observe one of the two norms of subjective or objective assertibility, but lack the idea of the third norm, that of truth itself. This shows up in the fact that by default, disagreements tend to be of a no-fault kind (in the way that expression of different preferences often are for us).

Think of the Mo’ans as speakers of a proto-language. How are we to understand what happens when these speakers hit upon the third norm of assertibility—the notion of truth—and the MOA becomes extinct? I think it is important to realise that there are two quite different possibilities. One possibility is that Mo’ans gradually come to realise that there is an important pre-existing property that the psychological attitudes they use MOAs to express may have or lack: perhaps the property of corresponding to how things are in the world, or—as we would put it—of being *true*. Perhaps this property is in some sense essentially normative. If not, then it might at least be such that the sensible Mo’ans come to recognise its

importance, and treat it as a norm—in other words, they come to adopt the convention that an expression of a belief is at fault, in so far as the belief fails to possess this property. Call this the *substantial* account of how the MOA becomes extinct.

The second alternative is quite different. Suppose there is no substantial, objective, property of this kind, which the Mo'ans' belief-like behavioural dispositions either have or lack. Nevertheless, it might turn out to be very much to the Mo'ans' advantage to behave as if there were such a property. As it turns out, it isn't difficult to adopt this pretence. The practice Mo'ans need to adopt is exactly the same as that required by the previous alternative. They simply need to ensure that when they believe that *p*, they be prepared not only to assert (in the old MOA sense) that *p*, but also to ascribe fault to anyone who asserts not-*p*, independently of any grounds for thinking that that person fails one of the first two norms of assertibility. In other words, the usage rule for this imaginary norm is exactly what the corresponding rule would be according to the first story—which is hardly surprising, for it is effectively the disquotational schema, transformed into the rule that one should be prepared to assert that *p* is *correct*, if and only if one is prepared to assert that *p*.

Why might the invention of such a norm be useful? Perhaps for the reason that it makes what would otherwise be no-fault disagreements into unstable social situations, whose instability is only resolved by argument and consequent agreement—and it provides an immediate incentive for argument, in that it holds out to the successful arguer the reward consisting in her community's positive evaluation of her dialectical position. If reasoned argument is generally beneficial—beneficial in a long-run sense—then a community of Mo'ans who adopt this practice will tend to prosper, compared to a community who do not.

There are a number of possible objections to this proposal, and I want to mention some of these, and provide some brief responses. (I consider such objections at greater length in Price 1988, ch. 7.)

The theoretical knowledge objection

First, it might be argued that even if there were this benefit available, it is implausible the Mo'ans would ever have realised it, or indeed been able to do anything about it if a rare individual did realise it, given that at that stage they lacked the means to engage in significant argument. This point is well-taken, and implies that if the story is to have any plausibility, it must explain how the third norm could arise in a linguistic community, without explicit planning or even awareness of its potential benefits. However, this doesn't seem to be an insurmountable obstacle. Even for the Mo'ans, not all disagreements involving expressions of (what we would call) belief are peaceful, no-fault affairs. In our own case, disagreements involving expressions of preferences often acquire an evaluative dimension, especially when the context requires agreement—for example,

when two or more people need to coordinate their choices. (“Shall we put the fence between us here, or there?”) In the same way, practical constraints would often, as it were, impose dialectical imperatives on the Mo’ans, without the help of the third norm. All our story requires is that these constraints should be gradually internalised in language, so that eventually any disagreement comes to be seen in fault-enhanced terms. The payoff, of course, lies in the fact that the benefits of reasoned agreement do not lie simply in cases in which the immediate context makes agreement imperative.

The truth-as-success objection

Second, it might be suggested that my story actually depends on a version of the first story. Surely the reason that argument is beneficial is that some beliefs are more *successful* than others, in behavioural terms—and doesn’t this notion of successfulness provide the key to the substantial property of belief required by the earlier story of how the MOA becomes extinct?

Two quick responses to this: First, it is notoriously difficult to analyse truth in terms of success in any plausible way. One problem is that the behavioural consequences of mental attitudes are a notoriously holistic affair, so that it is difficult to isolate the contribution to success or failure of a single belief. A related problem is that it is easy to find circumstances in which true beliefs are unhelpful, and false beliefs helpful. And these problems arise even for simple categorical beliefs about a speaker’s immediate surroundings. How are we to analyse *success* for, say, modal beliefs, or beliefs about the past? Second, unlike my story, this one does seem to require that the Mo’ans be aware of this property of beliefs, and this itself is problematic. If contemporary philosophers have a great deal of difficulty in spelling out what the successfulness of a belief consists in, is it really plausible that our linguistic ancestors had an intuitive feel for it?

The hindsight objection

A third argument against my proposal—or rather, in favour of the earlier account of how the MOA becomes extinct—might appeal to what we take ourselves to know about beliefs. Surely *we* know that beliefs are correct or incorrect in some world-determined sense, even if our Mo’an ancestors didn’t know it. If it is essentially the same mental states in both cases, then we know that the materials are in place for the first kind of extinction: the Mo’ans simply need to hit on the idea of a property which we know is there for the finding. (Compare: We know that Aotearoa [New Zealand] exists, and therefore that the Maoris could discover it, and didn’t need to invent it.)

But in fact, of course, we don’t know that truth is a substantial property of belief, in the relevant sense. Indeed, this is precisely the issue at hand. The story of the MOA and the Mo’ans is supposed to illustrate how our ancestors might have come to the view we now find so natural. The grip of the myth should not blind us to the possibility of such an explanation of its genealogy.

The irenic suggestion

Isn't there a plausible middle road? Perhaps truth equates with success in the case of simple categorical beliefs about a speaker's immediate environment—the kind of beliefs we would expect to be most familiar and salient to early speakers. In these cases, then, truth is a substantial property of belief, and hence, derivatively, of expressions of belief. In these cases, there is a substantial third norm of assertion. For the reasons sketched above, it then turns out to be beneficial to treat *all* expressions of belief in the same way—that is, to adopt the convention that all assertions are subject to such a norm, in addition to the first and second norms of subjective and objective assertibility.

I am not strongly opposed to this suggestion, provided we clarify a possible confusion. Even in cases in which there is a well-defined single-case notion of the behavioural “successfulness”, or utility, of belief, such a notion is not intrinsically normative. It only becomes normative in the minds of creatures who care about behavioural success. Moreover, until a third-degree norm is already in place in assertoric practice, speakers are unable to engage in fault-laden disagreements about the successfulness of particular beliefs. Without the third norm, disagreement about this is as frictionless as disagreement about anything else. Without truth, the dialectical cogs simply fail to engage.

This means that at best, the notion of success enters the picture as an important ingredient in the process whereby disagreement come to be seen in normative terms. One route might be altruistic, for example: We disagree about whether a particular fruit is poisonous. The disagreement matters to me, because I believe that your belief that the fruit is not poisonous will be spectacularly unsuccessful, and that troubles me. As a Mo'an, I can't say that you are *wrong*, that your belief is *false*. But my predicament, repeated countless times in our community, seems to be the kind of thing which would encourage the development of such norms.

The upshot is that a notion of (single-case) success may well play an important role in an account of this kind. It is a very indirect role, however, and we should certainly be cautious of declaring that to be true *is* to be successful, even for simple beliefs of this kind. It is better to say that success plays a role in the genealogy of an *insubstantial* notion of truth, than to say that in certain cases, truth reduces to the substantial property of successfulness.

7. Conclusion

In my view, then, the deflationary theory is right about the fact that truth is not a substantial property, but wrong that the key to our use of truth lies in the disquotational schema. As Wright has claimed, though in a different way, normativity does turn out to a central feature of our notions of truth and falsity, and a feature not explained by the disquotational view. What is needed is an alternative account of the genealogy of truth, an account in which normativity takes centre stage. In my view the only plausible candidate is a view which explains

truth in terms of the difference that the third norm of assertibility makes to disagreement—especially, the fact that it encourages reasoned argument—and in terms of the advantages of this difference to our linguistic ancestors.¹

Notes

1. I am grateful to Daniel Stoljar, Jakob Hohwy and Linus Broström for many helpful comments on previous versions of this paper.

References

- Horwich, P., 1990: *Truth*, Oxford: Basil Blackwell.
- 1993: “In the Truth-domain”, *The Times Literary Supplement*, July 16, 1993, 28.
- 1996: “Realism Minus Truth”, *Philosophy and Phenomenological Research*, LVI, 877–881.
- Price, H., 1988: *Facts and the Function of Truth*, Oxford: Basil Blackwell.
- Wright, C., 1992: *Truth and Objectivity*, Cambridge, Mass.: Harvard University Press.