Ratio (new series) X 1 April 1997 0034-0006

WRIGHT'S ARGUMENT FROM NEUTRALITY

Max Kölbel

Abstract

In the first chapter of his book *Truth and Objectivity* (1992), Crispin Wright puts forward what he regards as 'a fundamental and decisive objection' to deflationism about truth (p. 21). His objection proceeds by an argument to the conclusion that truth and warranted assertibility coincide in normative force and potentially diverge in extension (I call this the 'argument from neutrality'). This argument has already received some attention.¹ However, I do not believe that it has been fully understood yet. In this short paper, I shall assess the cogency of Wright's objection in some detail. My agenda is as follows. First, I give what I believe to be an adequate rendering of the objection. Secondly, I reveal the real force of the neutrality argument and say thirdly why it does not, as Wright thinks, refute deflationism. Finally, I argue that Wright's insistence that truth is a 'substantial property' is uncongenial to the overall project of his book.

I

In Wright's terminology, a deflationist holds

that, subject perhaps to certain provisos of context, the Disquotational Schema

(DS) 'P' is T if and only if P

is (all but) a complete explanation of the truth predicate—a contention from which he infers, dubiously, that truth is not a 'substantial property', whatever that means, of sentences, thoughts, and so on, but merely a device for accomplishing at the metalinguistic level what can be accomplished by an assertoric use of the mentioned sentence. (pp. 14–5)

The dubious inference that truth is not a substantial property is what Wright dislikes about this view. He set out to prove that holding the (DS) to be wholly explanatory of the truth predicate is incompatible with denying that truth is a property. The proof comes in three stages.

The first stage. Wright introduces notions of normativity for

¹ For example in Ian Rumfitt's 'Truth Wronged' (1995) and Neil Tennant's 'On Negation, Truth and Warranted Assertibility' (1995).

predicates the application of which guides the selection of 'moves' in a 'practice'. He argues that both 'is T' and 'is warrantedly assertible' register norms governing our assertoric practice.² Wright also remarks that 'is T' and 'is warrantedly assertible' coincide in normative force, which means that 'reason to suppose that either predicate characterises a move is reason to suppose that the other characterises it too' (p. 18). I shall not question this part of his argument.

The second stage consists in the following derivation from (DS):

- (i) 'not-P' is T iff not-P. (instance of DS))
- (ii) not-('P' is T) iff not-P. (from (DS) and the rule from "P iff Q" derive "not-P iff not-Q"")
- (iii) 'not-P' is T iff not-('P' is T). (from (i) and (ii))

At the third stage, Wright shows that 'is T' and 'is warrantedly assertible' cannot be registering the same norm, because the result of substituting the latter for the former in (iii):

(iii¹) 'not-P' is warrantedly assertible iff 'P' is not warrantedly assertible.

is false. It is false, because in a state of information neutral with respect to 'P', the right hand side of (iii¹) is true, while the left hand side is not.

The overall conclusion of the argument is thus

(C) truth and warranted assertibility, while normatively coincident, are potentially extensionally divergent. (p. 22)

For Wright, (C) is incompatible with, and thus provides a refutation of, deflationism because 'it is essential to deflationism . . .

² More accurately, a predicate F, is *descriptively normative of a practice*, just in case participants' selection of a move is as a matter of fact guided by whether or not they judge that move as F. A predicate is *prescriptively normative*, just in case the selection of moves ought to be so guided. Wright argues that participants in assertoric practice must 'for the most part' respect some norm of defeasible warrant, some distinction between justified and unjustified assertion, for otherwise their assertions will lack determinate content (p. 17). This descriptive norm is also, trivially, prescriptive, because the selection of assertions ought to be guided by whether they are justified or not. In Wright's terminology, the predicate 'is warrantedly assertible' registers this norm of defeasible warrant. The truth predicate is also normative of assertoric practice. Prescriptively, 'because any reason to think that a sentence is T may be transferred, across the biconditional [i.e. (DS)], into reason to make or allow the assertoric move which it expresses'. Descriptively, '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' (p. 17).

that "true" is merely a device for endorsing assertions, and hence can import non norms over assertoric discourse distinct from warranted assertibility' (p. 33–4).

II

It is crucial for an understanding of Wright's argument, that when he concludes that truth and warranted assertibility are distinct, he does *not* mean to rule out an identification of truth with some *idealised* or *absolute* norm of assertibility. The question of whether such an identification (e.g. of truth and 'superassertibility', as Wright calls it) is possible is not discussed until the subsequent chapter of *Truth and Objectivity*, and there he does not employ the neutrality argument to deny the identification—in fact, he argues in favour of it. In Wright's argument against deflationism, 'warranted assertibility is assertibility relative to a state of information' (p. 47).

In his objection to Wright, Neil Tennant (1995) evidently fails to realise that by 'warranted assertibility' Wright just means assertibility relative to a state of information. He argues that Wright's counterexample to (iii') (see p. 36 above) is not really a counterexample, for in the envisaged state of information, which is neutral with regard to 'P', one would not be entitled to claim that 'P' is not warrantedly assertible. He says on p. 103:

All that would be warranted, in such a case, would be the weak assertion

not-('P' is warrantedly assertible in I) [where I is the neutral state of information].

But it would be a grave error to infer from this weak assertion the much stronger assertion that

not-('P' is warrantedly assertible).

However, since for Wright assertibility just is assertibility relative to the relevant state of information, Wright never made, nor needed to make, more than the 'weak assertion' for his imagined neutral state of information. Thus, Tennant will have to admit that the counterexample is indeed a counterexample.

But if warranted assertibility is an implicitly relative notion, then what exactly is the force of the conclusion (C)? What is it for some implicitly relative predicate to potentially diverge in extension from another predicate? We can think of an implicitly relative predicate as a predicate with a hidden variable, and such a predicate's extension will then depend on which constant is substituted for the variable. Potential divergence of extension would then be divergence on some possible substitutions for the variable. In order to make this more transparent, let me introduce a symbolism that brings out the relativity.

Let \mathbf{A} be the set of all possible states of information. Furthermore, if $s \in \mathbf{A}$, then let 'W_s(p)' means that p is warranted relative to the state of information s. Then Wright's claim that truth and warrant are potentially extensionally divergent (one half of (C)) can be represented as follows:

(C1) $\exists s [s \in \mathbf{A} \& \exists p [W_s(p) \not\equiv T(p)]]$ (read: there is at least one state of information *s*, such that warrantedness relative to *s* and truth differ in extension with respect to at least one proposition)

Or equivalently:

 $\neg \forall s \ [s \in \mathbf{A} \to \forall p \ [\mathbf{W}_s(p) \equiv \mathrm{T}(p)]]$

(read: it is not the case that for all states of information *s*, warrantedness relative to *s* and truth are extensionally equivalent.)

Wright demonstrates (C1) by asking us to consider a particular state of information n which is neutral with respect to some proposition p. Neither p nor $\neg p$ are warranted with respect to n. That is, n is such that both $\neg W_n(p)$ and $\neg W_n(\neg p)$ hold.³ But since they both hold, truth cannot be coextensional with warrant relative to n, for $\neg T(p) \& \neg T(\neg p)$ contradicts (iii) and therefore contradicts the (DS). The envisaged stage of information n does therefore provide a perfectly valid proof of (C1)—because it presents a counterexample to its negation.

But what about the other half of (C), Wright's claim that truth and warranted assertibility coincide in normative force? He says that two predicates coincide in normative force only if 'reason to suppose that either predicate characterises a move is reason to

³ Contrary to Tennant's claim (1995), an appeal to the constructivist meaning of negation does not call this into question. For if we lack evidence regarding p in state of information n, then we are thereby warranted in asserting that neither p, nor not-p are warranted with respect to n. To suppose otherwise is to suppose that we were not in state of information n after all, which contradicts the assumption.

suppose that the other characterises it too' (p. 18). It sounds as if this is the claim that *whenever* we have reason to suppose that a move is warranted, we also have reason to suppose that it is true and vice versa. If 'having reason to suppose' identifies the same norm of warrant, then this translates into our new idiom as follows:

(C2)
$$\forall s [s \in \mathbf{A} \rightarrow \forall p [W_s(W_s(p)) \equiv W_s(T(p))]]$$

This amounts to saying that one cannot follow the norm of making a move just when it is warranted without following the norm of making a move just when it is true and vice versa. In this sense, truth and warrant are normatively coincident. However, this does not amount to the claim that a move is true whenever it is warranted, and vice versa, which would contradict (C1).

In order to show now that (C) does not, as such, touch on the possibility of identifying the notion of truth with some non-relative, absolute notion of warrant, let me define one such notion in terms of the relative one, by fixing the variable in one particular way: an assertion is *absolutely assertible*, if and only if it is warranted relative to particular state of information **I**. Now the Suggestion that truth *is* absolute warrant or absolute assertibility would be this:

(W)
$$\forall p [W_I(p) \equiv T(p)]$$

(W) is compatible with (C1) and therefore the potential extensional divergence of truth and relative warrant does not, as such, preclude one from identifying truth with some absolute warrant. However, it quite obviously follows from (W), and the above derivation (iii) from the (DS), that whenever some p is not warranted relative to \mathbf{I} , then not-p is warranted relative to \mathbf{I} , and vice versa:

(iii¹) 'not-P' is warranted relative to **I** iff 'P' is not warranted relative to **I**.

In other words, *if* truth is to be identified with warrant relative to information state **I**, then **I** must not be neutral with respect to any *p*. **I** must be complete.⁴

To sum up: Wright's argument to the conclusion that truth

⁴ It is not my business here to discuss the coherence of a state of information such as **I**. I introduced it only to demonstrate that something *like* Wright's super assertibility (defined on p. 48 of his book) can be identified with truth from the perspective of (C).

and *relative* warrant potentially diverge in extension, i.e. (C1), appears to be sound. Moreover, it is compatible with (C2) and (W)—with the coincidence in normative force of truth and warranted assertibility and the identification of truth with some absolute warrant.

Ш

Now, does the conclusion of the neutrality argument, (C), really show that deflationism is incorrect? As it is not easy to see why Wright thinks it does, let me quote at some length:

Since the defining thesis of deflationism is that 'true' is merely a device of disquotation—a device for endorsing assertions, which we need only for the purposes of indirect ('Goldbach's Conjecture is true') or compendious ('Everything he says is true') such endorsements—since that is the very essence of the view, [1] a deflationist must of course insist that the only substantial norms operating in assertoric practice are norms of warranted assertibility, and [2] that the truth predicate can indeed mark no independent norm. [3] For were it normatively independent, to predicate 'true' of a sentence would be to claim that sentence's satisfaction of a norm distinct from warranted assertion. [4] No room could then remain for the contention that 'true' is only grammatically a predicate, whose role is not to attribute a substantial characteristic. (p. 18; see also the formulations on pp. 16 and 21)

(a) Contrary to claims 1 and 2, the deflationist need not hold that relative assertibility is the only assertoric norm, nor need he deny that the truth predicate 'marks' some norm distinct from relative assertibility. As claim 3 shows, however, Wright thinks that admitting the existence of such a distinct norm (1), and moreover admitting that the truth predicate 'marks' that norm (2), would force the deflationist into the further admission that employing the truth predicate amounts to claiming that this norm is being complied with. He takes this further admission to be incompatible with the deflationist's doctrine of the role of the truth predicate (4).

In response, the deflationist will (ideally) point out that there is one sense of 'claiming a sentence's satisfaction of a norm distinct from warranted assertion', in which the conditional

claim 3 is acceptable, but in which so claiming does not amount to ascribing a property to that sentence. In another sense, so claiming does amount to the ascription of a property, but so interpreted, claim 3 is unacceptable.

In order to explain this, let me briefly summarise a deflationist view on the function of the truth predicate. According to deflationism, it is the truth predicate's function to allow the formation of sentences that are in the following way parasitic on other sentences: attaching the truth predicate to the name of a declarative sentence expressing a certain proposition (or to a name of that proposition) yields a sentential phrase which expresses the same proposition. Thus, when I assertorically apply the truth predicate to a declarative sentence *s*, then I am making the same assertion I could have made by simply uttering *s*.

Despite the seeming modesty of this equivalence function, however, the truth predicate is needed to turn a certain syntactic trick—what Wright calls its 'indirect' and 'compendious' uses. For example, if I want to back my accomplice in a police interrogation I can just 'indirectly' say 'What he said is true' without knowing what exactly he said. The truth predicate also allows me to say, 'compendiously', that everything the Pope says is true, thus sparing me from the cumbersome 'If the Pope says "Abortion is wrong.", then abortion is wrong; if he says "Elephants can fly.", then elephants can fly; . . . '5 The deflationist's idea is that the truth predicate exists *solely* for the performance of this trick, and that it might therefore be misleading to assume that there is some property, truth, the ascription of which is the function of the truth predicate and that we can analyse in the way we analyse, for example, the property of redness.

Back to Wright's claim 3: in so far as the truth predicate marks or registers a norm, it does this only in virtue of its equivalence function. If a speaker is guided, in selecting a sentence $\lceil p \rceil$ for utterance, by his judgement as to whether $\lceil p \rceil$ is true, then this just amounts to his being guided by his judgement whether p. Suppose that, more precisely, the norm is to utter $\lceil p \rceil$ only if $\lceil p \rceil$ is true, i.e. to utter $\lceil p \rceil$ only if p. Then, predicating 'is true' of a sentence $\lceil p \rceil$

⁵ Cf. Quine (1970, p. 11) and Horwich (1990, pp. 1-8).

⁶ Wright's definition of 'predicate F registers a norm' (see fn. 2 above) in terms of F's guiding the selection of moves, is vague, and might admit of different interpretations, e.g. the norm of asserting that the cat is fat *if and* only if the cat is fat. Cf. Horwich (1990, p. 65), and B. Williams (1995, pp. 231 f).

is to claim that $\lceil p \rceil$ complies with that norm only in the sense in which it amounts to a claim that p. It is not to claim that $\lceil p \rceil$ complies with that norm in any sense incompatible with the deflationist's doctrine about the sole function of the truth predicate. Now, a thinker's judgement whether p is based, at any one time, on the warrant available to that thinker at that time. Wright's argument shows that some things, however, are true without being warranted and perhaps vice versa. This however is not incompatible with the deflationist's claim that judging whether $\lceil p \rceil$ is true is the same as judging whether p. For whenever the one is warranted, the other is too, and vice versa (see (C2) above p. 4).

(b) Now, apart from what Wright literally says, there may be further, underlying worries about deflationism. In order to address one of these, let me look at the result of the neutrality argument from a slightly different angle. According to corollary (iii) above, all subscribers to the (DS), among them the deflationist, must accept that whenever a sentence is not true, its negation is true and vice versa. Now suppose (for the sake of argument) that the truth predicate had a unique extension. This extension, so much is fixed by (iii), could only coincide with the extension of W_s for those substitutions of s that denote complete information states, i.e. states which are not neutral with respect to any p (see p. 39 above). Or in other words, if truth had a unique extension, that extension would diverge from the extension of warrantedness relative to a given information state, whenever that information state is not complete. But as our actual information states are never complete, warrant can never actually coincide with truth.8

Perhaps this can help articulate a worry about the deflationist's doctrine of the function of the truth predicate. For how, we might ask, can a syntactical device that merely serves the modest function the deflationist claims it serves introduce constraints on its extension that make it impossible for the kinds of warrant that we actually employ in assessing assertions to have the same

 $^{^7}$ In fact, the sentence 'Not everything true is also warranted' beautifully illustrates the compendious use of the truth predicate. Without this device, we would have to say: 'Its not the case that [if abortion is wrong, then it's also warranted that abortion is wrong; if elephants can fly, then it's also warranted that elephant can fly; . . .]' or perhaps: 'There is a sentence $\lceil p \rceil$, such that $\lceil p \rceil$ is not warranted but p.'

 $^{^{8}}$ (iii) entails even more: *any* extension of the truth predicate (whether unique or not) must diverge in extension from W_{s} whenever s is not complete.

extension? Doesn't that just show that there is more to the truth predicate than the function of generating sentential phrases that are equivalent to already existing such phrases?

I do not think the deflationist could be forced to admit that this constraint is part of 'the nature of the property of truth'. He can demonstrate that it follows from the equivalence function and independent constraints alone. For if a given sentence can be used to 'make a certain move', then it is the truth predicate's function to allow us to make the *same* move by assertorically applying it to a name of that sentence. Moreover, if a given sentential phrase can be used to express a certain proposition, then it is the truth predicate's job to allow us to express the *same* proposition by applying it to a name of that sentential phrase, or to a name of that proposition. Our corollary (iii) follows from this function and the law of non-contradiction alone: suppose (iii) were false. Then it would have to fail either right to left or left to right, i.e. either

or

But since the truth predicate, in order to fulfil its function, must allow us to substitute 'P' and 'not-P' respectively for "'P' is T" and "not-P' is T", which would turn both (iv) and (v) into a contradiction, (iii) cannot be false.

(c) Finally, let me address yet another, perhaps more serious misgiving about the coherence of deflationism. Wright's argument succeeds in showing that the extensions of the truth predicate and any actual warrant must diverge. But does it, in addition, show that there is a 'substantial' assertoric norm over and above relative warrant, registered by 'is true'? If that were so, then this might be thought to undermine deflationism in the following way. What motivates the deflationist's view that the truth predicate's only function is the performance of the syntactic trick is the fear that we might otherwise feel compelled to look for a property of truth, where really there is no such thing. An admission that there is some ('substantial') assertoric norm distinct from relative assertibility which is registered by the truth predicate, might therefore be at odds with the *motivation* for deflationism. For it is to admit that there is an assertoric norm to look for, and so while the truth predicate has the disquotational function, it should *also* be viewed as picking out the mentioned norm, for since that norm is already registered by the truth predicate, no better name for it than 'truth'.

What would a sensible deflationist say about the norm registered by the truth predicate? He will insist that thinkers (speakers) do not just believe (assert) any arbitrary thing, but rather that they want their beliefs (and thereby their sincere assertions) to meet a certain norm. This norm is to believe (assert) that the cat is fat only if the cat is fat, to believe (assert) that Sam smokes only if Sam smokes, to believe that Greg drinks only if Greg drinks and so on for everything one might believe (assert). There is a linguistic device, namely the truth predicate, which allows us to capture the norm thus indicated 'compendiously' in the form of the following rule:

(R) Believe (assert) only what is true!

However, by making use of the truth predicate in characterising the envisaged norm more conveniently, we are just taking a sort of short cut. Thus (R) is no more than the norm of asserting (believing) that the cat is fat only if the cat is fat, that Sam smokes only if Sam smokes, and so on.

The question envisaged above was whether in showing that the assertoric norm compendiously captured by (R) is distinct from relative warrant, Wright has shown that the deflationist's worry about the futility of a search for 'the property ascribed by the truth predicate' is unfounded. In this case the deflationist's continued insistence that the norm captured by (R) ought not be called 'truth' would be mere quibbling. Indeed, it would seem part of the syntactic function of 'is true' and its cognates that this norm can conveniently be called 'truth'.

To think that the neutrality argument removes the deflationist's worries, however, is to overestimate its force. The argument shows that in any incomplete (and therefore any actual) information state *s*, the set of moves prohibited by (R) is different from the set of moves prohibited by this rule:

(A) Believe (assert) only what is warranted relative to s!

If s is reliable, most things permitted by (A) will also be permitted by (R), but still, it cannot be that the prohibitions of (R) coincide with those of (A), as long as s is not a complete information state. This, however, is all Wright's argument shows about the norm (R). It does not tell us whether there is, for all thinkers and

for all their information states, a unique, definite set of things they are permitted by (R) to believe (assert). In other words, while the neutrality argument shows that any extension of the truth predicate is constrained not to coincide with the extension of any incomplete warrant, it does not show that the truth predicate has a unique or a definite extension. For all Wright's argument tells us, truth might be utterly relative or vague. Therefore, the deflationist's worry that a search for a truth-property might be futile is not removed by the neutrality argument.

IV

Let me conclude with a different remark relating Wright's claim that truth is a 'substantial property' to the overall project of his book. The project is to develop a new framework for 'realist/antirealist debates'. One such debate may be the debate about whether something can really and objectively be funny, others whether there are moral, modal, mathematical or scientific facts. Within Wright's new framework, realists and anti-realists would no longer be debating whether statements about the funny, good, etc. can be true, as it has been within error-theoretic and expressivist implementations of anti-realism. Rather, Wright's minimal notion of truth is intended to be so 'metaphysically lightweight' (p. 13) that 'truth need not be the exclusive property of realism' (p. 12). The aim is that any anti-realist can happily concede that (syntactically characterised) declarative sentences of all sorts are truth-apt and sometimes true, without thereby being 'immediately saddled with domains of, for example, intrinsically moral, or comic fact' (ibid.).

Given this overall strategy, it would seem that Wright ought to sympathise with deflationists about truth, since they provide a metaphysically non-committal notion of truth, which would allow the intended shift of the debates. Instead, he spends almost one chapter trying to refute deflationism with the argument discussed in the present paper. In particular, he argues against the deflationist's claim that truth is 'not a substantial property'.

However, there are good reasons why, given his overall strategy, Wright should *not* be arguing that truth is a substantial property—even ignoring the fact that the argument does not succeed. For it can be argued, that as long as the minimal notion

⁹ Equivalently, it does not tell us whether there is, for all thinkers and for all their information states, one unique set of things they are prohibited by (R) to believe (assert).

remains one of a 'substantial property', most anti-realists will not be able to concede the truth-aptness of all declaratives. Consider an anti-realist about matters of taste (since most people have anti-realist inclinations in this area). He will deny that there is a fact of the matter as to whether haggis is tasty—tastiness is not, for him, a real property that things can objectively have. Accordingly, if one person believes that haggis is tasty, and another believes that it is not, then neither need be wrong, and there is no point in arguing about the matter. Now suppose the sentence 'Haggis is tasty.' and its negation were truth-apt in Wright's sense, i.e. apt for minimal truth which is nevertheless a substantial property. Then of two people one uttering the sentence, the other its negation, only one could be speaking the truth. Therefore there would be scope for reasonable argument, for argument might help detect who is not speaking the truth and this is useful for anyone wishing to conform to norm (R) above. But the anti-realist cannot concede this, for he insists that there is no point in arguing about taste.10

One way of avoiding this situation would be to regard truth as an implicitly relative notion. For then the admission of truth-aptness would no longer carry the unwanted implication of objectivity: 'Haggis is tasty.' could then be true relative to one thing, while 'Haggis is not tasty.' is true relative to another. However, whatever a 'substantial property' may be, it is not, I take it, a relation. Thus Wright excludes himself from this option by claiming that minimal truth is a substantial property. Perhaps, then, the question Wright ought to address most is not the question whether truth is a substantial property, but rather whether the norm governing assertoric discouse is relative to something or absolute (regardless of whether that norm is or is not to be called 'truth'). ¹¹

Instituto de Investigaciones Filosoficas Universidad Nacional Autonoma de Mexico DF 04501 Coyoacan Mexico

¹⁰ Put in Wright's own terms, this amounts to the following complaint: Treating minimal truth as a substantial properly leaves no room for a discourse that is assertoric (and therefore comprises truth-apt sentences) but does not exhibit cognitive command. For one of two sincere, contradicting disputants must believe something not true. But believing something that is not true is, given norm (R), a cognitive failure. So disagreement implies cognitive failure.

¹¹ I am grateful to Keith Hossack, Alan Thomas, an anonymous referee and especially to Mark Sainsbury for their comments.

BIBLIOGRAPHY

Horwich, P. (1990) Truth (Oxford, Blackwell).

Quine, W. V. (1970) *Philosophy of Logic* (Englewood Cliffs, Prentice-Hall). Rumfitt, I. (1995) 'Truth Wronged' *Ratio* 8 (new series), pp. 100–7.

Tennant, N. (1995) 'On Negation, Truth and Warranted Assertibility' *Analysis* 55, pp. 98-104.

Williams, B. (1995) 'Truth in Ethics' *Ratio* 8 (new series), pp. 227–42. Wright, C. (1992) *Truth and Objectivity* (Cambridge, Mass., Harvard UP).