

## Postscript to ADeflationist Views of Meaning and Content@. EXCERPT

**3. The generalizing role of >refers=.** I stressed in the paper (following Quine and Leeds) that the term >true= increases the expressive power of our language, by serving as a device for expressing infinite conjunctions and disjunctions (which could have been expressed by more obviously logical means like substitutional quantifiers). The idea obviously works equally well for >true of=. But one might doubt that the same holds for >refers= as applied to singular terms: we already have quantifiers that play a generalizing role in the name position, so it might seem that no expressive power is gained.

I think this argument incorrect: ordinary quantifiers don't allow for generalizations of names that appear *both inside and outside quotation marks*, but that is what we would need to express what is said by AEvery name that came up in Department X's discussion of who to hire referred to a male@ without use of >refers=.

**4. Untranslatable utterances (1).** I now think that my restriction of an agent's disquotational truth predicate to the sentences the agent currently understands wasn't quite right, and is in considerable tension with my remarks on our commitment to extending the truth schema as we expand the language. Those latter remarks suggest that once we come to regard an expression S as an acceptable declarative sentence, then even if we have no understanding of it (or virtually none) still we ought to accept the corresponding instance of the truth schema. Admittedly, the cognitive equivalence thesis implies that to the extent that we don't yet understand S, we don't understand the attributions of truth to it; this may seem to make the extension of the truth schema to it idle. But it is not *completely* idle, for we tie any future improved understanding of >true= as applied to S to the future understanding of S. One way to think of the matter is in terms of indeterminacy: a sentence I don't understand at all is maximally indeterminate in content for me, and so is an attribution of truth to it; but the indeterminacies are A tied together@ (see the discussion of Acorrelative indeterminacy@ in Chapter 7, and Chapter 9, secs 1 and 2) in such a way that the truth schema holds.

The main point of modifying my view in this way is to better accommodate talk of truth for untranslatable utterances, utterances that are in no sense equivalent to anything in our current language.

When I wrote the paper I was tempted to bite the bullet and say that we simply can't apply the notion of truth to such utterances; I wasn't happy with this, but could see no good alternative. Stewart Shapiro (unpublished) presented a nice problem case for this position. In Shapiro's example, we have a guru who makes pronouncements about set theory; a disciple who thinks that everything the guru says is true, but who doesn't understand set theory; and a logician who distrusts the guru's set-theoretic pronouncements, but likes to draw their number-theoretic consequences, which the disciple does understand. Shapiro argues that if in addition the disciple trusts the logician's acumen about logic (though not about set theory), then the disciple ought to be able to reason that since everything the guru says is true, and consequence preserves truth, and the logician truly says that the number-theoretic sentence  $\neg p$  is a consequence of what the guru says, then  $\neg p$  must be true, so  $p$ . But though the disciple understands this conclusion, the reasoning is blocked if we can't meaningfully apply  $\neg$  to the guru's utterances.

There are ways of handling this example without modifying the view of the Chapter, but they aren't entirely natural, and I'm not sure that they would extend to all alterations of the example.<sup>1</sup> But the modified view suggested two paragraphs back handles the example very neatly: the disciple regards the guru's sentences as in a potential expansion of the disciple's language, so that there is no difficulty in his carrying out the reasoning. It may seem mysterious that the reasoning can be carried out in this framework, since according to this framework the guru's utterances and the notion of truth for these utterances have quite an indeterminate content for the disciple. But the joint acceptance of the claims (i) that the guru's utterances are true, (ii) that they logically imply the number-theoretic claim  $\neg p$ , (iii) that logical consequence preserves truth, and (iv) the truth schema, give just enough determinacy to enable the reasoning to be carried out.<sup>2</sup> (The reasoning could be carried out even if the guru's language were disjoint from the disciple's, rather than an expansion of it; though in that case some might prefer to appeal to expansions of the disciple's language that *translate* the guru's utterances rather than that *include* them.)

Incidentally, the guru problem is a *prima facie* problem not only for the version of deflationism in the paper, but for versions of deflationism that take truth to be attributed primarily to propositions. The

latter, if they are to have any claims to the title of deflationism, must explain what it is for a sentence to express the proposition that  $p$  in terms of equivalence in content to our sentence  $\ulcorner p \urcorner$ , but this explanation is unavailable for the gurus' sentences. For those views as for mine, the idea that the attributions have a quite indeterminate content, constrained only by such beliefs as that they have specific number-theoretic implications, provides the way out.

**5. Untranslatable utterances (2).** The incorporation model for dealing with foreign utterances has wider application than my remarks in Section 8 of the Chapter suggest. Indeed, in the next to last paragraph of Section 8 I explicitly ruled out using it for sentences that we don't understand. As the discussion of Shapiro's example should make clear, I recant.

Consider our reaction when on a certain occasion  $Z$  we hear another speaker, Mary, use a proper name such as  $\ulcorner \text{George} \urcorner$ , and we take her to be talking about someone we have never heard of before (as we colloquially put it). If we are interested in what she says, we may pick up the name  $\ulcorner \text{George} \urcorner$  from her and start speculating with it ourselves. (I wonder if George lives in California.) When I say that we pick it up from her, I don't just mean that we start using the sound  $\ulcorner \text{George} \urcorner$ : we probably used that sound already, e.g. in connection with various U. S. Presidents. But we pick it up with a new use, not equivalent to these prior uses, and a use that is in some way deferential to Mary's: we base our own beliefs on hers (more accurately, on what we take her beliefs to be).<sup>3</sup> If you like, you can put this by saying that we mentally introduce a new term of form  $\ulcorner \text{George}_j \urcorner$  where  $j$  is an inner subscript that we haven't previously used in connection with  $\ulcorner \text{George} \urcorner$ , and we use  $\ulcorner \text{George}_j \urcorner$  to translate those of Mary's assertions that we regard Mary as having mentally subscripted by the same subscript that she mentally attaches on occasion  $Z$ .

I do not see any reason to think that understanding this incorporation of Mary's term requires an antecedent grasp of the notion of reference. Of course, once we have a disquotational notion of reference, we can use it in this connection: I can say  $\ulcorner \text{George} \urcorner$  (as I'm using it now) refers to George, using the term  $\ulcorner \text{George} \urcorner$  to incorporate Mary's use of it on the occasion in question. Or on the hidden subscript picture:  $\ulcorner \text{George}_j \urcorner$  refers to  $\text{George}_j$  (and to nothing else). Since we translate Mary's use of

›George= on occasion Z with ›George<sub>j</sub>=, we can also say that Mary's use of ›George= on occasion Z refers to George<sub>j</sub> (and to nothing else). In other words, I can conclude:

George<sub>j</sub> is identical to the referent of Mary's term ›George= on occasion Z.

This allows me to use the phrase *the referent of Mary's ›George= on occasion Z* instead of ›George<sub>j</sub>= . Since the inner subscripts are silent (and other people don't know my inner subscripts even were I to vocalize them!), doing so is sometimes useful: it is a very good way to make clear which way I'm using ›George= on this occasion. Of course, I could have introduced other means for doing the disambiguating: I could have just said *I'm going to introduce the new term ›George<sub>Mary, Z</sub>= (or ›Shrdlu=)* as an equivalent of Mary's use of ›George= on occasion Z; but since we have a notion of disquotational reference already, there is no need for such new terms.

The approach carries over straightforwardly to demonstratives. Suppose that on occasion Z Mary uses the term ›that=, and suppose that I have no idea how to pick out in English (either demonstratively or descriptively) what I take her to be talking about; alternatively, suppose I have little or no idea what to take her as talking about. Here it might be misleading to incorporate her use of ›that= into my language with my own use of ›that=. normal conventions about the use of ›that= might suggest that my term stands for something quite different. But I could always introduce some special term, say ›Oscar= or ›THAT<sub>Mary, Z</sub>=, as a device for incorporating her use of ›that= on this occasion. But if I did so, I could argue just as above that in a disquotational sense of ›refers= (extended by translation), Oscar is identical to the referent of Mary's term ›that= on occasion Z; so given that we have a notion of disquotational reference already, there is no need for these special terms.

The approach just suggested is quite similar to, and perhaps identical to, the proposal of Brandom 1984. What I've said is that we should view the phrase *the referent of Mary's ›that= on occasion Z* as functioning as a device for incorporating her use of ›that= on occasion Z into our language. What Brandom says is that it functions as a device of cross-speaker anaphora. I don't entirely like Brandom's way of putting things, because taken by itself it seems to suggest that there is a non-deflationary notion of reference applicable to Mary's pronoun, and that our anaphorically

dependent phrase gets its referent from it. However, Brandom says things incompatible with that understanding of anaphora, and in support of some sort of deflationary story, and I think the incorporation model is at least very close to what he intended.

The point that I've been making for reference can be made for truth as well. There are many reasons why a direct incorporation of another person's utterance into my language could be inconvenient (e.g. if I can't pronounce it) or misleading (e.g. if it contains indexicals that would shift their reference were I to use them). In such cases, we can use 'Her utterance is true' as a means of incorporating her utterance. This indeed is what Grover, Camp and Belnap (1975) see as the main function of 'true'. I prefer to view the main function of 'true' as the disquotational function. But as with reference, I can argue from the disquotational use of 'true' to the incorporation use. For if I were to introduce a new sentence, say 'UTT<sub>Guru, Z</sub>', to incorporate the guru's unpronounceable utterance on occasion Z, then the disquotational properties of truth give an equivalence between

(1) 'UTT<sub>Guru, Z</sub> is true

and

(2) UTT<sub>Guru, Z</sub> ;

but since 'UTT<sub>Guru, Z</sub>' is my translation of the guru's utterance, (1) is equivalent to

(1') The guru's utterance is true,

and so (1') must be equivalent to (2). That is, (1') is legitimized as a way of incorporating the guru's utterance, so we don't need the special sentence 'UTT<sub>Guru, Z</sub>' any more.

The discussion in this section (and in Brandom, and in Grover, Camp and Belnap) shows another way, besides as a device of infinite conjunction and disjunction, in which 'true' and 'refers' can increase expressive power: or at least, can increase the range of what can be expressed *easily and in a way not subject to confusions of ambiguity*. Again, this expressive function could have been achieved by other means ('George<sub>Mary, Z</sub>' or 'Shrdlu' in the case of 'refers'); but 'refers' and 'true of' are a convenient way to do it, and I have argued that if we did use another device for doing it we would

have no need for the other device once we had the disquotational use of  $\text{>refers=}$  and  $\text{>true=}$ .

**6. APure Disquotational@ and AQuasi-Disquotational@ Truth.** The central feature of APure disquotational truth@ is that  $\text{A>p= is true@}$  is cognitively equivalent to  $\text{Ap@}$  (at least, modulo the existence of the sentence). I say in the chapter that  $\text{>true=}$  in the pure disquotational sense is to be understood as something like  $\text{>true-as-I-understand-it=}$ , what I really meant was  $\text{>true-as-I-actually-understand-it=}$ . As I emphasized, this is just a heuristic, and I claim that the notion does not require that we take  $\text{Aways of understanding@}$  to be understood in terms of propositional content. I have a bit more to say about this in the next Chapter.

The paper recommends using this notion of pure disquotational truth where possible, but it of course recognizes that we use the notion of truth in connection with other people=s utterances and in connection with our own utterances in counterfactual circumstances, and it suggests ways of accommodating this fact. One way is to use what I call AQuasi-disquotational truth@, I downgrade it because it assumes a stand on Quinean issues about interpersonal synonymy that I prefer to remain neutral about. But the general idea behind quasi-disquotational truth could have been put in a way that is more neutral about interpersonal synonymy:

(\*)  $S$  is  $\text{true}_{\text{qd}}$  (at possible world  $v$ ) :  $S$  is to be translated by a sentence of mine (in the actual world) that is purely disquotationally true (at  $v$ ).

My method of Aexplaining away@ modal intuitions about the truth of our sentences at the end of Section 9 really amounts to suggesting that it is  $\text{true}_{\text{qd}}$  in the sense of (\*) rather than pure disquotational truth that we are operating with in many contexts.

I have come to think that it is unnecessary to use two distinct truth predicates, the purely disquotational and the quasi-disquotational. We can use a single truth predicate as long as we take the entities in its extension not as orthographic types but as computational types: equivalence classes of (potential) tokens under the relation of computational equivalence. I explain what I mean by computational equivalence in the next chapter, but an important feature of it is that it is defined only *within an individual  $X$  in a given possible world  $u$* : it doesn't make sense to ask if one of my

tokens is computationally equivalent to a token of yours, or to a token of a counterpart of me in another possible world. So we can rewrite (\*) with the individual and the world made explicit, as follows:

(\*\*)  $S_{X,u}$  is true<sub>qd</sub> (at possible world  $v$ ) :  $S_{X,u}$  is to be translated by a sentence of mine (in the actual world) that is purely disquotationally true (at  $v$ ).

But then *A*pure disquotational truth<sup>@</sup> can be viewed as just the special case where the individual is me and the world is the actual world (by assuming that in that case the *A*translation<sup>@</sup> is just the identity function); (\*\*) is merely a way to generalize from the special case of me in the actual world to others and to my counterparts. (Of course the heuristic *≠*true-as-I actually-understand-it<sup>=</sup> is appropriate only for the special case where translation drops out. It is worth emphasizing, though, that this special case plays a central role in the account: the other cases result from the special case by translation.) If sentences are typed orthographically, the same sentence can be used in distinct ways in distinct worlds and we need to distinguish between a truth predicate that *A*holds meaning constant<sup>@</sup> and one which doesn't; but if sentences only exist within a single world then no such distinction is required.

If ordinary or quasi-disquotational truth is defined either as in the text or by (\*) or (\*\*), it has the curious feature that utterances not translatable into our language are not true. It seems more natural to suppose it indeterminate (or maybe just undetermined) whether they are true. The matter may be somewhat academic: if the points made in the two preceding sections of this postscript are correct, the notion of an untranslatable utterance does not have entirely clear application. Still, I would now prefer to do things in the more natural way. The way to do so is clear: instead of defining ordinary or quasi-disquotational truth in terms of purely disquotational truth, we explain it by a schema:

(\*\*\*) If  $S_{X,u}$  is translatable as  $\langle p \rangle$  then  $\sim (S_{X,u}$  is true iff  $p$ ). [And if  $S_{X,u}$  is translatable as  $\langle p \rangle$  then  $\sim (S_{X,u}$  has the truth conditions that  $p$ ).]

In the paper I attached some importance to the idea that we ought to *define* quasi-disquotational truth, in terms of the more-or-less logical notion of purely disquotational truth together with translation. I must have thought that the motivation for demanding a substantive theory of truth and truth conditions and reference, like that advocated in Chapters 1 and 2, could only be undercut if these

notions are explicitly defined in terms of more-or-less logical notions plus translation. But this seems a mistake: the argument for a substantive theory of truth conditions and reference depends on taking truth conditions and reference as having a certain kind of *Acausal explanatory* role, as noted in the Postscript to Chapter 1; and introducing a notion of truth conditions by means of the schema (\*\*\*) does nothing to make the notion *Acausal explanatory*. Of course, using (\*\*\*) to introduce the notion of truth conditions doesn't *guarantee against* putting the notion to the relevant sort of *Acausal explanatory* use, but as long as one doesn't do this and also doesn't make unwarranted determinacy claims for these notions (see footnote to Postscript to Chapter 1) then one can maintain a deflationist position.

**7. Bigger issues.** There are two big areas in which much more needed to be said than I said in this Chapter. First, there is little discussion of the aspects of meaning that go beyond truth, reference, and related notions. Don't we need a deflationary account of those aspects of meaning too? The answer is that we do need a deflationary account of meaning generally, but that there is no special difficulty for getting one once we have a deflationary account of truth and reference; I hinted as much at the beginning of the chapter, and say more about it in the next chapter.

The other area in which a lot more needed to be said was the explanatory role of the assignment of truth conditions to mental states, both in the explanation of behavior and in the explanation of the extent to which behavior is successful in achieving various results. In a paper written some years earlier (Field 1986), I had flirted with deflationism but ended up tentatively arguing against it, on the ground that such explanations required a role for the assignment of truth conditions to mental states that the deflationist could not legitimize. The argument was so abstract and convoluted it couldn't possibly have convinced anyone; I was skeptical myself, but couldn't pinpoint what was wrong with it. I now think it made a number of mistakes that collectively deprived the argument of all force. Or as Jimmy Carter said of his attempt to rescue the American hostages in Iran (an attempt that had to be aborted with eight deaths before the rescue helicopters even reached Iran): it was *Aan incomplete success*. It isn't worth going through the argument in detail here, but I'll sketch the sort of considerations it involved and what I now think a deflationist should say about them. One key point (which was recognized in the '86 paper, but not made sufficiently salient) will be that there is nothing in deflationism that prevents the use of *≠true=*

in explanations as long as its only role there is as a device of generalization.

It is perfectly obvious that in explaining how a pilot manages to land a plane safely with some regularity, one will appeal to the fact that she has a good many true beliefs: beliefs about her airspeed at any moment, about whether she is above or below the glideslope, about her altitude with respect to the ground, about which runway is in use, and so forth. (None of these beliefs need be based very directly on observation; she might be flying in bad weather with some of her instruments not working, so that she must rely on complicated cues.) The deflationist obviously needs to grant this explanatory role for the truth of her beliefs, and will have to say that it is somehow licensed by the generalizing role of  $\text{true}$ .

Can this deflationist strategy be maintained? A detailed explanation of how the pilot functions would involve the existence of some class  $C$  of internal representations, intuitively representing airspeed, such that when she believes a representation in  $C$  that represents too high an airspeed she slows the plane and when she believes one that represents too low a speed she speeds it up; and the representations in  $C$  that she believes tend to be true. The last part can of course be rephrased as a generalization: she tends to believe a representation in  $C$  that represents too low a speed when the plane is in fact too slow, etc. But is this restatement of the claim that she has true beliefs about airspeed innocent from the deflationist viewpoint? What may well seem a problem is that it is put in terms of what the pilot's internal representations represent. The basic idea behind the tortuous argument of the '86 paper was that the deflationist is forced to understand the sense in which the representations represent airspeed in terms of their translation into the explainer's language, but that this is inappropriate since this should be an objective explanation in which the explainer plays no causal role.<sup>4</sup>

An obvious strategy for responding to this argument is to say that talk of representation is serving a merely heuristic role in the explanation of the pilot's ability; put without the heuristic, the explanation involves the existence of some class  $C$  of internal representations in the pilot and two subclasses  $C_1$  and  $C_2$  of  $C$  such that (i) when she believes a representation in  $C_1$  she slows the plane and when she believes one in  $C_2$  she speeds it up, (ii) there is a 1-1 function  $f$  from  $C$  to a certain set of real numbers such that (a)  $C_1$  is that subclass of  $C$  that is mapped into numbers above a certain

threshold and  $C_2$  is that subclass of  $C$  that is mapped into numbers below a certain (slightly lower) threshold, and (b) she tends to believe a representation  $r$  in  $C$  when the airspeed in knots is approximately  $f(r)$ . This makes clear that we have a perfectly objective explanation, *expressible without translating the pilot's representations or talking of their truth conditions or of what they represent*. Of course part (ii)(b) of the explanation uses an indication relation, but no deflationist could object to that. (See the postscript to Chapter 2.) The function mapping internal representations into airspeeds needn't even give the intuitive truth conditions of those representations in all cases: one could tell a story in which the pilot's beliefs about what she was doing were so weird that it would be natural to assign quite different truth conditions to her representations. (Perhaps she believes she isn't in an airplane at all, but is using the controls to direct US ground forces on a foreign mission.) The representations would be reliably correlated with airspeed but would represent something quite different that they were not reliably correlated with.

It is also important to realize that in normal cases (as opposed to cases where the pilot has very weird beliefs about what she is doing) we can explain the pilot's competence in less detail, by simply saying that she tends to believe truths about the airspeed. (We would have to rest content with this if we had no idea that the pilot needed to maintain airspeed within a certain range to safely land, but knew only that she needed to base her actions *somehow* on her airspeed.) The objective core of such an explanation is clear: it is that a functional correspondence between internal representations and airspeeds like that mentioned above is playing some sort of role that we may not be in a position to specify. But when we are not in a position to specify that role with enough precision, it is inevitable that we use the notion of truth in one way or another. One way to do so would be to just say: some explanation roughly like the one in the previous paragraph is true; here it is transparent that truth is playing simply a role in generalizing, and so clearly this way of proceeding is available to the deflationist. But this way of proceeding requires that we have a sample objective explanation, however crude, to use as a model, and it is not always convenient to provide one. But the alternative way of proceeding mentioned at the start of this paragraph does not require this: if we just say "She tends to believe truths about the airspeed, and this somehow enters into her landing successfully", we can be vague on the details without

even sketching a sample precise model.

Is this latter way of proceeding available to the deflationist?<sup>5</sup> When we give this sort of explanation, we say in effect (i) that she has certain internal representations that are similar enough in their role to our representations of form >The airspeed is  $n$ = to count as <Asaying the same thing>, (ii) that they tend to be disquotationally true under the translation scheme just alluded to, and (iii) that the fact that their disquotational truth conditions under this mapping tend to be satisfied plays an important role in explaining the safe landing. This is a <Aprojectivist> or <Asecond class> explanation, in that we make use of a similarity between the agent and ourselves to avoid the need to fill in explanatorily crucial details; but it is a projectivist explanation of a broader sort than those considered in Section 4 of the Postscript to Chapter 2. The projectivist explanations discussed there are ones that in effect assert the existence of fuller explanations *in the agent's computational psychology*, explanations which they leave only vaguely specified. Here, on the other hand, the projectivist explanations in effect assert the existence of fuller explanations *in the agent's extended psychology* which they leave only vaguely specified; where the extended psychology is the computational psychology *together with the assumptions about correlations between the agent's internal states and the external world*. But the central point is the same: assignment of truth-conditional content is something we need only when we don't know how to fill in the details of the explanation.<sup>6</sup>

The same basic point holds in cases where the agent is less reliable in his beliefs about the external world. Suppose I explain Kennedy's having avoided nuclear war over Cuba in terms of a lucky guess as to what Khrushchev would do. Again, the objective core of the explanation is that Kennedy's action tends not to lead to war if Khrushchev acts as he did but to lead to war otherwise, and that the action was crucially based on his having believed a certain representation; if we knew a lot of detail about Kennedy and his decision making process, we could say a great deal about this representation and the role in his psychology, but typically we don't know enough to do this. So here we specify the representation by its truth conditions; or what amounts to the same thing, by a translation into our own idiolect. That is, our explanation is that Kennedy believed (or anyway, assumed as a basis for action) some representation which had for him roughly the role that >Khrushchev will blink= has for

us, and his believing or assuming this was a salient cause of his actions, and that those actions didn't lead to war only because Khrushchev blinked. Again, it is a second class explanation, in that a comparison to our psychology is used in place of specifying the relevant details of the agent's psychology. As such, it is nothing that causes any problem for a deflationist.

## Footnotes

---

1. For instance, in the example as it stands, where the guru and the disciple share an underlying logic, the disciple has a notion of consequence that applies to the guru's language; one might argue that the disciple's faith in the guru is inadequately represented by the claim that everything the guru says is true, it should be represented by the claim that all consequences of what the guru says (indeed, all consequences of those together with other truths) are true. In that case, we would get the desired conclusion without the need to reason as in the previous paragraph.

2. It isn't essential to suppose that the disciple accepts the application of the truth schema to the guru's sentence prior to his knowing that an extraordinarily reliable guru uttered it and prior to knowing that it has the number-theoretic consequence. These pieces of information do give the disciple some *very minimal* understanding of the guru's sentence, so it isn't really necessary to hold that we can apply  $\text{true}$  to sentences of which we have *no understanding at all*.

3. This is so even if we know Mary to be unreliable. For instance, if we know her to have exaggerated beliefs, we might merely accept  $\text{George is over } 6=4$  when Mary accepts  $\text{George is over } 6=7$ . Or if Mary gushes over George, we might accept  $\text{Mary is infatuated with George}$  instead of  $\text{George is a wonderful guy}$ .

---

4. As noted in Leeds 1995, pp. 28-9, arguing in such a manner seems *prima facie* to be misguided: *of course* it is the correlation between the agent's states and external conditions that is of primary explanatory importance, but why should the fact that we use a correlation between an agent's states and our sentences as a means of setting up the correlation be thought to undermine this? Why should we think that there must be some means of specifying the relevant correlation that doesn't go via translation into our language but is given by a substantive theory of truth conditions? I would rather not press this line, however, because doing so leads quickly to complicated issues. (Indeed, the most confusing parts of the '86 paper concern precisely this.) Instead of arguing that the basic line of argument against the deflationist is misguided in very conception, I will argue merely that it cannot be carried out in detail. I believe that this way of proceeding has the additional advantage of making clearer the role that talk of representation plays in the explanation.

5. In probably the most crucial passage in my 1986 paper (p. 97 middle to 99 top), I assumed not, without any clear argument. [In the immediately preceding pages I had discussed avoiding mention of truth conditions in explanations, along the lines mentioned two paragraphs above. I then raised the question about what to do when we didn't know the full explanatory details. The discussion is obscure: it recognizes that it is legitimate for a deflationist to use a disquotational truth predicate as a way of giving an explanation sketch in absence of full knowledge of the details (and that even an inflationist would need to resort to this), but concludes for reasons that it doesn't make at all clear that we would need something more *inflationist@in addition*.] Stephen Leeds rightly focuses much of his critique of my paper on this crucial passage (Leeds 1995, pp. 17-20), and is understandably baffled as to what my argument could have been. For what it's worth, my own diagnosis is given in the next footnote.

6. I think the key problem in the 1986 paper was an ambiguity about >projectivist explanation=. In some

---

passages I used it in the more inclusive sense, as when I argued that the deflationist could appeal to the truth conditions of another agent's states only in projectivist explanations. But in other passages I used it in the narrower sense, as when I argued that the role to which we put the assignment of truth conditions in explanations of success phenomena like landing the plane with regularity can't be projectivist. The confusing structure of the paper disguised the ambiguity.