

## DISCUSIONES

CRÍTICA, *Revista Hispanoamericana de Filosofía*  
Vol. XXIX, No. 85 (abril de 1997): 95–119

### THE DEFLATION OF BELIEF STATES\*

ROBERT J. STANTON  
Philosophy Department  
Carleton University

Whistling in the dark is not the true method  
of philosophy.

W.V.O. Quine, *Word and Object*, p. 207.

In an earlier paper (Stanton 1996), I discussed an attempt to deflate *belief contents* (i.e. concepts, propositions and the like). In what follows, I will challenge an effort by Arthur Collins (1979, 1987, 1994, 1996) to deflate the *state of believing* (i.e. the neurological, or computational, or \_\_\_\_ state which *has* content). My critique will not, sad to say, be followed by a positive theory of belief states. Nor will I directly oppose belief-state deflation, broadly construed—though my remarks on Collins’ specific proposal should raise doubts about the larger project.

\* Earlier versions of this paper, or parts thereof, were presented at the 1997 Mid-South Philosophy Conference, and at the 1997 Canadian Philosophical Association. My thanks to those in attendance—and to Andy Brook, Ernie LePore, John Leyden and Daniel Stoljar for comments. Thanks also to an anonymous reviewer for *Critica*. Finally, I gratefully acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada.

## I. *A Preamble: Exegetical and Other Caveats*

The method of deflating belief that I want to discuss goes like this: belief states are completely exhausted by their success conditions. In the language game of believing, the agent *wins* when she is correct (i.e. when she believes truly) and she *loses* when she is mistaken (i.e. when she believes falsely). Success and failure of this kind is *all there is* to belief: no more “robust” internal state, be it physical, spiritual, or what-have-you, is required. This tradition, whose roots lie in the writings of Austin and Wittgenstein, is taken to be an alternative to a purportedly Cartesian-inspired view, according to which a belief is a substantial state of the agent, on a par with familiar physical states.

Discussing this approach in the abstract is difficult. One wants to know: how can there be states which aren’t “robust” and “substantial” in the sense that physical states are, but which are *states* nonetheless? Specifically, what exactly is the ontological status of belief states, understood as deflated? For that matter, what is the contrasting ontological status of inflated states? Or again: why isn’t the state which *grounds* the success-conditionally defined belief-state suitably inflated? (Compare: although winning the lottery may not be an inflated state, there is an inflated state which plays an important part. A “grounding” state, if you will: namely, holding a certain physical *ticket*.) The best way to deal with such general questions, in the present context, is to skirt them —by discussing one specific proposal.

I shall shortly turn, then, to Arthur Collins —who, in a series of recent writings (1979, 1987, 1994, 1996) provides a reasonably clear example of the success-conditional approach. He has argued for an account of belief in terms of *epistemic risk*. He has further suggested that, if his account of belief is correct, then there can be no “real constitution in the believer” (1996, p. 311) which ren-

ders belief-ascriptions true: having a belief cannot be, says Collins, a matter of being in a neurological, dispositional, spiritual, functional or other “inner state”. He insists that, “The crucial thing about all these [inflationary] theories is not the particular *constitution* that they give to the state of believing that *p*, it is that they give the state of believing that *p* any constitution at all”. (1994, p. 930) He continues, “States of believing cannot be identified with items constituted in any way”. (1994, p. 932) An exegetical/terminological note, however, before I present and challenge Collins’ views. In speaking of “deflating” belief states, I am introducing terminology that Collins himself does not use. What he repeatedly inveighs against is treating beliefs as having a “real constitution”, or as being “inner states of the agent”. His rejection of these tendencies merits a label. Since he is no fan of externalism (in the sense of Burge or Putnam) with respect to belief, the designation ‘anti-internalist’ seems inappropriate. In fact, he rejects Burge-Putnam style externalism precisely because externalists assign to beliefs a (wide) “constitutive state”. (See especially Collins 1994, p. 943.) One might, therefore, label Collins an anti-constitutionalist. But that’s too homely, and too uninformative. As I’ve said, Collins works in that tradition, in philosophy of mind and philosophy of language, which compares belief to such states as: having a responsibility, winning a game, being eligible to vote, etc. Like these, beliefs are not, it is claimed, intrinsic or spatially isolable states of the agent. Instead, if one can properly call them *states of the agent* at all, one should call beliefs *epistemic* states only. States like winning and believing are thus to be contrasted with “robust” physical states like: being square, having an electrical charge, and so on. Since the latter, whose nature is *not* exhausted by success conditions and the like, can reasonably be called ‘inflated states’, I choose to call Collins’ project deflation-

ary. This fits for another reason as well: if Collins' view is correct, there is no *thing*, "the belief state", about which one needs a scientific or philosophical theory. And that is the heart of deflationary theories.

I might note, finally, that I'm not sure that there are such things as correct analyzes; nor do I endorse the idea that a meaning-analysis of 'believes that', if possible, would have straightforward implications for the *metaphysics* of belief. To my mind, the nature of belief, and the meaning of 'believes', are two quite different issues. Nevertheless, I will suspend disbelief for the moment, in an effort to address Collins on his own terms.

## II. *An Argument from Meaning*

Here is how I reconstruct Collins' discussion. He provides two closely related arguments for the same conclusion, i.e. that belief is not a substantial state: one from the *meaning* of 'believes', a second from its *use*. I begin with the first of these, in which Collins uses his meaning-analysis of 'believes that' to show that beliefs cannot be inner states.

As an initial step, consider some desiderata for a theory of belief ascription, by way of motivating Collins' semantic analysis. The first and (for present purposes) most important requirement that Collins imposes on a semantics for 'believes' is that, to use his terminology, it must turn out that, in saying 'I believe that  $p$ ', the speaker puts herself at "epistemic risk". In particular, says Collins, the speaker must risk being mistaken: if  $p$  is false, someone who says 'I believe that  $p$ ' is wrong about  $p$ . The second requirement is this: that 'I believe that  $p$ ' must be capable of being true, even though  $p$  is false. (Put otherwise, people can believe falsehoods.) The final demand is that, even though 'I believe that  $p$ ' commits its utterer to  $p$ , whereas ' $S$  believes that  $p$ ' does not, no analysis of 'believes' should introduce

an ambiguity between first person and third person belief reports. That is, the meaning of ‘believe’ ought to be the same in ‘I believe that  $p$ ’ and ‘ $S$  believes that  $p$ ’.

Each demand is reasonable. Collins’ proposed analysis meets all three. Here is his idea: ‘I believe that  $p$ ’ is simply a means of asserting that  $p$ . The difference between saying  $p$  on its own, and saying ‘I believe that  $p$ ’, is that the latter explicitly leaves open the possibility that the speaker is mistaken. That is, prefacing  $p$  with ‘I believe that’ is a means of hedging one’s assertion of  $p$ . As Collins puts it, “‘I believe that’ functions as a means for weakening the claim that  $p$ , and not as a device for changing the subject”. (1996, p. 316) Or again,

If we ask what it is that the speaker actually commits himself to in “I believe that  $p$ ” we get the answer by merely deleting the prefacing words “I believe that” [...] these words are standardly selected by a speaker, not because he is asserting something other than the  $p$  that follows, but in order to introduce an element of guardedness or an expression of insecurity about what would otherwise be a mere assertion of  $p$ . (1994, p. 940)

As he suggests (1987, p. 31; 1987, p. 166; 1994, p. 937), his view may be roughly summarized as follows. (This “principle” will be amended slightly in what follows.)

1. *Collins’ Principle*: ‘I believe that  $p$ ’ amounts to ‘ $p$ , or I am mistaken about  $p$ ’

Let me illustrate with an example. On Collins’ view, someone who says ‘I believe that it snows in Montevideo’ asserts that it snows in Montevideo. She therefore makes it the case that, if it *doesn’t* snow in Montevideo, she is wrong. And just what is the difference between saying (2) and saying (3)?

2. I believe that it snows in Montevideo

### 3. It snows in Montevideo

Applied to these examples, Collins' claim comes to this: (2) and (3) are alternative means for making the same conversational move —both convey the speaker's epistemic risk with respect to the proposition that it snows in Montevideo. The only difference, thinks Collins, is that (3) commits the speaker more seriously —her risk is therefore greater. That is, as Collins explicitly says, adding 'I believe that' is a way of expressing mild uncertainty —and nothing more. That is the proposal. Consider now how it satisfies the three desiderata.

First, and most obviously, this analysis of 'believes that' explains how speakers manage to commit themselves to  $p$  in saying 'I believe that  $p$ '. The reason: uttering 'I believe that  $p$ ' is, according to Collins' Principle, very like saying ' $p$ , unless I'm much mistaken'. And this latter phrase clearly commits one to  $p$ . Consider example (2) again. According to Collins, a speaker who utters (2) asserts something *about Montevideo* (namely, that it snows there), though she reduces her "risk" by choosing this form of words over (3) —just like someone who says, 'It snows in Montevideo, unless I'm much mistaken' communicates, though with reservations, that Montevideo gets snow. It is, therefore, no surprise that an utterance of (2) commits the speaker to certain facts about Montevideo's weather.

Second advantage of (1): it makes room for cases where 'I believe that  $p$ ' is true, even though  $p$  is false. Here again, it's clear why: ' $p$ , or I am mistaken about  $p$ ' can be true, even when  $p$  is false —for the simple reason that the second disjunct can be true, while the first disjunct is false. As Collins puts it:

The words ' $p$ , or I am much mistaken' express belief that  $p$ , and they will be true when  $p$  is false, not because they make a claim about something else, but because they expressly

allude to the possibility of error which is an ineliminable part of the concept of belief. If  $p$  is false, my assertion ‘ $p$ , or I am much mistaken’ is true, for the unspectacular reason that I am mistaken and my disjunctive assertion has canvassed that possibility in advance. (1996, p. 318. See also Collins 1987, p. 32.)

Finally, one need not introduce two senses of ‘believe’—one for first person reports, the other for third person reports—in order to explain why belief attributions sometimes commit the speaker to the belief attributed (the first person case), and sometimes don’t (the third person case). (For example, ‘I believe Cobain is dead’ commits its speaker to Cobain being dead, whereas ‘He believes Cobain is dead’ *doesn’t* so commit the speaker.) Collins’ suggestion regarding the third person case can be easily extrapolated from (1): ‘ $S$  believes that  $p$ ’ means, in effect, that  $p$  is true, or  $S$  is mistaken about  $p$ . Given this, a speaker attributing the belief that  $p$  to  $S$  doesn’t herself assert that  $p$ , because she makes a *disjunctive* (i.e. a conditional) assertion: ‘If  $\sim p$ , then  $S$  is mistaken about  $p$ ’. To say *this* is obviously not to say that  $p$  in fact obtains.

That the proposal meets all three requirements is decidedly a blessing. Another advantage of (1) is still more interesting, however. As Collins explains, his meaning-analysis of ‘believes that’ provides an elegant solution to a “paradox” introduced by G.E. Moore (1944a, p. 207):

it’s perfectly absurd or nonsensical to say such things as ‘I don’t believe it’s raining, but as a matter of fact it is’ or (what comes to the same thing) ‘Though I don’t believe it’s raining, yet as a matter of fact it really is raining’.

Collins thinks he knows why ‘I believe that  $p$ , and  $\sim p$ ’, a variant on this paradox, is “absurd”. To understand his account, consider again sentence (2). According to Collins’ Principle (1), in saying (2), the speaker in effect asserts:

4. It snows in Montevideo or I'm mistaken about it's snowing in Montevideo

In which case, in saying the Moore's Paradox sentence (5) below, the speaker asserts (6):

5. I believe that it snows in Montevideo, and it doesn't snow in Montevideo

6. (It snows in Montevideo or I'm mistaken about it's snowing in Montevideo) and it doesn't snow in Montevideo

The thing is, (6) entails (7).

7. I'm mistaken about it's snowing in Montevideo

And anyone uttering *this latter sentence* simultaneously assigns both TRUE and FALSE to the proposition that it snows in Montevideo. How so? Well, *S* can be *mistaken* about *p* only if *S* holds *p* to be true, and *p* is false; or *S* holds *p* to be false, and *p* is true. Consider the first case, where *p* is false. If 'I am mistaken about *p*' is true, then I assign TRUE to *p*. However, any speaker who says 'S is mistaken about *p*' must assign FALSE to *p*, if *S* assigns it TRUE. So, if I utter 'I am mistaken about *p*' then, being the speaker, I must assign FALSE to *p*, given that the reportee (namely, me) assigns it TRUE. In which case, I assign *p* both truth values. Now consider the second case, where *p* is true. If 'I am mistaken about *p*' is true, then I assign FALSE to *p*. The person who reports my mistake (namely, me) must assign TRUE to *p*. So, here again, I assign to *p* both TRUE and FALSE. Hence Moore's Paradox sentences are "absurd" only in the way that ordinary self-contradictions are.

I have just argued that Collins' proposed meaning-analysis of 'believes that' brings with it several rewards. It has an associated cost, however: the centrality of epistemic risk to the concept of belief appears to falsify any "inflationary"



theory. The reason, says Collins, is that no statement of the form ‘I am in state  $N$ ’ puts the speaker at epistemic risk with respect to, for example, the proposition that it snows in Montevideo. This holds true whether  $N$  is taken to be a state of Cartesian mind-stuff, a neurophysiological state, a dispositional state, or any other “inflated” state: that is, Collins takes his analysis of ‘believes’ —articulated in (1)— to be incompatible with *any* view that assigns beliefs “some real constitution in the believer” (1996, p. 311).

He illustrates this point as follows. Let ‘(BSp)’ refer to some purported “neural reality” constituting (or realizing) the state of belief that  $p$  in  $S$ . (That it be neural does not matter. A similar argument could be run, he thinks, for any “robust” state.) Given this stipulation, (8) is true if and only if (BSp) exists *in the speaker’s own brain*. (This isn’t to say, of course, that everyone uttering (8) must *know* that it’s true if and only if (BSp) is present. But still, this state must be the “truth maker” of the sentence.)

8. ‘I believe that  $p$ ’

Now recall Collins’ claims about (8). He maintains that, in saying (8), the speaker *asserts that  $p$* ; her assertion is about whatever  $p$  is about; in saying (8), she commits herself to being mistaken about  $p$ , if  $p$  is false; and, as a result, ‘I believe that  $p$ , and  $\sim p$ ’ turns out to be contradictory. The thing is, if what makes (8) true is the same state of affairs which makes-true ‘(BSp) is present in my brain’, then none of these conditions are met. Collins concludes that the supposition that ‘(BSp)’ refers to some “inflated” something is false; that is, there *is no* constitutive state, neural or otherwise, whose presence makes (8) true.

The force of this can be brought out by considering example (2) again. The speaker does not, in tokening (2), say something about herself —if Collins’ analysis in (1) is right. In particular, she does not describe her own mental state.

Instead, a speaker of (2) commits herself to the proposition that it snows in Montevideo. That's precisely why (5) is absurd. This "basic fact" about epistemic success and failure, as Collins would have it, is purportedly missed by any semantic account of 'believes that' which introduces a belief state as the "truth maker" —whether proposed by Descartes, Armstrong, Fodor or any other philosopher of mind.

### III. *Problems with Collins' Semantic Analysis*

This first argument, for the deflation of belief states, may be summarized as follows:

*Premise One:* The semantic analysis of 'believes' in (1) is correct.

*Premise Two:* If the semantic analysis of 'believes' in (1) is correct, then beliefs are not inflated states.

One natural response to this argument is to reject Premise Two: one might well complain that this premise confuses issues about the nature of belief with issues about the meaning of 'believes'. But this isn't the only means of response. In what follows I present some fairly specific—even technical—problems with Collins' meaning-analysis. They effectively block his argument against inner belief states, by showing that *the first* premise is false. Before I introduce these specific worries, however, I should say a little about the general orientation within which they become important. It seems to me that Collins moves too quickly from the *use* of 'believes that  $p$ ', in first person reports, to the *meaning* of this predicate. To establish the semantic equivalence of 'I believe that  $p$ ' and ' $p$ , or I am mistaken about  $p$ ' it is not enough to show that these two sentences are frequently, or even typically, used in the same way. Uncovering the meaning of an expression

requires (at least) looking at the semantic relationships between it and other sentences —that is to say, an examination of its *logical form*. It is here, as I'll now argue, that Collins goes wrong. I don't mean to say, of course, that use is irrelevant to meaning: use constrains semantic theorizing in the sense that, given the postulated meaning *along with other relevant factors* it ought not be a mystery that the expression is used as it is. However, as I'll eventually suggest, one can explain the fact that 'I believe that  $p$ ' commits the speaker to  $p$ , while nevertheless denying that this is merely a guarded way of asserting  $p$ .

Let me begin by noting a problem with Collins' view, in order to put it aside. Collins' Principle allows for 'I believe that  $p$ ' to be true while  $p$  is false. But notice how this is achieved: 'I believe that  $p$ ' is, for Collins, essentially a weakened form of  $p$  itself. Now consider: it would seem, precisely because of this, that 'I believe that  $p$ ' will be true whenever  $p$  is. Put otherwise, if (8) and (9) were really *logically equivalent* then, because the truth of  $p$  is sufficient for the truth of (9),  $p$  should also be sufficient for the truth of (8).

8. 'I believe that  $p$ '

9. ' $p$ , or I'm mistaken about  $p$ '

Now, it's clearly not the case that one believes *every* true proposition. So, there appears to be a problem with Collins' Principle (1). At this juncture one might point out that, for Collins, (8) doesn't *literally* express a disjunctive proposition; or, at any rate, that 'or' here shouldn't be read as the logician's 'vel'. Saying 'I believe that  $p$ ' is, instead, like saying  $p$  in a soft, hesitant voice. But this doesn't really improve things, since it's even *more* manifest that whenever  $p$  is true,  $p$  (even said with hesitation) is true.

Collins suggests an escape, however: slightly recast (1), treating it as partially elliptical. The resulting (fully spelled out) principle is:

10. *Collins' Principle (Revised)*: 'I believe that  $p$ ' amounts to ' $p$ , and I am right about  $p$ ' or ' $\sim p$ , and I am mistaken about  $p$ '

Given this revised form, even when  $p$  is true, 'I believe that  $p$ ' needn't be. For, when  $p$  is true but 'I am right about  $p$ ' is *false*, both disjuncts in ' $p$ , and I am right about  $p$ ' or ' $\sim p$ , and I am mistaken about  $p$ ' come out false. Which makes (8) appropriately false. From here on, then, I will take Collins' Principle to be (10).

Having put aside the foregoing non-problem, I will now argue that, though it ought to be the case that, for any meaning analysis, the analysandum has the same logico-semantic properties as the proposed analysans, Collins' paraphrases actually exhibit different logico-semantic properties than the sentences they purportedly analyze.

First case: 'believes that' sentences exhibit ambiguities which Collins' paraphrases do not. Thus (11) has both a narrow and a wide scope reading.

11. John believes that a liberal will win

The narrow scope reading is paraphrasable as 'John believes that the winner, whomever it may be, will be liberal'; while the wide scope reading is better captured by 'There is a liberal whom John believes will win'. An adequate analysis of (11) ought to capture this ambiguity, or at least explain it away. But Collins' paraphrase does neither: (12) is not ambiguous, and offers no insight into why (11) is.

12. Either a liberal will win and John is right about a liberal winning, or a liberal will not win and John is mistaken about a liberal winning

There is, of course, a reason for this discrepancy between the ambiguous original and the univocal Collins-paraphrase: belief attributing sentences are paradigm cases of opaque contexts, and embedding a quantifier in an opaque context typically gives rise to a scope ambiguity. Hence the ambiguity of (11). Crucially, however, Collins' paraphrase of (11) has the quantifier phrase 'a liberal' in *transparent* positions. So no ambiguity is observed.

It's worth noting, by the way, that it is not only the first and third occurrences of 'a liberal' (i.e. the ones right after 'either' and 'or' respectively) in (12), but also the occurrences after 'is right about' and 'is mistaken about', which must be considered to be in transparent positions. Here's why. To insist, to the contrary, that 'is right about \_\_\_' and 'is mistaken about \_\_\_' be read as an inducing opacity is, in the present context, effectively to *assume* an understanding of propositional attitudes (including 'believes that'), and to use *that very understanding* to explain 'is right about \_\_\_'. Taking 'is right about' and 'is mistaken about' to be understood opaquely, Collins could not go on to analyze 'believes that' in terms of being correct or mistaken—that would be circular. (In contrast, an analysis of belief which relates the agent to a *sentence* explains the opacity of belief-ascriptions by assimilating it to something understood independently, namely non-substitutivity in *quotational* contexts. See Fodor (1978) for an example of this approach.) In a nutshell: one cannot understand what it is to be correct/mistaken, *in the sense non-substitutivity requires*, without already knowing what it is to have a true or false *belief*. In which case, 'right' and 'correct' must, for Collins, essentially be variants on 'true'—though they can be predicated of both persons *and* propositions. But 'is true' is transparent if anything is. In which case, 'John is

right about \_\_\_' and 'John is mistaken about a \_\_\_' must be read transparently.<sup>1</sup>

A second logico-semantic feature of 'believes that  $p$ ' is that it allows quantification over the complement position. For instance, existential generalization holds: (14) and (15) both follow from (13).

13. John believes that it snows in Montevideo
14. John believes something about Montevideo
15. John believes something

Now, a good analysis of belief should satisfy two requirements: first, it should be applicable to every sentence containing 'believes';<sup>2</sup> second, where a sentence  $s_1$  entails a sentence  $s_2$  in virtue of form, the paraphrase of  $s_1$  should entail the paraphrase of  $s_2$  in virtue of form. (Compare Davidson (1967) on the semantics of event sentences: the problem with 'Butter( $j$ )( $b$ )( $k$ )' as a translation of 'John buttered the bread with a knife' is that this predicate calculus sentence *doesn't* entail, in virtue of its form, 'Butter( $j$ )( $b$ )' —whereas the original English sentence *does* entail 'John

<sup>1</sup> A closely related problem: if '( $p$ , and I am right about  $p$ ) or ( $\sim p$ , and I am mistaken about  $p$ )' is true then, because the occurrences of  $p$  in this expression are transparent, '( $q$ , and I am right about  $q$ ) or ( $\sim q$ , and I am mistaken about  $q$ )' is also true —where  $q$  is *any* logical consequence of  $p$ ! And yet it is not the case that an agent who believes  $p$  inevitably believes *every* logical consequence of  $p$ ; at best, one believes only the *obvious* entailments of one's beliefs. So, 'I believe that  $p$ ' does not entail, for every entailment  $q$  of  $p$ , 'I believe that  $q$ '. In which case, '( $p$ , and I am right about  $p$ ) or ( $\sim p$ , and I am mistaken about  $p$ )' and 'I believe that  $p$ ' have different logical properties —and hence cannot, contra Collins, be synonyms.

<sup>2</sup> Including, since I'm on the topic of quantified complements, sentences like: 'John believes every line of *The Philosophical Investigations*'. Such sentences may pose a problem for Collins. One can *imagine* treating the above as 'For every line of *The Philosophical Investigations*, if it is false, then John is mistaken about it', but the ultimately plausibility of this move is unclear. I leave the issue open.

battered the bread' in virtue of form.) Unless both demands are met, one fails to capture the logical form of the target sentence. I want to focus on the second requirement, of capturing form-based entailments.

Applied to Collins' proposal, the following difficulty arises. Sentence (16) is clearly ill-formed, so *it* can't be Collins' paraphrase of (14).

16. \*\*Something about Montevideo and John is right about something about Montevideo, or it's not the case that something about Montevideo and John is mistaken about something about Montevideo.

Such ill-formedness isn't my concern, however: there are, I suppose, numerous alternative, and well-formed, paraphrases of (14). For instance, one could paraphrase (14) as 'John is either right or wrong about something to do with Montevideo'. The key point is rather this: no paraphrase of (14) is likely to come out true in virtue of the *form* of its Collins-paraphrase, 'It snows in Montevideo and John is right about it's snowing in Montevideo, or it doesn't snow in Montevideo and John is mistaken about it's snowing in Montevideo'. Which suggests that Collins' Principle (10) does not give *the logical form* of (8). Again, it is one thing to accurately give the truth conditions of 'S believes that  $p$ '; it's quite another to uncover its logical form—in the sense of highlighting the logical relations of this sentence in a formal way. It seems to me that, *even if* (10) succeeded in the first task, it fails at the second. So, Premise 1 is false: Collins' semantic analysis of 'believes' in is *not* correct.

A final objection to (10), this one by counter-example.<sup>3</sup> Consider Jason, a philosopher who embraces truth val-

<sup>3</sup> The idea for the counter-example came from Daniel Stoljar, in conversation. We subsequently collaborated on its elaboration and application to (10).

ue gaps. Suppose Jason takes (17) to lack a truth value—because he is convinced that Canada has no queen.

17. The present Queen of Canada lives in Ottawa

In fact (17) *does* have a truth value, namely FALSE: the present Queen of Canada (i.e. Elizabeth II) lives not in Canada, but in England. Furthermore, because Jason takes (17) to lack a truth value, he is *mistaken* about the Queen of Canada living in Ottawa.<sup>4</sup> Therefore, the following conjunction is true:

18. It's not the case that the present Queen of Canada lives in Ottawa, and Jason is mistaken about the present Queen of Canada living in Ottawa

The truth of (18) is, of course, sufficient for the truth of the disjunction in (19):

19. Either (the present Queen of Canada lives in Ottawa and Jason is right about the present Queen of Canada living in Ottawa) or (it's not the case that the present Queen of Canada lives in Ottawa, and Jason is mistaken about the present Queen of Canada living in Ottawa)

So, in the imagined case, (19) comes out true. But (19) has the form  $\lceil(p, \text{ and } S \text{ is right about } p) \text{ or } (\sim p, \text{ and } S \text{ is mistaken about } p)\rceil$ , the supposed analysis of  $\lceil S \text{ believes that } p\rceil$ . So, according to Collins' Principle, Jason believes (17)—i.e. that the present Queen of Canada lives in Ottawa. Yet this is not the case: by hypothesis, Jason considers (17) to lack a truth value. So, the revised version of Collins' Principle

<sup>4</sup> One might reply: in the proposed example, Jason is “mistaken” *in the wrong sense*. Maybe so. But the obvious attempt to say what the *right* sense is—i.e. Jason must be mistaken in that he believes something false—evidently presupposes an understanding of what it is to believe that  $p$ . And, as before, Collins cannot explicate being mistaken in terms of belief—without falling into circularity. (Thanks to Ian Gold for noting this possible rejoinder.)



admits of counter-example. Therefore, once again, Collins cannot argue for deflationism, from the correctness of his semantic analysis of ‘believes’.

#### IV. *An Argument from Use*

It may reasonably be suggested that the technical worries I have raised miss the central point: that someone who says ‘I believe that  $p$ ’ commits herself to  $p$ .<sup>5</sup> One way of explaining this fact is to adopt (10) as an *analysis* of ‘believes that’. This is the route I have been considering. As I said, thinking this way, Collins’ argument goes as follows: if (10) is true, then every inflationary theory is false. But then my reply is that (10) is not, in fact, true. However, Collins’ argument against robust, inner states of belief may require less than the truth of (10): maybe all that’s required is the basic fact that ‘I believe that  $p$ ’ commits its speaker to  $p$ —not a purported explanation of this fact, captured in (10). Putting (10) aside, then, Collins’ argument becomes:

*Premise 1'*: Saying ‘I believe that  $p$ ’ commits the speaker to  $p$ .

*Premise 2'*: If inflationism is true, then saying ‘I believe that  $p$ ’ does not commit the speaker to  $p$ .

In my view, Premise 1 is essentially correct. However, as I’ll now argue, Premise 2’ is false: positing “robust” belief states is perfectly consistent with the hypothesis that speakers who utter ‘I believe that  $p$ ’ place themselves at epistemic risk with respect to  $p$ . Before presenting my arguments against it, however, I’d like to consider what *supports* Premise 2’. Let me start with an example. Collins

<sup>5</sup> Collins would likely think that they *do* miss the point, precisely because they are technical problems. He writes (1994, p. 929), for example: “A lot of contemporary work in the philosophy of mind combines technical sophistication and inadequate thinking about preliminaries.”

asks us to reflect upon what, for example, ‘I believe that Cobain is alive’ would mean, if there really were “constitutive belief states”. He notes first that, whereas (20) is about Cobain, sentence (21) looks to be about the speaker herself—according to inflationary theories.

20. Cobain is alive

21. I believe that Cobain is alive

Moreover, Collins suggests that, according to inflationists, ‘I believe that Cobain is alive’ would be true of the speaker if and only if some substantial state  $BS_p$  were present in her. Now, Collins points out, the presence of this state in the speaker, and the truth of (20), “are simply two distinct and independent matters of fact” (1996, p. 312). So, he concludes, inflationists must maintain that someone who says (21) is not asserting anything about Cobain at all; in which case, someone who sincerely says (21) is not, according to inflationists, necessarily mistaken (i.e. about Cobain), even though Cobain is in fact dead!

But, Collins says, this is absurd. As a matter of fact, a speaker cannot affirm her belief state, vis-à-vis Cobain, without also taking a stand on Cobain himself. Yet, says Collins, she ought to be able to, if beliefs are “inflated states”. Generalizing, if inflationism about belief states is true, then ‘I believe that  $p$ ’ commits the speaker not to  $p$ , but rather to some fact about her own belief state; and, by introducing such a special subject matter for belief reports, inflationism makes “conceptual room for reports of belief that merely assert the presence of something in the believer” (1994, p. 902). It therefore seems to insure the truth of Premise 2’.

My response to this line of argument? If, contra Collins, there are robust belief states, then ‘I believe that  $p$ ’ may not be a means of *asserting* that  $p$ ; nevertheless, uttering this sentence (typically) *commits* the speaker to  $p$ . So Premise 2’

is simply false. In a word, being committed to  $p$ , by saying ‘I believe that  $p$ ’, is a *pragmatic* phenomenon.

This pragmatics-based approach is suggested, first of all, by the expressions below.

8. ‘I believe that  $p$ ’
22. ‘I sense that  $p$ ’
23. ‘I fear that  $p$ ’
24. ‘I suspect that  $p$ ’
25. ‘I presume that  $p$ ’

Each of the above puts the speaker at epistemic risk. In each case, the speaker is (typically) mistaken if  $p$  is false. Given this, it would seem to be a *general* feature of these kinds of verbs that, in uttering them, one commits oneself to  $p$ . Yet, as most of these examples attest, one does this *precisely by* describing one’s own mental state.<sup>6</sup> To take one example: (23) pretty clearly *is* about the speaker’s mental state, and yet saying it does commit the speaker to  $p$ . So, *pace* Collins, sentence (8) could be “about the speaker” —and yet still have an utterance of it commit the speaker to  $p$ .

What’s more, one can see *why* saying ‘I believe that  $p$ ’ (or related constructions) typically commits the speaker to  $p$ . Put in terms of Relevance Theory, ‘I believe that  $p$ ’ ordinarily has  $p$  as an implicature —that is to say, the speaker *communicates* the thought that  $p$ , without actually *saying* that  $p$ . The same will be true of ‘I fear that  $p$ ’ and all the rest: the implicature will normally be present whenever the thought ‘I  $\phi$  that  $p$ ’, taken alone, is insufficiently relevant

<sup>6</sup> A related point: how are we to distinguish between these expressions, meaning-wise? In particular, how are we to single out (8), with respect to what is predicated of the agent? The natural answer is: the truth of each depends, in part, upon *the mental state* predicated of the speaker. But this answer seems to fly in the face of belief-state deflation.

to warrant the processing involved in understanding the utterance. (See Sperber and Wilson (1986) for discussion. Note that their use of ‘implicature’ departs somewhat from Grice’s.)<sup>7,8</sup> At the same time, saying ‘ $\sim p$ ’ communicates (indeed “explicates”, as Sperber and Wilson put it) that  $p$  is false. So, someone who utters ‘I believe that  $p$ , and  $\sim p$ ’ communicates —though she does not assert— that  $p$  is both true and false. Indeed, what is communicated is doubly odd, since *saying* ‘ $\sim p$ ’ is the best possible evidence for the proposition that one believes that  $\sim p$ . Thus, in saying ‘I believe that  $p$ , and  $\sim p$ ’, a speaker implies, though she does not assert, that she *believes* both  $p$  and  $\sim p$ .<sup>9</sup>

<sup>7</sup> Collins suggests that this approach “won’t succeed because a conversational implicature is always expressly cancelable” (1994, p. 934). This merits two comments: first, not all unasserted-but-communicated information can be canceled by the speaker; second, commitment to  $p$  can be canceled by a speaker of ‘I believe that  $p$ ’. Consider the following example, due to Dan Sperber (p.c.): “Oh Doctor, I must be going mad. I somehow deeply believe that rain is always an illusion, even though I can see for myself that it’s raining. So you see, Doctor, I believe that it’s not raining, and yet it is raining!” Here, the patient’s claim ‘I believe that it’s not raining’ is clearly relevant enough when taken *only* as a claim about herself. It therefore lacks an implicature to the effect that it is not raining.

<sup>8</sup> I don’t mean to suggest, of course, that Relevance Theory offers the *only* pragmatics-based alternative to Collins. Indeed John Searle, responding to Malcolm (1991), offers another one. He writes: “Now, the reason that the sentences of the form ‘I believe that  $p$ ’ can often be used to make hesitant assertions, even though the sentence does not *mean* ‘I hesitantly assert that  $p$ ’, is that belief is a sincerity condition on the speech act of asserting, and one in general can, by familiar mechanisms of indirect speech acts, perform a speech act by simply asserting that one has the sincerity condition in question. Thus, ‘I believe that  $p$ ’ can be used indirectly to make an assertion that  $p$  in the same way that ‘I want you to leave the room’ can be used to order you to leave the room. . .” (Searle 1991, p. 187)

<sup>9</sup> Two exegetical notes. First, Moore held that, “by saying ‘he has not gone out’ we *imply* that we do *not* believe that he has gone out, though we neither assert this, nor does it follow from anything we do

In sum, one can explain why saying ‘I believe that  $p$ ’ commits one to  $p$ , without giving up the idea that what it asserts is the presence of some substantial, robust, “inflated”, state in the speaker. The idea, in a nutshell, is that someone speaking ‘I believe that  $p$ ’ communicates, though she does not assert, that  $p$ . A nice feature of this solution—in contrast with Collins’ own—is that it explains why a whole family of propositional attitude self-ascriptions commit one in this way. What they all share is the purely pragmatic property of (usually) being irrelevant to the hearer, when taken merely as statements about one’s mental state.<sup>10</sup>

Reflecting upon all this, by the way, one immediately sees an alternative means of satisfying Collins’ three desiderata for an analysis of ‘believes that’:

## 26. Collins’ Desiderata

a. In saying ‘I believe that  $p$ ’, the speaker puts herself at “epistemic risk”

assert. That we *imply* it means only, I think. . . [that] people, in general, would not assert positively ‘he has not gone out’, if they believed that he had gone out” (1944b, p. 204). On the other hand, Grice (1989) did *not* treat ‘I believe that  $p$ ’ as merely conversationally implicating  $p$ .

<sup>10</sup> That the phenomenon is essentially a matter of speaker’s meaning, rather than of semantics, is also suggested by a thought experiment. Imagine a language—call it Anticollinsian—in which ‘I credo that  $p$ ’ really *does* attribute a substantial mental state to the speaker. ‘I credo that  $p$ ’ is *not* a way of saying  $p$ , and it is *not* about whatever  $p$  is about. Having stipulated this semantics for the Anticollinsian word ‘credo’, one can ask whether a speaker of this language would ever commit herself to  $p$  in saying ‘I credo that  $p$ ’. The answer is that she might—if the explicit content of her self-ascription would be insufficiently relevant. Indeed, it might end up that ‘I credo that  $p$ ’ is *typically* used to communicate that  $p$ . Such usage would not alter the fact that ‘I credo that  $p$ ’ is not, in Anticollinsian, merely a way of asserting  $p$ . (Compare indirect speech acts, in the sense of Searle (1975)—‘Can you pass the salt?’ is most often used to make a request. But this isn’t its literal meaning.)

- b. Even when  $p$  is false,  $\lceil I \text{ believe that } p \rceil$  may be true
- c. The meaning of ‘believe’ is the same in  $\lceil I \text{ believe that } p \rceil$  and  $\lceil S \text{ believes that } p \rceil$

Assume that both  $\lceil S \text{ believes that } p \rceil$  and  $\lceil I \text{ believe that } p \rceil$  are about the agent’s “inflated” belief state, not about whatever  $p$  is about. This makes ‘believes that’ unambiguous—which satisfies desideratum number three. Moreover, it turns out that  $p$  and  $\lceil I \text{ believe that } p \rceil$  affirm rather different facts—the latter is about the believer’s belief state, while the former (generally) is not—hence it’s no surprise that their truth values may diverge. So (26b) is satisfied. Precisely because ‘believes’ is unambiguously about the speaker’s state, however,  $\lceil I \text{ believe that } p \rceil$  is no more a means of asserting  $p$  than  $\lceil S \text{ believes that } p \rceil$  is.  $\lceil I \text{ believe that } p \rceil$  can, however, be a means of *communicating* that  $p$ . And when it is, the speaker ends up committed to  $p$ . Which satisfies desideratum number one: the speaker of  $\lceil I \text{ believe that } p \rceil$  puts herself at epistemic risk with respect to  $p$ —though she does this not by asserting  $p$ , but by asserting something about herself.

## V. Conclusion

I have considered two arguments from Arthur Collins, against treating beliefs as inflated “constituted inner states”. Each relies on an appeal to success conditions. The first argument employed a success conditional *semantic analysis* of ‘believes’. I replied to it by showing that the proposed semantics was inadequate. The second argument capitalized on a feature of the *use* of ‘believes’—i.e. the fact that someone who says  $\lceil I \text{ believe that } p \rceil$  runs the risk of failure with respect to  $p$ . It was suggested that this use is inconsistent with the postulation of inflated belief states. I argued, however, that there is no such inconsistency between beliefs-as-robust-states and the use of  $\lceil I \text{ believe that } p \rceil$ .

$p$ <sup>1</sup>. In which case, Collins has yet to provide an argument against deflating belief states.

#### REFERENCES

- Collins, Arthur W., 1979, "Could Our Beliefs Be Representations in Our Brains?", *The Journal of Philosophy*, 76(5), pp. 225–243.
- , 1987, *The Nature of Mental Things*, University of Notre Dame Press, Notre Dame.
- , 1994, "Reply to Commentators", *Philosophy and Phenomenological Research*, 54(4), pp. 929–945.
- , 1996, "Moore's Paradox and Epistemic Risk", *The Philosophical Quarterly*, 46, pp. 308–319.
- Davidson, Donald, 1967, "The Logical Form of Action Sentences", in N. Rescher (ed.), *The Logic of Decision and Action*, The University of Pittsburgh Press, Pittsburgh. Reprinted in D. Davidson, 1980, *Essays on Actions and Events*, Clarendon Press, Oxford.
- Fodor, Jerry A., 1978, "Propositional Attitudes", *The Monist*, 61(4), pp. 501–523.
- Grice, H. Paul, 1989, *Studies in the Way of Words*, Harvard University Press, Cambridge.
- Malcolm, Norman, 1991, "I Believe that  $p$ ", in E. LePore and R. van Gulick (eds.), *John Searle and His Critics*, Blackwell, Oxford.
- Moore, G.E., 1944a, Untitled Manuscript, published as "Moore's Paradox", in T. Baldwin (ed.), 1993, *G.E. Moore: Selected Writings*, Routledge, London.
- , 1944b, "Russell's Theory of Descriptions", in P.A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, Northwestern University Press, Evanston, Illinois.
- Searle, John R., 1975, "Indirect Speech Acts", in P. Cole and J.L. Morgan (eds.), *Syntax and Semantics*, vol. 3, Academic Press, New York, pp. 59–82. Reprinted in A.P. Martinich (ed.), 1990, *The Philosophy of Language*, Oxford University Press, Oxford.

- , 1991, “Response: Perception and the Satisfactions of Intentionality”, in E. LePore and R. van Gulick (eds.), *John Searle and His Critics*, Blackwell, Oxford.
- Sperber, Dan and Deirdre Wilson, 1986, *Relevance: Communication and Cognition*, Harvard University Press, Cambridge.
- Stainton, Robert J., 1996, “The Deflation of Belief Contents”, *Critica*, 28(84), pp. 63–82.

*Recibido: 11 de agosto de 1997*



## RESUMEN

Recientemente, se han intentado dar explicaciones “deflacionarias” de las creencias: ya sea de los *contenidos de creencias* (i.e. conceptos, proposiciones y demás) o del *estado de creencia*. En este trabajo presento y critico una de estas explicaciones deflacionarias de los estados de creencia, a saber, aquella defendida por Arthur Collins, la cual se basa en comentarios hechos por Wittgenstein en sus *Investigaciones filosóficas*.

En primer lugar, argumento que el análisis semántico de Collins según el cual  $\lceil \text{Yo creo que } p \rceil$  simplemente se reduce a  $\lceil p \text{ o estoy equivocado acerca de } p \rceil$  concibe inadecuadamente la *forma lógica* de los reportes de creencia. Primero, el análisis deja fuera ciertas ambigüedades clave de alcance; segundo, no captura al menos una inferencia válida basada en la forma; y tercero, tiene contraejemplos cuando los agentes creen de manera incorrecta que una oración carece de valor de verdad. Por lo tanto, Collins no puede apoyar su conclusión deflacionaria utilizando este análisis.

En segundo lugar, niego que el *uso* (en contraposición con el análisis semántico) de  $\lceil \text{Yo creo que } p \rceil$  apoye una conclusión deflacionaria, pues el hecho de que preferencias de  $\lceil \text{Yo creo que } p \rceil$  comprometan al hablante con  $p$  puede explicarse de manera *pragmática*, aun asumiendo que lo que hace que  $\lceil \text{Yo creo que } p \rceil$  sea verdadero es un estado “inflado”.

[Traducción: Maite Ezcurdia]