

## **INTERPRETATION AND FIRST-PERSON AUTHORITY: DAVIDSON ON SELF-KNOWLEDGE**

**David Beisecker**  
**University of Nevada, Las Vegas**

It is a curious feature of our linguistic and epistemic practices that assertions about a speaker's own state of mind, and their beliefs in particular, enjoy a greater credence (or presumption of truth) than assertions about another's state of mind. What is even more remarkable is that we typically accept as credible the first-person reports of *changes* of mind, despite the fact that there might be precious little or even no publicly observable evidence to support such a claim. How, then, are we to account for this incredible credibility that speakers have with respect to their own mental states? Contemporary philosophers have been understandably reluctant to follow Descartes and conclude that knowledge of our own minds is knowledge about a non-physical, inner realm to which we have a direct and incorrigible access, which is different in kind from the access we have to the external world and to the minds of others. For if the manner in which we have knowledge of other minds is so different from the way we know our own minds, then it becomes hard to see how the mental concepts we apply to ourselves could be the same as those we apply to others. Still, even if we avoid elevating this epistemological gulf into an ontological one, the puzzling fact of first-person authority remains to be explained-- or explained away. Recently, Donald Davidson has suggested that an explanation of first-person authority naturally falls out of a proper understanding of interpretive constraints.<sup>1</sup> However, as is so characteristic of Davidson's work, these arguments are overly compressed and in desperate need of unpacking. His most

extensive account of first-person authority occupies only a couple pages. In this paper, I'd like to explain why I find Davidson's discussion of the matter less than compelling.

## I

At the very end of "First Person Authority," Davidson summarizes his argument as follows:

"There is a presumption - an unavoidable presumption built into the nature of interpretation - that the speaker usually knows what he means. So there is a presumption that if he knows that he holds a sentence true, he knows what he believes." (FPA, p. 111)

When we unpack the conditionals in this passage, we find that Davidson takes first-person authority to follow from two premises:

Premise (I): Speakers generally know which sentences they hold true.

and

Premise (II): Speakers generally know the meanings of their own utterances.

Now one might have thought that self-knowledge would be a straightforward consequence of Premise (I). For if speakers, but not interpreters, can reliably determine which sentences they hold true, then their self ascriptions of belief would seem to enjoy a greater presumption of truth. But Davidson avoids this route, realizing that it only backs up the explanation of first-person authority. Any asymmetry between the knowledge of speakers and interpreters built into Premise (I) would in turn have to be given its own explanation. Instead, the asymmetry that Davidson wishes to exploit in order to explain first-person authority lies in Premise (II). Thus, even if we assume that interpreters and

speakers have equally reliable information about which sentences speakers hold true, Davidson believes that speakers would *still* have privileged access into the contents of their own minds, because they know the meanings of their own utterances in a way that they don't know the meanings of another's utterances.

## II

In the wake of Putnam and Burge, the near consensus in the philosophy of language has been that subjects do not have complete command over the meanings of their own terms. Faced with what appear to be challenges to linguistic self-knowledge, Davidson's primary concern has been to defend the truth of Premise (II). Here his thought is that subjects do not need to engage in radical interpretation in order to assign meanings to their own terms. The gap that exists between an interpreter's idiolect and that of the interpreted subject simply doesn't arise when one tries to make sense of one's own utterances. The homophonic (or disquotational) translation scheme is all that speakers need in order to make sense of themselves. Davidson elaborates this thought in "Knowing One's Own Mind." There he remarks that "unless there is a presumption that the speaker knows what she means, i.e., is getting her own language right, there would be nothing for an interpreter to interpret." (KOM, p. 456) The whole interpretive enterprise depends upon the assumption that speakers do not regularly misapply their own terms, and that sincere utterances consistently express a subject's mental states. Fortunately, externalist theories of meaning appear to underwrite this assumption. For if "what a person's words mean depend in the most basic cases on the kinds of objects and events that have caused the person to hold the words to be applicable," (KOM, p. 456) then the

disquotational translation scheme, albeit trivial and uninformative, cannot go wildly wrong. Moreover, externalist theories of meaning also help to explain the privileged position that Premise (II) assigns to speakers, but not interpreters.

“An interpreter of another’s words and thoughts must depend upon scattered information, fortunate training, and imaginative surmise in coming to understand the other. The agent herself, however, is not in a position to wonder whether she is generally using her own words to apply to the right objects and events, since whatever she regularly does apply them to gives her words the meanings they have and her thoughts the contents they have.” (KOM, p. 456)

So if externalist theories of meaning are correct, then it would seem that Davidson has indeed identified an asymmetry between speakers and their interpreters concerning the knowledge of a speaker’s terms. Davidson thus concludes that “first-person authority, the social character of meaning, and the external determinants of thought and meaning go naturally together.” (KOM, p. 456).

### III

But does this asymmetry really explain first-person authority? While one could certainly challenge Davidson’s contention that speakers are not in a position to worry whether they use their own words correctly, the concern I’d like to raise here with Davidson’s account lies not in the truth of the second premise, but rather in the validity of his inference from the premises to his intended conclusion. For even if we grant, at least for the sake of argument, that Davidson’s interpretive assumption is true, it is by no means obvious that first-person authority follows. Now it’s no mystery why Davidson

would see a connection between the truth of Premise (II) and first-person authority. After all, it is a central tenet of Davidson's account of radical interpretation that meaning determination and belief specification go hand-in-hand. Assigning meanings to a speaker's utterances requires interpreters to make concomitant determinations of their beliefs. But at the same time, one cannot fix the contents of a subject's beliefs without identifying the meanings of their utterances. Given this connection, since Premise (II) asserts that speakers are presumed to have knowledge of the meanings of their own utterances, it would stand to reason then that they would have knowledge of the states of mind that those utterances express.

However, this justification is too quick. Davidson derives the connection between meaning determination and belief specification in the context of radical interpretation, where interpreters cannot get independent fixes on either belief or meaning. Davidson's defense of Premise (II), however, suggests that he does not really regard the first-person case to be an instance of radical interpretation, for the disquotational translation scheme provides oneself with the privileged fix on one's own meanings that is so sorely lacking in the third-person case. Is there, then, a plausible route from premises (I) and (II) to first-person authority? I think there is, but I fear it will turn out to be much less than what we want.

Consider a speaker S who sincerely asserts, "I believe that p." By premise (I), it is presumed that S knows that it is the sentence "p" that S holds true. How so? Since Davidson is reluctant to trace any asymmetry of knowledge between speakers and interpreters to premise (I), presumably the speaker knows this in the same way as his interpreters would, namely by its being displayed within the 'that'-clause of the self-

ascription. At this point, premise (II) tells us that our speaker does not need to engage in any radical interpretation to know that “p” means that p for him, as he would, if this were a third-person ascription instead. So it would seem that S’s assertion that “I (S) believe that p” does indeed have a special security for him that wouldn’t apply to cases in which S makes third-person belief ascriptions.<sup>2</sup>

#### IV

It might appear that this argument demonstrates that we are in a privileged position to know our own minds. It shows that one has a special assurance for their own first-person ascriptions of belief that they don’t have for third-person ascriptions. My self ascriptions of belief are especially warranted *for me* since I don’t have to engage in radical interpretation to discover the meanings of my own terms. At this point, however, I’d ask you to step back and notice just how limited this account of first-person authority is. Davidson sets out to explain “the difference in the sort of assurance you have that I am right when I say ‘I believe Wagner died happy’ and the sort of assurance I have.” (FPA, p. 109) But that’s *not* what we want from an account of first-person authority! Rather than just being assured that *one’s own* first-person ascriptions of belief are most likely to be true, one should rather demand an account of why sincere first-person ascriptions of belief *in general* are especially equipped to withstand challenge. That is, when *interpreters* recognize that another has made a sincere first-person ascription of belief, why should *they* be prepared to regard the other’s first-person ascription as enjoying a special sort of credence?

Far from delivering us an account of this more extensive authority, the argument I just presented on Davidson's behalf actually undermines it. For once again, let us suppose that a speaker S sincerely asserts, "I believe that p." Moreover, let us further suppose (and let's not pause to wonder how) that S's assertion is generally recognized as a first-person belief ascription. Premise (I) tells us that any arbitrary interpreter, including the speaker, will be in position to identify "p" as the sentence that S holds true. But to get from that to the conclusion that "S believes *that p*," an interpreter would have to be in a position to know that "p" *in the speaker's tongue* means that p *in the interpreter's tongue*. And this is precisely what Davidson's remarks concerning premise (II) seem to block, for the knowledge that "p" disquotes is secure only when one is trying to make sense of one's own terms. Indeed, when speakers need to make their meanings clear to others, they are well-advised to do much more than simply stutter (or disquote).<sup>3</sup>

## V

So far, I've argued that Davidson's attempt to link first-person authority to interpretive constraints is unsuccessful. So what are the alternatives? In the space remaining, I'll briefly outline three. Rather than appealing to interpretive constraints, Sidney Shoemaker attempts to link first-person authority to broad constraints on rationality.<sup>4</sup> For a creature to be rational, he claims that it must on the whole know the methods by which to test the truth of its beliefs. But as he sees it, this in turn requires the beliefs of rational subjects to be "self-intimating;" in order to know how to check the truth of their beliefs, they must be able to determine which beliefs they have. So to rationally believe that p, it follows that rational subjects must have the belief that they

believe that p. While self-intimation amounts to the mind's being *transparent* to itself, Shoemaker deftly observes that if the absence of beliefs is also self-intimating, then subject's beliefs about their beliefs will turn out to be reliable. Thus first-person authority follows from the extension of self-intimation to the *absence* of belief.

Unfortunately I cannot accept Shoemaker's route through self-intimation either. His picture of rational belief revision requires creatures to have an awareness of their beliefs, which suggests that for creatures to rationally hold any beliefs at all, they must possess the concept of belief. Such a view implausibly threatens to turn the concept of belief into a ground-level concept, the possession of which would be required for a creature to possess any concepts whatsoever.<sup>5</sup>

Another strategy would be to argue that first-person ascriptions of belief are especially secure from challenge, because they make relatively little substantial claims about how the world is. Arthur Collins adopts this general strategy in *The Nature of Mental Things*, where he suggests that first-person ascriptions of the belief that p register a modest acknowledgement that one would be mistaken if p turned out not to be the case.<sup>6</sup> Since such statements serve more to *withdraw* one's full commitment to a claim than to take a stance on how things are, there is little positive content remaining in the self-ascription for the subject to be mistaken about in the event the claim is challenged. Self-ascriptions of belief thus function like statements about how things "look" or "appear" to subjects. Obviously, one might object that "belief"-talk should not be assimilated to "looks"-talk, because subjects can believe that things are other than they appear to be. But on top of that, this proposed analysis suffers from the decisive disadvantage that unless we regard belief attributions as generally non-factual discourse,

then self-ascriptions of belief have surprisingly different contents from third-person ascriptions, making it hard to see how subjects could ever agree or disagree with the third-person belief ascriptions made of them. The proposal also looks unappealing when we consider *embedded* self-ascriptions of belief. For on the proposed analysis, one's attributing to oneself the belief that p would seemingly emphasize the possibility that one's belief that they believe that p could be mistaken. Far from explaining first-person authority, such an admission appears to wind up denying any presumption of first-person authority altogether.

## VI

My preferred strategy would be to tie first-person authority directly to logical and inferential peculiarities surrounding the concept of belief. If we subscribe to the broadly Sellarsian idea that the concept of belief has been introduced to refer to those functional states of subjects responsible for the production of intelligent behavior in general, and linguistic behavior in particular, then as long as creatures can report when they are disposed to assert or deny certain propositions, then they can easily be trained to report the presence of their beliefs. For the evidence that one has for one's believing that p is just the same evidence that one has for asserting "p."

So let's suppose one last time that S sincerely asserts, "I believe that p." And let's further assume that this assertion is generally regarded as a self-ascription of belief. This latter assumption is necessary in order to rule out certain conditions that would undermine any alleged first-person authority. Specifically, it allows us to presume that S understands that this ascription refers to himself, that he understands that "p" is the

content of the ascribed belief, and most importantly, that S grasps the concept of belief. But now observe that creatures with an appropriate grasp of the concept of belief would understand that sincerely ascribing to oneself the belief that p *practically commits* them to defending the truth of “p.” For speakers are obliged to defend the truth of their claims, and the best evidence for defending the claim that one believes that p is simply to assent to the sentence “p,” as well as to challenge the truth of any claim obviously incompatible to it. Clearly, third-person ascriptions of belief do not afford speakers this additional avenue of justification for one’s belief ascriptions. Herein lies the asymmetry between first- and third-person ascriptions of belief, which manifests itself in the appearance of first-person authority.<sup>7</sup> Given the assumption that S recognizes that his claim is a self-ascription of belief, interpreters may conclude that S presumably understands that his self-ascription practically commits himself to the truth of “p.” But on the conception of belief outlined above, one’s being committed to defending the truth of a claim is *prima facie* evidence that one does indeed believe that statement. So S’s sincere self-ascription provides evidence that S is disposed to assent to “p,” and so turns out to be evidence of its own truth. First-person “authority” is therefore secured.

Now one could well point out that this *prima facie* evidence of a self-ascription’s truth would be undermined if the speaker lacks sufficient understanding of the ascription’s content clause-- that is, if the speaker fails to recognize what further statements that the content clause entails as well as those it is entailed by. So it would seem that Davidson is correct to suspect that first-person authority relies upon a presumption that speakers possess sufficient grasp of the meanings of the terms in their language to accurately express the contents of their beliefs. Still, it would be wrong to

conclude that this assumption of linguistic self-knowledge ultimately *explains* first-person authority, since it doesn't place the knowledge speakers have of their language on any higher plane than the knowledge that their interpreters might have.<sup>8</sup> Interpreters might have a much greater command of a speaker's language, and yet we'd still regard speakers as having special authority over their first-person ascriptions of belief.

While I've tried to close on a somewhat positive note, I realize that this sketch is every bit as compressed as Davidson's, and woefully incomplete. In particular, I haven't shown how it could be extended to self ascriptions of non-doxastic attitudes such as intentions or desires. Fortunately, intuitions that speakers have authority over reports of those states are much weaker than in the case of belief. In any event, I hope I've shown that there are alternative explanations of first-person authority available, some of which might prove more satisfactory than Davidson's proposal.<sup>9</sup>

---

<sup>1</sup> Davidson's most extensive treatment of this topic can be found in "First Person Authority" *Dialectica* 38 (1984): 101-11 (hereafter FPA). He reaffirms his conclusions there in "Knowing one's own Mind" *Proceedings and Addresses of the American Philosophical Association* 60 (1987): 441-458 (hereafter KOM).

<sup>2</sup> Diagrammed, the argument looks as follows:

1. S sincerely asserts "I believe that p".
2. S presumably knows that S holds-true "p" [By premise (I)]
3. S presumably knows that "p" means that p. [By premise (II)]
4. S's first-person ascription is presumably true. [From 1., 2., and 3.]

<sup>3</sup> For comparison's sake, the parallel argument looks like this:

1. S sincerely asserts "I believe that p".
2. I (an arbitrary interpreter) presumably knows that S holds-true "p" [By premise (I)]
3. I presumably knows that "p" means that p. [**Not by premise (II)!!!**]
4. I will agree that S's first-person ascription is presumably true. [From 1., 2., and 3.]

<sup>4</sup> See in particular the first four essays of *The First-Person Perspective and Other Essays*, Cambridge University Press (1996).

<sup>5</sup> Ironically, Davidson would find this suggestion quite congenial. For more discussion, see the author's "Some More Thoughts About Thought and Talk," *Philosophy* 77 (2002): 115-124.

<sup>6</sup> See Collins, *The Nature of Mental Things*, Notre Dame, IN: University of Notre Dame Press (1987).

---

<sup>7</sup> This asymmetry also accounts for the pragmatically self-defeating, yet not logically inconsistent, character of Moore-paradoxical statements. Shoemaker (1996) also draws a connection between Moore-paradoxical statements and self-knowledge. Subjects who doubt their own first-person authority are liable to find themselves committed to pragmatically self-defeating “Moore-paradoxical” statements. While expressing a belief that they believe that p (for instance, by making a first person ascription of belief), they might accept that they do not believe that p after all. And so they put themselves in position to be committed to pragmatically self-defeating statements like “I believe that I believe that p, but I do not believe that p.” Now I submit that we’re apt to impugn the grasp of the concept of belief of anybody who makes apparently sincere Moore-paradoxical utterances such as this, rather than impugning (as Shoemaker does) their overall rationality.

<sup>8</sup> Mark McCullagh also suggests that this is the presumption from which first-person authority “derives.” See McCullagh, “Self-Knowledge Failures and First Person Authority,” *Philosophy and Phenomenological Research*, LXIV (2002): 365-380.

<sup>9</sup> This paper originated as a commentary for A. Minh Nguyen’s “Davidson on First-person Authority” at the 2001 Mountain-Plains Philosophy Conference. I’m much obliged to Nguyen for helping me distill Davidson’s account.