Philosophy and Phenomenological Research
Vol. LXIV, No. 2, March 2002

Self-Knowledge Failures and First Person Authority[1]

MARK MCCULLAGH
University of Guelph

Davidson and Burge have claimed that the conditions under which self-knowledge is possessed are such that externalism poses no obstacle to their being met by ordinary speakers and thinkers. On their accounts, no such person could *fail* to possess self-knowledge. But we do from time to time attribute to each other such failures; so we should prefer to their accounts an account that preserves first person authority while allowing us to make sense of what appear to be true attributions of such failures.

While the core idea behind Davidson's and Burge's accounts appears inadequate to this task, I argue that it can be deployed in such a way as to deliver the desired result. What makes this possible is that two attitude-types can differ as follows: the self-knowledge required for an utterance to be a Φing that *p* is different from the self-knowledge required for it to be a Ψing that *p*.

Semantic externalism has been charged with undercutting "first person authority": the *prima facie* authority of self-knowledge attributions. Davidson and Burge, aiming to rebut that charge, have claimed that the conditions under which self-knowledge is possessed are such that externalism poses no obstacle to their being met by ordinary speakers and thinkers. On their accounts, no such person could *fail* to possess self-knowledge of any thought she is having, or statement she is making. But we do from time to time attribute to each other failures of self-knowledge; so we should prefer to these accounts an account that preserves first person authority while allowing us to make sense of what appear to be true attributions of such failures.

It may be thought that the core idea behind Davidson's and Burge's accounts is inadequate to this task. Against this, I shall argue that that idea can be deployed in such a way as to deliver the desired results. What makes

[1] For financial support of the work that led to this paper I am grateful to the Social Sciences and Humanities Research Council of Canada for a Doctoral Fellowship. For helpful discussions I am grateful to Robert Brandom, Henry Jackman, John McDowell, Ram Neta and Jonathan Sutton; and to the audience and commentator (Susanna Nuccetelli) when I read a version of this paper at the 1999 meeting of the Central States Philosophical Association.

this possible is that two attitude-types Φing and Ψing can differ as follows: the self-knowledge required to Φ that *p* is different from the self-knowledge required to Ψ that *p*.

1.

On Davidson's (1984, 1987) account, one has self-knowledge concerning a statement one is making because one is able to specify its truth condition. For example, I have self-knowledge concerning my statement of the sentence "Wagner died happy" because I can specify its truth condition by means of the sentence

> My utterance of "Wagner died happy" is true if and only if Wagner died happy.[2]

Burge (1988), on the other hand, holds that one has self-knowledge concerning a thought one is having because one can think it "self-ascriptively."[3] To think one's thought that *p* "self-ascriptively" is to think that one is thinking that *p*.

Common to these accounts is the thought that one has self-knowledge concerning one's propositional attitudes *almost* entirely because one is able to have them. What is being exploited here is the truism that it is a necessary condition on having a propositional attitude, that one have something worth calling an understanding of its content.[4] That is why these accounts of the conditions under which self-knowledge is possessed do deserve to be called accounts of self-*knowledge*.

Moreover, these accounts support neat explanations of the *prima facie* authority of attributions of self-knowledge. For the only way in which one could lack self-knowledge concerning a statement one is making, on Davidson's account, is to lack either the quoting ability or the ability to frame biconditionals; and the only way in which one could lack self-knowledge concerning a thought one is thinking, on Burge's account, is to lack the ability to think it self-ascriptively. But no thinker or speaker, other than a borderline case, lacks any of these abilities.

---

2 In the context of Davidson's (1967) claim that a theory of truth can serve as a theory of meaning, such a specification is not one that we can object to as failing to get at the content of the sentence quoted on the left-hand side; for that claim is precisely that biconditionals can, in the right setting, specify contents.

3 The present progressive tense is important here. Burge restricts his account to cover only what he calls "basic self-knowledge," i.e. self-knowledge concerning thoughts one is "currently and consciously" thinking (1988, 653).

4 We might want to restrict the domain of this truism's application to *conscious* states; but that is a harmless restriction in the present context since the domain of first person authority is similarly restricted.

Moreover, on these accounts there is no reason to think that semantic externalism undercuts that authority. To do so, it would have to undercut our entitlement to presume that agents are able to quote their own words, frame biconditionals, or think their thoughts self-ascriptively. But it is very hard to see how any semantic doctrine, let alone externalism, could do that (Falvey and Owens 1994, 110-23; Boghossian 1994, 35). So Davidson's and Burge's accounts do appear to meet the requirement that they set themselves, namely, to explain the *prima facie* authority of self-knowledge attributions in a way that is consistent with semantic externalism.

2.

One objection to these accounts is that they don't completely declaw externalism because there is a different conception of self-knowledge, such that externalism *does* undercut the *prima facie* authority of attributions of *it*.[5] I shall have nothing to say about that sort of objection here. The objection I want to explore is that Davidson's and Burge's accounts fail to meet an important intuitive requirement: they don't allow for the possibility of true attributions of *failures* of self-knowledge to ordinary speakers and thinkers.

I shall not be discussing cases of the sort that interested Freud. An example of what I do have in mind was reported upon by Truman Capote in his nonfiction novel *In Cold Blood*. Two small-time criminals, Dick Hickock and Perry Smith, were discussing the possibility of their being seen while carrying out a certain robbery:

> "*No* witnesses," [Dick] reminded Perry, for what seemed to Perry the millionth time. It rankled in him, the way Dick mouthed those two words, as though they solved every problem; it was stupid not to admit that there might be a witness they hadn't seen. "The ineffable happens, things *do* take a turn," [Perry] said. (Capote 1962, 37)

Asked whether Perry knew what he was saying, I think we would ordinarily say that he didn't know it as well as he should. He seems to have thought that "ineffable" means something like "unexpected" or "unpredictable," and intuitively that seems like a shortcoming in his knowledge of his statement. Dick certainly seems to have seen things in some such way, for he described Perry as "always using hundred-dollar words he doesn't half know the meaning of" (334).

---

[5] Some philosophers have pointed out that Davidson's and Burge's accounts don't give us any reason to think that attributions of what they call "comparative" knowledge of content are *prima facie* authoritative (Falvey and Owens 1994, 110; Boghossian 1994, 35-36; Butler 1997, 782). They have differed on what to say about this. Falvey and Owens say that it is no problem (112), Boghossian argues that it challenges our conception of psychological explanation (39-40), and Butler says that it is a serious problem for externalism (795, 799).

Now there is an impulse many philosophers have, upon hearing a case such as Perry's described, to try to assimilate it to cases in one of the following three categories:

- those who utter words but fail thereby to make a statement;

- those who utter words knowing only that they are words in the language of their audience; or,

- those who speak their own, idiolectic, language.

Such assimilations carry little intuitive conviction in this case. Dick's description of Perry, for example, places him in none of these categories.[6] Taken at face value, Dick's description characterizes Perry as lacking knowledge of the meanings of his statements. Dick doesn't say that they're not even really statements; nor does he say that Perry knows nothing of his words other than the language they're in; nor does he say that Perry speaks his own special language. We can imagine ourselves in Dick's place and saying pretty much the same thing about Perry that Dick says. In our ordinary dealings with each other we can very easily (although not very *often*, for reasons to be explained below, §9) find ourselves wanting to describe someone as not knowing what he's saying—in a way that isn't captured by any of the assimilations just considered.

Those assimilations don't get at what Dick says about Perry; but neither, however, does Davidson's or Burge's conception of self-knowledge. Dick evinces no reluctance to characterize Perry's uses of hundred-dollar words as *statements*—he does describe him as "using" those words—and he doesn't have any reason not to credit Perry with the abilities to quote his own words or to frame biconditionals. So Dick does not attribute to Perry a lack of Davidsonian self-knowledge. Nor does he attribute to Perry a lack of Burgean self-knowledge (say, of the conscious occurrent thought he has while making the statement). For again, thinking thoughts self-ascriptively is something he has no reason to think Perry cannot do. Perry has his shortcomings, but Dick does not believe any of these to be among them. So whatever kind of self-knowledge he has in mind when apparently describing Perry as failing to possess it, it isn't Davidsonian or Burgean self-knowledge.

---

[6] Of course, there are many cases that do clearly fit into such categories. The imitative babbling of a child fits into the first. The user of a mostly-incorrect foreign-language phrasebook (as in one Monty Python sketch) fits into the second. (Were the phrasebook reliable, the user of it might know which foreign sentences translate which sentences in his language; and to know this is to know more about the foreign-language sentences than merely that they are foreign-language sentences.) The musician Lester Young, who had a highly idiosyncratic vocabulary that included words such as "oodastaddis" and "vout," would fit into the third, at least partly.

3.

The example illustrates something we should have thought anyway: that there's something entirely *too* easy about self-knowledge as Davidson and Burge conceive of it. Philosophers often claim—and non-philosophers can very easily be made to claim—that we know the contents of our own thoughts and speech acts, that we know what we *mean* when we have a thought, make an assertion, and so on. Behind these sincere assertions must be some intuitive conception of what self-knowledge is, what it consists in. But if self-knowledge is as Davidson and Burge conceive of it, it's difficult to see where we would have got such a conception. That is because it's difficult to see why we'd *bother* attributing Davidsonian or Burgean self-knowledge to each other. Such attributions would register a useful distinction only if we actually do come across people who lack the ability to think their thoughts self-ascriptively, quote their own words or frame biconditionals. But we don't. If we don't have occasion, then, to register any distinction by attributing self-knowledge to each other, then it's hard to see how we could nonetheless have acquired the intuitive notion of self-knowledge that figures in our sincere assertions that we almost always have self-knowledge. Self-knowledge as Davidson and Burge conceive of it, then, is unlikely to be the self-knowledge we're talking about when we make those assertions.

The general moral is that we can account for our intuitive conviction that we almost always have self-knowledge only if we explain what useful distinction we register in terms of the notion of self-knowledge we employ. (More on this point below, §9.)

4.

Back to our example. Our question is, What kind of self-knowledge might Dick have in mind, if not Davidsonian or Burgean self-knowledge, when he attributes to Perry a failure to possess it concerning his uses of the hundred-dollar words he doesn't half know the meaning of? Since Dick credits Perry with the ability to make *statements* with those words, it would have to be self-knowledge of a sort the possession of which is *not* ensured by possession of the ability just to make statements. We want to know what that sort of self-knowledge might be.

Toward this end, we will find it useful to have a term for the self-knowledge possession of which *is* ensured by the ability to make statements. One notion that answers very bluntly to that description may be defined as follows:

> An agent *S* has **statement-knowledge** of an event *e* iff *e* is a statement *S* is making.

The idea being very crudely captured here is that statement-knowledge is the sort of knowledge of an event that one is guaranteed to have in virtue of that event's being a statement. Its possession conditions seem even *more* trivially satisfied than the possession conditions of Davidsonian self-knowledge, since a speaker will have statement-knowledge of his statement even if he lacks the ability to quote words or the ability to frame biconditionals. But not much more: as I noted above (§1), such possibilities are *recherché*; so statement-knowledge is very near to the sort of self-knowledge whose possession conditions Davidson describes. Similarly we may define a notion that stands to Burgean self-knowledge as statement-knowledge stands to Davidsonian:

> An agent *S* has **thought-knowledge** of an event *e* iff *e* is a thought *S* is having.

At this point one might object that having defined the possession conditions of two sorts of self-knowledge that are even more easily attained than the corresponding sorts whose possession conditions Davidson and Burge describe, we are more vulnerable to whatever problem is raised for those accounts by the possibility of failures to possess self-knowledge. For we've eliminated even the *recherché* possibilities as ones in which a speaker or thinker might fail to possess self-knowledge. And the objection I made out in §3 is that our intuitive conviction that we have first person authority can't involve a notion of self-knowledge with which we can register no useful distinction.

I shall argue, however, that when we take account of a certain dimension along which attitude-types can differ, we see that this sort of characterization of self-knowledge is *less* vulnerable to this problem than it seems. In the next two sections I'll explain the sort of difference among attitude-types that I have in mind.

5.

Consider this schema, of which our characterization of statement-knowledge is an instance:

> An agent *S* has **Φing-knowledge** of an event *e* iff *e* is a current Φing by *S*.

Perhaps for different attitude-types, instances of this schema state the possession conditions of different *kinds* of self-knowledge. If so, then perhaps an agent can possess self-knowledge of one of these types while lacking self-knowledge of another of them.

The possibility is not an interesting one when the two attitude-types are disjoint, as asserting and believing are: nothing that is an assertion is also a belief. For we may adequately explain why someone lacks belief-knowledge of one of his assertions simply by pointing out that *no* assertion is a belief.

But the possibility is an interesting one when the two attitude-types are not disjoint. Recall our example. Dick seems to take it that Perry's utterance of "The ineffable sometimes happens" is a statement. As we noted, he doesn't take it that it's a statement in Perry's idiolect; he takes it that it's a statement in a public language. So he takes it that it's not only a statement but that it's a statement *that the ineffable sometimes happens*. But he thinks that Perry doesn't have much of a grip on that content: he "doesn't half know the meaning" of the words he uses. Now in this case, it seems, Dick would not attribute to Perry a *belief* that the ineffable sometimes happens. For the belief that Dick—and any reader of *In Cold Blood*, for that matter—takes Perry to be attempting to express with his statement is a belief that the *unexpected* sometimes happens (or something close to that). In short, then, Dick takes it that the content of Perry's statement doesn't match that of the belief he intends it to express.

That gives us all we need to describe one attitude-type, Φing, such that Dick would deny that Perry has Φing-knowledge of his statement of "The ineffable sometimes happens," even though there are other statements, Φing knowledge of which he *would* attribute to Perry. That attitude-type is *assertive belief-expression*, stipulatively defined as follows.

> An event *e* is an **assertive expression by *S* of a belief that *p*** iff (i) *e* is a statement by *S* that *p*; (ii) there is a belief of *S*'s which *S* intends that assertion to express; (iii) that belief is a belief that *p*.

It follows from clause (iii) of the definition that attributions of assertive belief-expressions license inferences to attributions of *beliefs* with the same content. Since in our example Dick would *not* infer from Perry's statement, that Perry *believes* that the ineffable sometimes happens, he would *not* characterize Perry's statement as also being an assertive belief-expression.[7] And this means, in turn, that Dick should say that Perry *lacks* assertive-belief-

[7] This might mean that Perry, as Dick conceives of him, is a counterexample to Kripke's (1979) "disquotational principle" licensing inferences from ascriptions of sincere, competent assertions to ascriptions of beliefs with the contents of those assertions. It depends on what "competent" means: for a speaker can be competent to perform speech acts of some types but not others. Ambiguous as it is, Kripke's principle is controversial anyway. Some think that it is a basic truth about our attitude-ascriptive practice (Bilgrami 1992, 1; David Sosa 1996, 383) while others think that it is falsified by those cases to which Kripke applied it to generate his "puzzle about belief" (Owens 1986, 380-881n.22, Loar 1987, 172-73; Brandom 1994, 576-79).

expression-knowledge of his statement.[8] (Of course there are many others of Perry's statements that Dick would agree *do* meet the conditions on being assertive belief-expressions.)

Were Dick called upon to explain why Perry lacks assertive-belief-expression-knowledge of his statement, he wouldn't explain it by claiming that no statement is also an assertive belief-expression—that these are disjoint attitude-types. Statements are ontologically suited to be assertive belief-expressions, since many of them are such. Rather, Dick would advert to Perry's lack of *understanding* of his statement: to the fact that "ineffable" is one of the "hundred-dollar words he doesn't half know the meaning of." Intuitively, that seems the right way to explain what disqualifies Perry's statement from being also an assertive belief-expression. It follows that *stating* and *assertively belief-expressing* are attitude-types by reference to which, in our schematic manner, we can characterize the possession conditions of different *kinds* of self-knowledge.

6.

The point may be generalized, for there are other attitude-types that also seem to differ in this way. Consider some claims that Gareth Evans makes about the following example.

> A group of people are having a conversation in a pub, about a certain Louis of whom *S* has never heard before. *S* becomes interested and asks: 'What did Louis do then?' (Evans 1973, 6)

Evans maintains that when *S* uses the name "Louis" in the pub, as a party to the conversation there, he speaks of the same person that the other participants in that conversation speak of—that in that sense, *S*'s in-pub utterances of "Louis" *denote* Louis; they are not mere noises. But he takes a different attitude towards *S*'s later uses of the name. We can imagine *S*, one week later, coming across someone *T* who was not a participant in the conversation in the pub, and recounting to him the things about Louis that he was told of in the pub. Evans maintains that in a case like this, since *S* is detached from the conversational context of the pub, in these uses of the name "Louis" he "is simply not sensitive [any longer] to the outcome of any investigations regarding the truth of what he is said to have said [to *T*]" (1973, 7), and therefore, it is "detached, useless" to maintain that those uses are *sayings*.

---

[8] The possession conditions of this sort of self-knowledge are given according to the schema. An agent *S* has **assertive-belief-expression-knowledge** of an event *e* iff *e* is an assertive belief-expression *S* is performing.

We may find it plausible to follow Evans in holding that *S* has less of that "sensitivity" when detached from the context of the pub than he has in that context. Now, what follows from that? Evans seems to think that it follows that we should not assign semantic properties to *S*'s utterances of "Louis" outside the original context of the pub conversation. He warns us against what he calls the "mouthpiece syndrome," an ailment "by which we assign sense and reference to a man's remarks only because we hear someone else speaking through him; as we might with a messenger, carrying a message about matters of which he was entirely ignorant" (7).

But we should be equally wary, I think, of another danger. For to deny sense and reference to *S*'s non-pub utterances of "Louis" is to leave it an open question whether there is anything that distinguishes *S*'s understanding of those utterances from the understanding that a parrot, say, might have of noises that it made in imitation of the same conversation. But that question is *not* open, because *S* does have *some* sensitivity to the outcomes of *some* investigations concerning Louis—even though as Evans points out, he does have *less* of that sensitivity than he did in the context of the original conversation. (For instance, he is still capable of recognizing the people he was talking with in the pub, and receiving updates on Louis from them.) Surely, there are *other* characterizations of *S*'s non-pub utterances of "Louis" that *don't* leave it an open question whether he has any such sensitivity, yet which don't entail the claim Evans rejects (that *S*'s non-pub utterances of "Louis" are *sayings*). Such descriptions would characterize those utterances as tokens of some speech-act-type that isn't such as to require exactly the same sort of sensitivity that Evans insists *saying* requires, yet which is such as to require *some* sensitivity of the same general sort. And there are, of course, speech-act-types of that sort: speech-act-types that require less, in this regard, than saying does. My intuitions have it that *passing along testimony* is such a speech-act-type. But we needn't argue over intuitions: what is important is the *possibility* of characterizing *S*'s non-pub uses of "Louis" in a way that simultaneously respects the requirement on *saying* that Evans insists upon, while not leaving it an open question whether *S* has any more understanding of those utterances than a parrot might have. If necessary, we can simply define new speech-act-types in order to do that discriminatory work.

I shall discuss one more example. In §5 I argued that *assertive belief-expression* is a speech-act-type that requires more, in the way of understanding, than *saying* does. There are other speech-act-types that require more still. Consider the charge that Frege made against the mathematician Weierstrass in a late essay (1914). Frege thought that because Weierstrass's *definitions* of the concept of number were so egregiously faulty, he didn't have a "clear grasp" of it—he "had a notion of what number is, but a very

hazy one" (221). Now we may define a sort of speech act that does require such a grasp of the notions involved: call it "*clear-asserting* that *p*." (I omit the obvious details.) Since that is possible, there are attitude-types Φing such that having assertive-belief-expression-knowledge of the utterance one is making (which Weierstrass obviously did possess concerning many of his utterances involving his word for number) *doesn't* entail having Φing-knowledge of it.

The general point is simply this: the sort of self-knowledge—the sort of *understanding* of what one is doing—that is required for one's utterance to be a token of a speech-act-type can differ from one type to another.[9] I consider that point almost a truism, once it is emphasized how great the range of speech-act-types is.

I noted above (§1) that the core thought behind Davidson's and Burge's accounts of self-knowledge is that one has self-knowledge of one's propositional attitudes *almost* entirely because one is able to have them. In the remainder of this paper I shall explain how that core thought—actually, the unqualified version of it—can be deployed in such a way as to reflect the point I've just made about differences among attitude-types. The result is an account of self-knowledge that has all the virtues of Davidson's and Burge's yet which allows us to make sense of the possibility of true attributions of *failures* to possess self-knowledge.

7.

My proposal concerning the contents of self-knowledge attributions is that each of them has the same truth condition as some instance of the schema:

S has Φ-knowledge of *S*'s Ψing that *p*,

where *S* is an agent, the schematic letters "Φ" and "Ψ" are replaced by expressions for propositional attitude-types, and the schematic letter "*p*" is replaced by a declarative sentence.

This is a generalization of the core idea behind Davidson's and Burge's accounts. For self-knowledge as they account for it is characterized almost entirely in terms of the subject's tokening some attitude-type. The important difference is that they *specified* particular attitude-types—statements and conscious thoughts—while on the proposal I'm making, self-knowledge attributions of the form "*S* has self-knowledge concerning *x*" are systematically ambiguous, depending on which attitude-type is specified in the truth-condition-preserving paraphrase.

---

[9] Similarly for mental states, although theorists will differ concerning what the entities are that stand to mental states as utterances of linguistic expressions stand to speech acts: brain-states, some will say; others, tokenings of sentences in the language of thought.

8.

One might object that this claim of systematic ambiguity makes a hash of our intuitions about self-knowledge, for our intuitions have it that there is some *one* kind of self-knowledge, attributions of which are *prima facie* authoritative. I have two replies to that.

First, the claim that self-knowledge attributions are systematically ambiguous entails only that there's a range of meanings that self-knowledge attributions *could* have; it does not entail that there is as great a range among the meanings that self-knowledge attributions have *actually* had. So the proposal is compatible with the possibility that there is only a very small number of things we've meant to attribute, when we've attributed self-knowledge to each other.

Secondly, and more to the point, I think that we should take seriously the possibility that this number is greater than one. For our self-knowledge intuitions can't plausibly be claimed to be so coherent and fine-grained as to determinately pick out only one conception of it. We don't *often* make such attributions; they don't play much of a role in our lives. Moreover, there is nothing resembling a canonical way of assessing such attributions, of giving and asking for reasons concerning them. There simply isn't a great mass of ordinary-usage data to constrain accounts of the truth conditions of self-knowledge attributions.

Indeed, by holding that there are several things we've meant by those occasional self-knowledge attributions we do make, we allow for a neat disjunctive explanation of both the *prima facie* authority of self-knowledge attributions and their possible falsity. The claim would be that the authority of self-knowledge attributions derives from the fact that the great majority of them are to be read in such a way that they cannot be false due to a failing in self-knowledge: namely, as having the same truth condition as some instance of the schema

S has Φ-knowledge of *S*'s Φing that *p*

where the schematic letter "Φ" is replaced in both occurrences by the same expression (for a propositional attitude-type). The remainder, on this line, are to be read in such a way that they *can* be false due to a failing in self-knowledge: namely, as having the same truth condition as some instance of the schema

S has Φ-knowledge of *S*'s Ψing that *p*,

where the schematic letters are replaced by expressions for *different* propositional attitude-types. (E.g. "Perry has assertive-belief-expression-knowledge of his statement that the ineffable sometimes happens" is an instance of this

latter schema that I've suggested is false.) Call the former, homogeneous attributions and the latter, heterogeneous attributions. Obviously, heterogeneous attributions can, and homogeneous attributions cannot, be used to attribute self-knowledge failures. Then the claim is that what explains our intuitions about first person authority is that those intuitions are about what is stated by homogeneous attributions; and that what explains the possible falsity of self-knowledge attributions is each self-knowledge attribution that is possibly false states what some heterogeneous attribution states.

9.

While I do not endorse that disjunctive explanation of the *prima facie* authority and fallibility of self-knowledge attributions, it's not because I see any compelling reason to reject it. It's because I think that an account can be given that is even more plausible as an explanation of our intuitions, such as they are, about self-knowledge, because it involves commitment only to there being *one* sort of self-knowledge at issue in our actual, ordinary self-knowledge attributions. (Again, this claim about the contents of our *actual* self-knowledge attributions is not at all inconsistent with the claim of §7, that self-knowledge attributions are systematically ambiguous.)

Here is the explanation that I find more plausible. The *prima facie* authority of self-knowledge attributions is derivative. It derives from the *prima facie* authority of certain presuppositions that we make about each other. Those presuppositions have that authority because by and large they are true, although some of them are false. When they're true, so are self-knowledge attributions of certain sorts; when they're false, the corresponding self-knowledge attributions are as well.

What are those presuppositions? I shall focus on statements, because I think that it is primarily in such cases that we are inclined to attribute failings of self-knowledge. These cases are, therefore, the best ones for illustrating how self-knowledge attributions can be *prima facie* authoritative *and* possibly false.

Regarding sincere statements, the presuppositions are that speakers possess assertive-belief-expression knowledge of those statements. The claim is that the default position, as it were, that we take towards each other when confronting each other as sincere statement-makers is that people understand their words well enough for their sincere statements to qualify also as assertive belief-expressions.

Why are these presuppositions true by and large, though not always? And why do we make them?

Assertive-belief-expression-presumptions are by and large true, because most speakers' sincere statements are made in expression of beliefs whose contents are the same as those of the statements. That fact about our sincere

statements and our beliefs may, in turn, be explained by reference to facts about how speakers learn their words. *Most* speakers, *most* of the time, learn their words well enough that the sincere statements they make have the same contents as the beliefs they intend those statements to express. Most people do not have Perry's half-baked notions of words they've come across in books they've half-comprehended; nor do they use proper names while having only the meager sensitivity to facts about their bearers that *S* has after leaving the pub. Most of us learn most of our words from explanations given by competent users; or from observation of the uses such speakers make of their words; or from dictionaries; and so on. Learning a word in one of those ways endows one with enough understanding of it to use it in assertive-belief-expressings. Perhaps if our speech community were one in which *everyone* learned words as Perry did, then assertive-belief-expression-presumptions would no longer be generally true of us. But that is not, nor has it ever been, our situation.[10] Those presumptions are by and large true. They are sometimes false. Some of us do have shortcomings like Perry's with "ineffable" or *S*'s with "Louis." But these cases are not common.

It follows from the claims just made, that self-knowledge attributions, understood as I've suggested, mark a useful distinction. They mark a distinction between statements involving words the speaker acquired in one of the ways most of us acquire most of our words, and statements involving words acquired otherwise—and therefore *not* acquired in such a way as to ensure that the speaker possesses assertive-belief-expression-knowledge of his uses of it in sincere statements. The suggestion being made thus satisfies the criterion of adequacy I discussed above (§3).

Why do we make assertive-belief-expression presuppositions? The answer is that *all of us believe* the claims just made to explain the general truth of these presuppositions. We all believe that most speakers learn most of their words in such ways that their sincere statements are made in expression of beliefs with the same contents as the statements.

As concerns our assessments of each others' sincere statements, then, the proposal is this. We presume each other to have assertive-belief-expression-knowledge of our statements. Those presumptions have a *prima facie* authority, because they are by and large true, but sometimes they are false. These presumptions support, in turn, the *prima facie* authority and possible falsity of the self-knowledge attributions we ordinarily make about each other's statements, because the kind of self-knowledge attributed in those attributions is assertive-belief-expression-knowledge. Analogous proposals can be formulated for the self-knowledge we presume each other to have when we issue commands, ask questions, and so on for other speech-act-

[10] One might mount an interpretation-based argument for the claim that in no possible speech community could that be the situation; but I shall not explore this possibility.

types; and for the self-knowledge we presume each other to have about our thoughts, desires, hopes, beliefs, and so on for other mental-state types. In this way, we have an explanation of the *prima facie* authority and the possible falsity of self-knowledge attributions that (unlike the proposal considered in §8) does not presume that there is more than one kind of self-knowledge we attribute when we speak of self-knowledge concerning statements.

10.

Davidson's and Burge's proposals about self-knowledge were meant to show that externalism does not threaten the *prima facie* authority of claims to possess it. It should be no surprise that my proposal, which rests on the core idea behind those accounts, makes self-knowledge no more vulnerable to externalism than theirs do.

On the explanation I've just offered of the *prima facie* authority of self-knowledge attributions, the only way in which externalism could be a threat to it is if it were a threat to the *prima facie* authority of our presumptions that the speakers we deal with have assertive-belief-expression-knowledge of their sincere statements. For it to do that, it would have to entail that fewer sincere statements are assertive-belief-expressings than we ordinarily think. The question we should ask about externalism and self-knowledge, then, is this: Does externalism threaten speakers' possession of assertive-belief-expression-knowledge of their statements?

No. Whether someone satisfies the conditions for assertively expressing a belief that *p* depends on whether they satisfy the conditions for (i) asserting that *p*, (ii) believing that *p*, and (iii) intending, of an assertion with the same content as some belief, that the former be an expression of the latter. These conditions are no doubt very difficult to spell out; there is as yet no consensus in the philosophy of language as to how they should be spelled out. But it is plausible to maintain that they are implicitly defined by our practices of attributing beliefs, assertions, and intentions to express beliefs with assertions.

But if the standard arguments for externalism—those of Kripke (1972, 1980), Putnam (1975) and Burge (1979)—are right, then we're *already* externalist in our pretheoretical commitments concerning the ascription of speech acts and mental states. For those arguments rely on thought experiments in which we imagine which content we'd *ordinarily* assign to a propositional attitude of some given type in some envisioned situation. That means that externalism supplies no reason to think that we are any *less* able to make assertions, have beliefs, or to form intentions, concerning assertion-belief pairs identical in content, that the assertion express the belief. Since our meeting those conditions is what's involved in our meeting the

conditions for having assertive-belief-expression-knowledge, externalism supplies no reason to think that we have any *less* assertive-belief-expression-knowledge of our sincere assertions than we ordinarily think we have.

In short, externalist commitments are part of the background of commitments against which we've *already* assigned a *prima facie* authority to presumptions that speakers possess assertive-belief-expression-knowledge of their sincere statements. All that the externalist arguments do with respect to the question of the *prima facie* authority of self-knowledge attributions, when self-knowledge is conceived of as I've suggested, is to make explicit the fact that our ordinary commitments to externalism extend not only to the assignment of truth conditions to each other's mental states and speech acts, but also to the attribution of the particular sort of self-knowledge that we presume each other to have concerning our sincere statements.

At that level of generality, of course, the same may be said of Davidson's and Burge's accounts. For those philosophers, we're *already* externalist in our implicit conception of the conditions that must be met in order to make a statement, or have a thought. With uninteresting qualifications (see §1), Davidson's and Burge's claim is that those conditions *just are* the conditions under which we possess self-knowledge. It follows that we're already externalist in our implicit conception of the conditions that must be met in order to have self-knowledge concerning a statement one is making, or a thought one is having. Where I've departed from their accounts is in using their *method* of characterizing the possession conditions of self-knowledge in such a way as to open a possible gap between being able to make a sincere statement (for example) and having the self-knowledge of that statement that we ordinarily presuppose each other to have (namely, assertive-belief-expression knowledge). Arriving at an account that allows self-knowledge attributions to be possibly false while being *prima facie* authoritative requires distinguishing different kinds of self-knowledge; but Davidson's and Burge's core idea is adequate to that task once we appreciate that some attitude-types constitutively involve different kinds of self-knowledge than others do.

WORKS CITED

Bilgrami, Akeel. 1992. *Belief and meaning*. Oxford: Blackwell.

Boghossian, Paul. 1994. The transparency of mental content. *Philosophical Perspectives* 8: 33-50.

Brandom, Robert. 1994. *Making it explicit*. Cambridge, MA: Harvard University Press.

Burge, Tyler. 1979. Individualism and the mental. *Midwest Studies* 4: 73-121.

_____. 1988. Individualism and self-knowledge. *Journal of Philosophy* 85: 649-63.

Butler, Keith. 1997. Externalism, internalism, and knowledge of content. *Philosophy and Phenomenological Research* 57: 773-800.

Capote, Truman. 1965. *In cold blood*. New York: Random House.

Davidson, Donald. 1967. Truth and meaning. *Synthese* 17: 304-23. Reprinted in *Inquiries into truth and interpretation*, 17-36. Oxford: Clarendon Press.

_____. 1984. First-person authority. *Dialectica* 38: 101-11.

_____. 1987. Knowing one's own mind. *Proceedings & Addresses of the American Philosophical Association* 60: 441-58.

Evans, Gareth. 1973. The causal theory of names. *Aristotelian Society Supplementary Volume* 47: 187-208. Reprinted in *Collected papers*, 1-24. Oxford: Clarendon Press.

Falvey, Kevin, and Joseph Owens. 1994. Externalism, self-knowledge, and skepticism. *Philosophical Review* 103: 107-37.

Frege, Gottlob. 1914. Logic in mathematics. In *Posthumous writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, 203-50. Chicago: University of Chicago Press.

Kripke, Saul. 1979. A puzzle about belief. In *Meaning and use,* ed. Avishai Margalit, 239-83. Dordrecht: Reidel.

_____. 1972, 1980. Naming and necessity. In *Semantics of natural language*, ed. Donald Davidson and Gilbert Harman, 253-55, 763-69. Dordrecht: Reidel. Reprinted as *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Loar, Brian. 1987. Names in thought. *Philosophical Studies* 51: 169-85.

Owens, Joseph. 1986. Synonymy and the nonindividualistic model of the mental. *Synthese* 66: 361-82.

Putnam, Hilary. 1975. The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7: 131-93. Reprinted in *Philosophical papers*, vol. 2, *Mind*, *language and reality*, 215-71. Cambridge: Cambridge University Press.

Sosa, David. 1996. The import of the puzzle about belief. *Philosophical Review* 105: 373-402.