

FIRST-PERSON KNOWLEDGE AND AUTHORITY¹

I

Let us call a thought or belief whose content would be expressed by a sentence of subject-predicate form (by the thinker or someone attributing the thought to the thinker) an 'ascription'. Thus, the thought that Madonna is middle-aged is an ascription of the property of being middle-aged to Madonna. To call a thought of this form an *ascription* is to emphasize the predicate in the sentence that gives its content. Let us call an '*x*-ascription' an ascription whose subject is *x*, that is, an ascription such that the subject of the sentence which would express its content is *x*. Let us call a 'self-ascription' an ascription whose subject is identical with the ascriber. Let us call a 'reflexive ascription' a self-ascription such that either (i) in the sentence the ascriber would use to attribute correctly the ascription to himself he would use the first-person pronoun to refer to himself both as the ascriber and as the subject of the ascription, as in a sentence of the form

I believe that I ϕ ; [1]

or, (ii), in case the individual is not a speaker, provided that we would attribute the ascription to him only in a sentence of the form:

X believes that he [or she or it] ϕ . [2]

Let us say that an ascription is expressed by an ascriber by the sentence that gives its content. Typical examples of sentences used to express *reflexive* ascriptions by the person making the ascription are the following:

I feel hungry. [3]

I am taller than you are. [4]

I am 4000 miles from the North Pole. [5]

I believe that Bill Clinton is the President of the United States. [6]

Reflexive ascriptions expressed by [3] and [6] have a different epistemic status for the ascriber than those expressed by [4] and [5]. Although I can ascribe to myself a height, and a distance from the North Pole, I am not a special authority on how tall I am, or how far I am from the Pole. I am perhaps better placed in some respects to ascertain my own height and my position relative to the Pole than most people, and this is not entirely accidental, but there is no difficulty in principle about someone else being as or more authoritative about my height and relative position than I am. In contrast, it seems hardly possible to know what could be meant by the suggestion that someone else might be as well or better placed to know that I feel hungry or that I believe that Bill Clinton is the President of the United States than I am. Evidently, reflexive ascriptions with this special epistemic status are restricted to ascriptions of mental states or properties.²

To the extent that we can make sense of someone else being better placed to know or say what one's mental states are than oneself, this is restricted to dispositional mental states, and we must tell an elaborate accompanying story. Even this accompanying story in the end depends for its confirmation on the subject's acknowledgement of its correctness from his perspective, and so ultimately the subject of the story is still specially placed to determine whether he has the ascribed state.³ When we restrict our attention to conscious mental states and occurrent attitudes, even these stories fail us. Still, in such cases we have an example of how someone may know at a time no better than someone else what he thinks, or not know as well. My psychiatrist might tell me, for example, as part of an elaborate story, that I want to kill my father, and on this basis I may come to reflexively ascribe to myself the desire to kill my father. This reflexive ascription does not differ epistemically from my ascribing to myself gum disease upon being so informed by my dentist. However, this seems to be a kind of ground we would accept for reflexive-ascriptions of a mental state only in the case of dispositional mental states, and then only reluctantly. In the case of reflexive ascriptions of conscious mental states or currently manifested propositional attitude dispositions, we are not inclined to take anyone else's evidence to be so much as relevant.

Despite this difference between reflexive ascriptions of conscious mental states and dispositional mental states, the special warrant of our reflexive ascriptions of conscious or occurrent mental states extends naturally to latter because necessarily these are in part dispositions to

have particular conscious or occurrent mental states. A gap remains because we may have a dispositional mental state without its currently being manifested in consciousness. It is the possibility that such dispositions may be blocked from manifesting themselves in consciousness that gives rise to the possibility of psychoanalytic explanations of our behaviour, and, more generally, of someone being in a better position to say what we dispositionally believe, etc., than we are ourselves. This is not the usual case, however. So we may say that generally we are specially placed to know both our own conscious and unconscious mental states.

Let us call reflexive ascriptions of mental states 'first-person ascriptions'. Reports of first-person ascriptions by their subjects we will call 'first-person reports'. These reports (in English) would be of the form [1], with a predicate that picks out a conscious mental state or event substituted for ' ϕ '. A report is an assertive speech act, and one may insincerely report something. So it is possible for a speaker to issue a first-person report although he does not believe that what he reports is so, and for first-person reports to be false.

First-person ascriptions can be contrasted with non-first-person reflexive ascriptions, such as those expressed by [4] and [5], with non-reflexive self ascriptions of non-mental states and properties, such as that expressed by,

The author of this paper is 4000 miles from the
North Pole, [7]

with non-reflexive self ascriptions of mental states, such as the one expressed by,

The author of this paper feels hungry, [8]

and with other person ascriptions, such as those expressed by,

Bill Clinton is the President of the United States, [9]

and

Bill Clinton believes that he is the President, [10]

when the ascriber is not identical with Bill Clinton. Of course, I am no more specially placed with respect to non-reflexive ascriptions of

non-mental properties than with respect to reflexive ascriptions of non-mental properties. More interesting is the fact that I am not unqualifiedly specially placed to determine the truth of [8]. That is because I am not specially placed to determine that I am the author of this paper. With respect to either [9] or [10], clearly, I am not in a special epistemic position; and so my ascriptions expressed by these sentences have no special warrant of the sort that attaches to those expressed by [3] and [6]. Whatever epistemic advantage over others one might have with respect to the ascription of a property to another, whether it is a psychological property or not, one has contingently. In contrast, the epistemic advantage one enjoys with respect to (reflexive) ascriptions of psychological states to oneself one has necessarily. There is thus a necessary asymmetry between first-person and other ascriptions of psychological states. No one can be in a better position than oneself to ascribe to oneself a conscious mental state. And no one else can generally be better placed than oneself to ascribe to oneself dispositional mental states. This is reflected in our according special evidential weight to sincere assertions speakers make about their own mental states. We can call this phenomenon 'first-person authority'.

Let us now set these facts down for future reference:

Necessarily, no one is in as good an epistemic position to ascribe conscious or occurrent mental states to a thinker as the thinker is himself. (E1)

Necessarily, no one is in general in as good a position to ascribe dispositional mental states to a thinker as the thinker is himself. (E2)

Necessarily, no one is in general in as good a position to ascribe mental states to a thinker as the thinker is himself. (E3)

Necessarily, a speaker's sincere first-person reports should be accorded more evidential weight than reports about the speaker's mental states by others. (E4)

(E1) and (E2) together entail (E3). (E3), together with some obvious epistemic principles, entails (E4). Clearly, (E1) is the central phenomenon.⁴

This difference between the epistemic position one is in with respect to one's own mental states and those of others corresponds to a difference between the evidence one goes on when making first-person ascriptions and when making other ascriptions. When ascribing mental states to others, one's evidence must include their past or present behaviour. Generally, in reflexively ascribing mental states to oneself, one does not go on any evidence, and does not consult one's behaviour. This is a difference between, on the one hand, the way that each of us knows our own mental states and, on the other, the way we know the mental states of others and the way others know our mental states. Let us call knowledge of our own mental lives of this sort 'first-person knowledge'. First-person knowledge and first-person authority are not equivalent. First-person knowledge does not necessarily require first-person authority, because a difference between the way one knows something and the way another knows it does not entail that there is a difference between the authority with which each one knows. And first-person authority does not by itself imply that there is a difference in the way one knows one's own mental states. Nonetheless, one should expect there to be some intimate connection between first-person knowledge and authority.

II

To draw attention to the asymmetry between our warrant for first and other person ascriptions of psychological states is not to explain it. In this paper, my primary aim will be to examine an important and strikingly original explanation by Donald Davidson for the special warrant of reflexive ascriptions of mental states. My primary source will be Davidson's article "First-person Authority", though I will draw occasionally on other articles.⁵ The topic is of central importance to our attempt to understand the nature of thought and communication. The special character of Davidson's explanation of first-person authority, furthermore, plays an important role in the defense of his relational account of the determination of thought content.⁶

Davidson aims to explain the asymmetry between our knowledge of our own mental states and our knowledge of the mental states of others (or, alternatively, the asymmetry between our own knowledge of our mental states and the knowledge others have of them) by explaining a closely related asymmetry: why there is a "presumption that a speaker

is right when he sincerely attributes a belief, desire, or intention to his present self, while there is no such presumption when others make similar attributions to him" (p. 101). This is a variant (E4) above. My main question will be how far Davidson has succeeded in explaining this asymmetry in authority, and to what degree it explains the related asymmetry in our knowledge of our own and others' mental lives, represented by (E3). I will argue that Davidson's explanation is unsuccessful on both fronts, and, further, that the hope for an illuminating explanation of the special epistemic status of first-person ascriptions is a philosophical *ignus fatuus*.

The thrift with which Davidson's argument is presented justifies some elaboration. This I undertake in section V. First, however, in section III, I lay out constraints on a successful explanation of the asymmetry of warrant between first-person and other person ascriptions. Next, in section IV, I identify a number of restrictions which Davidson places on the scope of his initial explanandum. In section V, I lay out what I call Davidson's 'master argument' for first-person authority. I provide an explication of two lines of argument for a crucial premise in the master argument which are suggested in Davidson's compressed discussion, and then explain the importance of Davidson's argument in the context of his larger philosophical position. In section VI, I criticize the argument. I argue first that Davidson's master argument, even if sound, is unsuccessful in explaining our primary explanandum, (E3). Then I criticize each of the two lines of argument for the crucial premise in the master argument. Finally, I argue that no argument of the form that Davidson presents can provide an understanding of the special role that consciousness must play in any explanation of the asymmetry in warrant between first and other ascriptions of mental states. Section VII is a short conclusion in which I argue (briefly) that no explanation of this asymmetry of the sort Davidson's seeks is likely to be found.

III

A successful philosophical explanation of the asymmetry between our warrant for first-person ascriptions of mental states and other ascriptions⁷ of mental states must meet the following criteria of adequacy:

The explanans must entail the explanandum. [11]

The explanans must not contain explicitly a statement of the asymmetry to be explained, or any asymmetry which stands in need of an explanation at least as much as the original asymmetry. [12]

The explanans must be conceptually prior to the explanandum. [13]

[11] is a minimal condition on the relevance of the explanans to the explanandum. [12] is a minimal condition on the informativeness of the explanation. It requires that the explanans not be a mere redescription of the explanandum or rely on an unexplained asymmetry at least as mysterious as the original one. [13] requires that the explanans be illuminating. To illustrate the force of [13], consider the following putative explanation of personal identity over time:

X at time *t* is identical with *Y* at time *t'* because *X* at time *t* remembers doing something at time *t'* that *Y* did at *t'*.

Intuitively, the explanans does not provide an adequate explanation of the explanandum. The reason is that an understanding of the explanans presupposes an understanding of the explanandum. The explanans is sufficient for the explanandum because no two people can perform numerically the same action, and states of memory are necessarily veridical, so that if *A* remembers doing *x*, *A* did *x*. If *B* did *x*, then *A* and *B* are identical. However, our understanding of what it is for *A* to remember something that he did in the past presupposes an understanding of what it is for an individual at one time to be the same individual as an individual at another time.

IV

The asymmetry that Davidson investigates is in three ways narrower than (E3). First, Davidson restricts his attention to propositional attitudes. Second, among the propositional attitudes, he restricts his attention to belief. Third, he restricts his attention to linguistic beings.

Given the third restriction, Davidson argues that we can legitimately reformulate the question why we have a special warrant for first-person ascriptions⁸ of mental states as the question why we speak with special

authority in reporting our own mental states: “if one can speak with special authority, the status of one’s knowledge must somehow accord; while if one’s knowledge shows some systematic difference, claims to know must reflect the difference” (p. 102). The latter question can in turn be reformulated as the question why an interpreter of a speaker should give the speaker’s first-person reports special evidential weight in attributions of attitudes to him, (E4). This question can without loss, Davidson says, be treated as the question, “what explains 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?” (p. 109). Thus, Davidson is committed to the following conditionals:

If we explain the difference in warrant we each have that I am right when I state what I believe, then we will have explained the difference in warrant between first-person ascriptions of beliefs and other ascriptions of beliefs. [14]

If we have explained the difference in warrant between first-person ascriptions of belief and other ascriptions of beliefs, we will be in a position to explain the difference in warrant between first-person ascriptions of propositional attitudes and other ascriptions of propositional attitudes. [15]

If we are in a position to explain the difference in warrant between first-person and other ascriptions of propositional attitudes, we will be in a position to explain the difference in warrant between first-person and other ascriptions of sensations and other non-propositional mental states. [16]

The success of Davidson’s overall strategy in explaining first-person authority depends upon the truth of these conditionals, and upon the claim that if we explain the asymmetry for linguistic beings, there will be no separate task required for non-speakers, or at least that if there is, it will not reflect adversely on the adequacy of the explanation for linguistic beings. While all of these assumptions deserve attention, and in partic-

ular the assumption that we can safely restrict our attention to linguistic beings in searching for an explanation of first-person authority, I will be restricting my attention to the first assumption, [14].

v

There are, I think, suggestions of at least two explanations of the special weight accorded first-person reports in Davidson's discussion. I will first lay out the general form of the argument, and then what seem to be the two principal explanations for the crucial premise in the argument which are suggested in the discussion.

The asymmetry Davidson aims to explain in the first instance, and upon which the original asymmetry is said to rest, is that between my warrant for thinking that I have said something true in making a first-person report and your warrant for thinking that I have said something true. As a preliminary stage in the explanation, we can observe that if anyone knows that I hold true a sentence *s* on a particular occasion of utterance, and what I mean by that sentence, he is in a position to know what belief I express with that utterance. Let us suppose, now, that we both know that I hold a sentence *s* true, and that we each know *that* I know what *s* means. We have in this, Davidson notes, assumed no asymmetry between the knowledge that you and the knowledge that I have. Now, however, "there is this difference between us, which is what was to be explained: on these assumptions, I know what I believe, while you may not" (p. 110).

Before continuing with the presentation of Davidson's argument, we must note a difficulty that arises at this point. That is that it seems *prima facie* possible (a) that I know that I hold a sentence *s* true, and know what *s* means, and yet not believe what is expressed by *s*, and (b) that I may believe something, without believing for any sentence *s* which expresses what I believe that *s* is true. In the first case, (a), this is because it does not follow either from my believing that *s* is true and my knowing the meaning of *s* that I believe that *p*, where 'p' is replaced by *s*. I may come to believe that a certain Latin sentence is true, without understanding it, and later learn its meaning, but not put these two pieces of knowledge together, and so fail to come to believe what is expressed by the sentence. If I can know that I hold *s* true and

know what *s* means without believing what *s* expresses, then it will not follow from our assumptions above that I know what I believe while you do not. The reason for (b) is that it does not follow from my believing that *p* that I have any beliefs about sentences. But if the assumptions cited above are to serve as a ground for the asymmetry in our knowledge of my beliefs, it must be the case that whenever I have a belief, I also hold true a sentence that expresses that belief's content. Thus, Davidson needs some additional assumptions that connect my holding true a sentence with my believing what the sentence expresses, and my believing something with my holding true some sentence which expresses what I believe. Davidson needs to assume that (i) for all *x*, if *x* holds *s* true, and *x* knows the meaning of *s*, then *x* believes what *s* expresses, and that (ii) for all *x*, if *x* believes that *p*, then there is a sentence *r* which expresses what *x* believes and which *x* holds true, and that (iii) both you and I know (i). In addition, in order to infer that I do know what I believe while you do not from these assumptions, Davidson needs to assume that (iv) from all of these assumptions, I will infer what I believe.

Let us grant all of this. Even so, we do not have an explanation of our asymmetry. Even if it follows from the fact that both *A* and *B* know that *C* holds true *s* and that *B* knows what *C* means by *s*, that *B* knows (or at least is in a position to know) what *C* believes, while *A* may not know (or be in a position to know) what *C* believes, this does not show that *B*'s pronouncements about *C*'s beliefs should in general be given special evidential weight in attributing beliefs to *C*. So far as anything we have said yet goes, *B*'s knowing what *C* means by *s* may be entirely fortuitous. Davidson recognizes this. "It remains to show why there must be a presumption that speakers, but not their interpreters, are not wrong about what their words mean" (p. 110). What we must show, then, is why there should be a presumption that *B* knows what *C* means by *s* when *B* is identical with *C*, while there is no such presumption when *B* is not identical with *C*.

Once we have shown this, on the assumption that both *A* and *B* know that *B* holds true the sentence *s*, we can show that there is a presumption that *B* knows (or can know) what *B* believes while there is no such presumption in the case of *A*. The argument, which we can call the 'master argument', can be laid out as follows, first for the case of a single sentence, then more generally:

Master Argument

- (1) *A* knows that *B* holds true *s*.
- (2) *B* knows that he holds true *s*.
- (3) There is a presumption that *B* knows what he means by *s*, while there is no presumption that *A* knows what *B* means by *s*.
- (4) For all *x*, if *x* holds true *s* and *x* knows the meaning of *s*, then *x* believes what *s* expresses.
- (5) For all *x*, for all *y*, if *x* knows that *y* holds true *s* and knows what *y* means by *s*, and *x* knows (4), then *x* knows or is in a position to know what *y* believes in holding true *s*, i.e., to know that *y* believes that *p*, where 'p' is replaced by *s*.
- (6) *A* and *B* both know (4).
- (7) Therefore, by (1)–(6), there is a presumption that *B* knows or is in a position to know what he believes in holding true *s*, while there is no presumption that *A* does.

To extend this result to every belief of *B*'s, we must assume,

- (8) For all *x*, if *x* believes that *p*, then there is a sentence *r* which expresses what *x* believes and which *x* holds true,
- (9) For all sentences *s*, if *B* holds true *s*, then *B* knows that he holds true *s*,

and then modify (3) to cover all sentences *B* holds true. Thus,

- (10) For all sentences *s*, if *B* holds true *s*, then there is a presumption that *B* knows what *s* means, while there is no presumption that *A* knows what *B* means by *s*.

Then, even assuming that *A*'s knowledge of which sentences *B* holds true is equal to *B*'s, this yields,

- (11) For all sentences *s*, if *B* holds true *s*, then there is a presumption that *B* knows what he believes in holding true *s* while there is no presumption that *A* does (by (4)–(6), (9), (10)).
- (12) Therefore, for any belief of *B*'s, there is a presumption that *B* knows what he believes while there is no presumption that *A* knows what *B* believes (by (8) and (11)).
- (13) Therefore, *B*'s sincere reports of his self-ascriptions of beliefs should be accorded special evidential weight in attributions of beliefs to him (by 12).⁹

Although a number of these premises are questionable, (10) (or in the more restricted argument (3)) is the crucial premise in the sense that it is the source of the asymmetry derived in the conclusion. What is the argument for it? The answer is supposed to lie in reflections on the different roles of the interpreter and speaker in communication. At first, Davidson seems to suggest that the asymmetry is to be explained by appeal to an asymmetry in the evidence that the speaker and interpreter go on:

A hearer interprets (normally without thought or pause) on, the basis of many clues: the actions and other words of the speaker, what he assumes about the education, birth-place, wit, and profession of the speaker, the relation of the speaker to objects near and far, and so forth. The speaker, though he must bear many of these things in mind when he speaks, since it is up to him to try to be understood, cannot wonder whether he generally means what he says (p. 110).

[Thus,] The asymmetry rests on the fact that the interpreter must, while the speaker does not, rely on what, if it were made explicit, would be a difficult inference in interpreting the speaker (p. 110).

To stop here, however, is not to explain the asymmetry, for all we have done is to invoke another asymmetry whose relation to the asymmetry in warrant remains mysterious. That the interpreter must go on behavioral evidence, while the speaker does not go on evidence at all, is properly part of the description of the puzzle, not its solution. As Davidson himself remarks in an earlier passage, "claims that are not based on evidence do not in general carry more authority than claims that are based on evidence, nor are they more apt to be correct" (p. 103). What

we minimally need, in addition to this asymmetry in evidence, is some reason to think that in this case the speaker's reports about his mental states, in addition to not being based on evidence, are "apt to be correct". It is this claim, I believe, that it is the main burden of Davidson's argument in the following passages to establish, in each of which I believe a different argument is suggested. First:

The speaker, after bending whatever knowledge and craft he can to the task of saying what his words mean, cannot improve on the following sort of statement: 'My utterance of "Wagner died happy" is true if and only if Wagner died happy'. An interpreter has no reason to assume this will be his best way of stating the truth conditions of the speaker's utterance (pp. 110–111). (I)

Second:

[imagine] a situation in which two people who speak unrelated languages, and are ignorant of each other's languages, are left alone to learn to communicate . . . Let one of the imagined pair speak and the other try to understand . . . The best the speaker can do is to be interpretable, that is, to use a finite supply of distinguishable sounds applied consistently to objects and situations he believes are apparent to his hearer . . . It is . . . obvious that the interpreter has nothing to go on but the pattern of sounds the speaker exhibits in conjunction with further events (including of course, further actions on the part of both speaker and interpreter). It makes no sense in this situation to wonder whether the speaker is generally getting things wrong. His behaviour may simply not be interpretable. But if it is, then what his words mean is (generally) what he intends them to mean . . . 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 (p. 111). (II)

It is not clear to me that Davidson intends to be giving two separate arguments in these two passages. Yet the first passage suggests a ground for the asymmetry which seems quite different from that suggested in the second passage, and which I believe is of sufficient interest to justify examination whether or not Davidson intends to be presenting a separate argument there.¹⁰ Therefore, I will treat each passage as offering a different argument.

In the first passage, our attention is drawn to the following asymmetry. If a speaker desires to state the conditions under which a sentence of his is true in virtue of what the sentence means, he can do no better than to employ disquotation. In the case of an interpreter, whose language must at least initially be assumed to be different from that of the speaker, even if the interpreter uses a sentence syntactically identical with a

sentence the speaker utters, there is no guarantee that disquotation in the interpreter's language will yield the correct truth conditions of the speaker's utterance. If in my language 'dog' means what 'cat' does in yours, and *vice versa*, then applying disquotation to your sentence, 'dogs chase cats', will yield the wrong interpretation in my language. No such possibility is open when the speaker uses disquotation for his own language, since the sentence he uses on the right hand side of the biconditional and the sentence he mentions on the left hand side are in the same language. Thus, we may say that a speaker who employs disquotation will not be able to make a mistake in stating the truth conditions of his sentences; he will state the truth conditions of his sentences in a way that ensures that the sentences he uses in stating the truth conditions provide interpretations of them. On the first line of argument, we conclude from this that the speaker is in general in a better position to state the truth conditions of his sentences in a way which exhibits their meaning than the interpreter, and therefore can be presumed to know what his words mean, while no such presumption attaches to the interpreter's assignments of truth conditions to the speaker's utterances in the interpreter's language. Let us call this argument I, which we can represent as consisting of the following steps: (1) A speaker can always correctly state the truth conditions of his sentences by using on the right hand side of the biconditional which he uses to state its truth conditions the sentence which he mentions on its left hand side. (2) An interpreter of the speaker cannot be sure that disquotation in the interpreter's language will yield a correct statement of the truth conditions of the speaker's utterance. (3) Therefore, a speaker is always in a position to correctly state the truth conditions of his sentences while the interpreter is not. (4) Therefore, there is a presumption that a speaker knows the meanings of his words, while there is no such presumption that his interpreter does.

In the second passage, we are asked to imagine two individuals who wish to communicate with one another but who share no common language. We consider the situation from the point of view of the interpreter. The aim is to show that an interpreter must, as a part of his project, assume that a speaker knows the meanings of his words.

We can note first that an interpreter must assume that a speaker is trying to communicate with him; otherwise, the interpreter will not have any reason to treat any of the speaker's actions as speech acts. Thus, the interpreter will assume that the speaker intends to be supplying

clues to the meanings of his utterances. The speaker can do no better than to talk about features of his environment he believes to be salient to his interpreter. Since the interpreter has only the speaker's utterances in the presence of conditions and events in their shared environment to go on, the interpreter must assume that the speaker is largely rational and has true beliefs about his environment, and in particular about features of the environment salient to the interpreter. "It makes no sense", Davidson says, "in this situation to wonder whether the speaker is generally getting things wrong" (p. 111). This assumption, which Davidson has elsewhere called "the Principle of Charity",¹¹ is required (Davidson argues) because the only strategy available to the interpreter in assigning interpretations to a speaker's utterances is to assume that the salient conditions in his environment that cause the speaker's hold-true attitudes provide their truth conditions.¹² In practice, of course, the interpreter's first access to hold-true attitudes is through the speaker's assertions about events and conditions in his environment. It is at this point that an assumption enters about the speaker's knowledge of the meanings of his words. For if one is to take the speaker's assertions to be a guide to his beliefs about his environment, one must assume that the speaker is correctly applying his words to events and conditions in his environment on the basis of his intention to say what he believes to be the truth about them. In the absence of the need to keep track of how other members of one's linguistic community use words, this comes to the speaker using his words consistently in expressing his beliefs. Thus, if interpretation is possible, the speaker must know in this sense what his words mean. The interpreter then has no choice but to assume that the speaker is mostly right about what his words mean, on pain of losing the subject of his interpretation. Thus, the assumption is forced on the interpreter in virtue of his commitment to carrying out his project, to treating his subject as a speaker; while, of course, no such assumption must be made about anyone else's having knowledge of the meanings of the speaker's terms. Let us call this argument II. Its crucial steps are these: (1) A speaker is interpretable only if he can apply words consistently to objects and events in his environment with the intention of stating what he believes about them. (2) If a speaker can apply words consistently to objects and events in his environment with the intention of stating what he believes about them, then he knows the meanings of his words. (3) By (1) and (2), if a speaker is interpretable, then the speaker knows the meanings of his words. (4) It is an assumption of the project

of interpretation, then, that the subject of interpretation knows the meanings of his words; no such assumption must be made about anyone else's knowledge of the speaker's words.

Arguments I and II each independently supports premise (10) of the master argument (with the additional plausible assumption that most of the sentences a speaker holds true will be sentences in his language). Thus, if either argument is successful, provided that the master argument is successful, we have a ground for an asymmetry between the weight attached to *B*'s sincere reports of his beliefs and that attached to *A*'s reports of *B*'s beliefs, and of the knowledge that *B* has of his beliefs and knowledge that *A* has of *B*'s beliefs.

If successful, Davidson's explanation of the source of the authority which attaches to first-person reports can play a central role in the defense of Davidson's account of the relational determination of thought content. By now, a standard criticism of accounts of thought content that hold that an individual's thought contents are partially or wholly determined by the individual's relations to his physical environment is that such an account is incompatible not just with our special authority to say what our own mental states are, but with our being in a position to know what they are at all.¹³ Our special warrant for saying what our beliefs are seems to be threatened by relational theories of thought content because the contents of our thoughts, according to these views, are determined by our relations to objects and events in our environment about which we are not authoritative. Surely, the complaint goes, we can be no more authoritative about the contents of our beliefs than we are about what determines those contents. Moreover, our having any knowledge of the contents of our thoughts seems threatened, for our knowledge of facts about our environments seems to rest on prior knowledge of the contents of our thoughts. But if our thought contents are determined by their relations to objects and events in our environments, it seems we must know first about our environment before we can know about our thoughts. Thus, it seems that if the relational account is correct, we could know either of these things only if we first knew the other, which must give rise to a thorough scepticism about the contents of our thoughts.¹⁴

Davidson's argument for the relational character of thought content, which is at the same time an argument for our being mostly right about our environment, depends crucially on the claims that every speaker is in principle interpretable, and that an interpreter has no choice but to

apply the principle of charity, that is, the assumption that those of a speaker's beliefs which are prompted by (salient) events and conditions in his environment are by and large correct.¹⁵ Without going into the details of this latter argument, we can see that if, in particular, the second explanation of first-person authority I have attributed to Davidson (the master argument plus argument II to support (10)) is successful, the fact that makes for the relational character of thought content on Davidson's view, and which secures the possibility of knowledge of the world around us against the sceptic, at the same time secures the speaker's authority over his utterances. If Davidson is right, then, far from the relational character of thought being incompatible with first-person authority, the two spring from the same source. The result is a unified philosophical account of our knowledge of our own and other minds and the world around us of great elegance and beauty.

VI

While I believe that Davidson's argument illustrates some important connections between the special authority of first-person reports and the possibility of a speech community, I do not think that we have in these considerations an explanation of that special authority.¹⁶ In the following, I will consider first whether what I have called Davidson's master argument succeeds in explaining the asymmetry in warrant between first and other person ascriptions of mental states and the corresponding asymmetry between the warrant a speaker has for first and other person reports. I will argue first that even if we grant the soundness of the master argument, it goes very little ways toward explaining any of the asymmetries we started with in section I, (E1)–(E4). Next, I will argue that if it were necessary to know what we believed in the way that the master argument represents us as knowing this, we would be unable to do so. After this, I turn to criticisms of arguments I and II for premise (10) of the master argument. Argument I fails in two ways. First, there is no difficulty in principle for someone other myself stating in my language the truth conditions of my sentences. Second, stating the truth conditions of sentences in my language cannot be a sufficient condition for an asymmetry in knowledge of the meanings of my sentences because it is not sufficient to know their meanings. Argument II fails for several reasons. As stated, it cannot explain why all speakers know the meanings of their words or speak with special authority about their

mental states because it does not entail this. To ensure this, we must assume that every speaker is interpretable. But even with this assumption, we do not explain the asymmetry in knowledge of the meanings of a speaker's words, but presuppose it. Furthermore, an examination of the assumptions that go into the argument for a presumption in interpretation that a speaker knows the meanings of his words reveals that we are presupposing in this that the speaker knows with special authority the contents of his thoughts, which renders appeal to an asymmetry in knowledge of meanings, when so supported, unsuitable for explaining the asymmetry in knowledge of attitudes. Finally, I raise the objection that Davidson's argument does not have the resources to explain why consciousness should play a special role in our understanding of the special epistemic position we occupy with respect to our own mental states.

To begin, a preliminary observation we should make is that Davidson's official explanandum, that

there is a "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", [17]

is neither sufficient to establish any special authority for first-person ascriptions we may have nor entailed by the asymmetry that Davidson explains. It is not sufficient to establish any special authority for our first-person ascriptions because it is compatible with my knowing that I am right when I say 'I believe Wagner died happy' that I am ignorant of the meanings of my words. If I am, then despite knowing that I am right, I will not know what I believe in virtue of knowing that I am right. (Thus, [14] is false.) By the same token, since the asymmetry Davidson explains is an asymmetry in knowledge of the meanings of our words, it could not by itself explain an asymmetry in our warrant for thinking I had said something true when saying that 'I believe Wagner died happy'.

In itself, this shows only that Davidson initially misstated the explanandum for the explanation he goes on to give. But it points to an important flaw in the master argument. Unless sentences such as [17] are true and Davidson has an explanation of this, he will not be in a position to explain the asymmetry we began with between our warrant for first and other person ascriptions, or that between first and other person reports.

Davidson's explanation of first-person authority rests on an explanation of an asymmetry between the knowledge that a speaker and interpreter have of meanings of the speaker's words. This asymmetry between the knowledge one has of one's own words and an interpreter's knowledge of the meanings of one's words is most striking in the case of an interpreter who is not a member of one's speech community. To test whether this asymmetry is the real or only source of first-person authority, we can consider the best possible case of another person knowing the meanings of one's words in order to see to what degree this reduces the asymmetry between first and other person ascriptions of mental states.

Let us consider then, not just people who are members of the same speech community, but people who are raised together, and are regular interlocutors. Let us suppose, for example, that I have a twin brother and that we were raised together, went to the same schools, have most of the same interests, and spend most of our time together. In this situation, our knowledge of each other's idiolect may be presumed to be almost perfect. While there is perhaps still some sense in which I may know better what I mean by my words than my brother does, he is in an excellent position to interpret my words. If the asymmetry between self-ascriptions and other ascriptions depends solely on an asymmetry between one's knowledge of the meanings of one's own words and others' knowledge of their meanings, in this situation, we should expect the asymmetry in knowledge or warrant to be almost completely eliminated. For all practical purposes, we should each know the other's thoughts as well as we know our own. But the asymmetry is not eliminated, nor even very much reduced. I know, for example, for each of my conscious thoughts, that I am thinking it, and what I am thinking; my twin brother, of course, although his position with respect to his own conscious thoughts is equally good, for the most part does not know or even have very many detailed beliefs about my conscious mental states. Most of what I think may never be indicated in any way in my behaviour. This does not prevent my knowing what I am thinking, but without behavioral evidence of some kind, my twin would not be able to know any of my thoughts. This same point, of course, extends to our dispositional states, which I can know about without consulting my behaviour, while my twin cannot. Thus, our nearly equal knowledge of the meanings of the words in my idiolect goes only a very little way toward accounting for the asymmetry in our knowledge of my thoughts.

Even when we turn to consider those of my thoughts to which I give verbal expression, we find that equal knowledge of the meanings of my words is not by itself enough to eliminate an important asymmetry between my knowledge and his knowledge of my thoughts. For even in this case there is an inference which my twin would need to make which I would not. For my twin would need to have reason, grounded in past experience, present circumstances, and behavioral evidence, to believe that my assertion is sincere, while I would need no such evidence. This asymmetry in our positions is not explained by an asymmetry in our knowledge of the meanings of the words in the sentence I utter.

Once we remind ourselves of these features of our authority and knowledge about our own thoughts, we can hardly expect an asymmetry in the knowledge we have of the meanings of words in my idiolect to be an adequate explanation of the asymmetry between the warrant we have for first and other person reports or first and other person ascriptions of mental states. An asymmetry in knowledge of meanings would seem adequate only if we restricted our attention to knowledge of thoughts which are expressed verbally, and which we are allowed to assume are sincerely expressed. Much of the asymmetry between our knowledge of our own and our knowledge of others' thoughts comes from our knowledge of thoughts to which we and they do not give verbal expression.

This problem is clearly reflected in Davidson's argument. For an asymmetry in the knowledge of the meanings of our words to be sufficient to account for the asymmetry in warrant between self-ascriptions and other ascriptions, there would have to be no need for an asymmetry in the knowledge expressed in (1) and (2) in the master argument.¹⁷ However, if there were no asymmetry in the knowledge expressed in (1) and (2), it would not necessarily always be the case that a speaker knew with greater warrant what he thought than his interpreter, because although he might know with better warrant what his words meant, he might not know with as great a warrant as his interlocutor what sentences he held true. Thus, without an asymmetry between one's own knowledge of which sentences one holds true and that of an interpreter, we are not guaranteed an asymmetry between one's knowledge of one's own and of others' attitudes, and consequently we are not guaranteed that a speaker's own reports of his mental states should be accorded more weight than those of others.

My first objection to Davidson's explanation, then, is that even if

we grant the correctness of the explanation of the asymmetry between our knowledge of the meanings of our own words and of the meanings of others' words, we have not explained the asymmetry in warrant between first-person and other person reports. We want to explain why, necessarily, I am always in a better position to know and say what I believe, etc., than anyone else. The explanans does not entail this explanandum. This violates our first criterion of adequacy. In order for the explanans to entail the explanandum, we must add to the explanans that the speaker is presumed to know what sentence he holds true, while there is no such presumption on the part of the interpreter. This, however, would violate our second criterion of adequacy, that we do not rely in our explanation on any asymmetry which stands just as much in need of explanation as the original one.

A second objection against the master argument is that it misrepresents what our knowledge of our own propositional attitudes comes to. For the argument to succeed, we must represent our knowledge of the content of our propositional attitudes as composed of two separate bits of knowledge: first, knowledge of which sentences we hold true; second, knowledge of the meanings of our sentences. From these two items of knowledge, we can infer the contents of our beliefs. However, it seems clear that our knowledge of the contents of our beliefs does not in this way consist of knowledge of sentences we hold true and knowledge of their meanings.¹⁸ First, it should be clear that no such inference takes place when we report our mental states. Second, if our knowledge of the contents of our attitudes depended on first knowing both of these things separately and prior to our knowledge of any attitude content, we could not know the contents of our beliefs. For to know that one believes that *s* is true is to know the content of one of one's beliefs. Thus, if one needed to know these two things prior to knowing what one believed, one could never know what one believed. Our knowledge of the contents of our beliefs then is not correctly represented as based on an inference from knowledge of what sentences we hold true and what those sentences mean. Someone else could be in that position with respect to us and so come to know what we believe on the basis only of such knowledge, but we could not be in that position in general with respect to ourselves.

Now I turn to some objections to Davidson's explanation of the assumption that a speaker knows better than his interpreter what the meanings of his words are, and his use of this argument in his expla-

nation of why a speaker's first-person reports ought to be accorded special evidential weight.

We identified two potentially different arguments. The first is in passage (I). This argument for an asymmetry relies on the fact that when a speaker gives truth conditions of utterances in his language, he uses the language for which he gives truth conditions, while the interpreter may use another language. This would explain the asymmetry in knowledge of the meanings of the expressions in the speaker's language, however, only if (i) it were not possible for the interpreter also to use the speaker's language to state the truth conditions of the speaker's sentences and (ii) stating the truth conditions of sentences in the speaker's language in his language were sufficient for the speaker to know the meanings of the expressions in his language. Neither of these conditions is met. In terms of argument I, this means that, first, (3) does not follow from (1) and (2), and, second, (4) does not follow from (3).

Suppose that *A* speaks language *L* and *B* speaks language *L'*, and that *L* and *L'* are different languages. Let us consider how *A* and *B* would typically state the truth conditions of a sentence *s* in *A*'s language *L*.

A: *s* is true in my idiolect iff *p* (*L*) [18]

B: *s* is true in *A*'s idiolect iff *p* (*L'*) [19]

We indicate in parentheses in what language the sentences uttered by *A* and by *B* are spoken. Provided that *A*'s idiolect is *L*, *A* is guaranteed to get the truth conditions of *s* right in [18] if he uses *s* in place of 'p'. By the same token, however, if *A*'s idiolect is *L*, *B* is not guaranteed to get the truth conditions of *s* right in [19] if he uses *s* in the place of 'p', given that *L* and *L'* are different languages, and *B*'s statement is in *L'*.¹⁹

While this is true, it is not sufficient to explain an asymmetry in *A*'s and *B*'s knowledge of what *A* means by his words. It would be sufficient only if it were not possible for *B* to state the truth conditions of a sentence *s* in *A*'s language, which we can represent as follows:

B: *s* is true in *A*'s idiolect iff *p* (*L*) [20]

But there is clearly in principle no reason why *B* cannot do this provided that *L* is a language which can be spoken by more than one person. To

deny this would be to accept that *A*'s language is a necessarily private language, which is certainly not an option that Davidson would be willing to accept. If *B* states the truth conditions of *s* as in [20] using *s* in place of 'p', then *B* is guaranteed to get the truth conditions of *A*'s sentence right for the same reason that *A* is in [18]. Thus, there does not seem to be any necessary asymmetry between the ability of *A* and *B* to state the truth conditions of *A*'s sentences in *A*'s language. It does not follow, then, from the fact that *A* can always state the truth conditions of his sentences in his language correctly by disquotation, while *B* cannot necessarily state the truth conditions of *A*'s sentences by disquotation in *B*'s language, that a speaker is always in a position to correctly state the truth conditions of his sentences while an interpreter is not.

Furthermore, since it is clear that *B* could state the truth conditions of *A*'s sentences in *A*'s language without understanding *A*'s language, *A*'s being able to use *A*'s language to state the truth conditions of his sentences is not sufficient for *A* to understand his own sentences. Thus, *A*'s always being able to state the truth conditions of sentences in his language could not be sufficient to explain an asymmetry between his and *B*'s knowledge of the meanings of the words in *A*'s language because it would not be sufficient for *A* to know what his words meant. This difficulty might be circumvented by stipulating that one cannot state the truth conditions of a sentence in a language unless one understands that language. But this is simply to acknowledge that being able to state the truth conditions of a sentence in a language is no explanation of one's knowledge of the meaning of the sentence, since it presupposes that knowledge; thus, any appeal to an asymmetry between the warrant for a speaker and interpreter in stating the truth conditions of the speaker's sentences simply presupposes an asymmetry in their knowledge of the meanings of the speaker's sentences.

Have we overlooked the asymmetry that is really doing the work here? It might be said that although *A* and *B* can both state the truth conditions of *A*'s sentences in *A*'s language, only *A* is guaranteed to be in a position to know that he is stating the truth conditions of his sentences in his language. Even granting the premise for the sake of argument, this is not enough to ground an asymmetry between *A*'s and *B*'s knowledge of the meanings of sentences in *A*'s language. There is, perhaps, a trivial sense in which *A* knows that he is stating the truth conditions of his sentences in his language: he knows he is stating them in a language he speaks, and thus needs no independent check on whether the

meanings of the sentences he uses are appropriate for giving the truth conditions of his sentences. However, this does not entail that he has any special authority about the meanings of the sentences in his language, for he could know he is stating the truth conditions of his sentences in his language, for the reason given, even if he did not know their meanings.

The second line of argument we identified holds that a speaker is interpretable only if he can consistently apply words to objects and events in his environment with the intention of stating his beliefs. Since in the absence of an external check by his linguistic community on the appropriateness of his use of his words, getting it right can come to no more than using his words consistently, it follows that if the speaker can be interpreted, he knows what he means by his words. From this it follows that an interpreter must assume that a speaker knows the meanings of his words, since it is a presupposition of his project that his subject is interpretable.

As it stands, this argument is incomplete, for all it shows is that if a speaker is interpretable, he knows the meanings of his words. While this may be enough to ensure that if an interpreter intends to interpret a subject, he is committed to assuming that the subject knows the meanings of his words, this does not show that every speaker knows the meanings of his words or is a special authority on what his words mean.

Even restricting our attention to speakers who are interpretable, we do not have an explanation of their first-person authority. On this view, the interpreter's project can succeed only if the speaker is authoritative about his attitudes. Thus, the interpreter must assume that the speaker is authoritative if he undertakes to interpret him. This explains why the interpreter must *assume* that the speaker is authoritative about the meanings of his words, but it does not explain why the speaker *is* authoritative. That the interpreter must assume that the speaker is authoritative about the meanings of his words could not explain why the speaker is authoritative, because we infer that the interpreter must assume that the speaker is authoritative from the fact that only such speakers can be interpreted. To suppose that the need for the interpreter to assume the speaker is authoritative in order to interpret him explains why the speaker is authoritative would be like supposing that the need to assume that one has a hand in order to make a fist explains why one has a hand.

This objection to the argument would not be a concern of Davidson's,

if, as I believe, he is committed to the claim that everyone who speaks a language is in principle interpretable.²⁰ If this claim is correct, then from that fact that to be interpretable a speaker must (for the most part) know the meanings of his words, in the sense explained above, we can conclude that everyone who speaks a language knows the meanings of his words. Thus, on this view, the presumption that a speaker knows the meanings of his words is grounded in his speaking a language at all.

I will not challenge the claim that every speaker is in principle interpretable. Appropriately qualified, I think this is correct.²¹ Nor will I challenge the claim that a speaker is interpretable only if the speaker is mostly right about what his words mean. This too is correct, I think. But these facts do not explain why every speaker knows the meanings of his words. The form of our argument is the following: one speaks a language only if one is interpretable; one is interpretable only if one is mostly right about the meanings of one's words; therefore, one speaks a language only if one is mostly right about the meanings of one's words. We have here two options for explaining why a speaker knows the meanings of his words. First, he knows the meanings of his words because he is interpretable. This is unsuccessful because it gets the order of the explanation backwards. It meets our first two criteria of adequacy, but not the third. As we have seen, a speaker knows the meanings of his words not because he is interpretable but is interpretable because he knows the meanings of his words. Second, he knows the meanings of his words because he is a speaker of a language. This looks initially more promising. But the difficulty here is again that this explanation provides no philosophical illumination of the explanandum. For our conception of what it is to be a being who speaks a language is that it is to be a being who uses his words consistently in his application of them to objects and events in his environment in communicating with other members of his linguistic community. To say this is to say that a speaker is someone who, *inter alia*, knows the meanings of his words. Suppose we say that an *X* is anything that is *A*, *B*, and *C*. We cannot then explain why *s* is *A* by saying that it is *X*, for this cites no fact deeper than that for which we sought an explanation. And so, in citing the fact that someone speaks a language, we cite a fact which is no deeper than that which we wanted explained.

Another, independent, difficulty arises in using argument II as part of an explanation of the special authority that attaches to first-person

reports, and of the special warrant of first-person ascriptions. The difficulty is that the argument presupposes and depends upon the asymmetry between the speaker's knowledge of his attitudes and the interpreter's. To see this, notice that in our description of the interpreter's assumptions in the interpretive project we treat the speaker as a rational agent who is trying to communicate with the interpreter. This provides us with an explanation of his consistent use of words to express his beliefs, which is a necessary condition of his being interpretable. The speaker typically uses, e.g., the same vocable 'dog' whenever he wants to say what he believes about a salient dog in his environment. This presupposes that the speaker knows what he believes about his environment, that he knows how he has used his words in the past, which requires him to know what he believed in the past, and that he knows what he intends. Otherwise, there would be no guarantee that he would be correctly using his words in saying what he believes about his environment. If the speaker were not a better authority on what he believed than his interpreter, then the speaker would not be in a better position to know what he means by his words. Thus, in order to explain the asymmetry between the speaker's knowledge of the meanings of his words and the interpreter's, we must presuppose an asymmetry between the speaker's knowledge of his attitudes and the interpreter's knowledge of them. To attempt to turn from this to explaining the asymmetry between the speaker's and interpreter's knowledge of the speaker's attitudes by appeal to an asymmetry between the speaker and interpreter's knowledge of the meanings of the speaker's words is to presuppose in one's explanans the explanandum. We cannot then expect to use this asymmetry between the knowledge that a speaker and interpreter have of the speaker's words to explain the asymmetry between their knowledge of the speaker's attitudes.

A final concern I have about the explanation that Davidson offers, which does not focus on the details of the argument, is that it does not give any role to consciousness in explaining the special warrant we have for our first-person reports and in explaining the special warrant of first-person ascriptions. This is a concern which I wish I could articulate more clearly. It seems to me that an understanding of the phenomenon of consciousness, however difficult it is to talk about, must play a central role in our understanding of our special epistemic position with respect to our own mental states. No argument or explanation which omits to mention consciousness can be correct because being conscious

is a necessary condition on knowing what our own mental states are. Yet, if Davidson's explanation were correct, it would seem that a being with propositional attitudes who was not conscious would know just as well as one who was conscious what the contents of its propositional attitudes were. However, it seems to me, such a being could not be said to know anything about its own mental states. Thus, a successful explanation of the asymmetry between first and other person ascriptions must entail that the knower is conscious. Davidson's explanation does not meet this requirement.

Another way to see why consciousness must play a central role in our understanding of our knowledge of our own mental states is to consider the difference between our knowledge of our mental lives when we are awake and when we are asleep and not dreaming. Clearly, there is a vast difference here in our knowledge of our mental states. This difference cannot be explained without reference to consciousness. But Davidson's argument does not have the resources to make any distinctions between periods of consciousness and unconsciousness. The knowledge that Davidson's master argument appeals to, knowledge that one holds certain sentences true, and knowledge of the meanings of one's terms, can be understood dispositionally. So understood, it does not distinguish between one's knowledge of one's mental life when one is conscious and when one is unconscious.

This complaint with Davidson's argument is connected with our identifying in section I (E1) as the primary source of first-person authority. Davidson's argument cannot explain (E1) because it cannot single out consciousness for a special or primary role in our knowledge of our own mental lives.

VII

The real source of first-person authority is first-person knowledge. We know about our own mental lives in a way that is different in kind from that of anyone else. This knowledge of our own mental lives is not based on observations or knowledge of our behaviour, and is non-inferential. Since facts about our behaviour do not entail that we have mental states or, if we do, what they are, someone who is in the position of having to infer from our behaviour to what our beliefs are is always at risk of getting it wrong. These facts are in themselves enough to explain why we are generally better placed to say what we believe, desire, intend,

etc., than are others. Whereas they must make an inference which puts them at epistemic risk, we need make no such inference in order to know. This is not, of course, a satisfying explanation. It rests the asymmetry in warrant for first and other person reports and ascriptions on an asymmetry in the way we know our own and others' mental states. To explain the former adequately we cannot just cite the latter, without having an explanation of it in turn.

With respect to first-person knowledge, however, I think there is little hope of a philosophically illuminating explanation. We can draw attention to its central role in our understanding of beings like ourselves who are rational agents. For knowing one's own mental states is of a piece with being a rational agent. (This, I think, is the ultimate source of the requirement that an interpreter assume a speaker to have knowledge of his meanings and attitudes, for the interpreter must assume that his subject, if the subject is a speaker, is a rational agent.) A rational agent is one who acts on the basis of his beliefs and desires in a way exemplified by rationalizing explanations of actions. This presupposes patterns of interaction among beliefs and desires that display an awareness of their contents. The only way to make sense of a failure of self-knowledge is to try to imagine that someone's actions fall into incoherence. Too much incoherence in someone's behaviour, however, convinces us not that they fail to know the contents of their attitudes, but that they are not, after all, agents. This is an insight we owe to Davidson. To notice this connection between rational action and self-knowledge is not an explanation of the latter, however. We invoke it in our explanation of what it is to be a rational agent; it is thus part of our description of what it is to be a rational agent. And so saying that we are rational agents does not constitute an illuminating explanation of that fact about us. I do not see that we can do better in striving for philosophical illumination than to recognize the role that non-inferential knowledge of our own mental states plays in our conception of the kind of beings that we are. This is not to explain first-person knowledge but to recognize it as something we can appeal to in explaining other things, as an end point of explanation.

*University of Florida
Gainesville, FL, U.S.A.*

NOTES

¹ I would like to thank my colleagues, John Biro and Greg Ray, for helpful comments on earlier versions of this paper.

² By mental states and properties, I will throughout have in mind pure psychological states and properties, i.e., those which are not analysable into a psychological and non-psychological component. I do not count, for example, knowing that Bill Clinton is the President of the United States as a pure psychological state. Clearly, one is not in a special position to determine whether one meets the non-psychological conditions for knowing something about the world around one.

³ I am, of course, thinking of Freudian accounts of the unconscious. According to Freud, crucial to the confirmation of hypothesis about the unconscious is that after therapy the subject should come to acknowledge that he or she has the ascribed attitude from his or her own perspective.

⁴ Note that this characterization of our special warrant for first-person ascriptions does not entail that they are either indubitable or incorrigible.

⁵ "First-person Authority", *Dialectica* 38(2), 101–111, 1984. All parenthetical citations to page numbers will be to this article. Davidson discusses problems involving our knowledge of our own thoughts in various other essays, e.g., in "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association* 60; and "What is Present to the Mind?" in *The Mind of Donald Davidson*. Johannes Brandl and Wolfgang Gombocz (eds.), Rodopi, Amsterdam, 1989. The latter two essays are mainly concerned with diagnosing a false picture of the nature of propositional attitudes which Davidson charges underlies the widespread belief that relational theories of thought content cannot properly accommodate first-person authority.

⁶ See "A Coherence Theory of Truth and Knowledge", and "Empirical Content", both reprinted in E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, New York, 1986.

⁷ The asymmetry can be stated in either of two equivalent ways: first, as the asymmetry between one's warrant for ascribing mental states to oneself and to others; second, as the asymmetry between one's warrant for ascribing mental states to oneself and others' warrant for doing so. These are equivalent for our purposes because the explanation of either will provide an explanation of the other.

⁸ Davidson does not draw attention to the difference between reflexive and non-reflexive ascriptions of mental states to oneself, but it seems clear that he has in mind reflexive ascriptions as the primary phenomenon of investigation.

⁹ One difficulty here which I will not be making much of is that this style of argument cannot explain why our special authority over self-ascriptions extends only to those which are reflexive self-ascriptions. As we have noted before, whatever special warrant I have for saying that Kirk Ludwig has a headache depends upon my knowing that I am Kirk Ludwig, a fact to which I have no special epistemic access, and my knowing with special authority that I have a headache.

¹⁰ That Davidson does intend the first passage to play an important role in the argument is suggested by this passage from "What is Present to the Mind?",

... it does not make sense to suppose I am generally mistaken about what my words mean; the presumption that I am not generally mistaken about what I mean is essential to my

having a language – to my being interpretable at all. To appeal to a familiar, though often misunderstood, point: I can do no better, in stating the truth conditions for my utterance of the sentence “The Koh-i-noor diamond is a crown jewel” than to say it is true if and only if the Koh-i-noor diamond is a crown jewel (p. 18).

In a footnote, Davidson refers the reader to “First-person Authority” for a more detailed discussion.

¹¹ See, e.g., “Radical Interpretation”, “Belief and the Basis of Meaning”, and “Thought and Talk”, in *Inquiries into Truth & Interpretation*, Clarendon Press, Oxford, 1984. Although I believe that there are some difficulties involved in the argument for the necessity and correctness of the Principle of Charity, I will accept it for the sake of argument. The criticisms I will make will not depend upon calling into question the Principle of Charity.

¹² We shift here to hold-true attitudes because this provides a non-question begging source of evidence for the interpreter. One can know that a speaker holds a sentence true without knowing what the speaker believes in holding it true or what he means by it, neither of which the interpreter could know prior to successfully interpreting the speaker. This introduces, however, the need to make the assumptions about the connection between hold true attitudes and beliefs we introduced in the master argument as assumptions (4) and (8). I do not believe that Davidson has ever discussed the need or justification for these assumptions.

¹³ See, for example, Brueckner, A. “Brains in a Vat”, *The Journal of Philosophy* 83 (March), 148–167, 1986; Boghossian, P. “Content and Self-Knowledge”, *Philosophical Topics* 17 (Spring), 5–26, 1989.

¹⁴ I do not endorse these criticisms. I do not think that externalism about thought content by itself entails that we do not have first-person knowledge. I think that it does, however, together with two independently plausible assumptions: (1) knowledge of our own thoughts supervenes on the subjective character of our conscious mental states; (2) the subjective character of our conscious mental states is not determined relationally. For a fuller discussion, see my “Brains in a Vat, Subjectivity, and the Causal Theory of Reference”, in *The Journal of Philosophical Research* XVII, 313–345, 1992; and chapter 6 of my dissertation, *Scepticism and Externalist Theories of Thought Content*, UMI, 1990.

¹⁵ See Davidson’s “The Method of Truth in Metaphysics”, in *Inquiries into Truth and Interpretation*; also “A Coherence Theory of Truth and Knowledge” and “Empirical Content”. For an explication and criticism of this argument, see my “Scepticism and Interpretation”, *Philosophy and Phenomenological Research* LII(2) (June), 317–339, 1992.

¹⁶ One difficulty here worth mentioning, though I will not dwell on it, is that the asymmetry between the warrant for self-ascriptions and for other ascriptions is not restricted to linguistic beings. The present argument holds that being interpretable entails that one can be presumed to know what propositional attitudes one has. But if we have the same phenomenon in non-linguistic beings, obviously this will not explain it there. Furthermore, on the assumption that the phenomena we find in non-linguistic animals is continuous with the phenomena we find in homo sapiens, we must look for a unified explanation for the asymmetry between first and other person ascriptions for linguistic and non-linguistic beings.

guistic beings. If this is right, then the offered explanation, since it applies only to linguistic beings, cannot be the right one.

It is unlikely that this objection would trouble Davidson, who offers an argument to show that non-linguistic beings do not have propositional attitudes in "Thought and Talk". The argument relies on the assumptions that (1) one cannot have any propositional attitudes unless one has beliefs, that (2) one cannot have a belief unless one has the concept of a belief, that (3) one cannot have the concept of a belief unless one understands the possibility of being mistaken, and that (4) one can understand the possibility of being mistaken only if one is a linguistic being, because the contrast between truth and error has utility only in the context of interpretation. For our purposes, the crucial premises are (3) and (4), for unless a non-linguistic being can have beliefs about beliefs in addition to beliefs, its existence would be irrelevant to our argument. In my view, the most problematic assumption is (4), but this is not the place to provide a careful evaluation of it.

¹⁷ This point was first made in my ken by Bruce Vermazen in a seminar at Berkeley in 1987. I believe Davidson was aware of this deficiency in the argument at the time he wrote "First-person Authority", for his conclusion is that, for any speaker, "there is a presumption that if he knows that he holds a sentence true, he knows what he believes" (p. 111). Clearly, the conditional conclusion cannot support our original explanandum, which is not a conditional. Davidson cannot have been unaware of this fact. I suggest, therefore, that he regarded his argument as only a partial explanation of first-person authority.

¹⁸ I am indebted here to Charles Siewert.

¹⁹ An initial difficulty is presented by the fact that in [18] *A* uses a first-person pronoun to refer to himself. Arguably, the proposition expressed by a sentence containing a first-person pronoun cannot be expressed by anyone else. If so, *B* cannot state the truth conditions for *s* in the same way that *A* does. It is not clear, however, why this fact about the sentence *A* asserts in [18] should make for a difference in knowledge of the meanings of *A*'s sentences; for this reason I will set this concern aside.

²⁰ Consider this passage from "The Inscrutability of Reference", in *Inquiries into Truth and Interpretation*:

The semantic features of language are public features. What no one can, in the nature of the case, figure out from the totality of the relevant evidence cannot be part of meaning (p. 235).

The assumption is also required, I believe, to derive a number of the other conclusions Davidson has argued for, such as the impossibility of radically different conceptual schemes ("On the Very Idea of a Conceptual Scheme", in *Inquiries into Truth and Interpretation*), and the impossibility of massive error ("A Coherence Theory of Truth and Knowledge").

²¹ I think this is correct in a form weaker than that to which I believe Davidson is committed. Roughly, the form in which it is correct is the following: every linguistic being is interpretable by members of his linguistic community in some environment. This leaves it open that not every speaker can interpret every other, and that in some environments, a speaker may be interpretable by no one (e.g., for obvious reasons, I may be

uninterpretable if I am a brain in a vat). As mentioned in note 16, I believe that Davidson is committed to the view that every speaker is interpretable by every other in every environment in order to secure the impossibility of radically different conceptual schemes and the impossibility of massive error. This need not be an issue in the present context, however, for the additional strength of the claim that Davidson needs for other purposes does not contribute to the argument for the claim that every speaker knows the meanings of his words.