

Meaning and Truth-Conditions

Richard G Heck Jr

Department of Philosophy, Brown University

1 Opening

Donald Davidson (1984g) famously proposed that what we want from a theory of meaning we may get from a theory of truth. What we want from a theory of meaning for a natural language, that being the case at issue, includes at least an account of the phenomena that give rise to the principle of compositionality: We want an explanation of how the meaning of a complex expression, such as a sentence, depends upon and is determined by the meanings of its syntactic parts. One central advantage of Davidson's proposal, as he saw it, is that we know from the work of Tarski (1958) how to construct theories of truth for languages with certain sorts of structures, and such a theory does help us to understand how at least some of the semantic properties of complex expressions are determined by those of their parts. Davidson's practice suggests he may have hoped that all sentences of natural language could be assigned logical forms of the variety familiar from mathematical logic. If so, then semantics would all but reduce to syntax: The only outstanding problem would be to uncover the 'logical forms' of the sentences of, say, English. This hope does not now seem reasonable. But the methods available for investigating logical form—that is, for investigating syntax—have become so powerful that it is not at all unreasonable to suppose that the syntactic part of the problem is solvable, at least in principle, and Tarski's work can be adapted to many of the resulting structures. The work that is left for the semanticist proper then becomes to solve the problems syntax and Tarski do not solve: These include, for

example, discovering what sorts of semantic values need to be assigned to novel semantic primitives, such as mass terms, and explaining the semantic significance of novel modes of composition, such as $[_{VP}V[_{CP}S]]$.¹

It is unfortunate that a proposal with such great advantages faces such grave difficulties. Indeed, it is probably fair to say that most philosophers of language, and perhaps even most working semanticists, regard truth-theoretic semantics as a non-starter, for the simple reason that it seems to have the wrong subject-matter. We will not get what we want from a theory of meaning from any theory that does not specify meanings: No theory that does not specify what the sentence 'snow is white' means, for example, can possibly explain how this sentence's meaning what it does is determined by its parts' meaning what they do. But a theory of truth for English does not specify what even one sentence of English means, since any such theory is consistent with infinitely many different suppositions concerning what any sentence of English means. For example, a theory of truth that proves this T-sentence² for 'snow is white':

¹ I do not mean to suggest that these are the only issues worth thinking about. And in actual research, semantics and syntax interact with one another, and work on semantic issues can inform work in syntax. It is, in fact, a nice question whether syntax is 'autonomous' in the sense my remarks somewhat sloppily suggest it is.

² By a 'T-sentence' here I mean any sentence of the form: S is true iff p . In particular, no relation between the sentence mentioned on the left-hand side and that used on the right-hand side is required for a sentence of that form to be a T-sentence. By a 'correct' T-sentence, I mean one that is true: No stronger relation is required for a T-sentence to be correct.

(1) ‘snow is white’ is true iff snow is white

is consistent with that sentence’s meaning anything, so long as it means something true. A monolingual speaker of German who believed that ‘snow is white’ was true iff snow was white could rationally believe as well that ‘snow is white’ meant that grass was green: If ‘snow is white’ did mean that grass is green, then, since snow is white iff grass is green, it would still be true iff snow was white.³ Since this sort of objection was first forcefully raised by John Foster (1976), it has come to be known as the ‘Foster problem’.

It is important not to confuse this problem with a closely related one, which I shall call the ‘many T-sentences problem’. One often finds people speaking of ‘the T-sentence’ that a given theory of truth proves for a given object-language sentence. Obviously, however, there isn’t any such thing as *the* T-sentence any such theory proves. Quite generally, if a theory proves $p \equiv q$ and also proves r , then it proves $p \equiv q \wedge r$. So any theory that proves (1) also proves:

(2) ‘snow is white’ is true iff snow is white and r

for any statement r that it also proves. This will include all logical truths and much else besides, though exactly what it includes will depend upon the details of the theory of truth.

The standard response to this problem is, in effect, to define what ‘*the* T-sentence the theory proves’ is supposed to mean, and the now standard way to do that is to invoke a conception of ‘canonical proof’: When we speak of ‘the’ T-sentence a theory proves for some given sentence S , we mean the T-sentence it proves in a certain special way, namely, by taking S apart, piece by piece, then applying the axioms that characterize the meanings of S ’s ultimate constituents, and then finally putting these together in accordance with the principles that characterize the significance of the various constructions that appear in S . It has sometimes been objected, however, that, while this may do the trick, formally speaking, the proposal is *ad hoc* and unmotivated (Soames, 1992, p. 34, n. 11).

³ This way of developing the point is due to Scott Soames (1992).

One way of motivating the restriction to canonical proof is psychological: We think of a competent speaker as tacitly knowing a theory of truth for her language, and the conception of canonical proof is then conceived as an empirical hypothesis concerning how her ‘language faculty’ goes about churning out T-sentences. Richard Larson and Gabriel Segal (1995, §2.2.1) present a form of this proposal that modifies the underlying notion of a ‘theory’, restricting the sorts of inferences that are permitted so that only canonical proofs are possible. I am sympathetic to that idea,⁴ but there seems to me also to be a way of motivating the restriction without appealing to psychology. The meaning of a sentence is supposed not just to depend upon but to be *wholly determined by* the significance of its constituents and how they are combined. Canonical derivations, as I have described them, are ones that appeal *only* to information about the significance of the constituents and the modes of construction. A proof of (2) will appeal to that information, but not only to it, and so will of course include more information than what ‘snow is white’ means.

In any event, the many T-sentences problem is not the one I want to discuss here. My goal here is to defend a form of Davidson’s proposal from the Foster problem. What most fundamentally distinguishes my view from Davidson’s is that I accept a ‘cognitive conception of understanding’: According to me, our actual ability to interpret the utterances of other speakers depends upon our consciously⁵ knowing what their utterances, as made on that occasion, mean. I shall not argue for this conception of understanding here,⁶ but I shall not simply assume it, either. What I am going to argue is that no solution to the Foster problem is available unless we adopt the cognitive conception and so that truth-theoretic semantics is committed to it. In particular, I shall argue in section 2 that Davidson’s own proposed solution rests, and rests essentially, upon his metaphysics of the mind, and I, at least, find that metaphysics unpalatable. In section 3, I shall consider an attempt to resolve the Foster problem based upon a suggestion

⁴ See, in particular, the discussion in my “What Is Compositionality?” (Heck, 2004).

⁵ Let me emphasize that element of the view. The claim is not that speakers must *tacitly* know facts about what their words mean, but that they must *consciously* know facts about what uttered sentences mean.

⁶ See my paper “Reason and Language” (Heck, 2005a) for the argument.

made in a related but somewhat different context by John McDowell. I shall argue that it too fails unless we embed it within the cognitive conception, towards which it therefore points, once properly understood. In section 4, I shall explain and further motivate the cognitive conception and then, in section 5, drawing on work of James Higginbotham and Ian Rumfitt, argue that the Foster problem can be solved from within the theoretical framework thereby established. Section 6 addresses an objection due to Scott Soames, namely, that the view I am defending implausibly ascribes a concept of truth to all competent speakers. The final section mentions two issues that still need addressing.

2 Radical Interpretation as a Solution to the Foster Problem

Davidson addresses versions of the Foster problem in several places. His first suggestion was that a T-sentence cannot, in the relevant sense, ‘give the meaning’ of the sentence it mentions on its left-hand side unless it is a consequence of a compositional truth-theory for the language to which the sentence belongs (Davidson, 1984g, pp. 25-6). That plausibly accounts for such cases as⁷

(3) ‘snow is white’ is true iff grass is green

since, as Davidson notes (1984g, p. 26, fn. 10), it is hard to see how this sentence could be a (canonical) consequence of a correct compositional truth-theory. By itself, however, this amendment is insufficient. As Foster observed, one can easily reformulate any compositional theory of truth that yields the desirable (1) so that it instead yields:

(4) ‘snow is white’ is true iff snow is white and grass is green.

If the theory makes use of the axiom

(5) ‘is white’ is true of x iff x is white,

⁷ Obviously, truth here needs to be relativized to a language, but I shall omit that variable to avoid cluttering the exposition.

for example, one need only replace it with

(6) ‘is white’ is true of x iff x is white and grass is green,

which is true if (5) is (Foster, 1976, p. 13).

In later work, Davidson insists that T-sentences that purport to be interpretive “are empirical generalizations about speakers, and so must not only be true but also lawlike” (1984g, p. 26, fn. 11).⁸ This suggestion, too, however, is insufficient. Soames (1988, pp. 187-8) observed that, if we replace (5) with

(7) ‘is white’ is true of x iff x is white and arithmetic is incomplete,

then the resulting theory delivers

(8) ‘snow is white’ is true iff snow is white and arithmetic is incomplete,

which is an empirical law if (1) is. As Soames notes, the same example disposes with Foster’s own suggestion that we need only strengthen the biconditional so that it requires the two sides to be necessarily equivalent: Clearly, (8) will then be true if (1) is.⁹

Davidson’s solution to the Foster problem must therefore be taken to lie elsewhere, namely, where it is usually taken to lie, in his discussion of radical interpretation. Davidson claims, at the end of “Radical Interpretation”, that

A T-sentence of an empirical theory of truth can be used to interpret a sentence. . . provided we also know the theory that entails it, and know that it is a theory that meets [certain] formal and empirical criteria. (Davidson, 1984c, p. 139)

The main formal criterion is just that the theory of truth in question should be compositional. The empirical criterion is that the theory should be verifiable by a radical interpreter.

⁸ The footnote from which this quotation is taken was added in 1982, the paper itself having been published in 1967. See also (Davidson, 1984d, p. 174).

⁹ I myself think that there is an even stronger objection to be made to Foster’s proposal, namely, that (1) itself is not true if “iff” is read as requiring necessary equivalence. But some philosophers think it is, if the relativization to a language is made explicit.

There are different ways we might understand this proposal. In the passage just quoted, Davidson says that one can use a theory of truth to interpret a language if one has certain sorts of knowledge about that theory. This proposal is oddly meta-linguistic, however, and that fact has troubled many commentators.¹⁰

But there is a better way to understand Davidson's suggestion, so that what is required is not that one should know something *about* a theory of truth but, rather, that one's knowledge *of* the theory should have a certain provenance. On this interpretation, Davidson is committing himself to the following thesis:¹¹

(RI) Someone who knew a truth-theory for a language *L*, and who knew that theory on the basis of the sort of evidence that would be available to a radical interpreter, could use T-sentences (canonically) derived from that theory to interpret speakers of *L*.

So far as I know, however, Davidson offers no serious argument for (RI) in "Radical Interpretation" or, for that matter, anywhere else. Davidson is of course aware of the need to say something in defense of (RI). But the discussion of this issue in "Radical Interpretation" occupies just the last page and a half Davidson (1984c, pp. 138–9), and Davidson does little more there than to assert that (RI) is true: "The present idea is that what Tarski assumed outright for each T-sentence"—namely, that the sentence used on its right-hand side translates the sentence mentioned on its left-hand side—"can be indirectly elicited by a holistic constraint", namely, "that the totality of T-sentences should... optimally fit evidence" available to a radical interpreter. If so—that is, if the holistic constraint really does

¹⁰ Foster (1976, §3) was perhaps the first to be troubled by it. Soames (1992, pp. 25–28) raises several problems of detail for how it is supposed to work.

¹¹ I think that (RI) is the most charitable way to read Davidson and that it sidesteps a lot of the issues about formulation that bother Foster (1976, §3). Davidson himself remarks, in "Reply to Foster", that he had "imagined the theory [of truth] to be known by someone who had constructed it from the evidence..." and that this is the key to his answer to the Foster problem (Davidson, 1984d, p. 173). But Davidson himself does not formulate (RI). Nonetheless, I shall not pause to consider the question whether Davidson really does accept (RI). I shall simply assume that he does.

constitute a workable alternative to Tarski's appeal to translation—then "each T-sentence will in fact yield an acceptable interpretation" (Davidson, 1984c, p. 139). But Davidson does not actually give any argument that the 'holistic constraint' is adequate to its intended purpose.

There is further discussion of the matter in "Belief and the Basis of Meaning", where Davidson draws well-known analogies with problems in decision theory. But there, too, the really pressing issue, the status of (RI), is only taken up in the last couple pages (Davidson, 1984a, pp. 152–4). And, so far as I can tell, Davidson says essentially nothing in defense of the view that, if a theory of truth optimally fits the evidence available to a radical interpreter, then it can be used to interpret. He does not, in particular, give any argument that might even seem to show that a theory of truth issuing in such T-sentences as (8) cannot meet the 'holistic constraint'.¹²

My old colleague Burton Dreben is reputed to have said that what a philosopher doesn't argue for is more interesting than what s/he does argue for. And in this case, at least, he's right. I think it just seemed obvious to Davidson that any theory of truth that was confirmable by a radical interpreter would *have* to be usable for interpretation, and the interesting question is why this seemed so obvious to him. To answer this question, we need to examine more closely the thought experiment in which Davidson is asking us to engage.

Radical interpretation is modeled on Quine's notion of radical translation (Quine, 1960, ch. 2).¹³ We are to imagine someone traveling abroad to a country whose language is utterly alien to her and attempting to construct and confirm a theory of truth for that language on the basis of her observations of native speakers. But why should this thought experiment be of any interest to a philosopher?¹⁴ Davidson is, of course, free to engage

¹² And again, in "Reply to Foster", Davidson writes: "By a course of reasoning, I have tried to show that if the [formal and empirical] constraints are met by a theory, then the T-sentences that flow from them will in fact" be interpretive Davidson (1984d, pp. 172–3). But I must confess to having no idea what "course of reasoning" Davidson has in mind here.

¹³ I sometimes think Davidson thought this was all that needed saying here. We've all read our Quine, right?

¹⁴ The importance of this question was made clear to me by Michael Dummett's (1978b) insistence upon a similar question regarding radical translation, and my remarks here are

in whatever speculation he wants, and if he wants to know what sorts of theories of truth could be confirmed by a radical interpreter, he is free to ask that question. But plainly, Davidson thinks the answer to this question matters: He thinks one can draw important conclusions from it. Davidson seems to think, for example, that one could establish a thesis with the same import as Quine's thesis of the indeterminacy of translation by showing that the evidence that is available to a radical interpreter will not, in general, support a unique theory of truth for a given language (Davidson, 1984c, p. 139). But that makes it all the more important to ask why this thought experiment should be supposed to illuminate questions about meaning.

One answer might begin with Davidson's claim that "[a]ll understanding of the speech of others involves radical interpretation" (Davidson, 1984c, p. 125). Surely, however, we do not actually confront other speakers from a standpoint of total ignorance about what their words mean and what their beliefs are: There is a clear sense, then, in which we are *not* all radical interpreters. Then again, in the very first paragraph of "Radical Interpretation", Davidson explicitly disclaims any interest in the psychology of language-use, so he cannot have intended radical interpretation to model the psychology of ordinary communication. Perhaps, then, the clue to Davidson's intentions is the sentence immediately preceding the one just quoted: "Speakers of the same language can go on the assumption that for them the same expressions are to be interpreted in the same way, but this does not indicate what *justifies* the assumption" (Davidson, 1984c, p. 125, my emphasis). Maybe what radical interpretation is supposed to model, then, is not the psychology of ordinary communication but its *epistemology*.

Although there is, as I've noted, some evidence for this reading, it is not, in the end, satisfying. There are two main difficulties. First, the epistemology to which Davidson would be committing himself is deeply problematic. It is all well and good to raise the question why we are entitled to assume, as we typically do, that what sounds like the language we know really is the language we know, but it simply does not follow that the epistemic situation of the ordinary speaker is comparable to that of a radical interpreter. If challenged to defend the assumption that, say, Ned Block

inspired by Dummett's.

speaks English, I would appeal to a great deal of evidence that Davidson would deny a radical interpreter. One might respond that this evidence could itself be called into question and that, ultimately, my evidence is only what would be available to a radical interpreter. But the occurrence of the word 'ultimately' here is a sure sign of a kind of foundationalism that has gone badly out of style and is anyway not the sort of epistemology one typically associates with Davidson.¹⁵

More importantly, the epistemological construal of the thought experiment does not fit the use Davidson makes of it. As noted, Davidson agrees with Quine that meaning is, in a certain sense, indeterminate, and his reason is that the evidence that is available to a radical interpreter does not suffice to restrict the interpretive theories of truth to one. No such conclusion can possibly be forthcoming on the epistemological interpretation. All we could conclude would be that the evidence available to a radical interpreter is inadequate to distinguish among the competing theories. In so far as the radical interpreter modeled the epistemic situation of ordinary speakers, one could then conclude that no ordinary speaker could justify one such interpretation over another. But it simply would not follow that there was no fact of the matter which interpretation was correct.

It seems to me, therefore, that the questions that are bothering Davidson are not epistemological but metaphysical. They are the questions Sir Peter Strawson asks at the beginning of his Inaugural Lecture (Strawson, 1971): In virtue of what are sentences meaningful? In virtue of what does a given sentence have the very meaning it has? Davidson argues in "Theories of Meaning and Learnable Languages" (Davidson, 1984e) that an answer to these questions must take due notice of the fact that the meaning of a sentence depends upon the meanings of its parts; he argues in "Truth

¹⁵ Tyler Burge (1999) makes some remarks along similar lines, though his are much better developed. My remarks here also have important points of contact with some ideas expressed by McDowell in his paper "Anti-realism and the Epistemology of Understanding" (McDowell, 1998a). McDowell would not apply these points to Davidson, however, as he seems to read Davidson somewhat differently from how I do. McDowell's target is Dummett. For what it's worth, however, my own view is that Dummett is not really an appropriate target after 1978—after, that is, the emergence of the views presented in "What Do I Know When I Know a Language?" (Dummett, 1993d) and the other papers of that period.

and Meaning” (Davidson, 1984g) that only a theory of truth can give a substantive account of this dependence. For that reason, he thinks, a theory of meaning must take the form of a theory of truth. If so, then we may be tempted to reformulate the initial question what it is for a sentence to mean what it does as: In virtue of what is a theory of truth correct or incorrect? As Foster was the first to emphasize, however, if our interest is in questions about meaning, that is not quite right: A theory of truth can be *correct* (in the sense that all its theorems are true) yet not throw much light at all on questions about meaning. So the question that needs asking, as Davidson sees things, is rather: In virtue of what is a theory of truth *interpretive*?

Davidson’s answer to this question, contained in the notion of radical interpretation, invokes a form of what I have elsewhere called the Use-Meaning Thesis (Heck, 2005b): The meaning of a sentence is determined by how it is used. Of course, until we are told what ‘meaning’ and ‘use’ are, the Thesis can be no more than programmatic. It is clear enough, however, how Davidson understands these notions: Meaning is to be understood in terms of truth-theories; use is to be understood in terms of Davidson’s notion of ‘holding true’ (or, more generally, in terms of the evidence that would be available to a radical interpreter). So, in Davidson’s hands, the Use-Meaning Thesis takes the following form: Whether a theory of truth for a given language is interpretive is determined by the circumstances under which sentences of that language are held true by its speakers.

Why does it seem obvious to Davidson that use (in his sense) determines meaning (in his sense)? Davidson is moved, I think, by much the same sort of thought that has attracted so many other philosophers to the Use-Meaning Thesis: If there is no difference in how two expressions are used, then there can be no difference in what they mean; nothing that does not somehow surface in the use of an expression can be part of its meaning. One might wish to quarrel with that idea, but one might well just concede it until more has been said about what use and meaning are supposed to be, since it is otherwise without content: As McDowell (1998a) notes, if it is a fact about how ‘snow is white’ is used that it is used to say that snow is white, then we can all quickly concede the Use-Meaning Thesis. The interesting question is why Davidson thinks use can, and should, be characterized as

he characterizes it.¹⁶ Why should the conditions under which a sentence is ‘held true’ by speakers of the language be included among the facts about how it is used? Why should semantic facts, such as facts about what expressions of the language are used to say, be excluded? Why should “detailed descriptions of the speaker’s beliefs and intentions” (Davidson, 1984c, p. 134) be excluded?¹⁷

Let me say as clearly as possible what question is at issue here: Why should we think that, if two sentences have different meanings, this difference must manifest itself in some difference in the conditions under which those sentences are held true by speakers?¹⁸ Why, in particular, should we

¹⁶ The dialectic here is thus essentially the same as that in Chomsky’s paper “Quine’s Empirical Assumptions” (Chomsky, 1969). Absent some reasonably precise characterization of the ‘behavioral dispositions’ on which facts about language are supposed to supervene, Quine’s position has no empirical content. But the precise characterization Quine gives, in terms of the responses the speaker would give in “a questionnaire circumstance, the circumstance of being offered a sentence for assent or dissent or indecision or bizarreness reaction” (Quine, 1972, p. 445), simply raises the question why we should think that facts about language actually do supervene on such a narrow base. This is what Chomsky has in mind when he remarks that “. . . we can, of course, raise the conditional probability of any sentence as high as we like. . . relative to ‘situations’ specified on *ad hoc*, invented grounds. . .” (Chomsky, 1969, p. 57). Quine is thus quite wrong to complain that his reply to Chomsky’s criticism of his use of ‘behavioral dispositions’ has been ignored (Quine, 1972, p. 445). It had been anticipated, and dismissed as unmotivated.

¹⁷ I have argued elsewhere (Heck, 2005b) that the answer to the second question—why semantic facts should be excluded—is just that their inclusion would trivialize the problem. With respect to the first, David Lewis (1985c) quarreled with Davidson’s inclusion of holding true. For all its differences, however, Lewis’s treatment of radical interpretation is still vulnerable to the objections I am bringing against Davidson. That is because Lewis’s view depends as heavily upon *his* metaphysics of mind—his analytical functionalism—as Davidson’s does upon his.

¹⁸ This is the same question, by the way, that Dummett presses against Davidson, though from a different perspective. The so-called ‘manifestation constraint’ simply *is* the principle we are discussing. And Dummett’s argument for anti-realism is then that, if we require all differences in meaning to be manifested by a difference in what is held true under which conditions, then there will be no way to defend the attribution of classical truth-conditions to, e.g., mathematical language. This point is especially worth remembering when reading Dummett’s writings from this period, especially the two papers titled “What Is a Theory of Meaning?” (Dummett, 1993e,f). I do not think these papers can properly be understood except as part of an on-going conversation with Davidson. In particular, the reason one

think that there could not be differences in meaning that make *no* difference to which sentences are held true when but that *do* make a difference to “the complex and delicately discriminated intentions with which the sentence[s] are] typically uttered” (Davidson, 1984c, p. 127)? The question is particularly pressing in light of the variant of the Foster problem due to Soames. How can Davidson’s radical interpreter can distinguish the case in which ‘Snow is white’ means that snow is white from the case in which it means instead that snow is white and arithmetic is incomplete? The sentence could of course mean either thing. But since it is a necessary truth that arithmetic is incomplete, it is not clear how such a difference in meaning could make any difference to the conditions under which ‘Snow is white’ is held true. The point, of course, is that Davidson understands ‘the conditions under which a sentence is held true’ to be *external* conditions. It is clear enough what kind of difference such a difference in meaning might make to the *internal* conditions under which the sentence was held true. As things are, many speakers who know nothing of logic hold ‘Snow is white’ true, and if that sentence meant that snow is white and arithmetic is incomplete, one would not expect that to be the case. Such differences of meaning are thus easily enough distinguished if we are allowed to characterize use in terms of the mental states of speakers. But Davidson explicitly denies such information to the radical interpreter.¹⁹

Davidson’s reply would surely be not so much to concede these points as to insist upon them. On his view, the radical interpreter is not just trying to interpret her subjects’ words but also to interpret her subjects’ *minds*. Davidson denies the radical interpreter access to prior information about her subjects’ minds precisely because he thinks that there is only the one, unified interpretive project:

The evidence [available to a radical interpreter] cannot consist in detailed descriptions of the speaker’s beliefs and intentions, since attributions of attitudes, at least where subtlety is required, demand a theory that must rest on much the same evidence as

searches in vain for any real defense of the manifestation constraint is that Dummett feels no need to argue for it: Davidson accepts it wholeheartedly.

¹⁹ Similar points are made by James Higginbotham (1991, pp. 277-8).

interpretation. (Davidson, 1984c, p. 134)

The difficulty is that, although this remark seems fair enough if it is intended as one about the epistemic predicament of a radical interpreter, the question why we should care about the epistemic predicament of a radical interpreter is still unanswered. If the radical interpreter were intended to model the epistemological predicament of an ordinary speaker, then that would provide a reason: Perhaps ordinary judgements about meaning would be vulnerable to skeptical attack if meaning were not determined by the evidence available to a radical interpreter. Sir Michael Dummett (1993a, p. 177) is clearly moved by just such a worry, but I have already registered my doubts about the epistemology that lies behind it.

The correct interpretation, once again, is to take Davidson to be making not an epistemological claim but a metaphysical one. On this interpretation, Davidson is denying the radical interpreter prior knowledge of what her subject believes and intends not because one can only *know* what someone believes and intends when one is able to interpret her words, but because the *facts* about what someone believes and intends are not independent of facts about what her words mean. The radical interpreter’s epistemological predicament, on this interpretation, is thus not a model of ordinary speakers’ epistemological situation but rather a reflection of the metaphysics of intentionality.

Some might suggest that we ought simply to have begun with the interpretation at which we have now arrived, and I certainly am not claiming to have discovered some hidden commitment of Davidson’s view. As is well-known, the metaphysical view this interpretation ascribes to Davidson—that meaning, belief, and intention are constitutively intertwined—is one to which he explicitly and enthusiastically subscribed (Davidson, 1984f). What I am doing here is simply emphasizing that Davidson’s metaphysics of meaning—his view that a whether a theory of truth is interpretive is determined by the circumstances under which sentences are held true—necessarily depends upon his metaphysics of intentionality. It is not just that Davidson happens to hold both views.

The question we have been discussing is whether every difference of meaning must manifest itself in a difference in the circumstances under

which sentences are held true. I have argued that, in so far as Davidson offers us reasons to believe that this is so, those reasons necessarily include the claim that meaning, belief, and intention are constitutively intertwined. But if we conjoin this thesis with the view it is being used to establish, then we thereby commit ourselves to a metaphysics of the mind that seems to me, at least, to be extremely unattractive. Radical interpretation, in Davidson's sense, is possible only if meaning supervenes on the sorts of facts that are available to the radical interpreter. But if meaning, belief, and intention are constitutively intertwined, then belief and intention must *also* supervene on the facts available to a radical interpreter: That is to say, what mental states one has, and with what contents, must be determined by (i) the conditions under which one would hold various sentences true and (ii) the facts about which sentences one actually does hold true. This view is but half a step beyond behaviorism, and, in my opinion, it deserves the same fate. But it is undoubtedly a view Davidson held (1984f, esp. p. 162): It is just an 'interpretive' view of the mind—what one's mental states are is a matter of how one would best be interpreted—coupled with a particular view about what the basis of interpretation is.

The argument here can perhaps be made a little clearer if we state it in the form of a challenge to Davidson. Davidson is claiming that what a sentence means is determined by the circumstances under which it and related sentences are held true. But suppose there were two situations in which the facts available to a radical interpreter—the facts about what is held true when—were the same, but in which the facts about speakers' mental states—and in particular the facts about the communicative intentions with which they uttered various sentences—were different. On Grice's view (1989b), meaning is determined not (just) by what is held true when but by facts about speakers' communicative intentions; on the view I hold (Heck, 2005a), meaning is determined by what speakers believe about the meanings of their words.²⁰ We would both be open to the possibility that the meanings of sentences could vary while the facts accessible to a radical

²⁰ Of course, neither Grice nor I would hold that *every* difference in the psychological states of speakers entails a difference of meaning: Supervenience is a one-many, not a one-one, relation.

interpreter remained the same. The question for Davidson, then, is what reason he can give us to suppose that no such possibility can be realized. His answer would be to insist that radical interpretation is simultaneously both of a speaker's words and of her mind. And Davidson's point, once again, would not just be an epistemological one: It is not just that it is hard to tell when one is in such a situation. Davidson's claim, rather, is that mentality supervenes on the kind of evidence available to a radical interpreter—on what is held true when—just as meaning does, so that there *could not be* such a situation: Mental states cannot vary while the facts about what is held true when remain the same.²¹

Davidson, then, may well have had a solution to the Foster problem that worked for him: If mentality supervenes on the kind of evidence that is available to a radical interpreter, as he thought it did, then, plausibly enough, a theory of truth known to be true on the basis of such evidence can always be used to interpret. But the price of using radical interpretation to solve the Foster problem is a commitment to Davidson's metaphysics of mind: A theory of truth known to be correct on the basis of the kind of evidence that is available to a radical interpreter can be guaranteed to be interpretive only if mentality supervenes on what is held true when. The mental doesn't supervene on what is held true when, however, at least not according to me, and so Davidson's solution is not one that will work for me.

3 Frege on Meaning and Truth-conditions

Davidson's suggestion that a theory of truth may serve as a theory of meaning was arguably anticipated by Frege. One might, therefore, seek a solution to the Foster problem in Frege's writings.

Part I of Frege's *Grundgesetze der Arithmetik* (1962)²² is devoted to the

²¹ In principle, Davidson could insist that, although mental states can vary while the facts about what is held true when remain the same, they can vary only within very narrow limits, so that such variations would not affect the meanings of sentences. But it is hard to see how one could argue for that position, and it amounts, in any event, to assuming that mentality *almost* supervenes on what is held true when.

²² The English is primarily taken from Montgomery Furth's English translation (Frege, 1964), with some minor changes. All further references will be made by section number and

exposition of the formal system in which he proposes to work. As part of that exposition, of course, Frege must explain his formal language. In Part I.1, he specifies the denotations of the language's primitive symbols. Concerning the identity-sign, for example, he writes:

- (9) ' $\Gamma = \Delta$ ' shall denote the True if Γ is the same as Δ ; in all other cases it shall denote the False. (Frege, 1962, §7)

Frege takes this clause to specify which function the identity-sign denotes: It denotes the function whose value for first argument Γ and second argument Δ is the True if Γ is the same as Δ and is the False otherwise. There is a similar clause for each of the other primitive symbols.²³ Frege then argues, in part I.2.i, that these stipulations are sufficient to assign denotations to all well-formed expressions of his formal language. The core of that argument is contained in sections 30 and 31: In the former, Frege argues that any complex expression correctly formed from denoting expressions denotes; in the latter, he argues that the simple expressions of the language denote. It follows that "the proper names [which, infamously, include the sentences], and names of first-level functions, that we can form. . . out of our simple names. . . always have a denotation" (Frege, 1962, §31).

Frege then writes:

. . . [N]ot only a denotation, but also a sense, attaches to all names correctly formed from our signs. Every such name of a truth-value [that is, every sentence] *expresses* a sense, a *thought*. Namely, by our stipulations it is determined under what condition the name denotes the True. The sense of this name—the *thought* it expresses—is the thought that this condition is fulfilled. . . . The

are to volume I. (The numbering of the sections is reset in volume II.)

²³ The stipulation concerning the horizontal is in §5; negation, §6; the first-order universal quantifier, §8; the definite article, §11; the conditional, §12; and the second-order universal quantifier, §20. The stipulation governing the smooth breathing, from which names of value-ranges are formed, is given in §9 and is, unsurprisingly, not only different from the other stipulations Frege makes but deeply problematic. I shall ignore the difficulties it poses here, as I have discussed them elsewhere (Heck, 1998, 1999, 2007). Those same papers also contain a defense of the interpretation of Part I that I am presenting, which is not universally accepted.

names, whether simple or themselves composite, of which the name of a truth-value consists, contribute to the expression of the thought, and this contribution of the individual component is its *sense*. (Frege, 1962, §32; emphasis in original)

That is to say, Frege takes the stipulations he has made regarding his primitive signs to determine the truth-condition of each sentence of his formal language, and we may take a sentence's truth-condition to determine, or perhaps just to be, its meaning. So, for example, the stipulations determine that the formula ' $0 = 1$ ' denotes the True if, and only if, zero is the same as one (these being, on Frege's view, the value-ranges of certain functions). The sense of this formula is thus the thought that zero is the same as one.²⁴

There is something obviously right about Frege's remark. Formal languages are introduced by philosophical logicians all the time. And when they are, an informal theory of reference is what is typically offered by way of explanation of the expressions of that formal language. Such explanations are typically intended to convey not just what the denotations of those expressions are but also what they *mean*, and the explanations often succeed: When I have properly appreciated such an explanation, I usually take myself to understand the formal language in question, that is, to know what its expressions mean. So I am, and am meant to be, reading the meanings of the expressions of the language off a theory of reference for it, just as Frege suggests his readers should.

Now, according to Frege, sense determines reference, but reference certainly does not determine sense (Frege, 1984, p. 159, op. 27), and the same plausibly holds for any reasonable notion of meaning that might replace the notion of sense. If so, however, it can easily seem mysterious how a theory specifying what the references of certain expressions are could possibly determine the meanings of those expressions: If reference does not determine meaning, then specifying the reference of an expression simply cannot suffice to specify its meaning. That line of thought, however, misses a crucial aspect of the case we are discussing, one nicely brought out by

²⁴ One might object that it ought to be the thought that: the condition that zero is the same as one is fulfilled. I take it, however, that Frege speaks of a condition's being fulfilled only because semantic ascent is required to state the general principle he is trying to state.

Dummett:

When we are concerned... with laying down... what the interpretation of a newly introduced symbol is to be, the particular specification of the reference may be taken as conferring simultaneously a definite corresponding sense. ... We may adapt here, in expounding a doctrine of Frege's, the famous distinction between *saying* and *showing* that Wittgenstein used in the *Tractatus*: the specification of the reference *says* what the reference is to be, and, by saying it in a particular manner, *shows* what the sense is to be. (Dummett, 1991, p. 149)

One's first reaction might be to wonder just how helpful any explanation can be that invokes the distinction between saying and showing. But Dummett's idea here is reasonable enough. In saying what the reference is, Frege has had to choose among different ways in which the reference itself might correctly be specified. So, for example, there are infinitely many ways Frege might correctly have specified the reference of the function-symbol '=' in his formal language. Actually to specify what its reference is, Frege has had to choose one such way, and the choice he has made may be taken to indicate—that is, to show—what sense he intends the symbol to have. Similar remarks apply to Frege's other primitive expressions, and to other formal languages: How its inventor chooses to state what the references of a new language's expressions are may reasonably be taken to indicate how s/he intends those expressions to be understood.

It is important to note that Dummett's remark explicitly concerns a case in which one is *stipulating* the meaning of a previously uninterpreted symbol. Dummett is not, I think, as clear about this point as he should be. In particular, the last line of the passage just quoted is importantly ambiguous. Dummett speaks of "the specification" of a symbol's reference as showing what the sense is to be, but a specification can be an act of specifying as well as a sentence used to make a specification. The *sentence* Frege uses to specify the reference of the identity-sign does not 'indicate' anything about what the sense of this symbol is: *Qua* sentence, it says what the reference of the symbol '=' is to be, and that is all. What indicates what sense the

symbol is to have is Frege's *using* a certain sentence to specify the reference of that symbol. It is not *what* Frege says, then, that shows what sense '=' is to have, but his saying what he says in a particular way or, perhaps, under particular circumstances. Showing, in Dummett's sense, is thus akin to implicating. (It may even be a kind of implicating.) Only speech acts, we may thus conclude, *show* anything, in the relevant sense.

This point does not affect Dummett's vindication of Frege, but it does affect any attempt to extend it to the semantics of natural language, where a similar problem arises. The problem I have in mind arises for anyone who accepts the apparently obligatory distinction between meaning and reference, but it arises particularly clearly for those who accept a distinction between sense and reference for proper names. McDowell (1998d) proposed some time ago that this form of the problem can be solved if we appeal to the idea of Frege's I have just been expounding.²⁵ I shall therefore explain the matter first for the case of proper names.

Consider the name 'Hesperus'. One could correctly say what its reference is in a variety of ways: One could say that it denoted Hesperus, or Venus, or Phosphorous. To say what its reference is, then, one must choose one such way, and which way one chooses will, under the right circumstances, show what one takes the sense of the name to be. One might therefore suggest that a semantic theory for English need say nothing about the meaning of a proper name other than what its reference is, even if one accepts the distinction between sense and reference for proper names: The theory will *show* what the sense of a name is even if it all it *says* is what its reference is.

It is clear that, if this solution works at all, then there is a similar solution to the Foster problem. Suppose we have a truth-theory for English that (canonically) proves the T-sentence

(10) *S* is true iff *p*.

It is true enough that, as far as the correctness of the truth-theory is concerned, one can replace '*p*' with any other sentence that has the same

²⁵ Following McDowell, Evans claims quite generally that Dummett's remarks "contain the answer to those who argue... that Davidsonian theories of meaning are inadequate because they do not state the meanings of sentences" (Evans, 1985, p. 295, n. 7).

truth-value. But to say what the truth-value of a sentence is, one must choose among the infinitely many different ways of specifying that truth-value, and which way one chooses will show what meaning one takes the sentence *S* to have. So, for example, in propounding a truth-theory that canonically proves (1) rather than (3) or (8), we thereby *show* what meaning we take the sentence ‘snow is white’ to have, even if the theory itself *says* nothing that distinguishes it from any other true sentence.

Dummett rejects this view: Although he thinks that a stipulation of what a new expression’s reference is to be can convey what sense it is to have, he insists that

... a meaning-theory is required to do more than merely show (to someone who understands the meta-language in which it is formulated) what the senses of the words of the object-language are. (Dummett, 1991, p. 149)

But Dummett does not, so far as I can tell, offer any reason for this claim, and so it might simply be rejected.²⁶ One might suggest that the parenthetical remark is the key to his thinking: The theory will convey what the the sense of the name is only to someone who already grasps that sense by associating it with the relevant word of the meta-language. But the theory will convey which object the name denotes only to someone already able to refer to that object by means of the relevant word of the meta-language. It is not clear what the distinction between these two cases is supposed to be.

The proper complaint is that *theories* do not indicate things. The fact that a particular theorist is proposing one theory of truth instead of some other may indicate what senses that particular theorist takes expressions in the relevant language to have. But it is, as we have seen, no part of what the theory actually *says* that a particular name has a particular sense. And the clauses of the theory, *qua* sentences, do not indicate anything about what the senses of expressions are: They say what those expressions denote, and

²⁶ If Dummett does offer a reason, it is equivalent to the insistence that a theory of meaning must, in his sense, be ‘full-blooded’ rather than ‘modest’ (Dummett, 1993e). But it seems likely that this issue and the one we are discussing here are effectively the same, and McDowell, in any event, would reject this requirement (McDowell, 1998b).

that is all; only speech acts indicate. If so, however, then, if what one wants a semantic theory for English to do is to specify the meanings of its various sentences, then a semantic theory for English cannot consist simply of a theory of reference for English, not if one accepts the distinction between meaning and reference. The point is clearest in the case of proper names: It is not enough that, by propounding this theory of reference instead of that one, I might thereby show what I take the sense of ‘Hesperus’ to be. One wants the *theory* to explain the phenomena, not the act of its being propounded by a theorist.

One way to appreciate this problem is to consider how a theory of truth, understood as McDowell suggests, might be used to explain the compositionality of meaning. For example, how does such a theory explain how the meaning of ‘Hesperus is a planet’ depends upon and is determined by the meanings of its parts? The theory derives a T-sentence like

(11) ‘Hesperus is a planet’ is true iff Hesperus is a planet

from clauses like

(12) ‘Hesperus’ refers to Hesperus

(13) ‘*x* is a planet’ is true of *x* iff *x* is a planet,

and, more generally, derives facts about the references (or semantic values) of sentences from facts about the references of their syntactic parts. In that way, the theory explains how the reference of a complex expression is determined by the references of its parts. But the theory seems to offer no explanation at all of how the *sense* of a sentence is determined by the *senses* of its parts. Certainly, in propounding a theory that has a consequence like (11), one indicates what one takes the meaning of ‘Hesperus is a planet’ to be; and since the theory also includes (12) and (13), one similarly indicates, by propounding the theory, what one takes the senses of ‘Hesperus’ and ‘is a planet’ to be. Since the theory derives (11) from (12) and (13), then, it derives a theorem whose (canonical) provability in the theory one is propounding indicates what one takes the meaning of ‘Hesperus is a planet’ to be from axioms whose presence in the theory one is propounding indicate

what one takes the senses of ‘Hesperus’ and ‘is a planet’ to be. But it simply does not follow that one has thereby explained why ‘Hesperus is a planet’ has the sense it has in terms of the fact that ‘Hesperus’ and ‘is a planet’ have the senses they do.

The problem can be seen from this diagram:

$$\begin{array}{ccc} \text{sense}(A) & \text{sense}(B) & \stackrel{?}{\Rightarrow} \text{sense}(C) \\ \uparrow & \uparrow & \uparrow \\ \text{ref}(A) & \text{ref}(B) & \Rightarrow \text{ref}(C) \end{array}$$

The double arrow ‘ \Rightarrow ’ symbolizes the relevant relation between the properties of the parts and the properties of the whole; the single arrow ‘ \uparrow ’ may be taken to mean *implies*, *indicates*, or *shows*, as one likes. No matter which of these readings one gives it, there is simply no reason to suppose there should be a double arrow in the top line, at least for all that has so far been said. Perhaps one could argue that for the special case in which ‘ \uparrow ’ means *shows*, the double arrow will always be present, but that very much needs to be argued. A good argument, it seems to me, would be one that replaced the metaphor of *showing* with something much more precise: It would need to be said just what relation holds between a specification of reference and a specification of sense. I conjecture that, if this could be done, then what replaced the metaphor of showing would be precise enough to be embedded in a theory of meaning, and there would no longer be any need for the sort of solution to the Foster problem we have been considering.

4 Understanding and Knowledge of Meaning

In our earlier discussion of radical interpretation, the central question was whether the ‘formal and empirical criteria’ Davidson specifies are sufficient to restrict the range of acceptable truth-theories to those that are ‘interpretive’.²⁷ Now, Davidson endorses Quine’s thesis of the indeterminacy of translation: He calls the suggestion that his criteria determine a unique

²⁷ As most commentators usually do, I tacitly conceded, for the sake of argument, that the criteria admit only correct truth-theories, that is, theories that assign the correct truth-values

interpretive theory “absurd”. Davidson nonetheless insists “that any theory that passes the tests will serve to yield interpretations” (1984c, p. 139). This remark is more puzzling, I think, than it is usually taken to be: In what sense is a truth-theory supposed to “yield interpretations”? What the theory yields would seem to be theorems in the form of T-sentences. Perhaps the right question to ask is therefore a different one: What does it mean to say that a theory of truth can be used to interpret a language, or a speaker? Suppose Fred utters a sentence, and I wish to interpret this utterance. Suppose further that I happen to know that Fred’s utterance is true iff Bedrock is cold. What am I supposed to do with this information?

One proposal would be that the radical interpreter is supposed to use her theory of truth explicitly to assign meanings to utterances. The radical interpreter would thus need to make something like the following inference:²⁸

S is true iff p

That fact is a consequence of a theory of truth I know to be correct on the basis of the sort of evidence available to me *qua* radical interpreter.

$\therefore S$ means that p

If interpreting a speaker requires one to assign meaning to her utterances, in

to sentences. One might want to question that assumption. It is not, in fact, obvious that Davidson himself would have endorsed it. Quine (1960) famously denied that the criteria for correct translation restrict the range of admissible translation manuals even to those that are materially equivalent. So far as I know, Davidson never says whether he thinks the truth-value of a sentence is uniquely determined by his criteria.

²⁸ In the spirit of the interpretation of Davidson offered in section 2, one could take the inference simply to be from ‘ S is true iff p ’ to ‘ S means that p ’. The inference is not, of course, always truth-preserving, but one might insist that it *will* be truth-preserving if the premise is canonically provable in a theory of truth known on the basis of the sort of evidence available to a radical interpreter. The radical interpreter will not go wrong, therefore, if she makes the inference only under these circumstances. Another way to put this point is to say that the second premise here need not be known but need only be true. So we would be thinking of the epistemology of our knowledge of meaning as being, in this respect, *externalist*. Though this point is important in other contexts, it is point is not terribly important in the present context. I shall therefore speak as Davidson himself does, as if the interpreter knows that her theory of truth satisfies the empirical and formal conditions.

this sense, however, then the theory of truth is simply an auxiliary device used to generate assignments of meaning (compare Soames, 1992, REF). It does not seem likely that Davidson wanted the notion of meaning to emerge so crucially here at the end of the story.

XXXXX

A different suggestion is that the T-sentence *itself* can be used to interpret the speaker. The difficulty is to say how, and part of the problem here is that the notion of interpretation itself is not particularly clear: What exactly does Davidson mean by ‘interpreting’ an utterance, a speaker, or a language, whichever the right target might be? I really don’t know, so I propose to reformulate the question. I take it that what Davidson has in mind is that someone who knew a theory of truth for a language, and knew that it met the formal and empirical criteria he specifies, could use that theory to speak and understand that language. So the question I am going to discuss is: How could one use a theory of truth, whatever its pedigree, to speak and understand the language it concerns? That, of course, depends upon what one means by ‘speaking and understanding the language’. As is common, I am going to concentrate here on what I take to be one fundamental aspect of speaking and understanding a language, namely, using the language to communicate, that is, to exchange information with other speakers: To be able to speak a language at least involves being able to use it to acquire beliefs from other speakers and to convey one’s own beliefs to them.²⁹

Our question, then, may be put as follows: In what sense, if any, can a theory of truth for a given language be used by a radical interpreter (or an ordinary speaker) to allow her to communicate using that language? The examples due to Foster’s and Soames’s may seem to suggest that the answer should be: It can’t. But there is reason for hope. I argued earlier that Foster and Soames are right, as regards Davidson’s actual position, that we have no good reason to believe that a theory of truth that is known to be correct on the basis of the sort of evidence available to a radical interpreter will always satisfy Tarski’s Convention T—that is, that the sentences used on

²⁹ Ian Rumfitt emphasizes in Rumfitt (1995) that there is no single reason for which we utter sentences, a point some of my own previous work—e.g., Heck (1995)—neglects. The present discussion is intended, however, to be compatible with Rumfitt’s observations.

the right-hand side of the T-sentences the theory (canonically) proves will translate the sentences mentioned on their left-hand sides. But suppose a radical interpreter were nonetheless irresponsibly to settle for one of the theories the criteria admit and set about using it to interpret the speakers of the language. Then it is tempting to say that, just as Frege’s choosing to specify the reference of a term in a particular way shows us what sense he takes it to have, so the radical interpreter’s using her chosen theory amounts to her choosing to regard sentences of the target language as having certain meanings rather than others. That is to say, the same line of thought we considered in the last section seems to lead to the conclusion that a radical interpreter who chose to interpret using a theory that proved (1) rather than one that proved (8) would thereby have chosen to regard utterances of the sentence ‘snow is white’ as meaning that snow is white rather than that snow is white and arithmetic is incomplete. But of course several questions remain: In what sense does using a particular theory of truth to interpret a sentence involve assigning a particular meaning to that sentence? Is there some analogue for ordinary speakers of our irresponsible radical interpreter’s decision to use one theory of truth rather than another?

Before I continue, I need to say something explicit about the overall structure of the considerations to follow. As I said earlier, I accept a cognitive conception of (occurrent) understanding: Understanding an utterance, according to me, is a propositional attitude Heck (2005a). That view leaves open the question what the content of that attitude is, however, and that is the question I mean now to discuss. How, though, should we evaluate the various proposals that might be made here?

The following conditional should be uncontroversial:

- (1) If B utters a sentence *S*, which in that context means that *p*, and if A understands B’s utterance and meets whatever conditions a correct theory of testimony would require us to include at this point, then A is in a position to come to know, or at least justifiably to believe, on a certain distinctive kind of ground, that *p*.

The conditional is uncontroversial because trivial. One might well say that the correct theory of testimony just is whichever theory makes this

conditional correct. It is controversial, of course, which theory that is, but we may ignore that issue here: We are not discussing the epistemology of testimony. What we are discussing is what constitutes understanding an utterance. And what I am suggesting is that, just as satisfying (1) is very nearly what it is for a theory of testimony to be correct, so satisfying (1) is very nearly what it is for a theory of understanding to be correct: To understand an utterance is to be in a position—*modulo* one's satisfaction of the sorts of conditions a correct theory of testimony would specify—to acquire a belief for which one would then have a certain distinctive sort of justification. If so, then theories of understanding may be tested by determining whether they satisfy (1). Particular examples, of course, may depend upon controversial claims about what the correct theory of testimony is. But not all examples need so depend. Particular examples may be neutral on what precise conditions a correct theory of testimony would require to be satisfied: All (reasonable) theories may agree that the relevant conditions are satisfied in that case.

There are some substantive assumptions about testimony to which I shall be appealing, but I take these to be relatively uncontroversial. I shall assume, for example, that, to acquire knowledge by testimony, one must be appropriately sensitive—whatever exactly that may mean—to whether one's informant is sincere and reliable. I shall also assume, slightly more controversially, that the distinctive sort of justification testimony makes available has a certain kind of structure. In particular, there is a way in which the justification one might have had for the belief that *p* can be undermined that leaves one with justification for the weaker claim that the speaker *believes* that *p*. In particular, one might be presented with strong evidence that the speaker is not reliable with respect to the question whether *p*: That would deprive one of justification for the belief that *p* but not for the belief that the speaker believes that *p*.³⁰

So let us now consider this proposal: To understand an utterance is to

³⁰ This last assumption entails that one can only have testimonial justification that *p* if one also has testimonial justification that one's informant believes that *p*. That could be, and has been, questioned, but it is weak enough, and intuitively plausible enough, that I am willing to rest my case upon it.

know its truth-condition.

Suppose that Fred asserts the sentence 'Bedrock is cold'. Suppose further that Barney meets whatever conditions a good theory of testimony would say he needed to meet: Let's just assume he has excellent reason to believe that Fred is sincere and excellent reason to regard Fred as reliable with regard to what he has said; any reasonable theory would regard that as sufficient. Then, *if* Barney understands Fred's utterance, he will be in a position to form a justified belief, indeed, a belief that might constitute knowledge, namely, the belief that Bedrock is cold. Now suppose that Barney knows

(2) 'Bedrock is cold' is true iff Bedrock is cold.

Intuitively, knowing (2) seems insufficient for Barney to garner knowledge, from Fred's utterance, that Bedrock is cold: The truth of (2) is, after all, compatible with the hypothesis that 'Bedrock is cold' means that Bedrock is broke, so long as Bedrock is broke iff Bedrock is cold. But if what the sentence 'Bedrock is cold' actually means is that Bedrock is broke, then it is not plausible that Barney could, in reaction to Fred's utterance, come to know that Bedrock is cold, at least not in the distinctive way in which one comes to know something by being told. One might want to object that if Barney has good reason to regard Fred as sincere and reliable, that constitutes good reason to suppose that Fred's utterance is true. That, together with (2), which we are assuming Barney knows, then gives him good reason to suppose that Bedrock is cold. But although Barney could come justifiably to believe, in this way, that Bedrock is cold, his belief would not be justified in the distinctive way in which beliefs acquired through testimony are justified. The reason is that, absent knowledge of what Fred's utterance actually meant, Barney's belief cannot be appropriately sensitive to the question whether Fred is reliable. To evaluate Fred's reliability, Barney must know what belief Fred is expressing. Barney may regard Fred as authoritative with regard to the weather, say, but not with regard to municipal finance, and these attitudes may be justified. So if Barney takes Fred sincerely to have said something about the weather, he may be prepared to believe it, and be justified in doing so, so that he might acquire knowledge

about the weather from Fred. But Barney may not be prepared to believe what Fred says about municipal finance, and he would not be justified in doing so if he did, whence he cannot acquire knowledge about municipal finance from Fred. That ‘Bedrock is cold’ is true iff Bedrock is cold does not, however, imply anything about what belief Fred is expressing.

That is the fundamental point: Absent knowledge of what Fred’s utterance means, Barney is in no position to determine what belief Fred is expressing.³¹ The point is worth making directly. Suppose Barney understands Fred’s utterance and justifiably regards it as sincere. Then even if he is not sure about Fred’s reliability, he may conclude, and even come to know, that Fred *believes* that Bedrock is cold. But if all Barney knows about Fred’s utterance is that it is true iff Bedrock is cold—if he does not know what it means—then he can draw no conclusion whatsoever about Fred’s beliefs. And so Barney’s knowing a correct T-sentence for the sentence Fred uttered cannot constitute his understanding it.

The natural counter-proposal is that understanding consists not in knowledge of an utterance’s truth-condition but in knowledge of what it means (Richard, 1992). Perhaps surprisingly, however, even knowledge of what an utterance means is insufficient to support attributions of belief on the basis of what someone has said. Consider again Fred’s utterance of ‘Bedrock is cold’ and suppose that Barney knows that ‘Bedrock is cold’ means that Bedrock is cold. Fred may not: Maybe he thinks it means that Bedrock is broke. Barney may even have reason to doubt that Fred thinks it means that Bedrock is cold. If so, then Barney cannot justifiably conclude that Fred believes that Bedrock is cold. What Barney needs to know if he is to attribute a belief to Fred is not just what Fred’s words mean but what they mean *to Fred*. The point should be obvious once stated: Communication with another speaker—in so far as it involves attributing beliefs to her and acquiring beliefs from what she says—depends not upon what her words mean nor even upon what one takes her words to mean oneself: It depends upon what *she* means by her words and upon what one takes her to mean by her words. Now typically, what one takes someone to mean by her words is

³¹ Thanks to Crispin Wright for helping me understand the structure of my thought on these matters.

what one takes them to mean oneself, and perhaps such a default assumption is in some sense necessary. Nonetheless, when I am responding to the words of another, the crucial question I must answer is what *she* means by her words.³²

I take it, then, that the following inference would be a reasonable one for me to make:

- (3) The speaker, *N*, has uttered the sentence *S*.
- (4) *N* takes her utterance of *S*, in the present context, to mean that *p*.
- (5) *N*’s utterance of *S* is sincere.
- (6) Hence, *N* believes that *p*.

What I am now going to argue is that the role played in such an inference by (4) may be played instead by something like:

- (7) *N* takes her utterance of *S*, in the present context, to be true iff *p*.

The motivation for the suggestion can be explained as follows. We began with the proposal that the work being done by (16) could be done instead by

- (8) *N*’s utterance of *S* is true, in the present context, iff *p*.

In moving from (4) to (8), we have made two changes: We have replaced the reference to the utterance’s truth-condition with a reference to its meaning, and we have replaced an absolute judgement about what the meaning of the utterance is with a judgement about what the speaker takes it to mean. My suggestion is that the latter change is the one that really matters and that it is, in fact, adequate by itself to resolve the problems faced by the initial proposal. Actually, that’s not quite right: (7) will not quite do. But something much like it will do. Before I explain how it needs to be amended, though, let me emphasize two features of the proposal I am making.

³² In fact, something similar is true even when I am speaking: The crucial question for me is how my words will be understood. See Heck (2006) for more on this matter.

First, one might wonder whether it is enough to defend Davidson's suggestion that a theory of truth may serve as a theory of meaning to defend the view that the knowledge of meaning on which ordinary linguistic action depends consists in knowledge of truth-conditions. Questions about meaning, one might say, are one thing, and questions about what, if anything, ordinary speakers know, in virtue of which they are able to communicate, are another Soames (see, for example, 1988). But what interesting questions would remain unanswered if it could be established—as I think it can be—that a speaker's ability to communicate depends upon, and is explained by, her knowledge of truth-conditions? I do not know of any. In any event, the view I am defending accepts a strong form of Dummett's oft-cited insistence that “a theory of meaning is a theory of understanding” Dummett (1993e, p. 3): I take the primary notion to be *how a speaker understands an uttered sentence*, not *what a word means in a language*. One can reconstruct the latter notion if one likes in terms of the former notion, but I do myself think it is of much theoretical interest.³³

Second, my view, as I've stated it, does not require that there be any sense in which speakers know *theories* of truth for their languages. If understanding an utterance involves knowing a T-sentence for it, and if knowledge of that T-sentence is to be among one's *reasons* for reacting to the utterance in a certain sort of way, then one's knowledge of the T-sentence must be conscious, in the sense that it is available for the rational control of action. It is agreed on all sides, of course, that ordinary speakers do not have conscious knowledge, in this sense, of semantic theories for their languages. So if speakers do, in some sense, know semantic theories for their languages, that knowledge is tacit, and tacit knowledge is, almost by definition, not available for the rational control of action.³⁴ That said, however, the question nonetheless begs to be asked what the source of our conscious knowledge of truth-conditions is, if we really do have such knowledge. It is a natural, though not inevitable, proposal that the truth-

conditions we consciously know are the upshot of subpersonal processes that draw upon the information contained in the axioms of a theory of truth and, indeed, that these processes implement an algorithm for deriving T-sentences from tacitly known axioms of the theory of truth. I think something along those lines is correct Heck (2004), but I will not defend the claim here.

It is unclear to what extent this view differs from Davidson's. On the one hand, Davidson insists, in “Radical Interpretation”, that the question that interests the semanticist is hypothetical: What *could* we know that *would* allow us to interpret the words of other speakers? Regarding this question, he writes that it

is not the same as the question what we *do* know that enables us to interpret the words of others. For there may easily be something we could know and don't, knowledge of which would suffice for interpretation, while on the other hand it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation. Davidson (1984c, p. 125, emphasis in original)

It is tempting to read this passage as rejecting the view I am defending. If, however, what Davidson means by ‘know’ here is *consciously know*, and if what he has in mind is not just knowledge of T-sentences but knowledge of a theory of truth, then there need be no disagreement between us: Davidson could consistently agree that understanding another speaker requires conscious knowledge of what she takes the truth-conditions of her utterances to be while remaining agnostic about the empirical question what the source of such knowledge is.

5 Understanding and Truth-Conditions

The view I am defending, then, is that, in so far as we rely, in linguistic communication, upon knowledge of meaning, knowledge of meaning may be taken to consist in knowledge of truth-conditions. More precisely, I am claiming that the role that is played by (4) in this reasoning

³³ This view is standard-issue among linguists but relatively uncommon among philosophers. See Heck (2006) for a defense of it.

³⁴ If speakers did consciously know semantic theories for their languages, then semantics would be easy: Just ask some native speakers what they know.

- (3) The speaker, *N*, has uttered the sentence *S*.
- (4) *N* takes her utterance of *S*, in the present context, to mean that *p*.
- (5) *N*'s utterance of *S* is sincere.
- (6) Hence, *N* believes that *p*.

can instead be played by something similar to

- (7) *N* takes her utterance of *S*, in the present context, to be true iff *p*.

I use the non-committal term 'takes' here because I mean to leave it open what the relevant psychological state is. I have said already that it must be conscious, in the sense that it is available for the control of rational action, but nothing in what follows will depend upon whether it is knowledge, belief, or something else still.

Embedding Davidson's proposal within the cognitive conception already gives us some headway with the Foster problem. On this view, after all, what one takes an utterance to mean is, in effect, a matter of what one believes the speaker takes its truth-condition to be, so not just any T-sentence one knows for the utterance will equally serve to interpret it. One might well know oneself, for example, that

- (9) 'Bedrock is cold' is true iff Bedrock is cold and arithmetic is incomplete,

but if one does not think that Fred knows (9), one will not use one's knowledge of (9) to interpret his utterances of 'Bedrock is cold'. But, of course, that's far from sufficient progress. Perhaps one met Fred at a logic conference and so presume that he does know (9). And even if Fred isn't a logician, maybe he was paying attention in his basic science courses, in which case he presumably knows

- (10) 'Bedrock is cold' is true iff Bedrock is cold and water is H₂O.

And even if he wasn't paying attention in science class, surely Fred must know

- (11) 'Bedrock is cold' is true iff Bedrock is cold and water is wet.

There are plenty of similar cases.³⁵

Higginbotham has suggested that, though the Foster problem shows that "meaning does not reduce to reference, . . . knowledge of meaning reduces to the norms of knowledge of reference" Higginbotham (1991, p. 274). These "norms" concern what a competent speaker of the language is supposed to know in so far as she is competent. So, very roughly:

- To know what a sentence *S* means is: (i) to know that *S* is true iff *p* and (ii) to know that knowing that *S* is true iff *p* is what, and all, competence requires a speaker to know about it.

For example, to know what 'Bedrock is cold' means is to know (2) and to know that (2) is what, and all, a competent speaker is supposed to know about it. If so, then since one is not, in general, expected to know (9), (9) does not, in any sense, 'give the meaning' of 'Bedrock is cold'.

It is unclear, however, what Higginbotham means by what one is expected to know 'in so far as one is a competent speaker'. Is one expected to know

- (12) "All bachelors are married" is true iff all bachelors are female

in so far as one is competent? On some views, the answer would arguably be that one is. "All bachelors are unmarried", the thought might be, is analytically false: Anyone who understands it must know that it is false, and so perhaps one must know that it is false if one is competent. But "All bachelors are female" is also analytically false, so maybe one is expected to know (12) simply in so far as one is a competent speaker. Perhaps not, but it is not obvious there are no similar counterexamples.

But even if Higginbotham's proposal is extensionally adequate, it is not clear how much explanatory value it has. Most competent speakers of

³⁵ Note that since (11) is presumably common knowledge if 2 is, it will not suffice to strengthen (7) by inserting a reference to common knowledge.

English do know (11), and they expect other competent speakers to know it, too. It is natural to say that they do not know (11) in the same way they know (2); that their knowledge of (11) depends somehow upon their knowledge of (2); and that their knowledge of (2) is ‘linguistic’ in a way that their knowledge of (11) is not. We might summarize these intuitions by saying that competent speakers of English know (2), but not (11), simply as competent speakers of English (Higginbotham, 1991, p.274). But talk of what one knows simply as a speaker is obscure, and one would like to have these intuitions explained, not just labeled. Such an explanation is what I shall now attempt.

Utterances are acts. As such, they are typically performed for reasons. Fred’s uttering ‘Bedrock is cold’ is an act, and one typical reason for uttering it might be to communicate, to Barney, that Bedrock is cold. But Fred’s utterance of this sentence is rational, as a means to this end, only if he is assuming certain facts about Barney, for example, that Barney will take him at his word. Of special interest to us, however, is the need for Fred to assume that Barney will understand him as having said that Bedrock is cold. If Barney instead took him to have said that Bedrock is broke, Fred’s attempt to communicate to Barney that Bedrock is cold would fail. So if, as I am assuming, Barney’s understanding Fred’s utterance consists in his knowing something about it, we may now ask: What is it that Fred is assuming Barney knows?

Earlier, I suggested that, if an irresponsible radical interpreter were simply to settle upon a theory of truth and set about using it to interpret the natives, her doing so would amount to her assigning certain meanings to their utterances. Suppose Fred regards Barney as similarly irresponsible: Despite the fact that Barney knows that Fred knows many correct T-sentences for ‘Bedrock is cold’, Barney has nonetheless decided to use a particular one of them to interpret Fred’s utterances. If Fred believed that the one on which Barney had settled was (2), then Fred’s uttering ‘Bedrock is cold’ in an effort to communicate to Barney that Bedrock is cold would make perfect sense: If Barney takes Fred at his word, that is precisely the belief he will acquire. It may well be that Fred also expects that Barney believes (11), at least implicitly. But unless Fred were prepared to suppose that Barney

would deploy that belief in interpreting his utterance, it would be irrational for him to utter “Bedrock is cold” in an effort to get Barney to believe, in response to his utterance, that Bedrock is cold and water is wet. Of course, Fred might suppose that, if he can get Barney to believe that Bedrock is cold, he can thereby get him to believe that Bedrock is cold and water is wet. But Fred’s reasons for uttering the sentence, in this case, are just different from the sorts of reasons he has if he assumes that Barney will deploy (2) in interpreting his speech.

The lesson that is supposed to emerge from this fantasy is that one way to distinguish interpretive T-sentences, such as (2), from non-interpretive T-sentences, such as (11), is simply by reference to the roles they play in the production and comprehension of speech: A speaker treats a given T-sentence as ‘giving the meaning’ of a sentence if she *deploys* that T-sentence in a certain distinctive way, both when speaking herself and when interpreting the speech of others. For example, we may reconstruct Fred’s reasons for uttering ‘Bedrock is cold’ roughly as follows:

- (13) I want Barney to know that Bedrock is cold.
- (2) “Bedrock is cold” is true iff Bedrock is cold.
- (14) Barney knows (2), and he will deploy that knowledge in interpreting my speech.
- (15) If I utter “Bedrock is cold”, in the present context, Barney will know (or believe) that I have spoken the literal truth.
- (16) If I utter “Bedrock is cold” in the present context, Barney will (be in a position to) come to know that Bedrock is cold. (From (14) and (15).)
- (17) I shall utter Bedrock is cold. (From (13) and (16).)

That (2) ‘gives the meaning’ of “Bedrock is cold”, for Fred, simply consists in his using it in the way here illustrated, and in the following sort of way when interpreting Barney’s utterances:

- (18) Barney has uttered “Bedrock is cold”.

- (2) “Bedrock is cold” is true iff Bedrock is cold.
 (19) Barney knows (2), and he expects me to deploy that knowledge in interpreting his utterance.
 (20) Barney is speaking the literal truth.
 (21) Bedrock is cold. (From (18), (20), and (19).)

One could obviously go on in this way for some time, exploring the role knowledge of (2), and mutual expectations about its deployment, play in the many different forms of reasoning in which speakers engage.

Note that Fred’s expectation that Barney will deploy the belief that his utterance of “Bedrock is cold” is true iff Bedrock is cold is an expectation that Barney will suppose that Fred himself has deployed the same belief in speaking. So Fred expects Barney to expect him to have deployed this belief in speaking, and we could continue the iterations. The expectations of the participants in a given communicative exchange are thus, at least when all goes well, *mutual*, not just in the sense that they are shared but in the stronger sense that they are intertwined: Each conversational participant’s expectations concern, *inter alia*, the expectations of the others. As Higginbotham notes regarding a similar feature of his own proposal (Higginbotham, 1991, p. 274), this fact reflects the same sort of phenomena that lead to Grice’s emphasis on the overttness of communicative intentions (Grice, 1989a, p. XXX) and to Lewis’s emphasis on common knowledge (Lewis, 1986, p. XXX).

The suggestion, then, is that linguistic knowledge is knowledge that is *used* in a certain distinctive way: It is the knowledge one deploys in speech and interpretation. The notion of how one deploys information in speech and interpretation is thus the crucial one. I take the notion to be explained implicitly by the role it plays in the sorts of reasoning we have just been discussing: A speaker has deployed the belief that *S* is true iff *p* in speaking if she uses it in the way Fred uses his knowledge of (2) in deciding to utter “Bedrock is cold”; an interpreter deploys her belief that *S* is true iff *p* in interpreting an utterance if she uses it in the way Barney uses his knowledge of (2) in interpreting Fred’s utterance of “Bedrock is cold”. But of course it would hardly be unreasonable to ask that the implicit be made explicit.

Unfortunately, I can’t seem to do that quite yet.³⁶ So let me instead suggest a way of approaching the issue from a slightly different direction.

As mentioned earlier, there is a strong intuition that knowledge of (9) and (11) depend in some way upon knowledge of (2), that is, that (2) is, in some sense, one’s most basic piece of semantic knowledge regarding the sentence “Bedrock is cold”. So one might hope to characterize linguistic knowledge in terms of this notion of a speaker’s most basic semantic knowledge. One way to do so would begin with the thought that our most basic linguistic knowledge has a distinctive *source*.

In any event, the question begs to be asked what the source of our conscious linguistic knowledge is, be it knowledge of truth-conditions or knowledge of meaning. The sorts of phenomena that are used to motivate the principle of compositionality suggest, though they do not of course imply, that, as Dummett puts it, “a speaker of a language derives his understanding of any sentence of that language from his knowledge of the meanings of the words” that compose it Dummett (1993f, p. 36). Of course, such a derivation is not conscious, so the proposal would have to be that our conscious linguistic knowledge is the product of subpersonal computational processes that draw upon the information contained in tacitly known theories of truth. If so, then the knowledge we deploy in speech and interpretation is, in the normal case, delivered by the language faculty. It is, of course, an empirical question what knowledge is so delivered, in any particular case. But it is not unreasonable to suppose that (2) is so delivered and (11) is not.

McDowell emphasizes, throughout his writings on such topics, that we do not simply hear the words other people utter and then add an interpretation in thought McDowell (1998a, pp. 331ff). Rather, in the normal case, our appreciation of what someone has said is a result of our perceiving not just the words she has uttered but also, in the strictest sense, perceiving what she has said. If so, then one’s interpretation of speech is not always a matter

³⁶ Crispin Wright suggested that, for the case of the speaker, one might appeal to some notion of the intention that is most immediately controlling the act of uttering the sentence. Once spelled out, that is probably close to a proposal of Rumfitt’s. Adapted to the present framework, it is that the T-sentences that constitute understanding are those that occur in what he calls the ‘lowest practical syllogism’ that occurs in practical reasoning of the sort we’ve been discussing Rumfitt (1995, p. 842).

of what information one *chooses* to deploy but is typically a matter of what information just does get deployed in one's perception of speech. One should not, however, suppose that such an observation shows that higher cognitive states, like beliefs, have no central role to play in speech and interpretation.³⁷ If one happens to know that Jones uses a particular word—say, 'livid'—differently from how one uses it oneself, then, although one may still hear Jones as saying that Smith was pale when he utters "Smith was livid", one may choose not to deploy that information in interpreting him. Rather, since one knows, *via* some other route, that, when Jones utters "Smith was livid", he is not himself deploying a belief that it is true iff Smith was pale but rather a belief that it is true iff Smith was flushed, one deploys that information instead. Of course, I usually just proceed on the assumption that Jones's words mean to him what they mean to me. But so to proceed is, in effect, to take appearances at face value and so to act on the perceptually justified belief that Jones's utterance is true iff Smith was pale.

Ultimately, or so I suspect, an adequate answer to the question what linguistic knowledge is will have to make reference both to facts about how that knowledge is put to use and to facts about its normal source. Unfortunately, I must confess that I am unable at present to say anything that is likely to satisfy someone skeptical that an adequate answer can be sewn from the materials we have been discussing. In my defense, I plead that the question is famously difficult one. The proposal I am making is at least not already known to fail.

³⁷ McDowell sometimes gives the impression that, epistemologically, it is of no particular interest that we do not *just* perceive what someone has said but also perceive her words. Tyler Burge has argued for a similar view, namely, that one's justification for the claim that someone has said that *p* need not involve any claim about what words she has uttered Burge (1993). But the possibility of the sort of re-interpretation I consider here depends upon our perceiving someone's words, and we do not perceive someone's words and what she has said as one can perceive an apple and an orange: We perceive what has been said *in* the words somehow, and so our perception of speech is itself 'semantic', in the sense that in it meaning binds to sound. Precisely what that might mean is a very nice question. See Heck (2005a) for more on this issue.

6 Truth and Semantic Knowledge

The foregoing constitutes a principled, if still incomplete, solution to the Foster problem: We do have reason to believe that a speaker who consciously knew the truth-conditions of sentences of a given language (as uttered, or potentially uttered, in given contexts), and who was prepared to deploy that knowledge in speaking and interpreting others, would understand, and be able to use, that language, if the truth-conditions she assigned were the same as those assigned by other speakers.

The arguments so far given do not, however, show that all competent human speakers *do* have this sort of knowledge. And Soames, for one, doubts that they do. In particular, he thinks we have no reason to suppose that all competent speakers possess a meta-linguistic concept of truth. Unless someone does grasp such a concept of truth, she will be in no position to have conscious beliefs about the truth-conditions of her sentences: One cannot believe that "snow is white" is true iff snow is white (or grass is green, or anything else) unless one grasps a meta-linguistic concept of truth, one that applies to utterances, sentences as (potentially) uttered in contexts, or some such thing. There is, of course, no reason to suppose that every competent speaker must understand a word synonymous with "true": Indeed, it is likely that not every competent speaker does. So the strongest defensible claim is that competent speakers *implicitly* grasp a meta-linguistic concept of truth.

Soames doubts that claim can be defended, however:

Must [young children] possess a metalinguistic concept of truth, even though they have no word for it? I don't see why. Perhaps it will be suggested that a person who lacked such a concept couldn't be a language user, since to use language one must realize that assertive utterances aim at truth. . . . But this suggestion is confused. The child will get along fine so long as he knows that 'Momma is working' is to be assertively uttered only if Momma is working; 'Daddy is asleep' is to be assertively uttered only if Daddy is asleep; and so on. . . . [I]t is not necessary to have [a metalinguistic concept of truth] in order to learn and understand a

language. (Soames, 1988, p. 189)

So the idea would seem to be that, rather than having to know the truth-conditions of a given sentence, all a speaker need know about a sentence, to understand it, is when it can appropriately be uttered. In particular, to understand the sentence “Momma is working”, it is enough to know what I shall call the *semantic injunction* for it: that it is to be uttered only if Momma is working. Since no meta-linguistic concept of truth is involved here, competent speakers need not have any such concept.

It is worth noting, before we continue, that Soames’s view is not an alternative to but a version of the cognitive conception of (occurrent) understanding. Like my view, his is committed to the claim that a speaker’s understanding of a sentence consists in knowledge or beliefs that she has about it. On Soames’s view, these beliefs are of the form: One is to utter *S* only if *p*, whereas on my view they are of the form: *S* is true iff *p*. If Soames is right, then, we need not abandon the view that to understand a sentence *S* is to know something about it nor even that it is to know what *S* means: We should say, rather, that to know what *S* means is to know the condition under which it may be uttered, either by oneself or by someone else. Knowledge of meaning, on this view, is thus knowledge of semantic injunctions, rather than knowledge of truth-conditions. And so, as said, Soames and I disagree not about whether competence with a language is constituted by linguistic knowledge but about what the content of that knowledge is.

Something similar is true of David Lewis’s view that the semantic properties of a language are fixed by conventions regarding its use (Lewis, 1985a, 1986). For it to be a convention in *P* that *S* is uttered only when *q*, members of *P* must intend to utter *S* only when *q*, expect one another to do so, know that others expect them to do so, and have such intentions for certain sorts of reasons. These conditions concern what members of *P* *intend* and what they *expect*. These intentions and expectations need not, of course, always or even often be occurrent. Nonetheless, on Lewis’s view, the ability to use a language in accord with the relevant conventions involves having certain sorts of psychological attitudes that are conscious in the sense that they are available for the control of rational action, such states including, for example, the standing intention to utter *S* only when *p*. One

could perhaps avoid this conclusion by abandoning the assumption that a regularity is a convention only if parties to the convention have certain sorts of intentions and so forth. To do so would, however, in effect be to abandon what, in Heck (2005a), I call the ‘propositional rationality of speech’, that is, the claim that speech is intentional under descriptions of the form *saying that p*. That claim is not optional.

Let me return, then, to the question whether every competent speaker must implicitly grasp a meta-linguistic concept of truth. I simply do not find it as peculiar as Soames apparently does to attribute such a grasp even to young children. Every competent speaker must have a conception of an utterance’s being correct or incorrect. Without such a conception, one could hardly be rational in one’s use of language: One would be unable even to conceive of the possibility of an utterance’s being incorrectly made; one therefore could not so much as make sense of the idea of evaluating what was said in order to decide whether one should believe it. Now, in a sense, of course, there are many ways in which an utterance can be incorrect: There is a very broad sense of incorrectness, for example, in which an utterance is incorrect if it is rude, inappropriate, or irrelevant. But it is hardly unreasonable to suppose that there is a more fundamental notion of an utterance’s being *objectively* correct that underlies the undifferentiated conception of correctness,³⁸ and even quite young children seem to understand that, although Daddy said that he has only one cookie, he may in fact have more. Of course, these claims could be questioned. But it is not hopelessly *ad hoc* to suggest that our implicit grasp of the concept of truth lies in the conception of an utterance’s objective correctness that competence presupposes.

³⁸ I am inclined to think, myself, that rational thought demands deployment of a corresponding conception of a belief’s being objectively correct. I shall not try to justify this metaphysical presumption here. For an attempt to do so, see Dummett (1993c). It is perhaps worth noting that the objectivity required here is very weak: It is not, in particular, distinctive of realism, in Dummett’s sense. All that is required is that there should be a difference between what is true and what is appropriate, relevant, or justified, as opposed to justifiable in principle. Dummett argues, in quite general terms, that at least this much objectivity is required of any notion of truth, be it realist or anti-realist. See Dummett (1978a, pp. XXX-XX) for the argument. A Dummettian anti-realist thus accepts that truth is objective in the weak sense that is relevant here. (Indeed, what I am arguing is, as Dummett would put it, precisely that truth must be the central notion of the theory of meaning in the ‘weak’ sense.)

Now, it is clear enough that no (meta-linguistic) notion of truth is needed to state semantic injunctions: Knowledge of semantic injunctions does not demand grasp of a concept of truth, but only of such concepts as that of an utterance and of normative notions that are central to practical reason in general. It is for this reason that Soames regards possession of the concept of truth as unnecessary for competence. But there is a subtle illusion here: Even if the *statement* of a semantic injunction—one should utter *S* only if *p*—does not require use of a notion of truth, its *application* does, and there is reason to suppose that the statement of the injunction actually involves such a notion, as well. The notion in question is, once again, that of an utterance's being objectively correct.

Consider again Fred's utterance that Bedrock is cold. How is Barney to reason to the conclusion that Bedrock is cold if Soames is right about what Barney, *qua* competent speaker, knows about Fred's utterance? If we replace (2), from our earlier discussion of such cases, with a semantic injunction, then we have the following:

(2*) In the present context, one is to utter "Bedrock is cold" only if Bedrock is cold.

(18) Fred has uttered "Bedrock is cold".

(19*) Fred knows (2*) and deployed this information is speaking.

(20) Fred is speaking the literal truth.

(21) Bedrock is cold.

But now (21) does not follow, since (2*) says nothing about truth. We need to replace (??) with something like:³⁹

³⁹ Exactly how (??*)—and the rest of this reasoning—should be formulated is presumably open to discussion, and one might worry that the discussion to follow turns on special features of how I have formulated it here. But it does not, as a little experimentation should show. For example, one might want to replace (??*) and (19*) with something like: Fred intends in uttering "Bedrock is cold" to utter it only if Bedrock is cold. Then (??*) would become something like: Fred's satisfied the relevant intention in uttering "Bedrock is cold". And now an utterance can be regarded as true if, and only if, the utterer satisfies the relevant linguistic intention in uttering it.

(20*) In uttering "Bedrock is cold", Fred has satisfied (what he takes to be) the semantic injunction associated with that utterance, viz., (2*).

Now (21) follows, for Fred has satisfied the injunction if, and only if, Bedrock is cold.

A speaker thus cannot make use of semantic injunctions in interpreting the speech of others unless she has and deploys the concept of *an utterance that satisfies the appropriate semantic injunction*. But such a concept just is a concept of an utterance's being objectively correct. Even though Bedrock is indeed cold, Fred's uttering "Bedrock is cold" may have failed to satisfy all sorts of other conditions, conventional and otherwise, regarding when one is to utter it: Fred's utterance may have been rude, irrelevant, or unjustified: The semantic injunction is not that one is to utter *S* if and only if *p* but only that one is to utter it only if *p*. That one has satisfied the semantic injunction for the sentence one has uttered implies that one's utterance is correct only in a certain limited sense, namely, that it was objectively correct.

Indeed, the concept of an utterance that satisfies its associated semantic injunction simply is a concept of (literal) truth for utterances. Let us make the following definition:

(22) N speaks the literal truth in uttering *S* \equiv *df* N utters *S* and in doing so satisfies the semantic injunction for *S*.

Now suppose the semantic injunction for *S* is to utter it only if *p*. Then if N does utter *S*, he will satisfy the injunction iff *p*. Hence:

(23) If the semantic injunction for *S* is that it should be uttered only if *p*, then, if N utters *S*, N thereby speaks the literal truth iff *p*.

Given the semantic injunction for a given sentence, we can thus derive a T-sentence for utterances of it from (22).

In a sense, then, semantic injunctions are equivalent to statements of truth-conditions: As we have just seen, statements of truth-conditions are easily derived from semantic injunctions. Conversely, semantic injunctions

are derivable from statements of truth-conditions. If S is true iff p , and if one intends to speak the literal truth when uttering S , then one should utter it only if p . One might suggest, then, that there is not much at issue between Soames and myself: We could debate whether semantic injunctions or truth-conditions are more fundamental, but what really hangs on who wins this debate? In fact, however, Soames's view faces a serious problem, one that is revealed by the need, two sentences back, to restrict attention to utterances intended to be literally true.

Understanding cannot consist in knowledge of semantic injunctions, because there are no semantic injunctions for English sentences. It just *isn't* a convention of English that one is to utter "Bedrock is cold" assertively only if Bedrock is cold: There is no convention of truthfulness, in Lewis's sense.⁴⁰ One can speak in such a way that one's audience will recognize one's communicative intentions only by recognizing one to have intended to speak literally falsely: Irony and hyperbole are like that. Perhaps more commonly, one can utter a sentence intending to communicate something other than what one thereby literally says without caring whether one's utterance is literally true: Metaphor is perhaps like that. Now, in so far as irony, hyperbole, and metaphor are themselves conventional forms of communication, one might plausibly argue that the semantic injunctions for the associated *sentences* shouldn't apply to them but rather other, related semantic injunctions: For example, one should utter "John is a fine friend" ironically only if John is *not* a fine friend (Lewis, 1985a, pp. XXX-XX). But the phenomenon to which I am drawing attention is not limited to such cases, and the familiar example of the letter of recommendation can be made to illustrate this point. When Prof. Smith writes only that Jones has always been on time for tutorials, thereby implicating and so communicating that Jones is a poor candidate for graduate school, it matters not at all whether what Smith *says* is literally true. Perhaps not even that much is true of Jones, but Smith figured he had to say something nice and that was the closest he could get. He doesn't intend to communicate anything about Jones's punctuality, and he knows his audience won't care to know about it. Are we really to believe that such an utterance violated the linguistic

conventions that give our words meaning? I see no plausibility whatsoever to that claim, and the alternative mentioned for the case of irony does not exist here: It is no *convention* that Smith's utterance means what it does in these circumstances.⁴¹

At best, then, there is conditional convention of *literal* truthfulness: If one intends to speak the *literal* truth, then one should utter "Bedrock is cold" only if Bedrock is cold. One might not think the difference would matter. Such 'literal injunctions', the thought might be, state the literal meanings of the sentences they concern, and what matters here is that speakers should know the literal meanings of their sentences. But that is mistaken. The literal injunction for "Bedrock is cold" is a hypothetical imperative: It tells one when the sentence may be uttered *if* one intends to speak the literal truth. It does not tell one anything about utterances of sentences that are not intended as utterances of literal truths. To know the literal injunction for "Bedrock is cold" is thus at best to know what it means when it is uttered with the intent to speak the literal truth. But that is not to know what its literal meaning is, since a sentence literally means what it does whether its utterer intends to speak the literal truth or not.

It is tempting to reply that, while of course not all utterances are intended to be literally true, the account we are considering just involves a harmless idealization. Of course, one might say, something needs to be said about speech that is not intended to be literally true, but that account can be built on top of an account of speech that is intended to be literally true. The fact that literal meaning is more fundamental than non-literal meaning makes it legitimate to focus, in the first instance, on literal speech. But this reply equivocates. As just noted, the idealization consists in focusing on *speech* that is intended to be literally true and temporarily setting aside *speech* that is not. To justify the idealization, one therefore needs to claim that

⁴¹ Examples of so-called "loose talk" may also be relevant here: Utterances of "It is three o'clock" often do not communicate that it is *exactly* three o'clock; but it is no violation of linguistic convention to have uttered a sentence that is literally false in order to communicate something that is true. It is controversial, in fact, whether loose talk should be explained in terms of implicature or rather in terms of pragmatic processes that affect what is said. See [REF] for discussion. My own view, however, is that implicature has not yet been shown inadequate to such cases.

⁴⁰ Similar remarks would apply, *mutatis mutandis*, to any alleged convention of trust.

literal *speech* is more fundamental than non-literal *speech*. And while I am certainly prepared to agree that literal *meaning* is more fundamental than any kind of non-literal *meaning*—be it the sort that derives from implicature, metaphor, irony, or what have you—it simply does not follow that literal *speech* is more fundamental than non-literal *speech*. I see no reason to believe this further claim.

Lewis's suggestion that there is a convention of truthfulness to which language-users are party is a development of an intuition that many philosophers share, namely, the intuition that, in some sense, language could not be used for communication unless one were entitled to rely upon one's fellows. It is important to understand that Lewis does not think it is a *convention* that one is to speak truly. Among the many objections Lewis considers in "Languages and Language" is that, by his own lights, part of what makes a regularity conventional is the existence of alternatives, and there is simply no alternative to speakers' generally being truthful: If they weren't, their words couldn't have any meaning at all (Lewis, 1985a, p. XXX). Lewis grants the argument in its entirety. But what he is claiming is conventional is not speakers' being truthful *simpliciter*: It is their being truthful in a particular language \mathcal{L} . It is not a convention that we should utter sentences that are true in whatever language we happen to speak: That, Lewis thinks, has to be so if our language is to be meaningful at all.⁴² But if \mathcal{L} is the language we happen to speak, then, Lewis insists, it *is* a convention that we should utter a sentence assertively only if it is true in \mathcal{L} : There is an alternative to that possibility, since we might have spoken \mathcal{L}' instead.

Lewis's strategy is, in effect, to use the *non-conventional* regularity he believes must obtain if language is to be meaningful at all as a kind of boundary condition and to solve for an unknown, namely, the language we happen to speak. We know that speakers must strive to utter truths in whatever language they speak: If we can figure out under what circumstances they regard it as permissible to utter a sentence S , then we can figure out when they regard S as being true and so what they take it to mean and hence what language they speak.⁴³ Note, however, that the strategy presupposes that

speakers strive to utter sentences that are *literally* true in whatever language they speak. Without that presumption, we cannot infer from the fact that speakers regard it as permissible to utter S only if p that S means that p in their language. And, it seems to me, this presumption should be rejected. To reject it, however, one need not reject the core intuition that communication is possible only if one is entitled to rely upon one's fellows. One need only deny—to use Grice's terminology—that the core intuition concerns what is *said* as opposed to what is *communicated* or *meant*. That is: Grant that there is some sort of requirement that people speak truly. Now does the requirement concern what they *say*? or what they *mean*? The answer seems obvious: It concerns what is meant. There is no obvious reason that one must, in general, be able to trust what one's fellows strictly and literally *say*. All that is necessary is that one should, in general, be able to trust what they *mean*. Of course, what one says is often part of what one means, and so one would expect speakers to be literally truthful much of the time, but there is no general requirement that speakers should be literally truthful all or even most of the time.

The issue we have been discussing concerns explanatory priority: The question is whether literal injunctions or statements of truth-conditions are more fundamental. Now, admittedly, I have not proven that one cannot regard literal injunctions as basic and somehow explain how they determine the literal meaning even of utterances that are not intended as literal truths. But I do hope to have done enough to make it seem unlikely that this can be done. The prospects seem bright only so long as one neglects certain distinctions, such as the distinction between what is said and what is meant, and the distinction between literal meaning and literal speech. Moreover, the converse derivation, of literal injunctions from truth-conditions, just seems like the right story: On that story, the literal injunction for a given utterance is a consequence of what the uttered sentence literally means and conventions governing speech that is intended to be literally true; corresponding injunctions for speech that is *not* (simply) intended to be literally true are in turn consequences of what the literal meaning of the uttered sentence is and the facts, whatever they might be, about how non-literal meaning is determined by literal meaning and the environment in which the

⁴² Note that Lewis's view actually entails this claim. CONTINUE.

⁴³ Davidson as well. Quine? Lots and lots of people, to be sure.

utterance is made.⁴⁴

If we can, then, we should distinguish facts about literal meaning from facts about the norms that govern literal speech: It would be best if we could say what a sentence literally means without adverting to speech intended to be literally true—or, for that matter, to any other specialized form of speech. That, of course, is precisely what the view I am defending allows us to do.⁴⁵

7 Conclusion

The view defended here has concerned only the semantic properties of sentences. As noted at the outset, however, part of what makes truth-theoretic semantics attractive is that it promises to make a substantive explanation of how the meanings of parts contribute to the meanings of wholes possible. For all that has been said here, however, natural languages might as well not be compositional. As I said earlier, however, it would certainly be in the spirit of this view to hold that what our conscious knowledge of the truth-conditions of utterances is the product of our tacit knowledge of a truth-theory. But what is tacit is, by definition, unconscious: Our knowledge of the semantic properties of whole sentences could be generated from tacit knowledge of the semantic properties of their parts even if we did not *consciously* recognize sentences as having parts at all. But we do consciously recognize sentences as having parts: Our knowledge regarding the meaning of a sentence thus includes but is not limited to knowledge that it is true iff *p*; somehow it also includes an appreciation of the sentence's articulation into parts and knowledge of what those parts mean. On the other hand, however, it cannot of course include knowledge of the sentence's syntactic structure, in anything like the sense in which a linguist would think of it. And so one question that is left open here is how we should conceive our conscious knowledge of meaning so as to take due notice of the fact that

⁴⁴ [[FIXME]] What I do NOT mean by that.

⁴⁵ Obviously, the truth-conditional aspect of my view is not what is responsible for this benefit. What is responsible for it is, rather, the idea that competence rests upon knowledge of meaning.

compositionality is not, as it were, entirely a tacit phenomenon.⁴⁶

References

- Burge, T. (1993). 'Content preservation', *Philosophical Review* 102: 457–88.
- (1999). 'Comprehension and interpretation', in L. Hahn (ed.), *The Philosophy of Donald Davidson*. Chicago, Open Court, 229–250.
- Chomsky, N. (1969). 'Quine's empirical assumptions', in D. Davidson and J. Hintikka (eds.), *Words and Objections: Essays on the Philosophy of W.V. Quine*. Dordrecht, Reidel, 53–68.
- Davidson, D. (1984a). 'Belief and the basis of meaning', in Davidson 1984b, 141–54.
- (1984b). *Inquiries Into Truth and Interpretation*. Oxford, Clarendon Press.
- (1984c). 'Radical interpretation', in Davidson 1984b, 125–139.
- (1984d). 'Reply to Foster', in Davidson 1984b, 171–179.
- (1984e). 'Theories of meaning and learnable languages', in Davidson 1984b, 1–15.
- (1984f). 'Thought and talk', in Davidson 1984b, 155–70.

⁴⁶ My work on these issues has benefitted enormously from conversations over the years with Jim Higginbotham, Michael Rescorla, and Jason Stanley. Material from this paper was presented during a series of seminars I gave at the University of St Andrews in January and February 2004 as a British Academy Visiting Professor and then again in a graduate seminar at Harvard University in Spring 2004. Thanks to all who attended and participated for their comments and encouragement, especially Jake Beck, Tony Corsentino, Bernard Nickel, Daniel Nolan, Agustin Rayo, and Crispin Wright. Thanks to the British Academy and to Arché, the AHRC Research Centre for the Philosophy of Logic, Language, Mathematics and Mind, for their support, which is much appreciated, and to Crispin for arranging the visit.

- (1984g). ‘Truth and meaning’, in Davidson 1984b, 17–36.
- Dummett, M. (1978a). ‘The justification of deduction’, in Dummett 1978c, 290–318.
- (1978b). ‘The significance of Quine’s indeterminacy thesis’, in Dummett 1978c, 375–419.
- (1978c). *Truth and Other Enigmas*. London, Duckworth.
- (1991). *The Logical Basis of Metaphysics*. Cambridge MA, Harvard University Press.
- (1993a). ‘Language and communication’, in Dummett 1993b, 166–187.
- (1993b). *The Seas of Language*. Oxford, Clarendon Press.
- (1993c). ‘The source of the concept of truth’, in Dummett 1993b, 188–201.
- (1993d). ‘What do I know when I know a language?’, in Dummett 1993b, 94–105.
- (1993e). ‘What is a theory of meaning? (I)’, in Dummett 1993b, 1–33.
- (1993f). ‘What is a theory of meaning? (II)’, in Dummett 1993b, 34–93.
- Evans, G. (1985). ‘Understanding demonstratives’, in *Collected Papers*. Oxford, Clarendon Press, 291–321.
- Foster, J. (1976). ‘Meaning and truth-theory’, in G. Evans and J. McDowell (eds.), *Truth and Meaning: Essays in Semantics*. Oxford, Oxford University Press, 1–32.
- Frege, G. (1962). *Grundgesetze der Arithmetik*. Hildesheim, Georg Olms Verlagsbuchhandlung. Translations of Part I are in Frege 1964, and of portions of Part III, in Frege 1970. Extracts appear in Frege 1997, 194–223.
- (1964). *The Basic Laws of Arithmetic: Exposition of the System*, ed. and tr. by M. Furth. Berkeley CA, University of California Press.
- (1970). *Translations from the Philosophical Writings of Gottlob Frege*, Geach, P. and Black, M., eds. Oxford, Blackwell.
- (1984). ‘On sense and meaning’, tr. by M. Black, in B. McGuinness (ed.), *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford, Basil Blackwell, 157–77. Also in Frege 1997, 151–71.
- (1997). *The Frege Reader*, Beaney, M., ed. Oxford, Blackwell.
- Grice, H. P. (1989a). ‘Meaning’, in Grice 1989b, 213–223.
- Grice, H. (1989b). *Studies in the Ways of Words*. Cambridge MA, Harvard University Press.
- Heck, R. G. (1995). ‘The sense of communication’, *Mind* 104: 79–106.
- (1998). ‘Grundgesetze der Arithmetik I §§29–32’, *Notre Dame Journal of Formal Logic* 38: 437–74.
- (1999). ‘Grundgesetze der Arithmetik I §10’, *Philosophia Mathematica* 7: 258–92.
- (2004). ‘What is compositionality?’, *Manuscript*.
- (2005a). ‘Reason and language’, in C. MacDonald and G. MacDonald (eds.), *McDowell and His Critics*. Oxford, Blackwells, 22–45.
- (2005b). ‘Use and meaning’, in L. Hahn (ed.), *The Philosophy of Michael Dummett*. Chicago, Open Court.
- (2006). ‘Idiolects’, in A. Byrne and J. J. Thomson (eds.), *Content and Modality*. Oxford, Oxford University Press, 61–92.
- (2007). ‘Frege and semantics’, *Grazer Philosophische Studien* 75: 27–63.
- Higginbotham, J. (1991). ‘Truth and understanding’, *Iyyun* 40: 271–288.

- Larson, R. and Segal, G. (1995). *Knowledge of Meaning*. Cambridge MA, MIT Press.
- Lewis, D. (1985a). ‘Languages and language’, in Lewis 1985b.
- (1985b). *Philosophical Papers Volume I*. Oxford, Oxford University Press.
- (1985c). ‘Radical interpretation’, in Lewis 1985b, 108–18.
- (1986). *Convention*. Cambridge MA, Harvard University Press.
- McDowell, J. (1998a). ‘Anti-realism and the epistemology of understanding’, in McDowell 1998c, 314–343.
- (1998b). ‘In defense of modesty’, in McDowell 1998c, 87–107.
- (1998c). *Meaning, Knowledge, and Reality*. Cambridge MA, Harvard University Press.
- (1998d). ‘On the sense and reference of a proper name’, in McDowell 1998c.
- Quine, W. V. O. (1960). *Word and Object*. Cambridge MA, MIT Press.
- (1972). ‘Methodological reflections on current linguistic theory’, in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*. Dordrecht, Reidel, 442–454.
- Richard, M. (1992). ‘Semantic competence and disquotational knowledge’, *Philosophical Studies* 65: 37–52.
- Rumfitt, I. (1995). ‘Truth-conditions and communication’, *Mind* 104: 827–62.
- Soames, S. (1988). ‘Semantics and semantic competence’, in S. Schiffer and S. Stelle (eds.), *Cognition and Representation*. Boulder CO, Westview Press, 185–207.
- (1992). ‘Truth, meaning, and understanding’, *Philosophical Studies* 65: 17–35.
- Strawson, P. F. (1971). ‘Meaning and truth’, in *Logico-Linguistic Papers*. London, Methuen, 170–89.
- Tarski, A. (1958). ‘The concept of truth in formalized languages’, in J. Corcoran (ed.), *Logic, Semantics, and Metamathematics*. Indianapolis, Hackett, 152–278.