REALISM AND ANTI-REALISM IN DAVIDSON'S PHILOSOPHY OF LANGUAGE, II

FREDERIC STOUTLAND
St. Olaf College

Davidson as Anti-Realist

1. As an anti-realist Davidson rejects the claim that our sentences are true or false *in virtue of* extra-linguistic objects. Although his theory of meaning (like the realist theory) presupposes a conception of truth, that conception is not realist. What is presupposed is, to put it crudely, simply the extension of "true" for the language we speak: the range of sentences which our linguistic community holds true —either generally if they are standing sentences, or paired with an occasion if they are occasion sentences.[1] Any account of that *in virtue of* which a sentence *is* true (and I do not mean merely an account of our evidence for *thinking* a sentence is true) will refer to other sentences held true (i.e. will refer to evidence), so that this notion of "in virtue of which" must be emptied of realist connotations. ". . . all the evidence there is is just what it takes to make our sentences or theories true. Nothing, however, no *thing* makes sentences and theories true: not experience, not surface irritations, not the world, can make a sentence true." (VICS, p. 16)[2] Although (most) true sentences are "about" extra-linguistic objects, they are not true in virtue of such objects, but in virtue of their fitting in (cohering) with other sentences held true by competent speakers of the language.

This *overall* anti-realism allows for *internal* realism, however, for having a Tarski-type analysis of the structure of our sen-

1 The terms, of course, are Quine's. Cf. *Word and Object*, Chapter II.
2 Abbreviations for Davidson's works are listed in the Bibliography.

tences enables us to filter out the referential apparatus of our language –the variables, quantifiers, and singular terms– and allows for the claim that insofar as we accept sentences which have this structure, we necessarily accept the objects referred to.[3] But, again, it is not in virtue of the objects referred to by referring expressions that our sentences are true (nor by the objects being related as they are); it is rather that having held true many sentences we *thereby* accept the objects required by our referential apparatus. Whereas for *overall* realism, extra-linguistic objects explain why any sentence is true or false (independently of which are held true), for *internal* realism it is the truth of most of our sentences which explains why *any* have truth relevant relations to extra-linguistic objects. As Davidson puts it, "Before some object in, or aspect of, the world can become part of the subject matter of a belief (true or false), there must be endless true beliefs about the subject matter." (TT, p. 20).

2. In working this out it is necessary to discuss what Davidson calls "radical interpretation."[4] In the first instance this aims at understanding the utterances of a speaker of an unknown language, but it gets extended to cover understanding of the utterances of *any* speaker: "All understanding of the speech of another," Davidson writes, "involves radical inter-

[3] In characterizing Davidson's position as an "*internal* realism," I am departing from Putnam's use of the term for someone who thinks of truth as relative to a scheme (cf. *Meaning and the Moral Sciences*, p. 125), an idea Davidson rejects. But I have not been able to think of a better term for someone who, like Davidson, rejects the realist claim that sentences in general are true in virtue of extra-linguistic objects (which leaves room for radical skepticism) and yet claims that particular bodies of discourse have to be understood as formulating true sentences about extra-linguistic objects.

[4] This term is deliberately modelled on Quine's "radical translation." For Davidson, however, "Translation is a purely syntactic notion" (RWR, p. 253); hence, if all we had was a translation of a sentence, we would not know its meaning unless we knew the meaning of the sentences into which it was translated. Given Davidson's conception of radical interpretation, this would presuppose radical interpretation of the translating sentences; but having that we could just as well interpret the original sentences directly without the detour of a translation. This is what the *T*-sentences of a Tarski-type theory of meaning do: they pair descriptions of sentences in the language to be translated not with *descriptions* (or mention) of sentences in the translating language (as a mere translation would do) but with sentences *used* to give the meaning of the translated sentences.

pretation." (RI, p. 313) Its goal is to interpret utterances without assuming that we know anything about what those utterances mean, without assuming any understanding of utterances of the speaker. A *theory* of radical interpretation, therefore, aims to "give the meaning of all sentences of a language, and can be verified by evidence available to someone who does not understand the language." (BBM, p. 309).

Davidson is not here setting the impossible task of developing a theory of interpretation which does not presuppose that the interpreter knows a language; the interpreter must, of course, know his own language. To say that radical interpretation is involved even in understanding speakers of one's own language is not to say that one should suppress knowledge of one's own language in interpreting others or that knowledge of one's own language is irrelevant for some such reason as that meanings are hidden in the minds of others. It is simply to say that similar considerations come into play in interpreting utterances in an unknown language —for example, trying to disentangle what a speaker means from what she believes. Interpreting speakers of one's own language is "radical," not because meanings are hidden in the mind, but precisely because the meaning of sentences is a matter of how they are used in a linguistic community, so that if, for example, a speaker deviates from the pattern of sentences generally held true, difficulties may arise, not merely as to what the speaker means by her sentences, but as to what her sentences mean. A speaker *may* mean by her sentences something other than what they mean, but this is necessarily a very isolated phenomenon, for meaning anything by a sentence requires that it be linked with other sentences whose meaning is determined by the structure of the language.

Davidson's theory of radical interpretation bears a number of similarities to Platts' theory of force (which is not surprising since the latter is modelled on the former). Both take as their starting point the utterances of a speaker and both have as their goal understanding what those utterances mean. Both the theory of radical interpretation and the theory of force *envelop* a truth-condition theory of meaning, and that in two

senses.[5] On the one hand, a truth-condition theory of meaning occupies in both theories the central position between the starting point—speakers' utterances—and the goal—interpreting the meaning of those utterances. On the other hand, both the theory of radical interpretation and the theory of force yield whatever *evidence* we have that a truth-condition theory is adequate to the language being interpreted: for both Davidson and Platts the enveloping theory provides the ultimate criterion for the adequacy of a theory giving the meaning of sentences in a language.

But these similarities mask the fundamental differences between the antirealism implicit in the theory of radical interpretation and the realism implicit in the theory of force. I am not referring here to Platts' realist construal of the truth-condition theory of meaning and Davidson's rejection of that (discussed in Part I). These differences are additional: in one sense they account for Platts being realist about the truth-condition theory and Davidson being anti-realist, in another sense they are a consequence of that difference.

The differences show up already in the next step beyond the starting point of identifying a speaker's utterances. For Davidson a speaker's utterances can play no role in radical interpretation unless we can identify which utterances express a certain attitude a speaker can take toward sentences, namely, the attitude of *holding true*.[6] The utterances themselves will, of course, be the main evidence that the speaker holds true a certain sentence, though other aspects of her behavior will be relevant. In any case Davidson thinks we can know a speaker holds true a certain sentence even if we cannot interpret her behavior: we do not need to *understand* the sentence she holds true (or any other sentences of her language) in order to know she is holding true a certain sentence.

[5] This feature of Platts' theory of force is discussed in Part I of this paper (Section 2).

[6] "There is no reason," Davidson writes, "to rule out other attitudes toward sentences, such as wishing true, wanting to make true, believing one is going to make true, and so on, but I am inclined to think that all evidence of this kind may be summed up in terms of holding sentences to be true." (RI, p. 322). I gather that Davidson's more recent view is that these other attitudes may play a more important role than this passage indicates.

This strategy of Davidson involves (at least) three differences from Platts. The first, which is the most crucial and which I will discuss at length shortly, is that for Davidson in order even to begin to interpret the meaning of a speaker's utterance on a given occasion, the interpreter must assume that he and the speaker agree in what they believe on that occasion – must assume that what the speaker holds true *is* true according to the interpreter's best judgment. Radical interpretation, in other words, rests on a general assumption that the beliefs ascribed are *true* – an assumption Davidson calls "the principle of charity." The theory of force does not require the general presumption that beliefs plausibly ascribed are true.

The second difference is that for Platts propositional attitudes are attitudes not toward sentences but toward language-independent contents *specified* by sentences (specified by the RHS's of the T-sentences of an acceptable theory of meaning). For Davidson, on the other hand, while knowing the sentence a speaker holds true is not sufficient for knowing the contents of her belief (since knowing that requires knowing not only what sentence is held true but also what the sentence means), nevertheless having a belief is simply holding true a sentence whose meaning one understands.

This difference is connected with the first difference, for Platts' view, unlike Davidson's, implies that the relation a language has to extralinguistic objects does not presuppose that any sentences held true by speakers of the language are true.[7] The relation is mediated, rather, by language-independent contents, which are expressed in common by sentences in any language which have the same (extra-linguistic) truth

[7] A realist, of course, need not deny that most sentences held true by speakers of a language are true, and he may have reasons for thinking so – for example, that the very survival of a race which holds sentences true shows that most of the sentences it holds true are true. In this sense a realist may hold that only if most sentences held true by speakers of a language are true can any be either true or false. But this is different from the anti-realist claim that it is only *because* most sentences held true by speakers of a language, are true, that *any* are either true or false or have truth-relevant relations to extra-linguistic objects. For the realist position implies that, however improbable it may be that most sentences held true by speakers of a language should be false, nevertheless if they were false they would be false in virtue of the truth-relevant relations they bear to extra-linguistic objects. It is this which the anti-realist denies.

conditions. The view, in other words, is the traditional realist view that it is the intentionality of *thoughts*—their being directed to objects, independently of whether they are true or false—which accounts for the relation of language to reality.

The third difference is that Davidson *begins* with sentences held true and the beliefs they express; for Platts the ascription of beliefs is the *end-point* of the theory, with the intepretation of a speaker's utterances tested by whether such ascriptions are plausible. The effect of this is that Davidson does without those elements in Platts' theory which produce the sentences which give "the strict and literal meaning" of utterances. Davidson does not worry about the mood of sentences, only (roughly) about the mode of utterances, identifying it (roughly) as asserting. (I say "roughly" because asserting is a speech act, holding true is not.)

This difference by itself is not fundamental. Platts is inclined to think that Davidson takes holding true as a kind of absolute foundation for his theory of meaning – a purely behavioral fact on which interpretation is supposed to rest (WM, p. 65) – whereas ascriptions of propositional attitudes, the goal of the theory of force, are always subject to modification. But Davidson can withdraw a judgment about a speaker's holding a sentence true if things do not work out. Moreover, identifying a speaker as holding true a certain sentence is no more difficult than identifying, as Platts' theory requires, the mode of an utterance or the mood of a sentence. In either case we can bring to bear all we know about the speaker's behavior, linguistic or otherwise, and all we know about languages in general, and if our original judgments do not work out, we can modify them.

This difference is important not by itself but because by it Platts hopes to avoid having to begin a theory of interpretation for a language with judgments about sentences in that language being *true*. Davidson's starting point – identifying sentences held true – requires that a theory of interpretation assume that most sentences a speaker holds true *are* true – that is, requires the principle of charity. Platts' strategy is to avoid

the principle of charity by making ascriptions of propositional attitudes the *end point* of the theory.

Platts' instincts at this point are sound, for unless Davidson's approach is rejected, there is no hope for a realist theory of meaning. Use of the principle of charity implies that a theory of interpretation can take account of what speakers believe only if it considers whether their beliefs are true. It implies, in other words, that a theory of interpretation for a language *must* take account not only of sentences *held* true by speakers of the language but of what sentences (in our considered judgment) *are* true in that language. This is inconsistent with the fundamental realist claim that the meaning of sentences consists of extra-linguistic truth conditions in virtue of which sentences are true or false, independently of any ability we have to determine if those truth conditions obtain. Realism holds that we can know what sentences in a language mean independently of any ability to determine whether they are true. The principle of charity implies that we cannot know what sentences in a language mean except by reference to our ability to determine that the beliefs they express are, for the most part, true.

But rejection of the principle of charity is no minor alteration in Davidson's theory; it is central to his theory of radical interpretation. Given that theory, there is no hope for realism in any but an internal sense.

3. The problem Davidson takes as central for the theory of radical interpretation is that, assuming we have identified the sentences a speaker holds true, in trying to interpret – i.e. understand the meaning of – those sentences (whether in another language or in one's own), we have to sort out what the sentences mean from what the speaker believes. Thus if someone standing in front of a bus says "That is a big truck", we can identify what sentence he holds true, but in trying to understand what he has said, we have a choice: we might assume he meant to say "bus" (for some reason he usually mixes up the words) or we might assume he believes falsely that the vehicle is a truck. His mistake may be one of meaning

or one of belief, and to understand him we will have to decide which.

In this case, the decision will be relatively easy; we could, for example, ask him a few questions. In the case of a language we do not understand, the problem is more difficult. Not knowing the language, we have no understanding of what the sentence means. But we do not know what the speaker believes either, for identifying a belief (such as his thinking that vehicle is a truck) is impossible unless we know his language: ". . .Making detailed sense of a person's intentions and beliefs cannot be independent of making sense of his utterances." (BBM, p. 312).

> The problem, then is this: we suppose we know what sentences a speaker holds true, and when, and we want to know what he means and believes. Perhaps we could crack the case if we knew enough about his beliefs and intentions, but there is no chance of this without prior access to a theory of interpretation. Given the interpretation, we could read off beliefs from the evidential base, but this assumes what we want to know. (BBM, p. 313.)

The theory of radical interpretation is designed to resolve this problem. Davidson has given a "crude outline" of such a theory, which I shall quote and then comment on.

> First we look for the best way to fit our logic, to the extent required to get a theory satisfying Convention T, onto the new language. . . The evidence here is classes of sentences always held true or always held false by almost everyone almost all the time (potential logical truths) and patterns of inference. The first step identifies predicates, singular terms, quantifiers, connectives, and identity: in theory, it settles matters of logical form. The second step concentrates on sentences with indexicals: those sentences sometimes held true and sometimes false according to discoverable changes in the world. The last step deals with the remaining sentences, those on which there is not uniform agreement,

or whose truth value does not depend systematically on changes in the environment. (RI, p. 323f.)

The first step is the theory of meaning based on Tarski's truth-theory, whose significance has already been discussed in Part I. The important thing to remember here is that, although this explains meaning in terms of truth-conditions, "truth-conditions" is not understood (realistically) in terms of extra-linguistic reality: it is understood neither in terms of states of affairs in virtue of which sentences are true or false nor in terms of properties which explain why objects are or are not satisfied by predicates. What Davidson's truth-condition theory of meaning does is to explain how the truth of complex sentences depends on the truth of simpler sentences and, most importantly, how the truth of sentences is determined by the words that make them up. "The intent of a theory of truth... is not that it tells us what truth is in general, but that it reveals how the truth of every sentence of a particular L depends on its structure and constituents" (RWR, p. 251). The output of this step, therefore, is the logical form, not only of individual sentences, but of the whole language (since the theory works only if it meets the holistic constraint of Criterion-T).

In his earlier papers ("Truth and Meaning" and "True to the Facts") Davidson sometimes wrote as if this first step were enough to interpret the meaning of sentences. This is true in the sense that *if* we had a correct truth-theory for a language, we would have a theory that gave the meaning of all sentences in the language. But we still wouldn't *know* the meaning of its sentences – i.e. we couldn't interpret its sentences – unless we had *knowledge* of a correct truth-theory for the language,[8] which implies that we *know* we have the correct truth-theory. The first step alone is not an adequate basis for claiming to

8 In one sense this is more than we need for interpretation. "In order to interpret a particular utterance, it is neither necessary nor sufficient to know the entire theory: it is enough to know what the theory says the conditions are for the utterance, and to know that these conditions are entailed by a theory of a required sort. On the other hand, to belong to a speech community. . . one does need to know much of the whole theory, in effect, and to know that it is a theory of the right kind" (TT, p. 13).

know we have the correct truth theory. For the first step imposes only *formal* constraints on a truth theory: it enables us to know that the T-sentences give the correct logical form of sentences to be interpreted (we will be able to identify which words function as predicates, singular terms, sentential connectives, etc.), but not to know if they give the correct *content*. (We will not know for example, if the truth condition is that the vehicle is a truck or that the vehicle is a bus, because sentences with these truth-conditions are formally identical.)

This is no problem if we abstract from the problem of radical interpretation and make use of our knowledge of languages (and speakers) we know well. For then our *evidence* for the truth of T-sentences will be our knowledge that the sentences described on the L.H.S. are *equivalent* to those used on the R.H.S. When the object language is our own, the sentence described on the L.H.S. will be the same as that used on the R.H.S. (or be related by known syntactical rules). When the object language is not our own, the R.H.S. will be known to be a *translation* of the sentence described on the L.H.S. In either case recognizing that T-sentences are correct will be straightforward.

Under these circumstances assumptions about the truth of the beliefs sentences in a language express need play no role in recognizing the truth of T-sentences, i.e., in recognizing what sentences in that language mean. And this suggests that perhaps we could make a clear and general distinction between what a speaker means and what he believes, without appeal to anything beyond a Tarski-based theory of meaning.

We have, however, not only limited ourselves to languages we understand, but we are making use of that understanding by assuming we know when we have a translation. Given that we understand the language and the speaker, we have already *made* a distinction between meaning and belief, so it is not surprising that we find such a distinction. But radical interpretation is not to use evidence which rests on an understanding of the language (or speaker) to be interpreted. It cannot, therefore, appeal to translation as evidence for the truth of T-sentences (i.e., as evidence that we have correctly interpre-

ted a sentence). Even in the case of our own language, the example of the mix-up about the truck and the bus shows that understanding is straightforward only when what we have assumed about the distinction between meaning and belief for a given speaker is straightforward. When a speaker says something deviant enough to draw us up short, that distinction becomes problematic, even in the case of a speaker of our own language.

As Davidson has stressed in his later papers, therefore, the problem of disentangling meaning and belief is present for any language, even our own, so that we cannot abstract from the problem of radical interpretation. Hence Davidson's statement that "all understanding of the speech of another involves radical interpretation." This means that we cannot assume we know a T-sentence is correct on the basis of our knowing that the R.H.S. is a correct translation of the L.H.S.

> If we knew that a T-sentence satisfied Tarski's Convention T we would know that it was true, and we could use it to interpret a sentence because we would know that the right branch of the biconditional translated the sentence to be interpreted. . . [But] in radical interpretation we cannot *assume* that a T-sentence satisfies the translation criterion. (RI, p. 326.)

> Our outlook inverts Tarski's: we want to achieve an understanding of meaning or translation by assuming a prior grasp of the concept of truth. What we require, therefore, is a way of judging the acceptability of T-sentences that. . . makes no use of the concepts of translation, meaning, or synonymy, but is such that acceptable T-sentences will in fact yield interpretations. (BBM, p. 318.)

4. To interpret the meaning of a speaker's utterance we must take the second step in the theory of radical interpretation, which "concentrates on sentences with indexicals, those sentences sometimes held true and sometimes false according to

discoverable changes in the world."[9] Beginning with a sentence held true, we want to determine the meaning of that sentence – that is, determine its correct T-sentence – without assuming we could recognize a translation of it. If we knew what belief the speaker was expressing, we could solve the problem, for then the meaning of this sentence would be given by a sentence which we, the interpreters, use to formulate that same belief. (If he holds true "zip a tup," and we know he is expressing the belief that snow is white, then we know that "zip a tup" means that snow is white.) Davidson's argument is that this is the *only* way we can solve the problem: the only way to get evidence for T-sentences (without begging the question) is to assume that we know, in general, what the speaker believes.

But how can we know that, starting only from the sentences he holds true, without assuming what we want to find out, namely what he means by the sentence? Davidson' proposal is that we must assume, in general, that we, the interpreters, and the speaker, whose utterances we want to interpret, *agree in our beliefs.* In particular we must assume, when interpreting a sentence held true because of discoverable changes occuring on a particular occasion, that we and the speaker share the same belief on that occasion (so that the sentence he holds true on this occasion is one I would hold true on this occasion if I spoke his language, and the sentence I hold true is one he would hold true if he spoke my language). The proposal, in other words, is "to solve the problem of the interdependence of belief and meaning by holding belief constant as far as possible while solving for meaning" (R1, p. 324).

The reason for the reference in the second step to sentences with indexicals, therefore, is that these are the best candidates

9 These are similar to the observation sentences of Quine's *Word and Object* (which are a sub-set of occasion sentences), except that Quine's notion that it is patterns of sensory stimulation which cause an observation sentence to be held true is rejected in favor of the idea that it is the objects that "correspond" to the sentence which directly cause the sentence to be held true. Here as elsewhere Davidson differs from Quine in a number of respects, but Quine remains the most important antecedent of Davidson's thought, a fact realist interpreters of Davidson tend to overlook. Quine in his recent book (*Theories and Things*) does claim to be a realist, but this is clearly realism only in an internal sense.

for sentences expressing beliefs which can be identified as ones which we and the speaker share on a particular occasion. For such sentences are uttered only when both interpreter and speaker are (reasonably) near features in the environment accessible to both.

Thus, to use Davidson's example (RI, p. 322), take this T-sentence:

(1) "Es regnet" is true-in-German when spoken by x at time t if and only if it is raining near x at t.

If we know (1) is true, we can understand a (non-deviant) utterance of the sentence "es regnet". But what is our evidence for (1), given that we disallow an appeal to translation? Davidson's proposal is that our evidence will consist of (many) sentences such as this one:

(2) Kurt belongs to the German speech community, and Kurt holds true "es regnet" on Saturday at noon, and it is raining near Kurt on Saturday at noon.

But sentences like (2) will not be evidence for (1) unless we assume that Kurt believes that it is raining near him on Saturday at noon, an assumption we can make only on the grounds that he believes just what we do. If he does not believe what we do, if, for example, he believes it is snowing, then by "es regnet" he must mean that it is snowing; he certainly cannot mean that it is raining.

What a speaker holds true, therefore, can be evidence for a corresponding T-sentence – i.e., can be evidence of what the sentence uttered means – only if we assume that he believes what we believe, only if we make his belief agree with ours. In practice, this will require indexical sentences, for indexicals, being tied to situations equally accessible to both of us, enable us to get beyond the too general idea that the speaker agrees with us on many things to the point where we can identify the belief we share on a particular occasion and which gives us the evidence we need for the relevant T-sentences.

The evidence for T-sentences at this point, then, depends on assuming that the speaker shares our beliefs. Obviously we cannot assume either that the speaker shares all our beliefs or that we are always in a situation where we can even begin to identify his beliefs. Hence the last step in the theory, which deals with sentences "on which there is not uniform agreement, or whose truth value does not depend systematically on changes in the environment." Having established T-sentences for many of the speaker's utterances in step two, the holistic account of meaning will allow us to piece out the rest, making various adjustments. We may even reject the results of the second step, for we may discover we were wrong in some beliefs attributed to the speaker. The method of radical interpretation does not depend on any infallible determination of a speaker's beliefs. "The method is rather one of getting a best fit. We want a theory that satisfies the formal constraints on a theory of truth and that maximizes agreement" (RI, p. 323). The method rests on the assumption of a general agreement between the speaker's beliefs and ours – the assumption Davidson calls the "principle of charity". But we can know we agree on most beliefs without knowing which beliefs we agree on.

5. A natural objection at this point to Davidson's proposal, based on the principle of charity, of "holding belief constant as far as possible while solving for meaning" is that it is merely *ad hoc*. It amounts to saying that since we have a problem which can be solved by adopting the proposal, we have sufficient reason for adopting it, an argument that would justify many bad proposals.

Let me first consider a slightly different form of the objection, namely, that Davidson's proposal simply *assumes* that radical interpretation *is* possible. Perhaps *radical* interpretation is not possible. Perhaps we can have at best guesses about what alien peoples believe and mean, so that there may be societies of rational beings who speak a language we cannot interpret at all.

The short answer to this is verificationist: if we cannot in-

terpret the speech of others, the claim that their utterances amount to speech or express beliefs is unverifiable and therefore empty. Davidson sometimes suggests this approach: "If we cannot find a way to interpret the utterances and other behavior of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything" (RI, p. 324).

But the short answer obscures something of great importance, namely, that for Davidson this kind of skepticism about radical interpretation rests on an unwarranted assumption. That assumption is the realist one that the meaning of sentences is the extra-linguistic truth-conditions in virtue of which they are true or false. From this it follows that we can interpret a sentence in a language only if the interpreting language (the meta-language) has a sentence we can use to state the extra-linguistic truth-conditions of the sentence to be interpreted. Given that assumption, a society of creatures radically foreign to us (because it is, e.g., from another planet, or from another historical epoch, perhaps another scientific epoch) may speak a language which is completely uninterpretable because we are unable to state its extra-linguistic truth conditions. Such a language would be unintelligible to us, and though we could not guess at the beliefs its speakers held (because we would have no way to formulate them), they will be true if their truth-conditions obtain.

To put the point in a slightly different way, sentences in such a language would be, though possibly true, untranslatable. For the realist theory of meaning implies an extra-linguistic criterion of correct translation: a sentence correctly translates another sentence (the two sentences have the same meaning) if and only if they both have the same extra-linguistic truth conditions. A sentence will be untranslatable, therefore, if it has truth conditions no sentence in our language has. If those conditions obtain, it will be true but not translatable.

The realist construal of Davidson runs flat up against his explicit rejection of this possibility.

> The criterion of a conceptual scheme different from our own now becomes: largely true but not translatable. The question whether this is a useful criterion is just the question how well we understand the notion of truth, as applied to language, independent of the notion of translation. The answer is, I think, that we do not understand in independently at all. (VICS, p. 16.)

To reject the possibility of a language whose sentences may be (largely) true but which we cannot interpret or translate just is to reject the realist theory of meaning.

Davidson's view is that it is not a relation to extra-linguistic *objects* of any kind – states of affairs, properties, material things, experiences, ideas, or what have you – that accounts for the meaning of an expression. An expression is meaningful not because it represents extra-linguistic truth-conditions but because and only because it has a place in the structure of a language. "Structure" means here the logical form analyzed by a truth-theory, which shows the relations both of the truth-conditions of simple sentences to the truth-conditions of complex ones and of the satisfaction-conditions of words to the truth-conditions of sentences generally. So to say that an utterance is meaningful is to say no more and no less than it is a sentence in a language which has a truth-structure of the sort analyzed by a Tarski-type truth-theory. Thus to say "zip a tup" is a meaningful utterance is to say it is a sentence which has a place in the truth-structure of a language: its truth-conditions are determined by the meaning of its constituent words and related in various ways to the truth-conditions of other sentences.

But if this is what it is for a sentence to *be* meaningful, then the sentence is necessarily interpretable and hence translatable. It is necessarily interpretable because, to put it roughly, to interpret it is to find a sentence which has the same place in the structure of our language that the sentence to be interpreted has in the structure of its language. We know that our language has all the structure a language requires: it has a finite supply of components (words) to generate a (potentially)

infinite number of sentences, and it has the necessary truth-relations among sentences. So there is no problem about the structure. The trick will be to find the particular sentence in our language which fits into the structure, but that presents no difficulty in principle (assuming the cogency of the proposal to hold belief constant).

What I have just put roughly is, of course, put more precisely in terms of truth-theory. To interpret a sentence is to formulate a T-sentence for it within the context of a theory which meets criterion-T.[10] This may not be easy. But we know we can (in principle) give a truth-theory for our language, and hence we know we have a meta-language sufficiently rich formally to do it for sentences in any language, since no language *requires* more structure than ours. Again, the trick will be, not to provide a theory which generates T-sentences which analyze the form both of the sentences to be interpreted and of the whole language, but to determine that the T-sentences give the correct *content*.

But if a sentence is interpretable, then it is translatable. Because if it is interpretable, then we know its T-sentence, and to find a translation, we find a sentence in our language whose T-sentence has the same R.H.S. That is to say, for a sentence to be a translation of a sentence from another language, we require, not that both represent the same extra-linguistic truth-conditions, but that T-sentences for both have the same R.H.S., in other words, that both have the same place in the truth-structures of their respective languages.[11] The constraint on translation, therefore, is holistic because structural: a translation of one sentence is not correct unless it fits in with the translation of all other sentences. In the language of truth-theory, the T-sentence for either the translating or the translated sentence is not acceptable unless each is part of a theory which meets criterion-T.

Davidson's account of meaning implies, therefore, that a

10 But cf. note 8.

11 Cf. Davidson's rejection of "a criterion of languagehood that [does] not depend on, or entail, translatability into a familiar idiom" (VICS, p. 130).

meaningful utterance is necessarily interpretable and translatable. This answers the objection that this proposal simply *assumes* that radical interpretation is possible: in arguing that meaning is inseparable from interpretation or translation, he *argues* for the impossibility of a non-interpretable language. But he also argues that *truth* is inseparable from interpretation or translation. In considering this argument we will deal with the objection that his proposal is *ad hoc* in simply laying it down that we should hold belief constant by applying the principle of charity.

6. To interpret a sentence is to find a sentence which has the same place in the structure of our language that the sentence to be interpreted has in the structure of its language. The trick is in finding precisely which sentence that is. Davidson's proposal amounts to saying that to find that sentence, we assume that the sentence the speaker holds true on a given occasion is a sentence we would hold true if we spoke his language, and that the sentence we hold true on this occasion is a sentence he would hold true if he spoke our language. This proposal is not *ad hoc;* it is necessarily connected with Davidson's conception of truth.

Davidson has remarked a number of times that his theory of meaning *assumes* but does not explicate a concept of truth. "In a theory of the sort I am describing, the truth predicate is not defined, but must be considered a primitive expression" (RWR, p. 248). What the theory does is explain how the truth-relevant properties of words (their satisfaction-conditions) are related to the truth-conditions of sentences, but it does this only by appealing in turn to the truth-pattern of sentences. This will be an explication of *truth* only if we already understand what it is for a sentence to be true.

What we need is some account of the conception of truth assumed in Davidson's theory of meaning – an account which does something other than analyze the truth-structure of a language. Such an account will have to say something about the relation of true (or false) sentences to extra-linguistic reality.

Davidson's account of truth can best be explained as con-

sisting of two parts.[12] Part one offers an account of that *in virtue* of which our sentences are true or false. A realist account is not available to Davidson: his theory of meaning does not allow that sentences are true in virtue of extra-linguistic objects.[13] What he offers in its place is not an alternative *theory* but an account of how "true" came to function, and how it now functions, in a linguistic community.

Part two offers an account of what it is for a sentence to *be* true, in terms of a relation of true sentences to extra-linguistic objects, a relation Davidson calls "correspondence" (TF, p. 748). Although this term is usually tied to (overall) realist theories of truth, Davidson does not use it in this sense: to say a true sentence corresponds to extra-linguistic objects is not to say the sentences are true *in virtue of* the objects. It is to say that for a sentence to be true *is* for it to correspond to the objects required by the referential constituents of the sentence. Sentences correspond to extra-linguistic objects, not because the sentences are true in virtue of the objects, but just because they *are* true. To speak of the relation of sentences to extra-linguistic objects, in other words, is not to *explain* why sentences are true (if they are), but to *presuppose* that they are in order to explain what it is for them to *be* true.[14]

12 This is a somewhat misleading way to put it. Strictly speaking, only what I call "part one" of Davidson's account of truth is *presupposed* by the truth-condition theory of meaning. Part two is a different way of stating what is accomplished by the theory of meaning itself. Putting it in this way seem to me illuminating, however.

13 See Part I of this paper. By saying a realist account is not available to Davidson I, of course, mean realism in the overall (not internal) sense.

14 This part of the theory, therefore, also rests on an extra-theoretical understanding of truth, which is explicated in what I called part one of Davidson's account of truth (cf. note 12). It should also be noted that Davidson can say, as he sometimes does (although I think it is misleading to do so), that it is extra-linguistic objects which *make* our sentences true or false – for example, the last sentence of VICS (p. 20): "In giving up the dualism of scheme and world, we do not give up the world, but establish unmediated touch with the familiar objects whose antics make our sentences and opinions true or false." This must be understood, not in terms of what I have called *overall* realism but in terms of internal realism. For Davidson's saying this presupposes that most of our sentences are *already* true, and hence are not (in the sense of overall realism) *made* true by extra-linguistic objects. But given that most of our sentences *are* true, individual sentences will be true if and only if they correspond to extra-linguistic objects, and in this sense the (antics of the) objects will make them true or false.

Another way of putting this point is to take the standard T-sentence:

"Snow is white" is true if and only if snow is white

and consider the interpretation of the "if and only if." (Overall) realism construes it as meaning "in virtue of" (and "only in virtue of"), it construes it as an *explanatory* notion. Davidson does not construe it as explanatory but takes it extensionally: his notion of "correspond" just puts a label to the extensional relation.

Let me briefly defend this interpretation of Davidson's "correspondence theory" of truth. The main defense is that, given what seems to me his clear rejection of an (overall) realist theory of truth, no other interpretation is possible. His account of that *in virtue of which* sentences are true or false *cannot* be in terms of the relation of sentences to extra-linguistic *objects*. But although his overall account is anti-realist, this does not prevent him from giving an internally realist account of (at least many) classes of sentences: an account which, not being reductivist, allows for objects which uniquely correspond to the referential terms in each true sentence (with, again, the anti-realist proviso that it is not in virtue of the objects being as they are that the sentences are true or false).

The other defense I would make rests on a brief exegesis of "True to the Facts" in which the term "correspondence theory" appears. The paper begins as follows (p. 748):

> A true statement is a statement that is true to the facts. This remark seems to embody the same sort of obvious and essential wisdom about truth as the following about motherhood: a mother is a person who is the mother of someone. The *property* of being a mother is explained by the *relation* between a mother and her child; similarly. . . the property of being true is to be explained by a relation between a statement and something else.

The analogy is with motherhood: being the mother of a child does not explain *why* someone is a mother – that would

require reference to deciding to have a child, to getting pregnant, etc. It explains *what* the property of being a mother is. So Davidson is not saying that being true to the facts explains why a sentence is true, why it has the property of truth – that would require reference to part one of the account of truth. He is saying that being true to the facts explains *what* the truth predicate is, namely, that to be true is to be true to – i.e. to correspond to – extra-linguistic objects.[15]

To return to what I called part one of Davidson's account of truth; this deals with the linkage of that in virtue of which sentences are true with extra-linguistic reality. This linkage is made not via properties or states of affairs – objects or concatenations of objects in virtue of which our sentences are true or false – but by the ways in which the sentences we hold true come to be held true because of their role in helping us achieve our (extra-linguistic) goals and further our interests.

> . . . The semantic features of words cannot be explained directly on the basis of non-linguistic phenomena. The reason is simple. The phenomena to which we must turn are the extra-linguistic interests and activities that language serves, and these are served by words only in so far as the words are incorporated in (or on occasion happen to be) sentences. (RI, p. 215.)

> Words have no function save as they play a role in sentences: their semantic features are abstracted from the semantic features of sentences, just as the semantic features of sentences are abstracted from *their* part in helping people achieve goals or realize intentions. (RWR, p. 252f.)

Davidson does not say much about this, and he should say more,[16] but given what I think he should say, there are reasons

[15] "Corresponding to the facts" is not, therefore, an informative definition of "being true," because understanding "corresponding to the facts" requires understanding "being true to the facts," and the latter is informative only if we already have some understanding of "true." In this sense "truth" remains a primitive predicate.

[16] The deepest account of all this has been given in the later works of Wittgenstein.

why he does not. For this dimension of language has to do primarily with its pre-historic development; Davidson's concern is with understanding its present structure and functioning. From that perspective *direct* consideration of how language helps people achieve goals or realize intentions is replaced by consideration of how the truth of sentences is determined by the many sentences which have already come to be held true in a linguistic community. In other words, when we are dealing with language as a developed structure, the determination of truth by purely pragmatic factors gives way to its determination by sentential evidence.

The reason for this is that once we come out of the prehistoric stage of language's development to its historical stage, we already have a developed language in which an enormous number of sentences are held true by the linguistic community. That community became a *linguistic* community as it *simultaneously* developed a language and came to hold true in common *many* sentences in that language. All this was done under pressure of "helping people achieve goals or realize intentions"; that ultimately explains why sentences were accepted or rejected. But consciousness of that pressure occurred, if it did occur, only in a context where many sentences were already held true, so that consciousness of the pressure also became a matter of holding true certain sentences about the development of the very language in which the sentences occurred. Our view of the relation of our own language to extra-linguistic reality is, therefore, determined by its relation to sentences already held true by speakers of our language.

Language developed as it did because of its role in helping people meet their (primarily) extra-linguistic needs and further their interests. But when we reached the historical stage, and conscious changes in language became possible (as in the development of science), changes and additions were not made (except in slogans and propaganda) to help people. They were made to adjust and connect sentences held true to others already held true in the linguistic community. They were

made, in other words, under pressure from the weight of evidence.[17]

We can, if we like, then, say that on this view a sentence is true in virtue of its helping people achieve goals and realize intentions. We can say this because this relation of language to extra-linguistic reality plays the *same role* for anti-realism that the realist notion of states of affairs (facts, situations) plays for realism: it provides an *ultimate extra-linguistic* account of that in virtue of which sentences are true. For Davidson that standard is how well language serves human interests, and that is the only overall extra-linguistic explanation he will offer of why sentences in general are true or false.

But saying that a sentence is true in virtue of its helping people achieve goals (and so on) is misleading unless we add immediately that for developed languages such pragmatic factors play a role only via the sentences generally held true in a linguistic community. A sentence is true in a developed language in virtue of all the other sentences held true in the linguistic community. Although truth is *ultimately* anchored in extra-linguistic reality – the goals, interests, and needs of humankind – there is no truth for a developed language except in relation to the sentences generally held true by the speakers of that language.

There cannot be, in other words, any such thing as an isolated true (or false) sentence, by which I mean, not only that there cannot be a sentence which does not belong to a language, but that there cannot be a sentence whose truth (or falsity) is independent of other sentences held true in a linguistic community. Sentences are true or false in virtue of their relation to other sentences held true – in virtue of their place in the truth-structure of a particular language. Hence

[17] Cf. L. Kolakowski, "Karl Marx and the Classical Definition of Truth" in *Marxism and Beyond*, p. 67. "From the moment. . . man begins to dominate the world of things intellectually – from the moment he invents instruments that can organize it and then expresses this organization in words – he finds the world already constructed and differentiated, not according to some alleged natural classification but according to a classification imposed by the practical need for orientation in one's environment."

the claim that we do not understand "the notion of truth, as applied to language, independent of the notion of translation." For if sentences are true in virtue of their place in the truth-structure of a language, we can understand what it is for those sentences to be true only if we can grasp the structure of the language, and this requires that we be able to interpret, and hence translate, it. A language cannot have true but untranslatable sentences, for a sentence is true just in virtue of its place in the structure of a language, and any language with a truth-structure is necessarily interpretable and translatable.

Before showing how Davidson's account of truth justifies his use of the principle of charity, I want to bring together the two parts I have distinguished. Part one is about that *in virtue of which* sentences are true; part two is about what it is for sentences to *be* true. This must not be confused with the realist distinction between what justifies us in accepting a sentence as true and that in virtue of which a sentence is true. Part one of Davidson's account is not (merely) about what justifies us in accepting a sentence as true; it is about that in virtue of which a sentence *is* true. Part two is about neither of these; it is about what we are committed to in accepting a sentence as true. Davidson's account differs from realism at this point, for realism collapses its account of what we are committed to in accepting a sentence as true with its account of that in virtue of which a sentence is true.

If Davidson were a realist, there would be a serious problem at this point, for there would be no guarantee that what it is for a sentence to be true and that in virtue of which a sentence is true would coincide. There might not be any suitable objects for true sentences to correspond to: the requirement that true sentences (true in virtue of other sentences held true) correspond to extra-linguistic objects might not be met.

This is no problem for Davidson because sentences which *are* true in virtue of their place in the truth-structure of a language will thereby correspond to those extra-linguistic objects to whose existence we are committed by the truth of the sentences. What accounts for this is just the Tarski-based theory of meaning; it links sentences which are true with the objects

to which they correspond. For a language in virtue of whose structure sentences are true has a truth-theory, which relates the truth-relevant properties of words to the truth-conditions of sentences, which means it links the objects which *actually* satisfy predicates (the objects predicates are true of) with the *truth* of sentences containing those predicates. A sentence cannot be *true* unless its predicates are *true of* objects.

For Davidson the correspondence of sentences with objects has to be made via the words which the sentences contain, for sentences taken as wholes (closed sentences) correspond, if they are true, to all (sequences of) objects (and, if they are false, to none). Sentences differ in what they correspond to only via their constituents. As Davidson puts it:

> . . . True sentences cannot be told apart in point of what they correspond to. . . [But] since different assignments of entities to variables satisfy different open sentences and since closed sentences are constructed from open, truth is reached, in the semantic approach, by different routes for different sentences. All true sentences end up in the same place, but there are different stories about how they got there; a semantic theory of truth tells the story for a particular sentence by running through the steps of the recursive account of satisfaction appropriate to the sentence. (TF, p. 759.)

This assures a necessary connection from the satisfaction-conditions of predicates to the truth-conditions of sentences, and hence from a predicate's actually being satisfied by (being true of) an object to the *truth* of a sentence.

But the crucial connection depends on the fact that a word has a meaning only in the context of a sentence: a predicate has satisfaction-conditions only insofar as it contributes to the truth-conditions of sentences. Predicate structure is entirely derivative from inter-sentential structure; the *only* justification, therefore, for an S-sentence like

Something satisfies "is red" if and only if it is red

is its role in a theory which, meeting criterion-T, enables us to derive all the T-sentences, which give correct *truth*-conditions for sentences.[18] This means that the *only* justification for thinking we have *correctly* assigned a predicate to an object is its structured role in sentences we think are *true*. This assures, therefore, a necessary connection from the *truth* of a sentence to a predicate's *actually* being satisfied by (being true of) an object. Sentences are true, therefore, if and only if predicates are actually satisfied by objects, which give us the required necessary linkage between the two parts of the account of truth.

Now to consider how this justifies Davidson's claim that radical interpretation *requires* that we hold belief constant by assuming the principle of charity. To hold belief constant in order to interpret a speaker's utterance means to assume that the sentence the speaker holds true on this occasion is a sentence I would hold true if I spoke his language, and that the sentence I hold true on this occasion is a sentence he would hold true if he spoke my language. Let us see what happens if we drop the assumption and try to interpret in such a way as to make the speaker differ massively from us in our beliefs.

Let us consider interpretation of the alien language in English, and, to simplify things, let us consider that the speaker has learned English and made the interpretation himself. What we would have then, if massive differences in belief are possible, is that the speaker uses English sentences but differs massively from us on the ones he holds true. In the rain he holds true that it is sunny, sitting down he holds true that he is flying, in a room full of people he holds true that there is no one around, and so on.

Our first reaction might be that he does not know English: though using English sounds he is speaking some other language. Or we might think that he is not speaking at all, only mouthing sounds. These are reasonable reactions, and they are consistent with Davidson's claim that an interpretation is not acceptable unless it puts us largely in agreement in our

18 See Part I of this paper.

beliefs. But they do not explain the connection of that claim with the account of truth.

These reactions can, however, be pressed. The assumption that is supposed to be operative is that these massive differences in the sentences we hold true express massive differences in belief – *real* differences in belief, so that for each pair of differing beliefs, one is true and one is false. The differences, however, will not be that something like half of our beliefs are true and half of the speaker's beliefs are false; Davidson's claim that sentences are true in virtue of other sentences held true ensures that either our beliefs will be massively false or the speaker's belief will be massively false. What has to be considered is whether this kind of systematically massive difference is intelligible, given Davidson's account of truth.

"We can make sense of differences all right," writes Davidson, "but only against a background of shared belief. What is shared does not in general call for comment; it is too dull, trite or familiar to stand notice. But without a vast common ground, there is no room for disputants to have their quarrel" (MTM, p. 244f). The reason should be clear: massive difference presupposes that what one of us believes is mostly true, what the other believes massively false. Since we are here assuming translation, and dealing with differences between speakers who are supposed to speak the same language, this becomes: massive difference presupposes that most of the sentences one of us holds true are true, while the sentences held true by the other are massively false. But this is inconsistent with Davidson's theory of truth.

In the first place, a sentence is true in virtue of the vast bulk of sentences held true in a linguistic community, and hence also false for the same reason. If the vast bulk of sentences a speaker holds true are false, therefore, he knows nothing of the practice of the linguistic community whose language he is trying to use, which is to say he does not understand the language. But not understanding the language, what appeared to be a difference in belief cannot be that.

In the second place, if a sentence is true in virtue of the vast bulk of sentences held true in a linguistic community,

then no competent speaker of a language can be massively wrong in the sentences he holds true, and, therefore, in his beliefs. Radical skepticism, in other words, is not a coherent, possibility for any language; in any language the sentences held true must be mostly true, and hence the beliefs of speakers of that language must be mostly true. Clearly, therefore, any translation that makes speakers mostly mistaken in their beliefs must be incorrect, which is, of course, just the proposal to hold belief constant.

Finally, as we have seen, for a sentence to *be* true is for it to correspond to extra-linguistic objects. But we have also seen that the correspondence of sentences to extra-linguistic objects requires, on the one hand, sufficient inter-sentential connections to determine the role of predicates in sentences, and requires, on the other hand, sentences which are *true*. Objects to which our sentences correspond (or fail to correspond), therefore, presuppose many sentences which are true.

This means that the idea of speakers who disagree massively in their beliefs is incoherent. For the one whose beliefs are massively false is holding sentences true, most of which are false and for which, therefore, there are no corresponding objects. This means not only that there is no common subject matter but that there are no objects to which the false beliefs of the massively mistaken speaker fail to correspond. "Just as too much attributed error," writes Davidson, "risks depriving the subject of his subject matter, so too much actual error robs a person of things to go wrong about" (MTM, p. 245). Without the correspondence with reality made by true sentences, there can be no false sentences either. "Before some object in, or aspect of, the world can become part of the subject matter of a belief (true or false), there must be endless true beliefs about the subject matter" (TT, p. 20).

There are many falsehoods in our language waiting to be rooted out and many true sentences yet to be formulated, and these are deeply significant to humankind. But compared to the truths we already know their number is small, which

shows – and in this we can take some consolation – how insignificant sheer number can be.[19]

BIBLIOGRAPHY

Davidson, Donald, "Truth and Meaning," *Synthese*, 1967 [TM].

_____, "True to the Facts," *The Journal of Philosophy*, 1969 [TF].

_____, "Radical Interpretation," *Dialectica*, 1973 [RI].

_____, "Belief and the Basis of Meaning," *Synthese*, 1974 [BBM].

_____, "On the Very Idea of a Conceptual Scheme," *Proceedings of the American Philosophical Association*, 1974 [VICS].

_____, "Thought and Talk" in S. Guttenplan, ed. *Mind and Language* (Oxford University Press, 1975) [TT].

_____, "The Method of Truth in Metaphysics," *Midwest Studies in Philosophy*, 1977 [MTM].

_____, "Reality Without Reference," *Dialectica*, 1977 [RWR].

Kolakowski, Leszek, *Marxism and Beyond* (London, Paladin Press, 1971).

Platts, Mark, *Ways of Meaning* (Routledge and Kegan Paul, 1979) [WM].

Putnam, Hilary, *Meaning and the Moral Sciences* (Routledge and Kegan Paul, 1978).

Quine, W. V. O., *Theories and Things* (Harvard University Press, 1980).

_____, *Word and Object* (MIT Press, 1960).

19 A number of persons gave me comments on Part II this paper – Vicki Harper, Gary Iseminger, Owen Jones, Curtis Brown, David Charles – but since I did not follow all their suggestions, they can hardly be held responsible for very many of its errors.

RESUMEN

La parte I de este artículo *(Crítica 41)* argumentaba que la teoría del significado de Davidson no puede utilizarse en la defensa de una posición realista. En esta parte se argumenta que la posición de Davidson es antirrealista no sólo en lo relativo al significado, sino también con respecto a la verdad. Aunque la teoría del significado de Davidson se formula en términos de condiciones de verdad, él no las concibe de manera realista como objetos extralingüísticos. Su posición consiste en que especificar las condiciones de verdad de una oración es especificar las relaciones de la oración con otras oraciones y con las condiciones que satisfacen las palabras que la constituyen. Aquello en virtud de lo cual las oraciones son verdaderas, no son, por lo tanto, objetos extralingüísticos, sino otras oraciones que se consideran verdaderas en una comunidad lingüística.

Esta teoría coherencial de la verdad no es inconsistente con postular que las oraciones corresponden a objetos extralingüísticos. Las oraciones que son verdaderas sí corresponden a objetos extralingüísticos, aunque sólo sea porque (ya) son verdaderas en virtud de otras oraciones consideradas verdaderas y porque la mayoría de las oraciones que se consideran verdaderas en una comunidad lingüística *son* verdaderas.

Defiendo esta interpretación de Davidson mostrando cómo se deriva de su teoría de la interpretación radical, la cual implica el principio de caridad —que la mayor parte de las creencias de la gente deben ser verdaderas— y la traducibilidad a cualquier lenguaje posible.

[*F. S.*]