SCOTT SOAMES

TRUTH, MEANING, AND UNDERSTANDING[1]

When theories of truth are taken to be theories of meaning a problem arises at the very outset that threatens to undermine the whole enterprise. Whereas we expect a theory of meaning to tell us what sentences mean, a theory of truth gives us only their truth conditions. But statements of truth conditions are weaker than statements of meaning. For example, instances of Schema M

M.   'S' means in L that p

together with analytic, apriori instances of Schema TM

TM.   If 'S' means in L that p, then 'S' is true in L iff p

entail the correponding instances of Schema T

T.   'S' is true in L iff p.

However, there is no true principle, let alone an analytic, apriori principle, which, together with instances of Schema T, allows one to derive instances of Schema M. Thus, instances of Schema M are stronger, and more informative, than instances of Schema T. This illustrates a problem inherent in taking theories of truth to be theories of meaning. If truth theories provide only statements of truth conditions, and these do not entail statements indicating what sentences mean, then theories of truth don't tell us what the sentences of a language mean. How then can one justify taking them to be theories of meaning?

An influential answer to this question, suggested by Donald Davidson in some of his early writings, is that a theory of truth for a language L can qualify as a theory of meaning for L if knowledge of that which it states is *sufficient* for understanding L. The plausibility of this criterion lies in the fact that if a theory tells us *everything* we need to know in order to understand a language, then it must be counted as specifying

all essential facts about meaning, even if it does not issue in theorems of the form '*S*' *means in L that p*, which state the meanings of individual sentences one by one.

Thus, we may accept the claim that a theory which provides information knowledge of which is sufficient for understanding a language would qualify as a theory of meaning. The difficulty lies in seeing how a theory of truth could satisfy this condition. Davidson's original idea was that a truth theory of the proper sort would give a "holistic" account of meaning, if it derived an appropriate statement of the truth conditions of each sentence from an account of its semantically significant structure, including the reference of its semantically significant parts. On this picture, the meaning of a sentence depends on the meanings, or, more properly, the referents, of its parts. These, in turn, are regarded as nothing more than abstractions from the contributions they make to the meanings of all the sentences in which they occur. Thus, meaning, that which a competent speaker grasps, is supposed to be a structure revealed only in the whole. Correspondingly, what we want from a theory of meaning is supposed to be a specification of the complex network of relationships mastery of which is sufficient to endow a speaker with semantic competence. This, I take it, was to be the excuse for regarding the appropriate sort of theory of truth to be a theory of meaning, even though it failed to provide theorems stating what any individual sentence means.

Davidson expresses essentially this view in several passages from "Truth and Meaning".

We decided a while back not to assume that parts of sentences have meanings except in the ontologically neutral sense of making a systematic contribution to the meaning of the sentences in which they occur. . . . One direction in which it [this insight] points is a certain holistic view of meaning. If sentences depend for their meaning on their structure, and we understand the meaning of each item in the structure only as an abstraction from the totality of sentences in which it figures, then we can give the meaning of any sentence (or word) only by giving the meaning of every sentence (and word) in the language.[2]

Davidson further elaborates this view in the well-known passage from the same article in which he discusses the non-translational, but true T-sentence (S).

(S)   'Snow is white' is true iff grass is green.

Davidson says that we ought *not* to think a theory that entails the standard T-sentence *'Snow is white' is true iff snow is white* is any more correct than a theory that entails (S),

> provided the theory gives the correct results for every sentence (on the basis of its structure, there being no other way). It is not easy to see how (S) could be a party to such an enterprise, [my emphasis] but if it were — if, that is, (S) followed from a characterization of the predicate "is true" that led to the invariable pairing of truths with truths and falsehoods with falsehoods — then there would not, I think, be anything essential to the idea of meaning that remained to be captured.

> What appears to the right of the biconditional in sentences of the form *s is true iff p* when such sentences are consequences of a theory of truth, plays its role in determining the meaning of *s* not by pretending synonymy but by adding one more brush-stroke to the picture which, taken as a whole, tells what there is to know of the meaning of *s*; this stroke is added by virtue of the fact that the sentence that replaces '*p*' is true iff *s* is.³

The view, here, I take it, is that there is nothing more to be known about the meanings of the sentences of a language than is stated by a theory of truth that systematically derives a true T-sentence for each object-language sentence on the basis of its structurally significant parts. It is suggested that the requirement that the derivations be systematic, and based on structure, may eliminate grotesque theories that issue in non-translational T-sentences like (S). The idea, it seems, is that in order to derive (S) from a compositional account of the structure of the sentence *Snow is white*, a theory would have to contain axioms specifying grass as the reference of *snow* and green things as the objects to which the predicate *is white* applies. But with such axioms one would end up deriving *false* T-sentences like *'Snow is grass' is true iff grass is grass* and *'The tree is green' is true iff the tree is white*, in addition to "accidentally true" T-sentences like (S). Thus, it is thought, truth theories which are both *true* and appropriately *structural* will end up deriving instances of Schema T in which the meta-language sentences on the right-hand side are close enough paraphrases of the object language sentences on the left that nothing essential to meaning would fail to be grasped by a speaker who knew the totality of that which is stated by the theory.

However, this view does not withstand scrunity. The central difficulty was noted in many places and forcefully expressed by J. A. Foster in his paper "Meaning and Truth Theory".⁴ First, Foster observed that the requirement that a theory of truth derive its T-sentences on the

basis of semantically significant structure does little to block the
derivation of non-translational T-sentences like (S). To derive such
sentences all one needs to do, given an extensional first order object
language, is to replace the clauses in an interpretational truth theory
that specify the reference and application of the non-logical vocabulary,
or even the clauses for the connectives and quantifiers, with arbitrary
extensional equivalents. If the theorems of the original interpretational
truth theory were true, then the theorems of its arbitrary extensional
equivalent will also be true, despite the fact that knowledge of that
which they state will not suffice for understanding a single sentence of
the object language.

For example, the original truth theory might include the theorem (1),
in which the English sentence on the right-hand side is a meaning-
preserving paraphrase of the Italian sentence on the left, whereas the
arbitrary truth theory might include the theorem (2), in which no
plausible relation of paraphase holds.

1.   'Firenze è una bella città' is true in Italian iff Florence is a
        beautiful city.

2.   'Firenze è una bella città' is true in Italian iff Florence is a
        beautiful city and arithmetic is incomplete.

Morever, a similar result might be derived for every sentence of the
language. Clearly, an arbitrary truth theory that yielded only non-
translational T-sentences like (2) could not be regarded as a theory of
meaning.

One effect of this observation was to stimulate the search for further
constraints to impose on truth theories to ensure that the meta-language
sentences appearing on the right-hand sides of T-theorems were proper
translations of the object language sentences on the left. We may put
aside the many difficulties in formulating such constraints, and simply
assume that the truth theories we are concerned with have translational
T-sentences among their logical consequences. Foster's next point was
that even if we have such a theory, knowledge of that which it states is
not sufficient for understanding the language. The problem is that one
may know *that which is stated* by a translational truth theory without
knowing that the theory is translational. Thus, one who knows a

translational truth theory that issues in theorem (1) may still believe that which is stated by (3), provided he believes that arithmetic is incomplete.

3.   'Firenze è una bella città' means in Italian that Florence is a beautiful city and arithmetic is incomplete.

Moreover, such a person might be in the same position regarding every other object-language sentence. This person would combine true beliefs about truth conditions with false beliefs about meaning, and so would not be a competent speaker. Consequently, knowledge of that which is stated by the most promising of truth theories does not suffice for understanding a language.

Having scotched this attempt to justify the claim that theories of truth are theories of meaning, we are back to our original problem. Given that truth theories do not make claims about meaning, how can they be taken to be theories of meaning? At this point it is natural to try to come up with some other criterion that will provide the needed justification. I will consider three proposals. All of them appeal to claims about what competent speakers believe, know, or understand. Thus, all of them can be seen as attempts to link the notions of truth and meaning by considerations having to do with what is involved in understanding a language.

The first suggestion is that what makes a translational theory of truth a theory of meaning is that knowledge of that which it states is *neces-sary* for understanding the language. The attraction of this proposal is that it avoids Foster-type worries. Truth theories that issue in deviant T-theorems, like (2), are non-translational. Moreover, they don't qualify as adequate theories of meaning because knowledge of that which they state is not necessary for knowledge of meaning. A competent speaker of Italian may be expected to know that which is stated by (1); however, such a speaker is not *required* to know that arithmetic is incomplete, and hence is not *required* to know that which is expressed by (2). Thus, by this criterion, non-translational theories of truth are correctly characterized as not being adequate theories of meaning.

Nevertheless, the criterion does not succeed in justifying the claim that *translational* theories of truth *are* theories of meaning. This can be seen by considering two different versions of the criterion.

Version 1: *T* is an adequate theory of meaning for *L* iff knowledge of that which is stated by all theorems of *T* is necessary for understanding *L*.

Version 2: *T* is an adequate theory of meaning for *L* iff everything necessary for understanding *L* is entailed by *T*.

Version 1 suffers from two different sorts of difficulties. First, there is no non-question-begging reason for thinking that translational theories of truth satisfy it. Thus, even if it were sound, appealing to it wouldn't help. Second, the criterion it states is too weak — so that theories that satisfy it have little claim to being considered theories of meaning in any case. I will take up these difficulties in turn.

First, is it necessary to know everything that is stated, or entailed, by a translational theory of truth in order to understand a language? Here it is important to distinguish knowledge of that which is stated by the translational T-sentences that are theorems of the theory from knowledge of that which is stated by the theoretical axioms that entail those theorems. It is one thing to suppose that all speakers of English know that 'Snow is white' is true in English iff snow is white, that 'Some philosophers are linguists' is true in English iff some philosophers are linguists, and so on. It is quite another to maintain that they must know that which is stated by the Tarski-like theoretical apparatus for dealing with quantified sentences, assignments of objects to variables, and the characterization of the truth of sentences in terms of the satisfaction of formulas by sequences. But this theoretical knowledge is precisely the sort that they have to have, if knowledge of that which is stated by the usual sorts of translational truth theories is to be *necessary* for understanding a language.

The problem is that there is no reason, short of appealing to the very views that we are trying to justify, for thinking that all speakers know this. The point may be illustrated by imagining what it would be like for them to have the knowledge in question. Presumably, to have knowledge of the content of an appropriate truth theory, speakers would have to be regarded as having some internal representation of it. The only way such an internally represented truth theory could help explain knowledge of the object-language is if the system of internal representa-

tions itself was antecedently understood. But once one grants this, there is no need to posit an internally represented truth theory in the first place. If we have an antecedently understood representational system, then understanding a natural language can be seen as involving pairing its sentences with appropriate internal counterparts. For this, any mechanical mapping between the two systems will do — truth theory or no. The point here is not to specify what our psychological model of understanding a language should be, but rather to indicate that there are many possible models, and no reason to give pride of place to those that incorporate Tarski-like theories of truth of the familiar sort. Thus, there is no reason to suppose that knowledge of that which is stated by these theories is necessary for understanding a language.[5]

A related problem with version 1 is that some entailments of translational truth theories are irrelevant to understanding the object language. Suppose, for example, that T entails some translational T-sentence of the form 'S' is true in L iff p. Suppose further that 'q' is any theorem of T, or any logical consequence of 'p'. Then T will entail a non-translational T-sentence of the form 'S' is true in L iff p & q. However, speakers need not know that which is stated by this theorem of T in order to understand the object language.

For these reasons we cannot accept the claim that translational theories of truth satisfy the criterion stated in version 1.[6] Moreover, it is worth noting that version 1 is too weak in any case. According to it a theory T will count as a theory of meaning for a language L iff knowledge of all that is stated, or entailed, by T is necessary for understanding, or knowing the meanings of, the sentences (and other expressions) of L. Suppose, however, that there are many things knowledge of which is necessary for understanding meaning, and that theory T entails some, but not all, of these things. Surely, T cannot then be regarded as an adequate theory of meaning for L. Thus, the criterion should be reformulated in the manner indicated in version 2. However, once this is done, the original Foster-type worries reappear.

To see this imagine that a person, Maria, knows that which is stated by a translational truth theory for an extensional fragment of Italian containing the sentence *Firenze è una bella città*. Suppose further that she believes theorem (1) and all the other translational T-sentences that follow from the theory. Let us grant, for the sake of argument, that all

of this is necessary for understanding the language. However, she believes, wrongly, the claim made by (3). Moreover, she is in a similar position with respect to all the other sentences in the fragment we are considering. Clearly, Maria does not understand the language in question.

Why not? Two natural answers suggest themselves. The first is that she fails to understand the language because she has false beliefs like (3) about what its sentences mean. The second is that she fails to understand the language because she lacks true beliefs like (4), which are necessary for understanding.

4.   'Firenze è una bella città' means in Italian that Florence is a
         beautiful city.

The second of these possibilities is supported by a slight variation in the case. As before, we imagine that Maria believes a true, translational truth theory, including T-sentences like (1), which follow from the theory. However, we need not suppose that she has the false belief (3) about meaning. It is enough that she be uncertain which of (3) or (4) is correct. If she is in this position with respect of every sentence, then she does not understand the language, even though she may have no false beliefs about it. This suggests that what prevents Maria from understanding the language is that she *doesn't know* that which is expressed by (4). If this is right, then knowledge of such claims is *necessary* for understanding. Since these claims do not follow from translational truth theories, such theories fail to state crucial facts needed to understand meaning, and so cannot be counted as theories of meaning.

We have now considered two attempts to justify the claim that appropriately constrained theories of truth are theories of meaning. According to the first, theories of truth provide information *sufficient* for understanding what sentences mean. According to the second, they provide information *necessary·* for understanding. As we have seen, neither attempt is successful in justifying the claim that theories of truth are theories of meaning. The reason these attempts are unsuccessful is that knowledge of truth conditions is too weak to explain linguistic competence. A person who understands the sentences of a language knows more than the conditions in which they are true; in addition, a competent speaker knows such things as the assertions that sentences

are used to make and the beliefs they are used to express. This suggests that we need to appeal to knowledge of propositional attitudes in order to explain linguistic competence.

This brings us to an interesting idea inspired by some remarks of Jim Higginbotham. The idea is that a theory of meaning should specify the information that competent speakers expect each other to possess, simply in virtue of being competent speakers. On this view, speakers have beliefs not only about the semantic properties of expressions, but also about the beliefs that anyone must have in order to qualify as a competent speaker. Thus, a theory of meaning for L is expected to specify a proposition p which is such that speakers believe that one must believe p in order to understand L. Clearly, speakers of Italian do not believe that one must believe that which is expressed by the non-translational T-sentence (2) in order to be competent in Italian. Thus, non-translational truth theories that issue in such T-sentences are correctly classified by the criterion as not being adequate theories of meaning. On the other hand, it seems natural to suppose that competent speakers of Italian do think that one must believe that which is expressed by the translational T-sentence (1) in order be a competent speaker. Thus, it might be thought, translational truth theories qualify as adequate theories of meaning.[7]

This is a mistake, as is shown by the fact that the present proposal suffers from the same problems as the previous one, according to which theories of meaning are theories that specify information knowledge of which is *necessary* for understanding a language. Indeed, the present proposal — according to which theories of meaning are theories that specify information that speakers expect each other to possess simply in virtue of being competent speakers — is tantamount to the claim that theories of meaning are theories that specify information knowledge of which speakers *believe to be necessary* in order to understand a language. Thus, it is not surprising that the same objections that undermined the previous proposal also undermine this one.[8]

The final proposal I will consider is a version of one discussed by Foster, and endorsed by Davidson in his reply to Foster.[9] According to it a proper, translational truth theory T for L is a theory of meaning for L because the result of embedding it in a slightly larger context results

in a theory knowledge of which is sufficient for understanding L. The
larger theory consists of the single sentence

> *Some truth-theory for L, meeting proper constraints, states
> that* . . .

where the dots are filled in by a sentence that formulates the truth
theory T. Here we are to understand the claim that the theory meets
proper constraints as guaranteeing that among its logical consequences
are translational T-sentences in which the meta-language sentence used
on the right-hand side expresses the same proposition as the object
language sentence mentioned on the left. In the case of a translational
T-sentence like (1), the idea is that we should know *both* that which it
states, and that the proposition so stated is expressed by a biconditional
whose right-hand side expresses the same proposition as the object
language sentence mentioned on the left — namely the proposition that
Florence is a beautiful city. From this we are supposed to be able to
deduce (4), and hence to understand the sentence.

It may be helpful to sketch the main stages in this imagined deriva-
tion:

Step 1   Some translational truth theory for Maria's fragment of
         Italian states that . . . (where the dots are filled in by the
         clauses of the theory itself).

Step 2   Some translational truth theory for the language states
         something that entails that 'Firenze è una bella città' is true
         in Italian iff Florence is a beautiful city. (Here entailment
         is understood as a relation between propositions. What is
         claimed is that the proposition expressed by some transla-
         tional truth theory entails the proposition expressed by (1).)

Step 3   Some translational truth theory for the language has as a
         logical consequence a T-sentence that states that 'Firenze è
         una bella città' is true in Italian iff Florence is a beautiful
         city. (Here logical consequence is a relation between sen-
         tences. The truth theory is regarded as a sentence, and it has
         T-sentences among its consequences.)

Step 4   Since the truth theory is translational, the T-sentence men-

tioned in Step 3 is one whose right-hand side means the same as the object language sentence, 'Firenze è una bella città,' mentioned on its left-hand side.

Step 5   Since the T-sentence states that 'Firenze è una bella città' is true in Italian iff Florence is a beautiful city, both the right-hand side of that T-sentence and the object language sentence mentioned on the left must state that Florence is a beautiful city.

Step 6   So, 'Firenze è una bella città' (states) means in the language that Florence is a beautiful city.

There are several points to notice about this derivation. The first is that the compositional character of truth theories, and even the claims made by the T-sentences that follow from them, play no direct role in the derivation. All that is needed for the derivation is that we be provided with sentences of the form

$$\text{'}S\text{' is } F \text{ iff } p$$

where the sentence replacing 'p' is guaranteed to be a translation of the sentence replacing 'S'. Beyond this, it is not important how these sentences are produced, what they say, or even whether they are true. One could, in fact, systematically replace the predicate *is true* in the translational truth theory with the predicate *is false*, without affecting the derivation. Thus the role of truth theories in specifying the knowledge that is supposed to be sufficient for understanding sentences is, on this proposal, essentially heuristic, and in principle dispensable.

The second point to notice about the derivation involves steps 2 and 3. At step 2 we derive a certain proposition from that which is stated by a translational truth theory. The use of indirect discourse here is crucial, and cannot be replaced by any locution that mentions, rather than uses, the truth theory — which is regarded as a sentence, or set of sentences. We may take it that what is stated by the truth theory is a proposition. Step 2 says that this proposition entails a certain other proposition. At step 3, we infer that the entailed *proposition* is expressed by some *T-sentence* that is a logical consequence of the truth theory. To take this step we need to appeal to an implicit principle

connecting the entailment relation holding among propositions with the relation of logical consequence holding among sentences. The required principle is something like the following: *If a sentence s expresses a proposition p, which entails a proposition q, then some sentence s' that expresses q is a logical consequence of s.* However, it is not clear that this principle is correct. For example, if substitution of coreferential proper names *a* and *b* in an extensional sentence preserves the proposition the sentence expresses, then an extensional *sentence* of the form *Rab* will express a *proposition* which entails that something bears *R* to itself even though *no sentence* expressing the latter proposition is a logical consequence of the *sentence Rab* that we started with. If any example of this kind turns out to be correct, then the principle relied upon at step 3 will be invalid, and the derivation will be blocked.[10]

The final point to notice about the deviation involves step 4. At step 3 it is supposed to be established that a certain T-sentence is a consequence of a translational truth theory — that is, of a truth theory that includes among its logical consequences T-sentences whose right-hand sides are translations of the object language sentences mentioned on the left. In order to get to step 4, we need to establish that the T-sentence which states that 'Firenze è una bella città' is true in Italian iff Florence is a beautiful city is translational in this sense. But this is problematic. The fact that the relevant truth theory is translational guarantees that one of its T-sentences for the object language sentence in question is translational, but it does not tell us which. If we knew that for each object language sentence a truth theory would provide *exactly one* T-sentence, we could be sure that the T-sentence we are interested in was translational, and we could move on to step 4. However, we cannot be sure of this, since, in general, truth theories provide *many* T-sentences for each object-language sentence.

For example, suppose that some T-sentence '*S*' *is true iff p* is a logical consequence of a truth theory, and that *q* is any theorem of the truth theory, or any logical consequence of *p*. Then the T-sentence '*S*' *is true iff p & q* is also a logical consequence of the truth theory. Since logically equivalent sentences may differ in meaning, many of these T-sentences will *fail* to be translational. But this means that we have no guarantee that the particular T-sentence which states that 'Firenze è una bella città' is true in Italian iff Florence is a beautiful city is translational. As a result, we cannot move to step 4, and the derivation is blocked.

Until this problem is solved, we cannot accept the final Foster-Davidson proposal for justifying the claim that familiar, Davidsonian-style truth theories may serve as theories of meaning.[11] Since this leaves us with *no* justification for the claim, I suggest that we reject it, and adopt a different view of theories of meaning. The view I advocate recognizes that sentences have meanings, and requires a theory of meaning to tell us what each sentence means by pairing it with its correct meaning, or meanings. However, it is *not* crucial that the theory specify the meanings of sentences in a way that would allow someone who did not know the language already to learn it. Rather, a theory may associate with each sentence a *theoretical* description of the proposition, or propositions, it expresses. It is enough that these descriptions in fact pick out the right propositions — whether or not they do so in a way that would allow someone who did not understand the language initially to become a competent speaker. If this is right, then the idea that a semantic theory *must* state facts knowledge of which would explain what it is to understand a language should *not* be taken as an apriori constraint on theories of meaning in general. Rather, it should be seen as a (so far) unsuccessful attempt to justify the claim that a theory that does *not* assign meanings to sentences may nevertheless serve as a theory of meaning. Once we stop insisting on theories of this sort, we no longer *need* to identify the theorems of a semantic theory with the beliefs that would suffice for semantic competence.

How might one construct a theory that pairs sentences with their meanings? For purposes of our discussion, we can set aside subtleties and technicalities such as those involving indexicality, and temporal modification. Given these simplifications, we may think of the meaning of a sentence as the proposition it expresses. Thus, our question becomes 'How might one construct a theory that pairs sentences with the propositions they express?' One familiar suggestion is that propositions are sets of circumstances assigned to sentences by a theory of truth with respect to a circumstance. Given such a theory, one can define the proposition expressed by a sentence as the set of circumstances in which it is true.

The most familiar theories of this sort are those in which cirucmstances are identified with possible worlds, and the proposition expressed by a sentence is identified with the set of worlds in which it is true. However, these identifications cannot be correct, for if they were,

then necessarily equivalent propositions would have to be identical, which they are not — as is shown by the many examples in which a person asserts or believes p without asserting or believing something necessarily equivalent to p. Elsewhere I have argued that the problem cannot be solved by invoking truth-supporting circumstances that are more finely grained than possible worlds.[12] So long as one maintains the usual recursive clauses in a truth theory, one can reconstruct essentially the same problem no matter how fine-grained one makes the truth-supporting circumstances.

Like a number of writers, I believe that in order to come up with a conception of propositions that is fine-grained enough to serve as objects of the attitudes, we need to think of them as encoding both the syntactic structure of the sentences that express them, and the semantic contents of subsentential constituents. If we do think of propositions in this way, then we can easily see how sentences with significantly different structures can express different propositions, even though they are true in the same circumstances. For present purposes, it is not important which of the many different ways of implementing this idea we select. For example, if we follow Russell's lead and build propositions out of objects, properties, and propositional functions, our semantic theory might issue in theorem (5), which provides a theoretical description of the proposition expressed by the sentence, 'Someone loves Nixon'.[13]

5.   'Someone loves Nixon' expresses the proposition which is the ordered pair whose first coordinate is the property of being sometimes true, and whose second coordinate is the propositional function g which assigns to any person p the proposition which is the ordered pair whose first coordinate is the relation of loving, and whose second coordinate is the ordered pair the first coordinate of which is p and the second coordinate of which is Nixon. A proposition of this form is true iff some object o is such that $g(o)$ is a true proposition.

We may suppose, for the sake of argument, that theorem (5) is correct, and hence that the description it gives of the proposition expressed by the sentence 'Someone loves Nixon' is accurate. Nevertheless, it does

not specify the proposition in a way that would allow someone ignorant of the language to understand the sentence. Moreover, knowledge of that which it expresses does not seem to be either necessary or sufficient for understanding the sentence.

It is instructive to compare (5) to (6).

6.   'Someone loves Nixon' expresses the proposition (means) that someone loves Nixon.

Both (5) and (6) specify the proposition expressed by the English sentence. However, the terms used to refer to the proposition are different in the two cases. In (6) the relevant term is the expression *the proposition that someone loves Nixon*; in (5) it is a complex description that identifies the proposition by specifying its structure and constituents. Since the two coreferential terms are not synonymous, (5) and (6) say different things. Thus, even if knowledge of meaning amounts to knowing (6), it does not amount to knowing (5). Although claims like (5) are genuine theorems of a semantic theory, semantic competence is not the result of knowing that which they express.

How serious is it that semantic theories of this sort do *not* issue in theorems of the kind illustrated by (6), and do *not* provide an explanatory account of semantic competence? In my opinion this is not a problem. It may be argued that it is necessary and sufficient to understand the meanings of sentences that one have propositional knowledge of the sort illustrated by (6). But even if this is true, attribution of such knowledge to speakers in no way *explains* semantic competence. The reason it doesn't is that, in many cases, the explanation of how it is we come to know or believe a semantic fact of the form '*s*' *means that s* must appeal, among other things, to the fact that we understand *s* and accept various sentences containing it. If this is right, then understanding meaning cannot be *reduced* to any conceptually prior notion involving propositional knowledge of semantic facts.[14] Thus, no account of such facts, and in particular no theory of meaning, whatever its form, can ever serve as an explanation of what is involved in understanding meaning. If one wants a deep, and revealing account of such understanding, one must look elsewhere.

NOTES

[1] This is a revised version of a talk written for a symposium of the American Philosophical Association in San Francisco in March, 1991 with James Higginbotham and Mark Richard. Special thanks are due to Julius Moravcsik, who read my talk for me at the symposium, after it became impossible for me to attend.

[2] Donald Davidson, "Truth and Meaning," *Synthese* vol. 17, No. 3, 1967; reprinted in *The Philosophy of Language*, A. P. Martinich (ed.), Oxford University Press, New York 1985, p. 75. All page references will be to the Martinich volume.

[3] Page 77.

[4] J. A. Foster, "Meaning and Truth Theory," in *Truth and Meaning*, Gareth Evans and John McDowell (eds.), Oxford University Press, Oxford, 1976.

[5] This point is made in my "Semantics and Psychology," in *The Philosophy of Linguistics*, J. J. Katz (ed.), Oxford University Press, 1985.

[6] If *all* is replaced by *some* in version 1, translational truth theories would satisfy the modified criterion. However this is of no real help since the modified criterion is clearly too weak. In this connection, it is worth noting that a variant of the objection just given to version 1 could be recreated so as to apply to the *some-variant* of version 1. Let T be a translational truth theory with clauses of the form (i) 'P' *applies to o iff o is F*, together with the usual clauses for quantifiers and connectives. Let T' be just like T except than in place of (i) T' has (ii) 'P' *applies to o iff o is F & Q*, where 'Q' is either a theorem of T or a logical consequence of *o is F*. (Similar variations could be made in the clauses for quantifiers and connectives.) Since the theorems of T = the theorems of T', both equally satisfy the *some-variant* of version 1. But clearly T' is not an adequate theory of meaning. Thus the criterion is too weak.

[7] When I wrote the original version of this talk, I assumed that Higginbotham's view was (i) that a theory of meaning is one that specifies the linguistic information knowledge of which speakers' believe to be necessary for semantic competence, and (ii) that translational theories of truth do just this. However, after reviewing the revised version of Higginbotham's paper (which incorporates some explicit responses to my original talk) I am no longer convinced that he regards theories of truth as theories of meaning at all. Instead, he seems to regard them as specifying information that is *incorporated* into theories of meaning, and he seems to suggest that theorems of such theories include those of the following form, where '. . .' is filled in by a translational T-sentence.

(i)   A speaker of L is expected to know that . . . solely in virtue of being a competent speaker.

One might argue that such theorems provide the information needed to understand the meanings of object language sentences by reasoning as follows: One who understands a sentence s knows that which is stated by the appropriate instance of Schema M, i.e. *s means in L that p*. Since one also knows the corresponding instance of Schema TM, one who understands s will be expected, simply in virtue of that understanding, to know that which is stated by the corresponding instance of Schema T, *s is true in L iff p*. Of course, one may know that which is stated by other claims of the form, *s is true in L iff q*. But one will never be *expected* to know that which is stated by these additional T-sentences, *solely in virtue of being a competent speaker.* Indeed, the *only* T-sentences concerning s which state things one is *required* to know, by virtue of linguistic competence alone, are translational T-sentences. Thus if one is given a theorem of the form (i), one will be able to draw the correct conclusion of the form, *s means in L that p*.

Since I don't have space to discuss this view in detail, I will confine myself to two

general comments. First, this conception of a theory of meaning has very little to do with theories of truth. In effect, truth theories become mere heuristics used to come up with the translational T-sentences to be substituted into the blank in (i). But if all we need is a theory that provides meta-language translations of object language sentences, we could obtain the needed substituends by dispensing with a theory of truth and employing an explicit theory of translation from object language to meta-language. (Given '*S*' in *L means the same as* '*p*' in *M*, we could stipulate that the relevant instance of (i) is either '*S*' *means in L that p* or '*S*' *is true in L iff p*.) Second, one might strengthen the connection between the theory of meaning and theories of truth by stipulating that the theorems of the theory of meaning are not limited to instances of (i) involving translational T-sentences, but also include instances obtained by replacing '. . .' with axioms of the truth theory. That this is Higginbotham's view is suggested by the following remark. "*What I have offered is the view that one will understand Gianni* [a speaker of the language] *when one knows what he, Gianni, knows and is expected to know about reference and truth.* **The general principles** [my emphasis] *and certain theorems of a theory of truth for Gianni will figure in one's knowledge about him.*" But if this is the view, then the usual objections to the claim that speakers know the theoretical apparatus employed by truth theories will come into play. (See in particular the discussion in the following footnote.)

A different way of using some of Higginbotham's observations to defend the semantic importance of translational theories of truth is expressed by the proposal (ii).

(ii)   A theory of truth *T* is an adequate theory of meaning for a language *L* iff some subset *C* of propositions expressed by theorems of *T* satisfies the following condition: *It is sufficient to understand L that for each sentence s of L there is a member p of C such that one knows that p is the only proposition stating truth conditions of s knowledge of which is necessary in order to understand L.*

It is at least arguable that translational truth theories satisfy the italicized condition. Nevertheless, there are potential problems with this proposal. First, it is not obvious that for each sentence there is exactly one proposition stating its truth conditions that one must know in order to understand the language. Second, there is reason to doubt that the condition stated in the proposal is strong enough to justify characterizing a theory that satisfies it as an adequate theory of meaning. Even if a truth theory *T* satisfies the condition, nothing in it specifies that any of its theorems are such that knowledge of them are necessary for understanding *L*. Further nothing tells us that if some are necessary, which those are. Third, if a truth theory *T* satisfies the above condition, then any truth theory *T'* whose axioms are logically equivalent to *T* will satisfy it as well. Nevertheless, there are cases in which we would not be willing to regard both *T* and *T'* as adequate theories of meaning for *L*.

[8] If anything, the objections to the present proposal are even stronger. This is seen by considering the version of the proposal which claims that knowledge of each theorem of an adequate theory of meaning is necessary in order to understand the language. As before, we must distinguish between two claims: (i) that speakers believe that in order to be competent one must know the contents of those theorems of the theory that are translational T-sentences; (ii) that speakers believe that in order to be competent one must know the contents of the truth theoretical axioms used to derive such theorems — for example the contents of the usual Tarski-like theoretical apparatus. It seems obvious that it is *not* the case that speakers of English typically believe that knowledge of such technical, truth-theoretic apparatus is necessary in order to understand the language. Even if — what seems highly unlikely — some compelling theoretical argument could be given to establish that they have, in fact, unconsciously internalized such apparatus, this

would still not show that they *believe* that the internalization of this apparatus is necessary in order to be a competent speaker. Thus it seems evident that translational theories of truth do not satisfy the criterion given in this version of the present proposal.

In addition, this version of the proposal is too weak in precisely the way that version 1 of the previous proposal was too weak. To rectify this problem it would have to be restated along the lines of version 2 above — so that an adequate theory of meaning is required to entail *everything* that speakers expect one to know in order to be competent. But then the Maria example can be invoked to show that translational truth theories do not satisfy this criterion.

⁹ Foster, pp. 20—1. Donald Davidson, "Reply to Foster," in *Truth and Meaning*, p. 36.

¹⁰ It is important to emphasize that counterexamples to the principle connecting the entailment relation holding among propositions with the relation of logical consequence holding among sentences do *not* depend on the assumption that substitution of coreferential names preserves the proposition expressed. For example, any account that treats names (indexicals) as rigid designators, and that takes the propositions expressed by extensional sentences differing only in the substitution of coreferential names (indexicals) to be modally equivalent, will provide examples in which a proposition B is a necessary consequence of a proposition A, even though no sentence expressing B is a logical consequence of any sentence expressing A. More generally, conceptions of propositions as sets of truth-supporting circumstances can be expected to have this result, as will many conceptions of propositions that abstract away from the notation used to express them.

¹¹ One possible strategy for dealing with this problem is to incorporate within truth theories some restrictive characterization of the notion of a cannonical derivation. The idea is to specify a procedure for picking out, for each object language sentence s, a single T-sentence as being appropriately translational. A translational truth theory would be one whose cannonical T-sentences are translational. The claim would then be that it is sufficient for understanding L that one know that which is stated by *Some translational truth theory for L, with cannonical proof procedure __, states that* . . . (where the blanks are filled in by the proof procedure specified by the theory, and the axioms of the theory, respectively.)

Although this strategy may be worth pursuing, there are three potential difficulties that should be addressed. First, in order to get from Step 1 to Step 3 in the above derivation one would have to specify a notion of a *proposition Q* being a cannonical consequence of a *proposition P* in a way that parallels the cannonical derivation relation among sentences of the truth theory — something it is not immediately clear how to do. Second, there is a problem of motivation. The appeal to cannonical derivations is something imposed on truth theories from without. It is not needed for truth theories to be true, but rather is a mechanism for specifying translations. But if translations (and disquotation) are the crucial things for theories of meaning and understanding, why isn't it enough for a theory of meaning to consist in a compositional translation theory from L into M that issues in infinitely many instances of '*S*' *in L means the same as* '*P*' *in M*, plus a single axiom schema '*S*' *in L means that P*, where an instance is obtained by pairing the sentence replacing '*S*' with its translation into M? Third, there is a problem of implementation. If one looks at truth conditional analyses of particular constructions, for example propositional attitudes, one often finds that though the resulting T-sentences may give the truth conditions of object language sentences, they do not provide acceptable translations. This is not a problem so long as the truth theories are only required to be true. However, it may turn out to be an intractable problem if their status as theories of meaning require them to be translational.

¹² See my "Lost Innocence," *Linguistics and Philosophy*, Vol. 8, No. 1, 1985, 59—71; "Direct Reference and Propositional Attitudes," in *Themes from Kaplan*, J. Almog, J.

Perry, and H Wettstein, (eds.), Oxford University Press, 1988; and "Direct Reference, Propositional Attitudes, and Semantic Content," *Philosophical Topics*, Vol. 15, No. 1, 1987, 47—87, reprinted in *Propositions and Attitudes*, N. Salmon and S. Soames, (eds.), Oxford University Press, 1988.
[13] This example is taken from my "Semantics and Semantic Competence," *Philosophical Perspectives*, Vol. 3, *Philosophy of Mind and Action Theory*, 1989, 575—596.
[14] See "Semantics and Semantic Competence," pp. 587—91.

Department of Philosophy
Princeton University
Princeton, NJ
USA