

Truth and Meaning Redux: Reply to Soames

1

Forty-two years ago, Donald Davidson, in “Truth and Meaning” (1967), suggested that insight into meaning might be gained by reflection on how to construct and confirm a Tarski-style truth theory for a natural language. The suggestion has been influential, but it has not been well understood.

In a recent paper, Scott Soames argues that this project has now run its course (2008). He argues that Davidson’s two rationales for his project were unsuccessful and that what he sees as the most promising recent attempt along the same lines, by Jim Higginbotham, is likewise unsuccessful. It should be replaced, he argues, with an older style theory that assigns entities to expressions, and which aims to explain the meaning of complex expressions by showing how to recursively construct appropriate entities to assign to them on the basis of assignments to their significant parts. Davidson’s own animadversions against such theories, he argues, while not without force, can be avoided by appeal to some ideas from Wittgenstein’s *Tractatus Logico-Philosophicus* (1961).

This paper argues (a) that Soames has misunderstood Davidson’s project, (b) that once we clarify the project, we can see it escapes his objections unscathed, and (c) that the appeal to meanings as entities does no work in Soames’s alternative approach to semantics and is no advance over truth-theoretic semantics.

The program of the paper is as follows: we begin with Soames’s characterization of Davidson’s project, and a brief indication of how he mischaracterizes it. We then turn to Soames’ evidence for his interpretation. This involves both an account of the historical context, and a reading of a key passage in “Truth and Meaning.” We show that his reading of the historical context does not fit Davidson’s project and that placing the key passage in its proper context in “Truth and Meaning” as well as the rest of Davidson’s corpus shows Soames’s interpretation to be incorrect. To this end, we look closely at how Davidson sets the stage for his suggestion that constructing and confirming a truth theory for a natural language is a way to pursue the project of devising a compositional meaning theory. We consider several stages in his rejection of the appeal to assigning meanings, construed as entities, to words or sentences as a way of providing a compositional meaning theory for a language. We pay particular attention to an illustration that Davidson gives using a simple reference theory to show why assigning entities to expressions is not in itself useful and how they can be dispensed with, and we also show how this illustration presages Davidson’s own suggestion for how to use a truth theory in pursuit of a compositional meaning theory. We look at why he rejects assigning meanings as entities to sentences, and note a key observation he makes

about what the effective goal is of a compositional meaning theory. In the light of this, we review the passage Soames cites and his criticisms of Davidson's project. Lastly, we consider Soames's own positive proposal, and argue it is no advance over Davidson's, and is instead a step backwards.

There are two issues to separate at the outset, one interpretive, one not. First, is Soames right about what Davidson's project is and that as characterized it cannot be carried out? Second, is Soames right that if the project he characterizes cannot be carried out, no broadly Davidsonian program of truth-theoretic semantics is tenable? We will defend an interpretation of Davidson's project significantly different from Soames's. But even were our interpretation mistaken, we still maintain it represents a Davidsonian project that escapes Soames's criticisms.

2

Soames begins with a mischaracterization of Davidson's project which informs the rest of his paper, i.e., that it was an attempt "to explain knowledge of meaning in terms of knowledge of truth conditions" (p.1). He continues, "[f]or Davidsonians, these attempts take the form of rationales for treating theories of truth, constructed along Tarskian lines, as empirical theories of meaning" (p.1). Both claims are mistaken.

The first characterization is not one Davidson ever gave nor is it one he would have endorsed. Davidson's project was to devise a satisfactory theory of *meaning*, ultimately, as he states it in the preface to *Inquiries into Truth and Interpretation* (2001), to answer the question: "What is it for words to mean what they do?" This project decomposes into two parts. One part concerns how the meanings of complex expressions, and ultimately sentences, depend on the meanings of words. The other part concerns understanding what it is for primitive expressions to mean what they do. Davidson suggested progress could be made on the first project by *indirection*, by reflection on what we learn from constructing a truth theory for a language subject to constraints designed to ensure it meets Tarski's Convention T, or its analog for natural languages with context sensitive elements (henceforth, 'Convention T' abbreviates this disjunction). Progress on the second project was to be attained in turn by reflection on how one could empirically confirm a truth theory that, meeting appropriate constraints, would have identifiable theorems from which one could read off what each sentence of the object language means, or, more precisely, for any utterance of a sentence in the language, what it means.

Clearly, neither part of this project is about explaining *knowledge* of meaning, and so neither is about explaining knowledge of meaning in terms of knowledge of truth conditions. In addition, it is clear even from this brief and general characterization that it is not *just* knowledge of what a truth theory states that is to provide insight, but also the

knowledge that arises from reflection on how to construct and confirm a theory that meets Convention T. We will return to this point at greater length.

With respect to the second part of his project, it is, first, not easy to see how giving such a rationale would amount to *explaining knowledge of meaning in terms of knowledge of truth conditions*. Second, it turns out that it is not the truth theory *per se* that is an empirical theory of meaning. This is clear by Davidson's middle period, and if we can take his retrospective remarks at face value, it was how he was thinking about it even in "Truth and Meaning," for which there is also evidence in how he develops the project there.

3

Why does Soames think Davidson intended to explain knowledge of meaning in terms of knowledge of truth conditions? He provides a just-so story about the origins of Davidson's project in the theory of meaning. Soames suggests Davidson's idea fell out of a combination of Quinean skepticism about meaning and reflection on Tarski's work on the semantic conception of truth:

- [i] For those laboring under the Quinean legacy of skepticism about analyticity, synonymy, and meaning, the idea afoot was that extensional notions from the theory of truth and reference were respectable, whereas intensional ones from the theory of meaning were not.
...
- [ii] Since Tarski's seminal work in the 1930s, it has been common place to view an interpreted formal language as the result of adding a model, plus a definition of truth-in-a-model, to an uninterpreted formal system, thereby arriving at an assignment of truth conditions to every sentence. But if truth theories can be used in this way to *endow* sentences with meaning, then surely, it seemed, they can also be used to *describe* the meanings of already meaningful sentences—provided, in the case of natural language, that we are clever enough to find the requisite logical forms to which to apply them. (p. 2)

His suggestion is that Davidson, like Quine, was a nihilist about meaning — not just about the utility of introducing entities called meanings, but also about the intelligibility of the concept of meaning. Yet at the same time, he was to have thought that Tarski's work provided a way of endowing sentences with meaning by adding a model to a formal system and defining truth-in-a model. But if one thought the first, why on earth think the second? The second can be true only if the first is false.

Davidson was not in fact a nihilist about meaning, even if he held that meanings, construed as entities, have no useful role to play in the theory of meaning. Perhaps it is a conflation of Davidson's rejection of *meanings* with a rejection of meaning that has led

some commentators—e.g. (Chihara 1975; Stich 1976; Katz 1982, pp. 183-5; Soames 1992; Cummins 2002; Glock 2003, pp. 142ff.) and now (Soames 2008)—to think Davidson was rejecting the very project he announces at the beginning of “Truth and Meaning” (p. 17; page citations for Davidson are to (Davidson 2001) when essays are reprinted there) rather than its pursuit by indirection.¹

[a] It is conceded by most philosophers of language, and recently by some linguists, that a satisfactory theory of meaning must give an account of how the meanings of sentences depend upon the meanings of words.² Unless such an account could be supplied for a particular language, it is argued, there would be no explaining the fact that we can learn the language: no explaining the fact that, on mastering a finite vocabulary and a finitely stated set of rules, we are prepared to produce and understand any of a potential infinitude of sentences. I do not dispute these vague claims, in which I sense more than a kernel of truth. Instead I want to ask what it is for a theory to give an account of the kind adumbrated.³

Later in “Truth and Meaning,” Davidson says, “the task of a theory of meaning as I conceive it is not to change, improve, or reform a language, but to describe and understand it” (p. 29).⁴ Quine’s project in *Word and Object* (Quine 1960), in contrast, was to provide a scientifically respectable replacement for the ordinary notion of meaning, one constructed out of the concept of stimulus synonymy. Davidson is not engaged in this sort of project, though he was importantly influenced by Quine, particularly in regarding the empirical content of theories of meaning as exhausted by the evidence available to an interpreter, and, specifically, to a radical interpreter. But what he took from Quine he transformed for his uses in pursuit of an understanding of what it is for words to mean what they do, where ‘mean’ is used in its ordinary sense.

¹ A different misunderstanding is expressed in (Horwich 2005, p. 4 & ch. 8) who suggests Davidson aimed to analyze sentence meaning in terms of truth conditions. Why this is a mistake will become clear in the course of our discussion. This misunderstanding is also widespread. Burge gives it as if it were the standard account of Davidson in (Burge 1992, pp. 20-1).

² In his initial characterization of the project, Davidson talks of meanings in the plural. What emerges from his discussion is that the kernel of truth in the following claim does not require the assignments of meanings to sentences or words.

³ The same project is his theme in “Theories of Meaning and Learnable Languages” (1966), published the year before “Truth and Meaning,” and in “Semantics for Natural Languages” (1970) read first in 1968, the year after publication of “Truth and Meaning” (1967).

⁴ In “Semantics for Natural Languages (1970, p. 62), Davidson says, “Making a systematic account of truth central in empirical semantics is in a way merely a matter of stating old goals more sharply.” He goes on to say, “Some problems that have dominated recent work on semantics would fade in importance: the attempt to give ‘the meaning’ of sentences...”; but here the inclusion of ‘the meaning’ in quotation marks serves to indicate he has in mind producing entities that will serve as meanings and a standard for translation; it is also Davidson’s view that the standard for translation emerges only in the context of radical interpretation.

We also note that Davidson was not thinking of truth-in-a-model in invoking Tarski, but instead Tarski's work on absolute theories of truth (1983). There is no suggestion in Tarski that we endow sentences with meaning by defining a truth predicate for the language of which they are a part. The idea went the other way: if we know the definition matches object language sentences with meta-language sentences that translate them, we know it is materially adequate, if it is formally correct. Davidson would have had to be seriously confused about Tarski to have reasoned in the way outlined in passage [ii].

4

It is no accident that Davidson is often interpreted as pursuing what might be called a Replacement Strategy. There are passages in "Truth and Meaning" which strongly suggest this, and Soames quotes one of the most striking (p. 24).

[b] There is no need to suppress, of course, the obvious connection between a definition of truth of the kind Tarski has shown how to construct, and the concept of meaning. It is this: the definition works by giving necessary and sufficient conditions for the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence. To know the semantic concept of truth for a language is to know what it is for a sentence—any sentence—to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language. This at any rate is my excuse for a feature of the present discussion that is apt to shock old hands; my freewheeling use of the word "meaning," for what I call a theory of meaning has after all turned out to make no use of meanings, whether of sentences or of words. Indeed, since a Tarski-type truth definition supplies all we have asked so far of a theory of meaning, it is clear that such a theory falls comfortably within what Quine terms the "theory of reference" as distinguished from what he terms the "theory of meaning." So much to the good for what I call a theory of meaning, and so much, perhaps, against my so calling it.

Is this not evidence that what Davidson calls a theory of meaning is one *in name only*, isn't he proposing to *replace* the old pursuit of a theory of meaning with the new, more tractable and intellectually respectable project of constructing a truth theory for a language? Does this not establish, as Soames puts it, that "Davidson thought that systematic knowledge of truth and reference could do all legitimate work for which we need a notion of meaning" (p. 3)? Does it not show Davidson's "strategy was to embrace Quine's rejection of analyticity, synonymy, and our ordinary notion of meaning, substituting knowledge of truth and reference for knowledge of meaning—whenever there was something genuine to be captured" (p. 3)?

To answer these questions, we must first place the passage in its context.

Here's "Truth and Meaning" in a nutshell: Davidson begins with the project of how to provide a compositional meaning theory for a natural language. He rejects the utility of an appeal to meanings as entities. He offers an initial argument against any such appeal that rests on the observation that assigning meanings to expressions and even concatenation leaves the difference between a list and a sentence unexplained. He develops the point by drawing some lessons from reflection on a simple reference theory, viz., that it is not necessary to assign an object to every significant part of a complex referring term to provide an assignment of referents to any of an infinite class of recursively generated referring terms. He extends this idea to the rejection of assignments of meanings as the referents of sentences (so that, although not all words are assigned meanings, every sentence is) primarily through an appeal to an argument widely known as 'The Slingshot'. Then he despairs of an appeal to a theory that generates theorems utilizing 'means that', if neither the sentences following it nor terms of the form 'that p' are referring terms. He observes that the desired work would be achieved if we can match each object language sentence with a meta-language sentence that 'gives its meaning', for that is in effect what generated theorems of the form 's means that p' do. He then proposes that a truth theory can be used effectively towards this end. We examine these stages in some detail before returning to [b].

5

In passage [a] above, Davidson states his ostensible goal in "Truth and Meaning," namely, *to inquire into what it is to give an account of how the meanings of sentences depend on the meanings of words*. In the first few pages he considers and rejects the appeal to assigning meanings, construed as entities, to words as a way of explaining sentential meaning in terms of word meaning.

Suppose the meaning of 'Theatetus' is Theatetus and the meaning of 'flies' is the property of flying. So far as this goes, 'Theatetus flies' might as well be interpreted as a list: Theatetus, the property of flying. Since concatenating 'Theatetus' with 'flies' is semantically significant, to be consistent, we must assign to concatenation itself an entity, say, a relation of instantiating. But this only increases the list: Theatetus, the relation of instantiation, the property of flying.⁵

Davidson's critical point emerges clearly in the context of a theory of complex singular terms. The expression 'The father of David' concatenates 'The father of' (treated as primitive here) with 'David'. If 'The father of' refers to a function, how can this help

⁵ Alternatively, assign an individual concept to 'Theatetus' and read 'Theatetus flies' as the list: the individual concept of Theatetus, the property of flying, and instantiation. This only renders the problem more evident.

determine what the complex refers to? We might say: 'the father of' refers to that function such that its value for argument x is the father of x . Clearly, reference to the function plays no role in the explanation; the same information is captured by: 'the father of' concatenated with a singular term t refers to the father of what t refers to, except that here we have a rule that determines, relative to assignments of referents to primitive expressions, what any expression formed from concatenating 'the father of' with a singular term refers to.⁶ This rule uses 'the father of', of course, but it turns out that this is unavoidable. As Davidson says,

[c] ...the task was to give the meaning of all expressions in a certain infinite set on the basis of the meaning of the parts; it was not in the bargain also to give the meanings of the atomic parts. On the other hand, it is now evident that a satisfactory theory of the meanings of complex expressions may not require entities as meanings of all the parts. It behooves us then to rephrase our demand on a satisfactory theory of meaning so as not to suggest that individual words must have meanings at all, in any sense that transcends the fact that they have a systematic effect on the meanings of the sentences in which they occur. (p. 18)

In the case of our theory of reference, we can state a simple criterion of success: a theory should entail every sentence of the form ' t refers to x ,' where ' t ' is replaced by a structural description of a singular term and ' x ' by that term itself. This condition should look familiar; it is an analog of Tarski's Convention T, but for a theory of reference, restricted to cases where the object language is embedded in the meta-language. Generalizing, we seek a theory that satisfies Convention R:

An adequate theory entails every sentence of the form ' t refers to x ,' where ' t ' is replaced by a structural description of a singular term and ' x ' is replaced by a term in the meta-language that translates it.

Having gotten this far, we are only a tiny step away from stating explicitly what each singular term in our fragment means. This is not a step Davidson took, but it is

⁶ It has been suggested to us that the problem raised here applies to the approach we sketch below. The line of thought seems to be this: "You say that the theory that assigns the function and the argument to expressions does not say how to put them together to get the meaning of 'Theatetus flies'. But the sorts of theories you recommend consist of axioms and rules and the theory doesn't say how to put them together to get theorems. So it's the same thing!" The *tu quoque* charge misunderstands the point of the argument, which is that (i) the information we need to understand is not given by the assignment of entities to expressions; and (ii) knowledge of those assignments, excepting in the case of ordinary referring terms, is not included in the information that is needed. Furthermore, it is not, as we will explain, part of our view that an axiomatic reference or truth theory contains all the information that is needed, though it is part of our view that information about assignments of meanings to all expressions in a language is not needed.

instructive to follow out its implications. Restricting attention for the moment to a theory the language of which embeds the object language, a syntactic criterion suffices for identifying theorems that satisfy (R). Given an adequate theory, it will entail, *inter alia*,

'The father of' \wedge 'David' refers to the father of David.

Since the expression used on the right of 'refers to' translates the one mentioned on the left, we can replace 'refers to' with 'means' to derive:

'The father of' \wedge 'David' means the father of David.

Thus, omitting for now an account of ' \underline{x} means \underline{y} ', we have a way of generating, for an infinite class of expressions built up systematically from a finite vocabulary, an explicit statement of what each means.

Looking at languages we do not already know will allow us to register further points that pave the way for understanding Davidson's proposal to use a truth theory in pursuit of the project of understanding how the meanings of complex expressions depend on the meanings of their significant parts.

Consider an object language fragment of French consisting of singular terms 'Marie', 'Jean', and any concatenation of 'La mère de' with a singular term. Call this infinite fragment 'L'. Theory S for L in an English meta-language consists of axioms (1)-(3):

1. 'Marie' refers in L to Mary
2. 'Jean' refers in L to John
3. For any singular term \underline{t} in L, 'La mère de' \wedge \underline{t} refers in L to the mother of what \underline{t} refers to in L.

Its (informally stated) rules of inference are: (I) From (3), any instance may be inferred; (II) from any sentences of the form ' \underline{t} refers in L to \underline{y} ' and any sentence of the form $S(\text{what } \underline{t} \text{ refers to in L})$, $S(\underline{y})$ may be inferred. A *canonical reference theorem* of S is any theorem derived from (1)-(3) using (I) and (II) which is an instance of ' \underline{t} refers to \underline{y} ,' in which ' \underline{y} ' is replaced by a meta-language referring term that does not include 'refers to'. Since in each base axiom the term used on the right of 'refers to' translates the one mentioned on the left, and since in (3) 'the mother of' in the meta-language translates 'La mère de' in L, in each canonical reference theorem the meta-language expression used on the right of 'refers to' translates the expression mentioned on the left.

Any theory of reference which satisfies this requirement on its axioms we shall say satisfies Convention A. That S satisfies Convention A, given rules (I)-(II), suffices for it to satisfy Convention R. At this point, we add another valid rule of inference: (III) From a canonical reference theorem of the form ' \underline{t} refers in L to \underline{y} ' infer ' \underline{t} means in L \underline{y} ' — call any theorem so derived a *canonical meaning theorem*.

If we know [KS]

- (i) the axioms of S, as stated in (1)-(3),
- (ii) what each axiom of S states and that each states what it does,⁷
- (iii) in base axioms the term used on the right of 'refers in L to' translates the one mentioned on the left, and in (3) 'the mother of' in the meta-language translates 'La mère de' in L,
- (iv) rules of inference (I) and (II),
- (v) what a canonical reference theorem is,
- (vi) rule of inference (III).

we are in a position to know for any expression of L what it means in this sense: we are thereby in a position to know for an arbitrary singular term t of L its corresponding canonical meaning theorem. Moreover, we can come to know this on the basis of being able to derive this canonical meaning theorem from axioms that we know give the referents of object language terms using expressions in the meta-language that translate them. Thus, we see how the meanings and referents of the contained terms contribute to determining the meanings and referents of the complex expressions, given how they are combined in them.

For primitive names, the theory *states* their referents and *shows* their meanings (relative to the knowledge that it satisfies Convention A). For the primitive recursive term 'la mère de', the theory *gives* a rule that *says* what the reference of the expression is *in terms of* the referent of another, and this rule *shows* what the expression means *by* using a expression that translates it, in the same grammatical role to the right of 'refers to'.

In S we find in microcosm Davidson's suggestion for how to pursue the project of giving a compositional *meaning* theory using materials drawn from the theory of reference. We need to unfold a bit more of the story before extending the idea to the whole of a language, and not just a fragment involving complex referring terms.

5

Reflection on S shows us that we do not need to assign a meaning to every word in order to fix the referents for complex referring terms. This suggests we do not need to assign meanings to every word to assign meanings to sentences, though we might retain the idea that each sentence has a meaning, construed as an entity, associated with it. A simple way of implementing this idea is to treat sentences as referring terms

⁷ We need both (i) and (ii) because (ii) does not ensure that we know the syntactic forms of each axiom, which is required if we are to use them to derive canonical reference theorems, as opposed to knowing what they are by way of names or descriptions. We need to know what the forms of the axioms are to use the canonical proof procedure to isolate the right set of theorems.

that refer to their meanings, and then extend a reference theory to cover sentences so understood. This is to take predicate terms generally to be functional terms and sentences to be complex referring terms gotten from supplying an argument term for the functional term which refers to the meaning of the sentence. In line with the observations above, however, we need not take seriously the idea that the functional terms refer to anything (anymore than we did with 'the father of'). All we need is a rule that takes us from an expression mentioned to a term used that refers to what the first expression refers to. For 'rouge' we give the axiom:

For any referring term t , t 'est rouge' refers to red(what t refers to).⁸

Davidson sought to scotch this proposal as well with an argument later dubbed 'The Slingshot' (by (Barwise and Perry 1981). This argument (we have argued elsewhere (Lepore and Ludwig 2005, pp. 49-55)) is guilty of equivocation, and ultimately relies on an assumption that is supposed to be its conclusion. But the proposal should be scotched anyway. What does the work in the proposal is that meta-language referring terms are sentences alike in meaning to their corresponding object language sentences. The only role reference plays in the theory is to facilitate matching object language sentences mentioned with used meta-language language sentences that translate them. Since this effect can be achieved without positing gratuitous meanings, it is clearly preferable.

6

All this is still background for the proposal Davidson eventually makes, and keeping it in mind renders transparent that he is *not* abandoning pursuit of the project of explaining how the meanings of sentences depend on those of their significant parts — though he abandons this particular way of expressing it. The trouble, as he sees it, is that assigning entities to words and sentences simply does not advance the project.

[d] What analogy [with our simple theory of reference] demands is a theory that has as consequences all sentences of the form 's means m' where 's' is replaced by a structural description of a sentence and 'm' is replaced by a singular term that refers to the meaning of that sentence; a theory, moreover, that provides an effective method for arriving at the meaning of an arbitrary sentence structurally described...My objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use. (pp. 20-1)

⁸ A clear potential problem with this is that it assumes the meaning of t is fixed by its referent. The Slingshot in fact exploits this in one of its assumptions.

The transition to Davidson's novel proposal occurs in the following passage, which we quote in full as it is crucial to understanding how he thinks a truth theory is relevant to the project of "Truth and Meaning."

[e] ... having found no more help in meanings of sentences than in meanings of words, let us ask whether we can get rid of the troublesome singular terms supposed to replace 'm' and to refer to meanings. In a way, nothing could be easier: just write 's means that p', and imagine 'p' replaced by a sentence. Sentences, as we have seen, cannot name meanings, and sentences with 'that' prefixed are not names at all, unless we decide so. It looks as though we are in trouble on another count, however, for it is reasonable to expect that in wrestling with the logic of the apparently non-extensional 'means that' we will encounter problems as hard as, or perhaps identical with, the problems our theory is out to solve.

[f] The only way I know to deal with this difficulty is simple, and radical. Anxiety that we are enmeshed in the intensional springs from using the words 'means that' as filling between description of sentence and sentence, but it may be that the success of our venture depends not on the filling but on what it fills. The theory will have done its work if it provides, for every sentence s in the language under study, a matching sentence (to replace 'p') that, in some way yet to be made clear, 'gives the meaning' of s. One obvious candidate for matching sentence is just s itself, if the object language is contained in the metalanguage. As a final bold step, let us try treating the position occupied by 'p' extensionally: to implement this, sweep away the obscure 'means that', provide the sentence that replaces 'p' with a proper sentential connective and supply the description that replaces 's' with its own predicate. The plausible result is

(T) s is I if and only if p

What we require of a theory of meaning for a language L is that without appeal to any (further) semantical notions it place enough restrictions on the predicate 'is I' to entail all sentences got from schema I when 's' is replaced by a structural description of a sentence of L and 'p' by that sentence.

[g] ... it is clear that the sentence to which the predicate 'is I' applies will be just the true sentences of L, for the condition we have placed on satisfactory theories of meaning is in essence Tarski's Convention I that tests the adequacy of a formal semantical definition of truth. (pp. 22-3)

In [e], when Davidson writes, "it is clear we have found no more help in meanings of sentences than in meanings of words," he is thinking about the project of assigning entities, called 'meanings', to words and sentences, and his complaint, as passage [d] makes explicit, is that introducing these entities does *nothing* to explain how the

meanings of sentences depend on those of words. It surely does not help us to arrive “at the meaning of an arbitrary sentence structurally described.” This hardly rejects the project of providing a compositional *meaning* theory, as Soames claims, for Davidson’s complaint is that an appeal to meanings (as entities) fails to do the work required for that project!

Davidson goes on to ask whether we might simply refuse to treat the position after ‘means’ as referential, casting the target theorems in the form ‘ \underline{s} means that \underline{p} .’ The difficulty with this proposal, he says, becomes how to formulate a theory that starts with axioms attaching to words and recursively generates theorems of this form, for it would seem we cannot invoke the apparatus of quantificational logic (objectual quantification, at any rate) to reach into the position occupied by ‘ \underline{p} ’, as it would seem we must in order to utilize the axioms attaching to significant parts of ‘ \underline{p} ’.

It is clear, however, that if such a theory were available, it *would* allow us to match each object language sentence \underline{s} with a meta-language sentence the same in meaning, but in something like a use position, for ‘ \underline{s} means that \underline{p} ’ is true iff the sentence that replaces ‘ \underline{p} ’ translates \underline{s} . Once we relinquish the goal of assigning entities to words and sentences that are their meanings, we can see that the work essentially comes to this.

This is Davidson’s point in the third sentence of passage [f], when he writes, “The theory will have done its work if it provides, for every sentence \underline{s} in the language under study, a matching sentence (to replace ‘ \underline{p} ’) that, in some way yet to be made clear, ‘gives the meaning’ of \underline{s} .” He hedges, first, because it is obvious he is not giving the meaning of \underline{s} by introducing an entity as its meaning, and the description literally would require that, and, second, because there is more to giving the meaning of a sentence than matching it with one in the meta-language the same in meaning (or use profile): it is also a goal to show how the meaning of the sentence is determined by, or depends on, the meanings of its significant parts and their mode of arrangement.

The design problem, then, is this (paraphrasing a sentence in [d]): formulate a theory that has as consequences all sentences of the form ‘ \underline{s} ---- \underline{p} ’, where ‘ \underline{s} ’ is replaced by a structural description of sentence and ‘ \underline{p} ’ by a metalanguage sentence that gives the meaning of that sentence; a theory, moreover, that provides an effective method for arriving at the meaning of an arbitrary sentence structurally described.

Davidson’s great insight was to see that there was a form of theory ready to hand that fits the bill, namely, a Tarski-style truth theory that satisfies Convention T. He arrived at this suggestion *indirectly*: what we want is a replacement for ‘-----’ in ‘ \underline{s} ---- \underline{p} ’ that permits us to exploit the familiar tools of logic on the position occupied by ‘ \underline{p} ’ so as to formulate a theory that has as consequences, for each sentence \underline{s} of the object language, a theorem of the relevant form in which the sentence that replaces ‘ \underline{p} ’ *gives*

the meaning of \underline{s} . We can achieve this straightforwardly if the meta-language embeds the object language, and the sentence that replaces 'p' is \underline{s} itself. Given that we want the mentioned sentence matched with a used one alike in meaning, and given that we demand an extensional filler, we must supply a predicate for s so as to have a sentence on the left hand side and a truth-functional connective between left and right. The simplest connective that does the job is 'if and only if', as exhibited in (T) above. But if the goal is to be able to derive for each sentence of the object language a theorem of this form in which 'p' is replaced by \underline{s} , what we will have (assuming consistency) is a theory meeting Tarski's Convention T for the language.

This is Davidson's point in [g], an observation we hope it is becoming clear was his aim all along. Note that he speaks of Tarski's Convention T as being *the* condition "we have placed on satisfactory theories of meaning." With this wording, he clearly is *not* limiting himself to vocabulary from the theory of reference, for generalizing Convention T to an arbitrary meta-language requires the theory to entail each instance of (T) in which the metalanguage sentence that replaces 'p' translates the object language sentence \underline{s} .⁹

In passage [f], a line that has drawn extensive exegetical and critical attention is Davidson's statement of what we require of a theory of meaning, namely, that "without appeal to any (further) semantical notions it place enough restrictions on the predicate 'is \underline{T} ' to entail all sentences got from schema \underline{T} , when ' \underline{s} ' is replaced by a structural description of a sentence of \underline{L} and 'p' by that sentence." Why does he say, "without appeal to any (further) semantical notions"?

This expresses a natural ambition, once one sees how the trick is to be turned, to see whether we may gain insight not only into how what complex sentences mean depends on their contained semantical primitives and mode of combination, but also into connections between concepts in the family of meaning and others, and, in particular, since we see we have in effect a truth theory, between the concept of meaning and the concept of truth. But this doesn't preclude that the goal is to match an object language sentence with a meta-language one that gives its meaning, in the straightforward sense that is guaranteed when the used sentence is the same as, and used in the same sense as, the object language sentence.

Davidson is clear that this is not a shift in the nature of the project, nor is it a rejection of the goal of solving the original problem:

⁹ It is obvious that for a context sensitive language the syntactic criterion won't work; we cannot give context relative truth conditions for 'I am hungry', e.g., using a sentence with context sensitive features. This requires a generalization of Tarski's Convention T. We want theorems to provide context relative truth conditions which interpret utterances of object languages sentences relative to those contexts.

[h] The problem, upon refinement, led to the view that an adequate theory of meaning must characterize a predicate meeting certain conditions. It was in the nature of a discovery that such a predicate would apply exactly to the true sentences. (p. 24)

In summary: (i) Meanings, construed as entities of words or sentences are no help in advancing the project of understanding how the meanings of sentences depend on the meanings of words — that is, how we manage to be in a position to understand a potential infinity of non-synonymous sentences on the basis of grasp of a finite vocabulary of semantical primitives. (ii) The project of formulating an intensional logic for handling ‘ \underline{s} means that \underline{p} ’, once we stop treating ‘that \underline{p} ’ as a referring term, looks to encounter “problems as hard as, or perhaps identical with, the problems our theory is out to solve” (p. 22). (iii) But the end result looks to be to match a sentence \underline{s} in our target language with a sentence ‘ \underline{p} ’ in the meta-language the same in meaning; and this goal, upon reflection, may be pursued without getting “enmeshed in the intensional,” for the matching may be achieved if we are able to provide a formally correct definition of a predicate “is \underline{I} ” that entails for every sentence of the form (T), where ‘ \underline{p} ’ is replaced (to put it generally) by a sentence that translates \underline{s} . (iv) But this requires the definition meet Tarski’s Convention T (if the language is non-context sensitive), and thus we discover a connection between our primary goal and the definition of a Tarski-style truth predicate for a language. (v) This is not to replace the original goal with another, or to eschew giving a compositional meaning theory in favor of giving a truth theory, but to discover a way of getting at what we wanted with a bit of indirection which turns out to be helpful.

7

It is in this context that we should understand Davidson when he says “the definition works by giving necessary and sufficient conditions for the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence. To give the semantic concept of truth for a language is to know what it is for a sentence — any sentence — to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language” (see full passage [b] in §4). His idea is not that truth conditions *are* meanings, nor is he *reifying* truth conditions, for this expression appears only in the phrase ‘give the truth conditions’.¹⁰ For Davidson, the locution ‘truth conditions’ is syncategorematic. His hope was that a theory, for a context sensitive language, that matches object language sentences with meta-language ones—the utterances of which are true or false together—would *ipso facto* meet a suitable analog for Convention T, and it would do so in a way that reflected the rules attaching to words

¹⁰ See (Cummins 2002) for an example of an author who takes Davidson to be reifying truth conditions and assigning them as meanings to sentences.

governing their contributions to what sentences mean, since what sentences mean determines under what conditions they are true.

When he says,

This at any rate is my excuse for a feature of the present discussion that is apt to shock old hands; my freewheeling use of the word 'meaning', for what I call a theory of meaning has after all turned out to make no use of meanings, whether of sentences or of words.

his point is that pursuing the project of saying how what sentences mean depends on what their words mean need not in the end invoke meanings *as entities*. Soames is quite mistaken to conclude Davidson has thereby rejected the project of giving a theory of meaning in favor of some distinct project. Rather, Davidson has merely discovered a novel way to pursue it which dispenses with meanings *as entities* and which does not use even 'means' in the theory itself (though we'll explain presently how to reintroduce it and how to clarify the manner in which the truth theory functions *in* a meaning theory).¹¹ Davidson means no more than this when he says,

Indeed, since a Tarski-type truth definition supplies all we have asked so far of a theory of meaning, it is clear that such a theory falls comfortably within what Quine terms the 'theory of reference' as distinguished from what he terms the 'theory of meaning'.¹²

He is not in this passage embracing a theory of reference to the exclusion of a compositional meaning theory; he is not replacing one project with another. Even if he eschews meanings as entities in the theory of meaning, he does not discard the concept of meaning as too obscure or confused for scientific work.¹³ Quine's agenda was *never* Davidson's despite whatever inspiration he drew from Quine.

¹¹(Church 1951, p. 102) seems to have had essentially this insight, as noted by (Wallace 1978, p. 54).

¹² Quine describes the contrast he has in mind as follows, "The main concepts in the theory of meaning, apart from meaning itself, are *synonymy* (or sameness of meaning), *significance* (or possession of meaning), and *analyticity* (or truth by virtue of meaning). Another is *entailment*, or analyticity of the conditional. The main concepts in the theory of reference are *naming*, *truth*, *denotation* (or truth-of), and *extension*. Another is the notion of *values* of variables" (Quine 1953, p. 130).

¹³ In this context we can also treat a passage in a later paper which seems to echo the theme here. In "In Defense of Convention T," Davidson says, "In the previous paragraph, the notion of meaning to which appeal is made in the slogan 'The meaning of a the sentence depends on the meanings of its parts' is not, of course, the notion that opposes meaning to reference, or a notion that assumes that meanings are entities" (1973, pp. 70-71). He goes on to say, "The slogan reflects an important truth, one on which, I suggest, a theory of truth confers a clear content. That it does so without introducing meanings as entities is one of its rewarding qualities" (p. 71). There are two things going on here. First, the truth theory itself employs only semantic concepts drawn from the theory of reference; the second is that it makes no use of assignments of meanings as entities. This is not to disclaim the aim of giving the

In this light, we can see clearly that Soames's replacement interpretation is mistaken. Davidson was doing something much more interesting than jettisoning the project that his article starts out describing. He was offering a novel approach, one reached by reflections which led him to conclude assigning entities to expressions does not illuminate how we understand complexes on the basis of their parts, and by noticing that in effect the goal was to arrive at a matching of mentioned object language sentences with used (and so understood) meta-language sentences. It is therefore a significant error to say the "grand Davidsonian-cum-Quinean theme presupposed that truth and reference can be retained while meaning is rejected" and then to go on to criticize Davidson on the grounds that "it is not clear that our ordinary notions of truth and reference can be separated from our ordinary notion of meaning" (p. 4).

Davidson exploits a truth theory for two purposes: first, to illuminate how the meanings of complexes depend on the meanings of significant parts, reframed as the project of illuminating how complexes are to be understood on the basis of understanding their significant parts; second, to illuminate meaning by showing how constraints on a truth theory stated using, ideally, non-semantic vocabulary suffice for the theory to meet Contention T. We will return in §10 to the relation of these projects to explaining knowledge of meaning in terms of truth conditions, after we have looked in more detail at how the first of these two tasks can be achieved.

8

We divided Davidson's project into two in the previous section. The first, providing a compositional meaning theory, we shall call *the initial project*; the second, providing a deeper understanding of meaning in general, we call *the extended project*. We gain greater clarity into his project if we first describe how a truth theory can be used to pursue the initial project. For this purpose, recall the example of a reference theory S for fragment L. We observed that if we had certain knowledge of S, i.e., that described in [KS], we would be in a position to interpret every expression in L, and not merely to say what its referent is. One condition we must know is that its axioms use meta-language expressions to give the referents of primitive terms that are the same in meaning as object language expressions and that those of its axioms which give recursively the referents of recursive devices of the language do so using a meta-language expression in a sentence of the same form which translates the recursive

meaning of each sentence of the language by way of identifying for it a theorem in which the truth conditions for it are given using a sentence alike in meaning (that is required by Convention T!), or by doing so on the basis of axioms that use expressions alike in meaning to those that give their truth, reference and satisfaction conditions (for that is, after all, the idea about how it is to be done). It is only that what the theory shows it does not say.

device in the object language. If we knew this, we would be in a position to understand complex expressions in the fragment on the basis of understanding their significant parts, and we would also be in a position to state explicitly the meaning of each expression of the object language. What puts us in this position is *not* knowledge of what S states, but rather knowledge *about* it.

If we identify the meaning theory with what we know that enables us to understand any expression of the object language, then the reference theory is not a meaning theory. The meaning theory is, rather, what we have to know *about* the reference theory to position ourselves to understand complex expressions on the basis of their parts. Note that the theory does *not* explain what primitive expressions mean, though that information can be gleaned from its axioms, given that we know they satisfy Convention A. But, as Davidson remarks, “the task was to give the meaning of all expressions in a certain infinite set on the basis of the meaning of the parts; it was not in the bargain also to give the meanings of the atomic parts” (p. 18).

This model for a compositional meaning theory carries over more or less directly to sentential meaning, replacing reference with truth. We will use a language without context sensitive features and quantifiers, though the lessons generalize to more complex languages in a relatively straightforward way. We give a theory T (‘refers to’ and ‘is true’ are to be understood as restricted to the object language), and assume rules of formation given in the natural way.

[T]

1. ‘Claudine’ refers to Claudine
2. ‘Robert’ refers to Robert
3. For any name N, $N \wedge \text{‘dort’}$ is true iff what N refers to is sleeping.
4. For any names N_1, N_2 , $N_1 \wedge \text{‘aime’} \wedge N_2$ is true iff what N_1 refers to loves what N_2 refers to.
5. For any sentence S, $\text{‘Ce n'est pas le cas que’} \wedge S$ is true iff it is not the case that S is true.
6. For any sentences S_1, S_2 , $S_1 \wedge \text{‘et’} \wedge S_2$ is true iff S_1 is true and S_2 is true.

‘N’ ranges over names and ‘S’, ‘ S_1 ’ and ‘ S_2 ’ over sentences of the object language. To see how the meanings of the complexes depend on those of the parts and to interpret any sentence of the language we need to know *about* T only the following:

- (i) What the axioms are, as stated in (1)-(6).
- (ii) What each axiom states and of each that it states what it does.

- (iii) That in each reference axiom the name used on the right of 'refers to' translates the name mentioned on the left.
- (iv) That in each predicate axiom the predicate used in the meta-language in giving truth conditions for the object language sentence translates the object language predicate.
- (v) That in each recursive axiom the logical connective used in the meta-language to give the truth conditions for the object language sentence translates the logical connective in the object language.

Given (iii)-(v), it is clear that, provided an adequate logic, T has as theorems all sentences of the form,

(T) \underline{s} is true iff \underline{p} ,

where \underline{s} is replaced by a structural description of an object language sentence and ' \underline{p} ' is replaced by a meta-language sentence that translates \underline{s} . We also need a way of identifying *just* those theorems which draw only on the content of the axioms. We call any proof procedure which draws only the content of the axioms and whose last line is of form (T) and in which no semantic vocabulary of the meta-language remains (i.e., 'is true') a *canonical proof procedure*. As rules of inference, we allow:

Universal Instantiation (UI): from a universally quantified sentence any instance may be inferred.

Substitution (S): from any sentences of the form ' \underline{t} refers to \underline{y} ' and any sentence of the form $S(\text{what } \underline{t} \text{ refers to in } \underline{L})$, $S(\underline{y})$ may be inferred.

Replacement (R): for any sentences \underline{x} , \underline{y} , $S(\underline{x})$, $S(\underline{y})$ may be inferred from $\underline{x} \wedge \text{'iff'} \wedge \underline{y}$ and $S(\underline{x})$.

Replacement is limited; the aim is to so restrict the rules that any proof of a sentence of form (T) without semantic vocabulary on its right hand side is a canonical proof. Its last sentence we call a *canonical theorem*. Since in a canonical theorem the sentence on the right of 'iff' translates the one mentioned on the left, we can validly infer the sentence we get from replacing 'is true iff' with 'means that' (Davidson 1970, p. 60). We will call this inference rule *Transference*.

Transference (T): from any canonical theorem, ' \underline{s} is true iff \underline{p} ', ' \underline{s} means that \underline{p} ' may be inferred.

In addition to (i)-(v), we need to know:

- (vi) The rules of inference UI, S, R, and T
- (vii) That any proof of a sentence of form ' \underline{s} is true iff \underline{p} ' without semantic vocabulary on the right hand side from the axioms on the basis of the rules UI, S, and R is a canonical proof, and, hence, the sentence is a canonical theorem.

This puts us in a position to infer for each sentence of the object language a meta-language sentence that explicitly states what the object language sentence means, on the basis of a proof that traces out, at each step, the contribution of each object language expression to fixing the truth conditions of the sentence to which it contributes, on the basis of reference or truth conditions given using a term the same in meaning with it. We can thus see what the contribution is of each semantical primitive to the interpretive truth conditions of the sentences in which it occurs on the basis of what it means.

Any theory for a language like this that satisfies conditions (iii)-(v) satisfies Convention A, which requires that the axioms for primitive expressions meet an analog of Convention T, namely, that they provide reference and truth conditions (and satisfaction conditions when we move to a language with quantifiers) using terms in the meta-language that translate the object language terms for which they are used to give reference and truth (and satisfaction) conditions. Convention A is a condition on a truth theory that, relative to an adequate logic, ensures it satisfies Convention T. Call any truth theory that meets Convention A *interpretive*.

Let us take stock. We have shown how a truth theory may be exploited in pursuit of the goals of a compositional meaning theory. However, we have not claimed the truth theory *is* a compositional meaning theory. That would be a mistake. Rather, we have claimed that having certain information about a truth theory for a language positions us to interpret each object language sentence and to see, through a canonical proof of a canonical theorem, how its parts contribute in virtue of their meaning to fixing the truth conditions of the object language sentence in a way that provides its interpretation.¹⁴

¹⁴ Interestingly, if we identify a meaning theory with what we must know, then we need not know the truth theory: we need to know what it is and what its axioms mean, but knowledge that its axioms are true is not part of what we must know to do the work that needs to be done. This is important because it might enable us to apply this technique to a language with vague terms and semantically defective terms without worrying about the truth theory not being true. At the same time, though, relative to the knowledge that the terms in the object language are not semantically defective, we can infer that the truth theory is true, and so, relative to this additional bit of knowledge, come to know what the truth theory states.

Davidson never cleanly separated the initial from the extended project. He aimed to illuminate how understanding complexes depends on understanding their parts at the same time that he illuminated meaning generally. To this end, he treated the truth theory as empirical. His goal was to show what it would take to confirm a truth theory for a language which met a suitable analog of Convention T. In this way, the theory and its concepts are related to evidence that does not presuppose their application. This illuminates theoretical concepts by showing how they organize data on the basis of which they are applied.

Initially, Davidson thought that the mere extensional adequacy of a truth theory would suffice, if the targets were natural languages with context sensitive elements (indexicals, demonstratives, tense, and the like). His hope was that bi-conditionals such as (S),

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

would be eliminated, given the need to secure the right truth conditions for utterances of 'This is snow', 'This is grass', 'This is green', 'This is white', etc.¹⁵ Davidson's hope was that, for natural languages, satisfying extensional adequacy alone would suffice for the theory to meet Convention T (more properly, Convention A). But, while context sensitivity helps, extensional adequacy is not enough; we can always add to the satisfaction conditions for a predicate axiom a condition that does not affect truth, but does affect meaning. Extensional adequacy would not distinguish between (i) and (ii):

- (i) Any function f satisfies ' \underline{x} is red' iff $f(\underline{x})$ is red
- (ii) Any function f satisfies ' \underline{x} is red' iff $F(\underline{x})$ is red and $2 + 2 = 4$.

The latter generates non-interpretive canonical theorems.

Davidson acknowledged the problem and tried to avoid it by insisting on *confirmability by the Radical Interpreter*. His hope was that any truth theory confirmable from the standpoint of a Radical Interpreter would satisfy Convention T.

¹⁵ Soames declares (p. 6) that Davidson thought "truth theories that are both true and compositional will end up deriving only those statements ['S' is true iff P]" in which what replaces 'P' translates 'S'. Again, he is mistaken. Davidson's idea was that context sensitive elements in natural languages would play a crucial role. As he himself notes (footnote 10 on page 26 of *Inquiries into Truth and Interpretation* (Davidson 2001)), "[c]ritics have often failed to notice the essential proviso mentioned in this paragraph. The point is that (S) could not belong to any reasonably simple theory that also gave the right truth conditions for 'That is snow' and 'This is white'. (See the discussion of indexical expressions below)."

It is noteworthy that Davidson's shift shows rather convincingly that his goal was *not, contra* Soames, to replace a meaning theory with a truth theory. For the objection Davidson is responding to is that truth is insufficient to secure that the theory meets (a suitable analog of) Convention T. If he were pursuing a replacement strategy, this objection would have provided no reason to change course.

For present purposes, it doesn't matter whether Davidson's new suggestion is adequate.¹⁶ Our current concern is to defend a semantic program that sees a truth theory as providing a vehicle for a compositional meaning theory. It is this more limited project Soames is attacking. To meet his attack, it suffices to show how a truth theory can be used in pursuit of a compositional meaning theory.

10

Soames says (p.1) that what Davidson aims to do is

... to explain knowledge of meaning in terms of knowledge of truth conditions [and that f]or Davidsonians, these attempts take the form of rationales for treating theories of truth, constructed along Tarskian lines, as empirical theories of meaning.

What would it be to explain "knowledge of meaning"? *Whose* knowledge? *What kind*? It is most natural to interpret the project so described as explaining the knowledge of speakers of a language. But, then, Soames' suggestion would seem to be that Davidson was proposing speakers know their languages by way of knowing a truth theory—how else to interpret "in terms of knowledge of truth conditions," which would seem to be propositional knowledge. Davidson, however, *explicitly* denies this.¹⁷ He is not offering a psychological theory about the mechanism by which speakers

¹⁶ We are convinced it is not. See (Lepore and Ludwig 2005, ch. 11).

¹⁷ In "Radical Interpretation" (Davidson 1973, p. 25), Davidson says, in motivation for asking the question "What could we know that would enable us to understand any potential utterance of a speaker?", that "... it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation." In "A Nice Derangement of Epitaphs" (Davidson 1986, p. 438), he says, "To say that an explicit theory for interpreting a speaker is a model of the interpreter's linguistic competence is not to suggest that the interpreter knows any such theory...They are rather claims about what must be said to give a satisfactory description of the competence of the interpreter." In "The Structure and Content of Truth" he says that the aim of the theory is, *inter alia*, to describe a certain complex ability: "it at once describes the linguistic abilities and practices of the speaker and gives the substance of what a knowledgeable interpreter knows which enables him to grasp the meaning of the speaker's utterances. This is not to say that either speaker or interpreter is aware of or has propositional knowledge of the contents of such a theory" (Davidson 1990, pp. 311-12). In "The Social Aspect of Language" he says, "let me say (not for the first time): I do not think we normally understand what others say by consciously reflecting on the question what they mean, by appealing to some theory of interpretation, or by summoning up what we take to be the relevant evidence" (Davidson 1994, p. 3).

understand. He has no commitment about whether “knowledge of meaning” is explicable in terms of propositional knowledge.

His goal is, rather, to provide a general framework for providing a compositional meaning theory which would enable us to show how what complexes mean depends on what their significant parts do and their mode of combination, and, thereby, to show how to illuminate the structure of a complex practical ability.

We start with semantical primitives; these are expressions the understanding of which is not derived from an understanding of their components. A speaker’s understanding of each will involve a disposition to use it in accordance with a semantical category into which it falls and the specific meaning it has within that category that sets it apart from other items in the same category. Grasp of its category amounts to knowing how it is to be combined with other terms given their categories and how being so combined helps determine what the complexes it helps to form mean. This does not entail propositional knowledge of rules. Rather, we know *how* to put terms together to mean various things and *how* to interpret the results of putting them together in appropriate ways. The upshot involves propositional knowledge, but of what utterances mean and the conditions under which they are true.

The compositional meaning theory concerns knowledge of meaning only in the sense that it aims to capture the structure of the language of which speakers have knowledge—in the sense of being competent speakers of it. It shows us something about what it is that speakers have knowledge of, and so something about what the structure of their dispositions with respect to words in it have to be. No further account is on offer of what semantic competence consists in or how it is realized.¹⁸ No explanation is being offered of how speakers learn the languages they speak, or of why they speak the languages they do. An analogy: if you describe the rules of chess, you say what it is that someone who knows *how* to play chess has to know, in the sense of knowing what to do in playing. You don’t explain how he knows it, or what his knowledge consists in, or how or why he learned chess.

Explaining knowledge of meaning, in any sense other than the rather attenuated sense of describing the structure of what is known, is no part of Davidson’s project. So it could

¹⁸ This serves as a response to another objection we have heard: that the account of how to provide a compositional meaning theory would mean that a child learning its first language would have to have the concept of truth, which is implausible. This would be so only if we were making the claim that the way the competence that the theory represents is realized in speakers is by way of their having propositional knowledge of the theory itself. But we have no such commitment, and it cannot be taken to be a general commitment of giving a compositional meaning theory because every compositional meaning theory that aims at an explicit statement of the meaning of every sentence will deploy concepts which a child learning a first language will not plausibly have, such as the concept of meaning itself.

not have been his aim to explain knowledge of meaning in terms of knowledge of truth conditions. But suppose it were. Then how would treating theories of truth as empirical theories of meaning have helped to explain knowledge of meaning in terms of knowledge of truth conditions? It is not obvious. For to treat a theory of truth as an empirical theory of meaning does not entail that knowledge of meaning is to be explained in terms of knowledge of truth conditions. And showing that a theory of truth can be empirically confirmed and used as a theory of meaning would not show that speakers of the language possess knowledge of a truth theory.

There is some reason to believe Soames does not intend 'explaining knowledge of meaning' to be about explaining speaker competence in terms of knowledge of a truth theory, for he treats the psychologizing of Davidson's program as a last gasp defense of the justificatory project. An alternative, less natural reading of 'explaining knowledge of meaning' is as the project of explaining what knowledge of truth conditions would suffice for knowledge of meanings. This is closer to what Davidson had in mind, but, as we hope has now become clear, it was not his (considered) view that knowledge of what a truth theory states *ipso facto* suffices for knowledge of meaning, even for a theorist.¹⁹ It is rather knowledge *of*, and *about*, a truth theory that puts us (theorists who know it) in a position to interpret its subject.

¹⁹ Soames says, "Davidson originally held that a truth theory for L qualifies as a theory of meaning, if knowledge of what it states is sufficient for understanding L" (2008, p.5). It is fair to call a body of knowledge sufficient to interpret any sentence of a language on the basis of rules governing its semantical primitives a meaning theory for the language, and it is fair to say this is what Davidson sought in an account of how the meanings of sentences depend on the meanings of their contained words. It would then follow that if knowledge of what a truth theory states sufficed to understand the language for which it was a theory, then it would be a meaning theory. This conditional thesis is trivial and nonsubstantive, and there is no reason to think Davidson would not have been happy to endorse it. But since it is trivial and nonsubstantive, it is also not a thesis that he would have had to give up. The substantive question is what knowledge one *could* have (propositional knowledge) that would suffice to interpret a speaker's language. And the exegetical question about "Truth and Meaning" is what Davidson's position on that was. In "Truth and Meaning," he does speak as if the meaning theory is the truth theory. But even there it seems clear he was thinking that the required knowledge went well beyond what was stated by the truth theory, for he offers an argument to show that a true truth theory would meet an appropriate analog of Convention T for a context sensitive language *because* it would track correctly the use (actual and potential) of predicative terms in combination with demonstratives. Isn't it transparent that that is something one would have to know in order to use the truth theory for interpretation? In a retrospective remark in "Reply to Foster" Davidson says, "That empirical restrictions must be added to the formal restrictions if acceptable theories of truth are to include only those that would serve for interpretation was clear to me even when I wrote 'Truth and Meaning'. My mistake was not, as Foster seems to suggest, to suppose that *any* theory that correctly gave truth conditions would serve for interpretation; my mistake was to overlook the fact that someone might know a sufficiently unique theory without knowing that it was sufficiently unique. The distinction was easy for me to neglect because I imagined the theory to be known by someone who had constructed it from the evidence, and such a person could not fail to realize that his theory satisfied the constraints" (Davidson 1976, p. 173). If we take Davidson at his word (why wouldn't we?), he misexpresses his underlying thought in "Truth and Meaning" when he speaks as if the truth theory were a meaning theory itself.

Soames's final criticism is lodged after he reviews various things we have to know about a truth theory to use it to interpret another. It follows his observation that a theory with a reasonably robust logic will have T-theorems which are not interpretive.

The natural response is to add a definition of *canonical theorem* to truth theories, picking out, for each S, a unique T-theorem as translational. However, it is doubtful that this would provide the needed justification. Once information about canonicity is added, the only role played by knowledge of the canonical truth theorem (CTT) is that of allowing one to identify a claim in which S is paired with a certain content, which is then stipulated to be the content expressed by a translation. *Neither the truth of CTT, nor the fact that it states the truth conditions of S, plays any role in interpreting S.* All it does is to supply a translational pairing, which could be supplied just as well in other ways. One could get the same interpretive results by replacing the truth predicate in a translational truth theory with *any arbitrary predicate F whatsoever*. Whether or not the resulting theory is true makes no difference. To interpret S, all one needs to know, of the canonical F-theorem, is that it links S with *the content expressed by a translation of S*. No one would conclude from this that translational F-theories count as theories of meaning. Why, then, suppose that translational truth theories do? So far, we have been given no answer.

The collapse of this justificatory attempt should not be lamented. Prompted by Foster's objection into invoking a notion of paraphrase beyond anything in the truth theories themselves, we face the dilemma of appealing to a notion strong enough to overcome the objection, at the cost of robbing truth and reference of their cherished roles in explicating meanings, or of not overcoming the objection at all. (p. 7)

Soames's objection rests on the assumptions that Davidson, as well as Davidsonians, are committed to saying the truth theory itself *is* a meaning theory and that truth and reference are supposed to explicate meaning. We hope it is transparent to you by now that neither assumption is correct.

The goal is to exploit a truth theory *to do the work of* a meaning theory. We start with axioms for primitive expressions and arrive at theorems which match object language sentences with meta-language sentences that in some sense give the their meanings. This is exactly what Soames describes the theory as doing.

Although in "Truth and Meaning" it may have looked as if Davidson hoped extensional adequacy would suffice for a truth theory for a natural language to satisfy an appropriate analog of Convention T, so that we would show important connections between truth

and meaning, his later work clearly did not seek to do so but rather to illuminate meaning by showing how to confirm an interpretive truth theory for a speaker from evidence presupposes neither knowledge of meanings nor knowledge of contents of the speaker's propositional attitudes. So the "cherished goal" was not central to the project as a whole and was a feature perhaps only of its initial phase. It is not something Davidson was committed to for more than a brief period, and it certainly is not something contemporary Davidsonians need be committed to in pursuing truth-theoretic semantics.

In response to Soames' claim that all the theory does is match object language sentences with meta-language sentences that translate them, we agree that it does that, but that is not *all* it does. Soames says the theory need not be true to issue in theorems which match object language sentences with meta-language sentences that translate them. This too is correct, as we noted above. Any predicate could play the role. However, as Davidson observed, if the theory is true, the predicate has the extension of the truth predicate. And if the predicate is the truth predicate and the axioms meet Convention A and the primitive expressions are not semantically defective, then theory *is* true, and we can retrieve more information from the theory than what is expressed in its theorems of form:

p means that p.

Proofs of the canonical theorems *exhibit* how parts of sentences, in virtue of their meanings, contribute to fixing the truth conditions of these sentences, by way of using terms the same in meaning. We see exhibited in the proof *the semantic structure of the sentence* and how it fixes truth conditions. This is not what the proof says, but it can be culled from the proof. Someone in possession of such a theory and appropriate knowledge of it is in a position to understand the compositional structure of the language. That is more than being able to pair object language sentences with meta-language ones that translate them. So, this objection misses its target, and it misses where some of the real utility of a truth theory lies in showing us how languages work.²⁰

Soames offers an alternative approach to semantics; we end with a few critical remarks about it. His proposal has two parts. First, we should reintroduce meanings as entities.

²⁰ Davidson makes this point in explaining why a theory of truth consisting just of the infinite class of (T) biconditionals in which the right hand side translates the sentence mentioned on the left is inadequate: "Such a theory would yield no insight into the structure of language and would thus provide no hint of an answer to the question how the meaning of a sentence depends on its composition" (Davidson 1970, p. 56).

He has in mind a Russellian picture on which we have particulars and universals assigned to words as appropriate. Obama is assigned to 'Obama', being red to 'is red'; and so on. He then imagines a theory that generates appropriate assignments of complexes of these to complex expressions. This is the picture Davidson faulted in the first pages of "Truth and Meaning." Soames admits there are problems. Take (i)-(ii).

- (i) a is larger than b
- (ii) $\langle \underline{a}, \underline{b}, \textit{is larger than} \rangle$

(ii) is an object and seems not intrinsically representational. How can it be the meaning of (i)? (i) has truth conditions, (ii) does not. Soames rejects the proposal that it is a *sui generis* fact about the proposition.²¹ His idea is that we interpret the relation R in which a, b, and *is larger than* stand in (ii) as representing that a is larger than b (or more properly: the fact that a, b and *is larger than* stand in the relation R represents that a is larger than b). It is not that their standing in the relations they stand in intrinsically represents that a is larger than b; it is rather that *we treat it as representing that*. Soames says: "The answer rests not on anything intrinsic to R but on the *interpretation* placed on R by the way that we use it" (Soames 2008, p. 18). What this comes down to is that we treat that fact that a, b, and *is larger than* stand in the relations they do in (ii) as meaning that a is larger than b.

Suppose a theorem of our theory assigned to (i) the meaning (ii).

The meaning of 'a is larger than b' in L is $\langle \underline{a}, \underline{b}, \textit{is larger than} \rangle$

To enable us to understand the object language sentence, we need to assume that what we treat the proposition as representing is what the sentence represents. Suppose we have so constructed our theory and notation for propositions that we know that among its theorems are all sentences of the form 'The meaning of s in L is p' where 's' is replaced by a structural description of an object language sentence and 'p' is replaced by a structure description of a proposition that (relative to our conventions) represents what s does. Suppose further we have a formal way of identifying theorems in the relevant class, and that we have identified the theorem above as one of the relevant class. Then we have a rule for reading off from the propositional notation in the metalanguage what the proposition represents—by way of being able to generate from the notation for the proposition a sentence in the metalanguage that represents what it does. In the present case, we know that the fact that a, b, and *is larger than* stand in

²¹ Soames says, "One could ... take propositions to be inherently and intrinsically representational, and so *sui generis*. However, this is a council of despair. Davidson would not accept such obscuritism, and we should not either. If we posit structured propositions as meanings of sentences, we ought to explain what they are, and how they are able to play the roles we assign to them" (Soames 2008, p. 17).

the relations they do in <a, b, *is larger than*> means that a is larger than b (i.e. represents what 'a is larger than b' represents). We can therefore infer that:

'a is larger than b' means that a is larger than b.

What essential work does assigning the proposition to the sentence do? None whatsoever! It turns out merely to be another way of matching an object language sentence with a meta-language sentence that translates it. For in effect all we have done is to associate with our object language sentence an expression, '<a, b, *is larger than*>', which we treat as coding for a metalanguage sentence we understand that, relative to an additional body of knowledge, we know to translate the object language sentence. For all the work that it does, the proposition might just as well drop out of the picture altogether. In what way, then, is this supposed to constitute an advance over Davidsonian semantics? We submit it has no advantage and that *except insofar as we can extract out of the theory an interpretive truth theory, it carries less information.*²²

12

In conclusion:

- 1) Soames gets off on the wrong foot by taking Davidson's project to be that of explaining knowledge of meaning when in fact it is the project of explaining meaning.
- 2) He compounds this mistake by supposing both Davidson and Davidsonians are committed to claiming a truth theory is a meaning theory. This is a view Davidson might have held in "Truth and Meaning," though even there a relatively close reading uncovers how much more he helps himself to than what is stated by the truth theory. In any case, it need be no part of truth-theoretic semantics to claim a truth theory, even one meeting various constraints, is *ipso facto* a meaning theory. More pertinently, Davidson explicitly revised his views between the publication of "Truth and Meaning" in 1967 and "Radical Interpretation" in 1973 – over 35 years ago. In "Reply to Foster," he *explicitly* writes, "A theory of truth, no matter how well selected, is not a theory of meaning" (p. 179).²³

²² For all we have said, natural language may involve terms that (putatively) refer to or denote propositions (just as it may contain terms that (putatively) refer to or denote unicorns). In giving the semantics, then, we may need terms in the metalanguage that correspond to those object language terms. Our point is that the machinery of the compositional meaning theory need not itself introduce such terms or endorse the existence of any entities of that type.

²³ Davidson says, "nothing strictly constitutes a theory of meaning," because the truth theory is not a theory of meaning and the "statement that a translational theory entails certain facts is not, because of the irreducible indexical elements in the sentences that express it, a theory in the formal sense" (Davidson 1976, p. 179). Davidson says this because of his paratactic analysis of 'that'-clauses; if we do not follow him, we need not shy from saying there is something that constitutes the meaning theory.

- 3) The role of a truth theory vis-à-vis a compositional meaning theory is clarified once we separate the initial project from the extended one of providing a deeper illumination of meaning by showing how a theory sufficient to interpret a speaker may be confirmed from an evidential base that excludes intensional and intentional notions. The theory does its job if it meets an analogue of Convention T for its axioms; Convention A, as we have dubbed it, requires meta-language expressions which translate object language ones in their reference and satisfaction conditions. Meeting Convention A suffices to meet Convention T, though not *vice versa*. It is a stronger requirement. If a theory meets Convention A, we can cull from its axioms what the object language expressions mean. We can track their contributions to fixing truth conditions of object language sentences using a sentence that means the same as the object language sentence through a canonical proof of the canonical theorem for the sentence (with straightforward modifications for context sensitive languages).
- 4) The project of truth-theoretic semantics construed as providing compositional meaning theories for natural languages is not committed to explaining meaning in terms of concepts drawn from the theory of reference.
- 5) The project of truth-theoretic semantics is not committed to the success of Davidson's extended project; in particular, it is not committed to the success of the project of Radical Interpretation.
- 6) Soames's criticisms of truth-theoretic semantics evaporate once we appreciate it is not committed to the theses he criticizes: namely, that the truth theory as such is a meaning theory and that it seeks to explain meaning in terms of truth and reference.
- 7) Truth-theoretic semantics is much more informative than a recursive translation theory, for it informs us about how truth conditions are fixed by the contributions of the components of sentences and their mode of combination.
- 8) Soames's suggestion for an alternative semantics, introducing entities to serve as meanings, only manages to match object language sentences with codes for meta-language sentences that translate them, and so fares no better than truth-theoretic semantics, and unless it includes all the information contained in a interpretive truth theory, it carries even less information.

References

- Barwise, Jon, and John Perry. 1981. Semantic innocence and Uncompromising Situations. *Midwest Studies in Philosophy* 6:387-403.
- Burge, Tyler. 1992. Philosophy of Language and Mind: 1950-1990. *The Philosophical Review* 101 (1):3-51.
- Chihara, Charles S. 1975. Davidson's extensional theory of meaning. *Philosophical Studies* 28:1-15.
- Church, Alonzo. 1951. The need for abstract entities in semantic analysis. *Proceedings of the American Academy of Arts and Letters* 80.
- Cummins, Robert. 2002. Truth and Meaning. In *Meaning and Truth: Investigations in Philosophical Semantics*, edited by J. K. Campbell, M. O'Rourke and D. Shier. New York: Seven Bridges Press.
- Davidson, Donald. 1966. Theories of Meaning and Learnable Languages. In *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science.*, edited by Y. Bar-Hillel. Amsterdam: North Holland Publishing Co.
- . 1967. Truth and Meaning. *Synthese* 17:304-323.
- . 1970. Semantics for Natural Languages. In *Linguaggi nella Societa e nella Tecnica*. Milan: Comunita.
- . 1973. In Defence of Convention T. In *Truth, Syntax and Modality*, edited by H. Leblanc. Dordrecht: North-Holland Publishing Company.
- . 1973. Radical Interpretation. *Dialectica* 27:314-328.
- . 1976. Reply to Foster. In *Truth and Meaning: Essays in Semantics*, edited by G. Evans and J. McDowell. Oxford: Oxford University Press.
- . 1986. A Nice Derangement of Epitaphs. In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by E. Lepore. Cambridge: Blackwell.
- . 1990. The Structure and Content of Truth. *The Journal of Philosophy* 87 (6):279-328.
- . 1994. The Social Aspect of Language. In *The Philosophy of Michael Dummett*, edited by B. McGuinness and G. Oliveri. Dordrecht: Kluwer.
- . 2001. *Inquiries into Truth and Interpretation*. 2nd ed. New York: Clarendon Press. Original edition, 1984.
- Glock, Hans-Johann. 2003. *Quine and Davidson on language, thought, and reality*. Cambridge, UK ; New York, NY, USA: Cambridge University Press.
- Horwich, Paul. 2005. *Reflections on Meaning*. Oxford: Oxford University Press.
- Katz, Jerrold. 1982. Common Sense in Semantics. *Notre Dame Journal of Formal Logic* 23 (2):174-218.
- Lepore, Ernest, and Kirk Ludwig. 2005. *Donald Davidson: Truth, Meaning, Language and Reality*. New York: Oxford University Press.
- Quine, Willard Van Orman. 1953. Notes on the Theory of Reference. In *From a Logical Point of View*. Cambridge: Harvard University Press.
- . 1960. *Word and Object*. Cambridge: MIT Press.
- Soames, S. 1992. Truth, Meaning, and Understanding. *Philosophical Studies* 65 (1-2):17-35.
- . 2008. Truth and Meaning: In Perspective. *Truth and Its Deformities: Midwest Studies in Philosophy* 32:1-19.
- Stich, Stephen. 1976. Davidson's Semantic Program. *Canadian Journal of Philosophy* 6:201-227.
- Tarski, Alfred. 1983. The Concept of Truth in Formalized Languages. In *Logic, Semantics, Metamathematics*. Indianapolis: Hackett Publishing Company. Original edition, 1934.
- Wallace, John. 1978. Logical Form, Meaning, Translation. In *Meaning and Translation*, edited by M. Guenther-Reutter. London: Duckworth.
- Wittgenstein, Ludwig. 1961. *Tractatus Logico-Philosophicus*. Edited by D. F. Pears and B. F. McGuinness. London: Routledge & Kegan Paul Ltd.