Journal for the History of Analytical Philosophy
Volume 3, Number 4

Was Davidson's Project a Carnapian Explication of Meaning?

Kirk Ludwig

There are two main interpretive positions on Davidson's project in the theory of meaning. The Replacement Theory holds that Davidson aimed to replace the theory of meaning with the theory of truth on the grounds that meaning is too unclear a notion for systematic theorizing. The Traditional Pursuit Theory, in contrast, holds that Davidson aimed to pursue the traditional project with a clever bit of indirection, exploiting the recursive structure of a truth theory to reveal compositional semantic structure and placing constraints on it sufficient for its canonical assignments of truth conditions to be used for interpretation. This paper responds to a recent defense of a sophisticated version of the Replacement Theory by Gary Ebbs according to which Davidson was engaged in a Carnapian explication of meaning, intending to preserve only those aspects of the usage of 'means' most important to us. I argue the Explication Interpretation cannot be sustained in the face of a detailed look at the passages that principally motivate it in "Truth and Meaning," when we take into account their local context, their context in the paper as a whole, and the context of that paper in Davidson's contemporaneous work, and later work which tries to improve the formulations which he advanced in his early papers.

Was Davidson's Project a Carnapian Explication of Meaning?

Kirk Ludwig

1. Introduction

In "Truth and Meaning" (1967) Davidson famously proposed to exploit a Tarski-style truth theory to make progress on understanding, as he put it in the introduction to *Inquiries into Truth and Interpretation* (2001d), "what it is for words to mean what they do" (p. xiii). The proposal has undeniably been influential, but there is still controversy about exactly how Davidson intended his proposal to be taken.

There are two main interpretive positions on Davidson's project. The first, which has had considerable currency, is that Davidson was suggesting that we replace the theory of meaning with the theory of truth on the grounds that the concept of meaning is hopelessly obscure or muddled, and that the best successor project (or at least a reasonable successor project), once we have absorbed this lesson, is providing a systematic, empirically grounded account of the truth conditions of sentences in natural languages. I will call this interpretation of Davidson the Replacement Theory (Chihara 1975; Cummins 2002; Glock 2003, pp. 142 ff.; Katz 1982; Soames 1992; Soames 2008; Stich 1976). The second main position is that Davidson aims to pursue the project of understanding what it is for words to mean what they do in roughly the ordinary sense. I will call this the Traditional Pursuit Theory, according to which Davidson took himself to be pursuing the traditional question about the nature of meaning with a novel approach. (A third less prominent strain is that Davidson aimed to reduce meaning to a special sort of "strong" truth condition (Burge 1992, pp. 20-1; Horwich 2005, p. 4 & ch. 8). If what I say about the errors of the Replacement Theory is correct, nothing additional need be said about the reduction interpretation.)

There are in turn two main sources of the Replacement Theory. The first derives from the difficulty of seeing how what Davidson proposes could intelligibly be seen as addressing the traditional question. If it is a non-starter as an account of meaning, then it is implausible that Davidson could have intended to be giving a theory of meaning. The second derives from certain passages in Davidson's work, especially in "Truth and Meaning," which seem to encourage the replacement view. A third supplementary source of support sometimes given is appeal to historical context: the idea that in the historical context in which he was writing it was natural for him to be undertaking just such a project of replacement of the ordinary notion of meaning with a new more scientifically respectable notion. Specifically the idea is that we should see Davidson's work as continuous with Quine's in seeking to replace the obscure notion of meaning with a scientifically respectable substitute (e.g. (Soames 2008, p. 2)). Davidson is, on this view, Quine's disciple.

In *Donald Davidson: Meaning, Truth, Language and Reality*, I argued with Ernie Lepore for the Traditional Pursuit Theory. In this paper, I use the occasion of a recent critique of that argument by Gary Ebbs, in "Davidson's Explication of Meaning" (2012), which is the most thorough and careful presentation of a sophisticated Replacement Theory I am aware of, to strengthen the case for the Traditional Pursuit Theory and to develop more fully than in the book the textual basis for it. Ebbs argues that Davidson proposes that "we replace [traditional and commonsense ideas of meaning] with a notion of meaning characterized holistically in terms of a empirically testable truth theory for a speaker's language" (2012, p. 76). Ebbs's overall argument has two main components. First, he argues that our interpretation does not make good sense of all of Davidson's writings about meaning and the relationship of those writings to work by Carnap, Tarski, and Quine. Second, he argues that on our interpretation we miss one of Davidson's central contributions to the philosophy of language. On Ebbs's view, Davidson's work aims to "*explicate* the problematic terms 'translates,' 'means that', or 'interprets'—to articulate, in ways that we find clear, those aspects of our uses of these terms that are most important to us" (2012, p. 103). Davidson's work on this view is "an invitation to adopt his proposal, not an implicit assertion that the proposals are correct relative to a theory-neutral standard of the sort that the s-means-that-p requirement [the requirement provided that the theory satisfy Convention T] is supposed to" (2012, p. 103). The importance of this "invitation" is "to challenge us to face" certain "fundamental methodological questions"—such as "whether we need an explication of these problematic terms" ('meaning', 'translation', 'interpretation', etc.) "and if we do, which of our ordinary uses of" them "an explication … should preserve" and how this "should be related to our understanding of truth"—and "to offer us ingenious and fruitful answers to them" (2012, p. 103). The first part of the argument itself has three stages. First, Ebbs argues that our interpretation of the passages that seem to favor the Replacement Theory at most show them to be compatible with our view. Second, he argues that the passages that we cite to support our interpretation do not support it, but instead that a "sophisticated version of the explicational reading of Davidson … makes better sense of [those] passages" (2012, p. 78). Third, he argues that "if we view Davidson's Convention T as methodologically similar to Tarski's Convention T, as Davidson himself did, then we have additional grounds to prefer the explicational reading" (2012, p. 78). The Explication Interpretation of Davidson is a sophisticated version of the Replacement Theory because while it does maintain that Davidson was not simply aiming to capture the ordinary notion, it does not claim that Davidson was simply writing off the concept of meaning altogether, but instead was trying to retain certain aspects of it (though as I will note below, Ebbs never explains what aspects he thinks Davidson is aiming to preserve) while jettisoning others in pursuit of a more adequate framework for theorizing about language.

In the following, I respond to this critique by taking a close and detailed look at what Davidson *actually* says in "Truth and Meaning," the epicenter of the interpretive debate, and the *context* in which he says it, specifically, (i) the context provided by "Theories of Meaning and Learnable Languages" (1965) which sets the stage for "Truth and Meaning," (ii) the immediate context of the passages whose interpretation is contested in "Truth and Meaning" and the overall context in "Truth and Meaning," (iii) the context provided by papers written for conferences in the immediate aftermath of "Truth and Meaning," specifically "Semantics for Natural Languages" (1970, first read at a conference in 1968), and to a lesser extent "In Defence of Convention T" (1973a, first read at a conference in 1970), and (iv) in the context of his later work, especially "Radical Interpretation" (1973b, first given at a conference in May 1973), and "Reply to Foster" (1976, first given at a conference in 1974), in which Davidson is looking back at and refining his initial project. I concentrate on my reading of Davidson. I remark along the way, however, about how it impinges on Ebbs's interpretation of the relevant passages, both to show how the reading responds it, and to use Ebb's interpretation as a foil to bring out underlying issues. In an appendix, I discuss some passages in "Belief and the Basis of Meaning" and the relation between Davidson's work and Quine, from whom Davidson drew inspiration, to supplement and defend further the line of interpretation pursued in main text.

For the reader wishing to track just the main line of argument, I recommend reading sections 2-4 and 7, returning for detail about difficult passages in "Truth and Meaning" in section 5, and then to interpretive issues connected with work contemporaneous with "Truth and Meaning" in section 6.

The two interpretive principles that guide the following discussion are (1) that one should interpret a philosopher as doing what he says he is doing insofar as possible and (2) that one should take into account context in interpreting what a philosopher writes, both the immediate context, the context of the piece within which it appears, the context of contemporaneous work, and the context of the philosopher's work as a whole, including later remarks on earlier work.

2. The Immediate Context of "Truth and Meaning"

"Theories of Meaning and Learnable Languages," read in 1964, and published in 1965, sets the stage for "Truth and Meaning," which was read in 1966 and published in 1967, though Davidson writes that the "main theme traces back to a paper delivered at the Pacific Division in 1962" (2001d, p. ix). The two papers are clearly meant to be part of a single project.

The first urges on philosophers and linguists the importance a constructive account of the meanings of sentences in natural languages, on the grounds that, given that we are finite, we must understand languages with infinite expressive resources on the basis of a finite number of semantical primitives and rules for their combination. Davidson makes the proposal in [1].

> [1] … I propose what seems to me clearly to be a necessary feature of a learnable language: it must be possible to give a constructive account of the meaning of sentences in the language. Such an account I call a theory of meaning for the language. (p. 3; unless otherwise indicated citations to Davidson's essays will be to reprints in (2001d))

Davidson cites a number of analyses of natural language constructions, analyses that are supposed to represent their meaning, and concludes they are incorrect. This is important context for "Truth and Meaning" because the project presented in "Theories of Meaning and Learnable Languages" is clearly intended to be one on which philosophers and linguistics have both traditionally been working. The paper also mentions the possibility of using a truth theory for the job. Here is how it is put in "Theories of Meaning and Learnable Languages" (letters in brackets added):

> [2] … we are entitled to consider in advance of empirical study what we shall count as knowing a language, how we shall describe the skill or ability of a person who has learned to speak a language. One natural condition to impose is that we must be able to define a predicate of expressions, based solely on their formal properties, that picks out the class of meaningful expressions (sentences), on the assumption that various psychological variables are held constant. This predicate gives the grammar of the language. [a] Another, and more interesting, condition is that we must be able to specify, in a way that depends effectively and solely on formal considerations, what every sentence means. With the right psychological trappings [here he means knowledge of psychological conditions sufficient to identify some product of an agent as a speech act using a sentence of the language], our theory should equip us to say, for an arbitrary sentence, what a speaker of the language means by that sentence (or takes it to mean). Guided by an adequate theory, we see how the actions and dispositions of speakers induce on the sentences of the language a semantic structure. [b] Though no doubt relativized to times, places, and circumstances, the kind of structure required seems either identical with or closely related to the kind given by a definition of truth along the lines first expounded by Tarski, for such a definition provides an effective method for determining what every sentence means (i.e., gives the conditions under which it is true). I do not mean to argue here that it is necessary that we be able to extract a truth definition from an adequate theory (though something much like this is needed), but a theory that meets the condition I have in mind if we can extract a truth definition; in particular, no stronger notion of meaning is called for. (p. 8)

In [2a] Davidson describes the project as that of providing a formal theory that specifies what every sentence of the language *means*. This is what it would take to satisfy the requirement that one give a constructive account of *the meanings of sentences of the* *language*. In [2b], he says that "a definition of truth along the lines first expounded by Tarski … provides an effective method for determining *what every sentence means*" (p. 8). It may not be initially clear how this could be so, but on the face of it, Davidson is proposing that a theory of truth along Tarski's lines suffices to "give a constructive account of the meaning of sentences in the language" (p. 3), and thus it appears that Davidson holds that properly conceived giving a truth theory for a natural language suffices for giving a meaning theory. If it turns out that there is a straightforward way of understanding how this could be so, then we can rest content with the straightforward reading of this passage.

There are two things that follow immediately in this passage that might give one pause. First, he adds to "what every sentences means" the tag "i.e., gives the conditions under which it is true." Second, he ends the passage with the claim that "a theory that meets the condition [he] has in mind if we can extract a truth definition" and then adds that "no stronger notion of meaning is called for." One way of interpreting these remarks is that he is being disingenuous about the claim that he is really aiming to give a constructive theory of *meaning*, at least if he is using 'meaning' in the usual sense. That is, one could interpret him here as saying "what every sentence means, at least if you interpret 'means' in a sense other than is usual to mean just conditions the obtaining of which suffice for the sentence to be true." This is not consistent, however, with the tone of the rest of the paper, in which Davidson takes others who are engaged in the traditional project to task for failing in it, or even in the rest of [2], in which he says "our theory should equip us to say, for an arbitrary sentence, what a speaker of the language means by that sentence (or takes it to mean)." The remark about what the speaker takes it to mean would not be to the point if it were some technical ersatz notion that was at issue. The principle reason for resorting to the disingenuous interpretation is the difficulty of seeing how what he says could be taken to be really pursuing the project of specifying what every sentence means. But, again, if it turns out that there is a straightforward way of understanding how this could be so, then we can rest content with the straightforward reading of the passage. And, it turns out, there is a straightforward way of understanding how this could be so.

"Truth and Meaning" begins with *exactly the same project*, in [3].

> [3] 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. 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 [here Davidson cites "Theories of Meaning and Learnable Languages" in a footnote]. Instead I want to ask what it is for a theory to give an account of the kind adumbrated. (p. 17)

That is, it picks up where "Theories of Meaning and Learnable Languages" leaves off, with the project of giving a constructive account of the meanings of sentences in natural languages.

In saying "it is *conceded* by most philosophers of language" that "a satisfactory theory of meaning must give an account of how the meanings of sentences depend upon the meanings of words," Davidson is clearly committing himself to the requirement. In fact, the next sentence just restates the central thesis of "Theories of Meaning and Learnable Languages." And here note that what is to be explained is how we understand utterances of sentences in the language. To understand a sentence is to understand what is meant by it on an occasion of use, given what it means in the language of the speaker. When Davidson says he senses *more than a kernel of truth* in "these vague claims," he cites in a footnote "Theories of Meaning and Learnable Languages": for it was just the central thesis of that paper. Then he states his project: what would it take to given an account of the kind in question? If we take him seriously, then we take "Truth and Meaning" to be aiming to at least sketch how to "give an account of how the meanings of sentences depend upon the meanings of words." There is the caveat about "vague claims" which signals that the goals of the project need to be precisified. The point of this emerges in the first six pages of the paper in which Davidson canvasses traditional ideas about how to carry out the project. At no point in the essay, however, does he suggest that this project should be abandoned or that he is not in fact engaged in it. In fact, in the penultimate paragraph, [4], he says:

> [4] In this paper I have assumed that the speakers of a language can effectively determine the meaning or meanings of an arbitrary expression (if it has a meaning), and that it is a central task of a theory of meaning to show how this is possible. (p. 35)

And he states, in [5], as he did in "Theories of Meaning and Learnable Languages," in no uncertain terms, that he thinks that a Tarski-style truth theory goes further towards satisfying this goal than anyone has realized.

> [5] I have argued that a characterization of a truth predicate describes the required kind of structure, and provides a clear and testable criterion of an adequate semantics for a natural language. No doubt there are other reasonable demands that may be put on a theory of meaning. But a theory that does no more than define truth for a language comes far closer to constituting a complete theory of meaning than superficial analysis might suggest; so, at least, I have urged. (p. 35)

He does not here say he is offering "an invitation to adopt his proposal" about what to substitute for the original project, but rather represents himself as having tried to persuade the reader that he has shown how to make progress on the original project. If we can interpret what he does in a way that is consistent with what he says he is doing, then clearly that interpretation that is to be preferred over interpretations that have him abandoning the project he announces at the outset of the paper and which he says he tried to carry out in the paper at its end. The interpretive principle at work here is this: do not attribute to a philosopher a project at odds with the one that he announces he is engaged in if there is a reasonable way to construe him as engaged in it.

What does Davidson mean in saying that "there are other reasonable demands that may be put on a theory of meaning"? This might be construed as an indication that he is engaged in the explication project, as Ebbs claims, and not aiming to capture everything in the ordinary concept of meaning. But there are other at least equally plausible explanations. One is that it is just a caveat about the range of issues that arise in the theory of meaning such as the treatment of illocutionary force. Moreover, there are passages in "Truth and Meaning," such as [6], that speak directly to the point:

> [6] … we have recognized that a theory of the kind proposed leaves the whole matter of what individual words mean exactly where it was. Even when the metalanguage [the language of the theory] is different from the object language [the language the theory is about], the theory exerts no pressure for improvement, clarification, or analysis of individual words, except when, by accident of vocabulary, straightforward translation fails. Just as synonymy, as between expressions, goes generally untreated, so also synonymy of sentences, and analyticity. (pp. 33-4)

Since lexical semantics and analyticity are properly subjects of the theory of meaning, it is clear why Davidson should note that more might be demanded of a theory of meaning. In addition, he notes, in the final paragraph in [7].

> [7] … finally, there are all the sentences that seem not to have truth values at all: the imperatives, optatives, interrogatives and a host more. A comprehensive theory of meaning for a natural language must cope successfully with each of these problems. (p. 36)

There is, therefore, no need to attribute to Davidson the ambition to treat something other than the common concept of meaning in his acknowledgement that more might be demanded of a theory of meaning his proposal provides. How much his proposal provides and how it connects with the theory of meaning traditionally conceived we turn to in the next two sections.

3. The Critique of Traditional Approaches in "Truth And Meaning"

I turn now to the argument of "Truth and Meaning." In this section, I discuss Davidson's critique of traditional approaches to the theory of meaning. In the next section, I turn to the positive proposal.

Most interpreters of Davidson ignore the first five and a half pages of "Truth and Meaning". They write as if the paper begins where Davidson makes his proposal for how to pursue the project he announces at the beginning of "Truth and Meaning" by way of a truth theory. But the dialectical context of the proposal is essential for understanding its point.

"Truth and Meaning" has the following structure. The first paragraph announces the project, which is to sketch how to give a theory of meaning in the sense of an account of how the meanings of sentences depend upon the meanings of words. I'll call a theory of meaning in this sense a compositional meaning theory. The next five pages are taken up with considering and criticizing various proposals for how to provide a compositional meaning theory. This in turn sets the stage for the proposal that constructing a Tarski-style truth definition for a language can do the job. This is followed by a defense of the proposal in the context of a discussion of how such a theory can be empirically tested. The final stage consists in a long discussion of the prospects of giving a Tarski-style theory for a natural language.

The crucial parts for understanding Davidson's proposal are the discussion leading up to it, the terms in which the proposal is introduced, and the discussion immediately following it. I will argue that if we pay close attention to what is going on in the first five pages of "Truth and Meaning," we get a very different picture of what Davidson is up to, when he proposes that we can solve the problem he set himself by appeal to a Tarski-style truth theory, than that suggested by the Replacement Theory, and that remarks that follow the proposal which have caused commentators so much trouble fall into place when they are read in the light of the first five pages of "Truth and Meaning."

Davidson begins with the suggestion that we assign entities, which we might as well call meanings, to expressions, as a first step in trying to give a compositional meaning theory for a natural language. He notes that this doesn't help with the project because it doesn't tell us how to go from the meanings of the parts of a sentence to the meaning of a sentence (p. 17). It doesn't help to label some of the entities 'saturated' and others 'unsaturated'. What is needed, as he notes (p. 18), is a rule that tells us how to go from expressions in different semantic categories (names and predicates for example) to an expression in yet another that has a different purpose (a declarative sentence, for example).

Davidson illustrates the point in connection with complex singular terms. He asks, in the case of 'The father of Annette', "how does the meaning of the whole depend on the meaning of the parts?" (pp. 17-8). On a Fregean view, we must say something of the following sort: the meaning of 'the father of' is such that when 'the father of' is prefixed to a singular term the result refers to the father of the referent of the singular term. Is there any part in this played by the unsaturated entity Frege would assign to ‘the father of’? The answer is ‘no’, since the same result is obtained without reference to the meaning of ‘the father of’ by dint of the rule that for any singular referring term *t*, the referent of ‘the father of ’$\frown t$ = the father of the referent of *t*. Then we can give an account of the referent of every expression by adding to this rule a specification of the referent of ‘Annette’ as Annette.

To the objection that this rule uses ‘the father of’, Davidson tellingly responds that “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). And he adds: “it is now evident that a satisfactory theory of the meanings of complex expression 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). And he says, in [8], that for the case at hand we can state the criterion of success in a way that satisfies this requirement:

> [8] what we wanted, and what we got, is a theory that 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 that term itself. Further, our theory accomplishes this without appeal to any semantical concepts beyond the basic ‘refers to’. Finally, the theory clearly suggests an effective procedure for determining, for any singular term in its universe, what that term refers to. (pp. 18-9)

This tiny example has many lessons. First, if the goal is to give a compositional meaning theory, the point is not to give the meanings of the parts but to give the meanings of the complexes on the basis of the parts. Second, this requires giving rules that take us from the parts to something that counts as a specification of the meanings of the complexes. Third, given this, we may use a primitive in the rule without assigning it a meaning as an entity or implying that it has a meaning in any sense other than its making a systematic contribution to our understanding of the sentences in which it occurs (the point is that compositional syntax is syncategorematic). Fourth, at least in this simple case, we can give a precise criterion for success in giving the meanings of the complexes on the basis of the meanings of the parts by specifying the classes of consequences of the theory which count as giving the meanings of the expressions. Fifth, in this case, the theory can accomplish its goal without using semantical concepts beyond those from the theory of reference.

But in what sense here can we speak of a theory of meaning as opposed to reference? To see how it functions as a compositional meaning theory, we need only note that if we understand the axioms, and understand them to be giving the intended function of the words in the language, then we are in a position to read off the recursive function of ‘the father of’ from the recursive rule and the function of ‘Annette’ from the base reference rule. In doing this we exploit our knowledge of ‘the father of’ and ‘Annette’, but again the point was to understand how we get to the meanings of the complexes on the basis of the parts, not to explain the meanings of the parts.

This clearly foreshadows, as we will see in detail, what Davidson does in introducing a truth theory later in the paper. Note here that there is no suggestion that we are not pursuing the project we started out with. The point is that the introduction of *entities* for *each* expression in a language is not needed in order to carry it out. And the point of introducing an alternative criterion of success is to show how to state the goal without presupposing that we must do so.

The connection with giving the meaning of an expression can be made explicit by noting that the criterion for success requires that you specify the referent of any given term by using it, which is to use a term synonymous with the term for which the referent is being given. The effect of requiring that the theory entail all sentences of the form '*s* refers to *t*' where '*t*' is replaced by the sentence that *s* refers to is to require an analog of Tarski's Convention T (which we discuss further below) for a simple theory of reference where the metalanguage embeds the object language. For the general case, in which the metalanguage does not embed the object language, the requirement is that the theory entail all sentences of the form '*s* refers to *t*' in which the term that replaces *t* refers to what *s* does. This in turn is guaranteed if '*t*' translates *s*. If we have a theory that meets this last constraint (and some additional knowledge the nature of which we will develop in the context of a truth theory below), then we can in effect read off what each object language expression means from the metalanguage expression used to give its referent. In effect we can replace 'refers to' with 'means' and preserve truth, since it suffices for '*s* means *t*' to be true that '*t*' translates *s*.

This is, in a nutshell, Davidson's strategy for pursuing the project he announces in the opening paragraph in "Truth and Meaning" and in the earlier "Theories of Meaning and Learnable Languages." The idea is basically this. To give the meanings of an infinite set of expressions on the basis of a finite set of primitive vocabulary items, in the sense of providing a theory knowledge of which would put one in a position to interpret each object language expression, it is neither necessary nor sufficient to assign entities to every expression of the language. It is not sufficient for two reasons. First, assigning entities doesn't by itself tell you how to interpret complex expressions *absent a rule*. Second, even if you give a rule that assigns an entity to a complex on the basis of assignments to its parts, this by itself is no guarantee that you will be in a position to understand it. What you have to do is to use an expression in the metalanguage that translates or codes for something you understand and which you know to translate the object language expression in assigning the entity to it that you do. Once that is clear, it is also clear that what is doing the work is not the entity but the matching of object language expression with metalanguage expression in use in a way that systematically tracks the contributions of the parts to the meaning of the whole. This in turn shows that it is not necessary to assign entities to every expression in order to provide a theory that enables one who understands it to interpret every expression in a language with a potential infinity of nonsynonymous expressions.

In reviewing proposals for giving a compositional meaning theory for the whole of a language, Davidson next considers a neo-Fregean strategy of treating sentences as singular terms that refer to their meanings and open sentences as functional expressions that take objects to sentence meanings (p. 19). This would be to model a theory for the whole language on the simple reference theory involving referring terms and functors. He scotches this with (what has come to be called (Barwise and Perry 1981)) the slingshot argument, which seeks to show that if sentences refer to their meanings, on two plausible assumptions, all sentences alike in truth-value have the same meaning. We need not go into the argument here, since we are interested not so much in its success as in what it says about how Davidson is thinking about his own proposal (see (Lepore and Ludwig 2005, ch. 3 sec. 4)). But a point to note here is that his complaint with it is not that it appeals to meanings as entities but that it yields "an intolerable result" (loc. cit.), since it requires all sentences alike in truth-value to be synonymous, which is plainly false.

We could step back from talk of the referents of terms and of sentences and talk instead of meaning as distinct from reference. But it is not clear how this helps. Given the meaning of 'Theatetus' as argument, the meaning of 'flies' yields the meaning of 'Theatetus flies' as value. But this doesn't tell us what 'Theatetus flies' means, *in the sense of enabling us to understand it*, even if it in some sense assigns it a meaning. Davidson characterizes what is wanted in [9].

> [9] What analogy 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. Clearly some more articulate way of referring to meanings than any we have seen is essential if these criteria are to be met. (p. 20).

The difficulty is twofold. First, the term that refers to the meaning of a complex must somehow be constructed out of terms that refer to their meaningful parts. And second it must manage to convey to us what the sentence means in a way that enables us to understand it. But it is obscure how to do this. And so Davidson concludes famously in [10].

> [10] 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*. (p. 21; emphasis added)

And here the point is not to say that the concept of meaning is incoherent or obscure, but that the introduction of *entities* for each expression *which we call their meanings* has not been shown to do any real work in the project of giving a compositional meaning theory. This point is reinforced by the next proposal that Davidson takes up and rejects, which is that "syntax [in the sense of an effective method for telling for an arbitrary expression whether it is meaningful] … will yield semantics when a dictionary giving the meaning of each syntactic atom is added" (p. 21). The trouble is that "knowledge of the structural characteristics that make for meaningfulness in a sentence, plus knowledge of the meanings of the ultimate parts, does not add up to knowledge of what a sentence means" (p. 21). The objection is not to the concept of meaningfulness, or to sentences and words having meanings in the sense of being understood and understood differently. The target hasn't shifted. The objection is to the utility of *assignment of entities* in pursuit of the task of giving a compositional meaning theory.

Davidson says further,

> [11] While there is agreement that it is the central task of semantics to give the semantic interpretation (the meaning) of every sentence in the language, nowhere in the linguistic literature will one find, so far as I know, a straightforward account of how a theory performs this task. (p. 21)

When Davidson says in [11] there is agreement here on the central task of semantics, he clearly includes himself in the group in agreement on it. And the task he has set himself is to say "how a theory performs this task." When he goes on to propose a truth theory for the job, it is for precisely this job, the central task of semantics, a point he reiterates at the end, as noted above. And our job as interpreters is to understand how he sees the truth theory as accomplishing this task.

Prior to the introduction of his proposal to use a truth theory, there are several more clues to how he is thinking about it. The first is his remarking that there was (at the time he was writing) no clear criterion for a successful semantic theory, in contrast to a successful theory of grammaticality ("What clear and analogous task and test exist for semantics?" pp. 21-2). He is thus looking for a way of stating the goals of a compositional meaning theory that will give a clear criterion for success and make clear what the task of a meaning theory is. (See the discussion of his remarks in an interview in 1988 in the appendix also.) The second is his remark that having eschewed the assignments of meanings as entities to parts of sentences in favor of thinking of them as having meanings rather "in the ontologically neutral sense of making a systematic contribution to the meaning of the sentences in which they occur," we should think of explicating meaning in a language in a holistic way—that is, by thinking about the systematic contributions of expressions to how we understand sentences. The third occurs in the paragraph immediately preceding the proposal, which I quote in full in [12].

> [12] This degree of holism was already implicit in the suggestion that an adequate theory of meaning must entail *all* sentences of the form '*s* means *m*'. But now, 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*' in '*s* means *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. (p. 22)

In other words, we want a theory that provides an effective method of specifying the meaning of each object language sentence on the basis an understanding of its parts. Once we have given up the idea that sentence meanings, construed as entities, are going to be any help, we might as well characterize the project as that of providing an effective method of generating for each sentence *s* of the object language a true theorem of the form '*s* means that *p*'. So if we could formulate a theory like that, it would do exactly the job we wanted it to do. The objection to this is not that this is outright nonsense or employs a muddled or incoherent notion. It is rather 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 aims to solve." (Why does he say "apparently"? Already at this time he had worked out his paratactic account of indirect discourse ("On Saying That" was published in 1968), and so had proposed an extensional solution to the semantics of that-clauses.) What does Davidson mean by this? In what sense would we encounter problems as hard as or identical to those we set out to solve? The problem is that we need rules that tell us under what conditions we can substitute in the complement of sentences of the form '*s* means that *p*' if we are to develop a recursive definition of 'means that' that starts with semantic axioms about parts of sentences. And in the general case, this will require figuring out when we can substitute for one complex expression another. And it looks as if the condition for substituting salva veritate (without change of truth value) will be sameness of meaning. And part of what is involved in judging sameness of meaning as between semantically complex expressions is judging sameness of semantic structure. But this is precisely what a compositional meaning theory is supposed to give us insight into. And so formulating the proper rules of inference for a theory that issues outright in theorems of the form '*s* means that *p*' seems to require us already to have a systematic account of the compositional structure of sentences in the metalanguage we use to give the theory for the object language. We must then do for the metalanguage what we want to do for the object language, and in a way that doesn't generate a regress. That is the problem Davidson envisages in what otherwise looks like the most straightforward way of carrying out the project, in the aftermath of giving up on the utility of meanings as entities in the theory of meaning. It looks as if working out an adequate logic for dealing with 'means that' presupposes a solution to the problems that the theory is to deal with. I will call this the Presupposition Problem.

4. Davidson's Positive Proposal

It is with this in mind that we need to understand the next three paragraphs in "Truth and Meaning," in which Davidson suggests that the sort of theory we are looking for can be found in an axiomatic truth theory that meets Convention T. I will deal with each of these in turn. The solution is given, in a nutshell, in [13].

> [13] [a]The only way I know to deal with this difficulty is simple, and radical. [b] 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. [c] 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*. [d] One obvious candidate for matching sentence is just *s* itself, if the object language is contained in the metalanguage; otherwise a translation of *s* in the metalanguage. [e] 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. [f] The plausible result is
>
> (T) *s* is *T* if and only if *p*

I discuss this pivotal passage in detail. In [13a], it is important to note that Davidson implies that he will deal with the difficulty, namely, the Presupposition Problem, for carrying out the project aiming at 'means that' theorems. Calling the way of dealing with it simple and radical does not imply that he is abandoning the project. In [13b] he explains what his idea is: the problem stems from using 'means that' between the description of a sentence *s* and a sentence in the complement '*p*', but he suggests that "the success of our venture," namely, the task of giving a compositional meaning theory, "depends not on the filling," that is, on 'means that', connecting the sentence described and the one we match it with, but "on what it fills," that is, *getting the right pairing of object language sentence with metalanguage sentence*. In [13c] he elaborates on this idea. The work—the real work— of the theory will be done if "for every sentence *s* in the language under study," we provide "a matching sentence (to replace '*p*') that, in some way yet to be made clear, 'gives the meaning' of *s*." That is to say: we want an appropriate pairing of a sentence *s* in the object language described as constructed out of its semantically significant parts, with a sentence which plays a role like that of '*p*' in '*s* means that *p*' that in some similar way can be said to 'give the meaning' of *s*. Why 'in some way yet to be made clear'? Because, of course, it has yet to be made clear how we are to think of the sentence '*p*' functioning to do what it would in '*s* means that *p*' when we have dropped 'means that' from the picture. Why put 'gives the meaning' in scare quotes? For the same reason, and for the further reason that we are not going to be adverting to meanings as entities. Now notice what light [13d] sheds on these locutions. Davidson says that an obvious candidate for matching sentence is just *s* itself, or, if the metalanguage does not include the object language, *a translation of it in the metalanguage.* The sentence itself would suffice because that suffices for it to meet the condition specified in the second clause, namely, the sentence itself is a translation of the sentence. So the idea is that the crucial thing is that we have a theory that pairs a sentence described as constructed out of its significant parts with a sentence in the metalanguage, not mentioned but used minimally in the sense of being understood as it would be in use in the metalanguage, which translates it. And this, of course, is precisely the relation between *s* and '*p*' in '*s* means that *p*' which obtains if '*s* means that *p*' is true. So when we look back to [13b], we can see that the idea is that the crucial work done by the theory lies in the pairing of an object language sentence with a metalanguage sentence in use that translates it in a way that reveals the semantic structure of the sentence. This is just a very general way of describing what the design specification is of a theory, abstracting from the particular filler between object language expression and metalanguage expression. And that is the sense in which '*p*' will 'give the meaning' of *s*. Anyone who understands '*p*' in the metalanguage and understands that it translates *s*, that is, knows the instance of the pairing meets the relevant condition, will be in a position to understand *s*. [13e] takes the final step: as we want to pair *s* with a use of '*p*' that translates it, but do not want '*p*' to be in an intensional position, we should look for some "filler" that treats the position of '*p*' extensionally, and supplies a predicate for *s* so as to provide a sentence involving it, and then think of some extensional connective which we can place between them. (Why does he say "the *obscure* 'means that'"? This is just a reference to the difficulty mentioned in the previous paragraph of getting a handle on a logic for the intensional context it creates.) It is perhaps not fully dictated by the description of the problem what extensional connective to choose, but the obvious connective is 'if and only if', and, as Davidson says in [13f], the plausible result is: (T) *s* is T if and only if *p*.

Davidson sums it up in [14].

> [14] 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 *T*' to entail all sentences got from schema *T* when '*s*' is replaced by a structural description of a sentence of *L* and '*p*' by that sentence.

Thus, the suggestion in [14] is that we will have an adequate theory of meaning for a language L if it defines a predicate 'is T' for sentences of L, in terms of course of its semantically primitive vocabulary, which has as consequences all sentences of the form (T) in which '*s*' is replaced by a description of a sentence of L as built up out of its semantically primitive vocabulary in which '*p*' is replaced by *s*, or, in line with [13d], which we must not forget, *a translation of it*, if the metalanguage does not contain the object language, which, in the general case, it will not.

Thus, the goal of the characterization of the predicate is to match an object language sentence *s* as appropriately described with a metalanguage sentence in use that translates it. It gives the meaning of the object language sentence in pretty much exactly the same sense in which '*s* means that *p*' does, given that we do not take 'that *p*' or '*p*' to be referring terms, but we do take '*p*' to be in use. For in that case we glean what *s* means not by grasping any object it is related to, but by the use of '*p*' as paired with it, and given that the truth of '*s* means that *p*' requires '*p*' to translate *s*, that is just a matter of knowing that '*p*' translates *s* and understanding '*p*'. But now we see, in general outline, a way of achieving the same effect with an extensional theory.

Of course, as Davidson notes immediately in [15], and as he obviously had in mind all along, the conditions placed on the predicate 'is T' in [14] are just those Tarski placed on an adequate characterization of a truth predicate for a language (modulo the intended application here to a natural language).

> [15] Any two predicates satisfying this condition have the same extension, so if the metalanguage is rich enough, nothing stands in the way of putting what I am calling a theory of meaning into the form of an explicit definition of a predicate 'is T'. But whether explicitly defined or recursively characterized it is clear that the sentences to which the predicate 'is T' 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 T that tests the adequacy of a formal semantical definition of truth. (p. 23)

As a reminder, here is the way Tarski puts it in "The Concept of Truth in Formalized Languages" (1983, pp. 187-8), which Davidson here cites in a footnote.

> **Convention T.** A formally correct definition of the symbol "Tr," formulated in the metalanguage, will be called an *adequate definition of truth* if the deductive system of the metatheory proves the following:
>
> a. all sentences which are obtained from the expression "Tr(*x*) if and only if *p*" by substituting for the symbol "*x*" a structural-descriptive name of any sentence of the language in question and for the symbol "*p*" the expression which

forms the translation of this sentence into the metalanguage.

b. the sentence "for any x, if $\mathrm{Tr}(x)$ then x is a sentence of LCC."

The key point to notice is that it requires in (a) that '*p*' is to be replaced by *a translation into the metalanguage* of the sentence picked out by the structural description of the object language sentence.

We can see the parallel with the earlier theory of reference. Our generalization of Davidson's criterion of adequacy was that the theory entail each theorem of the form '*s* refers to *t*' where '*t*' translates *s*. Knowing the language of the theory and that it meets the condition of adequacy puts us in a position to interpret each object language referring term. We can in fact, as we noted, replace 'refers' with 'means' in the relevant theorems *salva veritate*. The same goes for the truth theory. The relation that Convention T requires between *s* and '*p*' in relevant instances of (T), as we have noted, is exactly that required between them in

(M) *s* means that *p*,

and we can therefore infer (M) from (T). Once we see this, we can see that this affords an alternative way of stating Convention T: we want a truth theory that entails for every object language sentence *s* a theorem of the form (T) such that the corresponding instance of (M) is true. Davidson makes basically the same point, namely, that we can infer (M) from (T) if we know (T) satisfies the requirement specified in Convention T, in "Semantics for Natural Languages" (1970; first read in 1968—the year after "Truth and Meaning" was published) as expressed in [16] (see also [48] below from "Reply to Foster").

> [16] A theory of truth entails, for each sentence *s*, a statement of the form '*s* is true if and only if *p*' where in the simplest case '*p*' is replaced by *s*. Since the words 'is true if and only if' are invariant, we may interpret them if we please as meaning 'means that'. So construed, a sample might then read '"Socrates is wise" means that Socrates is wise'. (p. 60)

We have thus found a promising approach for avoiding the two horns of the dilemma in the preamble, the inability of assignments of meanings (as Davidson held) to put us in a position to interpret object language sentences, on the one hand, and the difficulties of formulating a logic for non-extensional contexts, on the other.

Davidson sums it up in [17].

> [17] The path to this point has been tortuous, but the conclusion may be stated simply: a theory of meaning for a language L shows 'how the meanings of sentences depend upon the meanings of words' if it contains a (recursive) definition of truth-in-L. And, so far at least, we have no other idea how to turn the trick. It is worth emphasizing that the concept of truth played no ostensible role in stating our original problem. That 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. (pp. 23-4)

Importantly, when Davidson says that the concept of truth played no ostensible role in stating *the original problem*, he has in mind *the problem of giving a compositional meaning theory*. The refinement he refers to is the observation that the trick is to find an effective method (which, moreover, reveals semantic structure), based on the primitive vocabulary of the language, to pair object language sentences structurally described with metalanguage sentences that translate them. This is a way of putting the project that eschews the formulation in terms of meanings, that is, it is the ontologically neutral way of formulating the problem. It is important also to note that he says it "was in the nature of a discovery" that the predicate characterized would apply to just the true sentences of the language. He says "in the nature of" no doubt because he got the idea originally from thinking about Convention T. But the point of saying it is in the nature of a discovery is that one can develop a description of the problem that leads to a proposal that turns out to guarantee that a predicate that meets certain constraints also has exactly the true sentences of the language in its extension. That means that whatever else he is doing here he not just suggesting, as many have thought, that we should replace the concept of meaning with the concept of truth! Not at all. We have an independent characterization of what it would take to satisfy the requirements of a theory of meaning. The relation to the concept of truth drops out of that. It is not as if he said: so much for meaning, why don't we try truth instead! No, what he says is: meanings as entities do no work, and the real work lies in matching object language sentence, via a effective (and revealing) method starting from claims about its parts, with a used metalanguage sentence that translates it. *And it turns out a truth theory does that job.*

Thus, to sum up where we have got so far, taking these passages at face value, it is clear that the proposal that Davidson advances here is meant to meet the challenge he takes up at the beginning of the paper, as refined specifically by a reformulation that drops appeal to meanings as entities, on the grounds that they make no contribution to solving the problem, and, hence, that they should have no role in stating it either.

This is not a rejection of the concept of meaning as confused. It is not the substitution of a different project (giving a truth theory) for the one he announces at the outset. It is not a selection of certain features of meaning for preservation (extensional properties, or "strong truth conditions," or anything else). It is in fact a straightforward response to that problem that shows how, surprisingly, a truth theory meeting Convention T meets all the reasonable demands we might place on a meaning theory. (Here I have to add that so far we have not got on the table all of the requirements, but the immediate aim is to say how Davidson is thinking of it, not to assess the full adequacy of the account he gives in "Truth and Meaning.")

Let us turn briefly to the intersection of these passages with Ebbs's argument for the Explication Interpretation. Ebbs boldly takes [13] to support his reading of Davidson (2012, pp. 79-80). The primary support he takes from the passage, however, rests on the observation that Davidson did not state the requirement that T-sentences ('s is true iff p') are to meet in terms of the corresponding M-sentences ('s means that p') being true (that was a way that Lepore and I expressed what Convention T amounts to in (Lepore and Ludwig 2005, pp. 83-4)). Yes, but *this is equivalent (as noted above) to the condition he does require, namely, that the sentence that replaces 'p' translate the sentence that s describes.* For 's means that p' is true just in case 'p' translates s. So it is difficult to see what weight this has, and Davidson, having already remarked that the logic of 'means that' is unclear, and having identified the point of that locution as to match object language sentence with metalanguage sentence, could hardly feel himself under any pressure to restate Convention T in this way. Furthermore, as noted above, Davidson explicitly makes the connection both in "Semantics for Natural Languages" and in "Reply to Foster" [48]. There are other passages that Ebbs cites, and we will come to them in due course, but [13] is a passage that *needs to be overcome* to make sense of the Explication Interpretation, not one that supports it.

Let me take up briefly a phrase in [14] that I haven't touched on, namely, the requirement that the theory "without appeal to any (further) semantical notions" place enough restrictions on the predicate to ensure that it meets Convention T. This is a phrase that Ebbs identifies later in his discussion as significant, and it is one Lepore and I also discussed in (Lepore and Ludwig 2005, p. 98). Ebbs takes this to be a rejection of the idea that the theory is to meet Convention T. But how can that be if the constraint is to ensure that it meets Convention T? Ebbs may be misled by Davidson's giving the syntactic criterion for meeting Convention T in [14]. But it is clear that he is not thinking that our goal in general is to state a truth theory in an extension of the object language. As he says in the previous paragraph [13d], "[o]ne obvious candidate for matching sentence [in order for the theory to have done its work in matching mentioned object language sentence with used metalanguage sentence so as to give the meaning of the object language sentence] is just *s* itself, if the object language is contained in the metalanguage; *otherwise a translation of s in the metalanguage*" (my emphasis). (Ebbs in fact elides the phrase just emphasized in his quotation of the passage on p. 79). Reversion to the syntactic criterion in [14] is a simplification, but the general condition (he has in mind here) is that the sentence that replaces '*p*' translate *s*, and the syntactic criterion suffices when the object language is embedded in the metalanguage.

In later work, specifically in [18] drawn from "Belief and the Basis of Meaning" (1974, first read at a conference in March 1973), Davidson notes that in empirical applications the syntactic test is in fact useless,

> [18] … since such a test would presuppose the understanding of the object language one hopes to gain. The reason is simple: the syntactical test is merely meant to formalize the relation of synonymy or translation, and this relation is taken as unproblematic in Tarski's work on truth. Our outlook inverts Tarski's: we want to achieve an understanding of meaning or translation by assuming a prior grasp of the concept of truth. What we require, therefore, is a way of judging the acceptability of T-sentences that is not syntactical, and makes no use of the concepts of translation, meaning, or synonymy, but is such that acceptable T-sentences will in fact yield interpretations. (p. 150)

Even translation is an idealization in the case of natural languages because we must, as Davidson notes later in "Truth and Meaning," relativize the truth predicate to contextual parameters like speaker and time, and so add argument places to the right hand side of the T-sentence bound by quantifiers over those parameters, so that what we want is not translation but something that gives the meaning of the object language sentence relative to a context of use. Think of this in connection with [27] below. (In "Reply to Foster," Davidson says, "In natural languages indexical elements, like demonstratives and tense, mean that the truth conditions for many sentences must be made relative to the circumstances of their utterance. When this is done, the right side of the biconditional of a T-sentence never translates the sentence for which it is giving truth conditions. In general, an adequate theory of truth uses no indexical devices, and so can contain no translations of a very large number and variety of sentences" (p. 175).)

Therefore, there is no room for interpreting [14] as rejecting the requirement that the theory satisfy Convention T (or a generalization for context sensitive languages), as Ebbs suggests.

But why is this phrase "without appeal to any (further) semantical notions" in here? Insight can be gleaned from noticing that this echoes something that Davidson says about the simple reference theory he discusses earlier in [8], repeated here:

> [8] what we wanted, and what we got, is a theory that 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 that term itself. Further, our theory accomplishes this without appeal to any semantical concepts beyond the basic 'refers to'. Finally, the theory clearly suggests an effective procedure for determining, for any singular term in its universe, what that term refers to. (pp. 18-9)

I suggest there are two things going on here. The first and simpler point is just that Davidson does have in mind that the constraints give you a truth theory, and that a truth theory that had the right output could be stated so that it did not appeal to any semantic concepts beyond truth and satisfaction. *This is just the point he makes about the reference theory.* It does *the job*, the job of matching object language sentence with used metalanguage sentence that translates it, without appealing to any notion beyond that of reference (broadly construed).

The second is something that emerges in the next few pages, and which looks back to his remarks about holism: it is the idea that in the case of a natural language, given that the theory is basically a recursive characterization of a truth predicate for the language, and that the theory has to assign correct truth conditions for a potential infinity of context sensitive sentences, any theory that is not false will *ipso facto* meet Convention T. This represents a broadening of the ambition of the paper from just giving an account of how sentences are to be understood on the basis of understanding their parts to providing a deeper illumination of the meaning even of the parts. The idea is that if a true truth theory sufficed for meeting Convention T (for a natural language), then we would have specified a condition on the theory, without using the concept of meaning, that would fix simultaneously the meanings of sentences and words, as the meanings of words are abstractions from their roles in contributing to the meanings of sentences. (This is essentially the task described in [18] from "Belief and the Basis of Meaning," where Davidson goes on to put it explicitly: "Our problem is to find constraints on a theory strong enough to guarantee that it can be used for interpretation" (p. 150); see the discussion in 7 and the appendix.)

That this is Davidson's idea is supported by the following passages from the last pages of "In Defence of Convention T," which, though published in 1973, was first read at a conference in 1970. The purpose of these passages is given in [19].

> [19] If a semantic theory claims to apply, however schematically, to a natural language, then it must be empirical in character, and open to test. In these concluding pages I should like to sketch my reasons for thinking that a theory that satisfies Convention T is verifiable in an interesting way. (p. 73)

Davidson remarks after this that when we treat a theory of truth for a language as an empirical theory we cannot assume that the object language is contained in the metalanguage, even if the same expressions appear in both. Then:

> [20] … what becomes of Convention T? How is a T-sentence to be recognized, let alone recognized for true?
>
> I suggest that it may be enough to require that T-sentences be true. Clearly this suffices uniquely and correctly to determine the extension of the truth predicate. If we consider any one T-sentence, this proposal requires only that if a true sentence is described as true, then its truth conditions are given by some true sentence. *But when we consider the constraining need to match truth with truth throughout the language, we realize that any theory acceptable by this standard may yield, in effect, a usable translation manual running from object language to metalanguage.* The desired effect is standard in theory building: to extract a rich concept (here something reasonably close to translation) from thin bits of evidence (here the truth of sentences) by imposing a formal structure on enough bits. If we characterize T-sentences by their form alone, as Tarski did, it is possible, using Tarski's methods, to define truth using no semantical concepts. If we treat T-sentences as verifiable, then a theory of truth shows how we can go from truth to something like meaning—enough like meaning so that if someone had a theory for a language verified in the way I propose, he would be able to use that language in communication. (pp. 74-5; emphasis added)

[20] makes it is clear that Davidson is suggesting that any truth theory that is simply true for a natural language will satisfy Convention T. In the lead up to this discussion, Davidson states that the "inevitable goal of semantic theory is a theory of a natural language" (p. 71). As I will remark below (see the discussion of [25]-[27]), it is because natural languages contain context sensitive elements that he thinks this is plausible, for then the theory has to deal with the potential application of predicates of spatiotemporal objects to potentially any such object through the use of demonstratives. Davidson himself remarks in the next paragraph but one

that "[o]ne important, indeed essential, factor in making a truth theory a credible theory of interpretation is relativization to speaker and time" (p. 74).

Should there be anything to concern us in the parenthetical remark "something reasonably close to translation" and the qualifier "something like meaning"? Given that he goes on to say "enough like meaning so that if someone had a theory for a language verified in the way I propose, he would be able to use that language in communication," it is easy to see why the result would be reasonably close to translation and something like meaning, for to use it for communication would mean that one could use it to enable one to understand what speakers of the language meant when using it. Why any hesitation at all then? The point to bear in mind is that he envisions a theory that departs from Tarski's by being outfitted to deal with context sensitivity, as just noted, and in that case we do not expect that the sentences that give interpretive truth conditions relative to context to mean the same as or to translate the context sensitive sentences for which they specify an interpretation relative to context, for they strip away the context sensitivity which is clearly a component of their meaning.

As it turns out, Davidson was mistaken in "Truth and Meaning" and in "In Defence of Convention T" both in thinking that what is stated by a truth theory that meets Convention T is enough to interpret another, and in thinking that a truth theory of a natural language merely being true (and even counterfactual supporting) suffices for meeting Convention T (or the analog for a natural language). In point of fact, we need to appeal to more than what is stated by the reference theory or the truth theory to be in a position to use either to interpret. And the bare requirement that the theory be true is not adequate for it to meet Convention T. These are points that Davidson concedes in later work, as we will see, and this reinforces the point that the target is a theory that generates interpretive T-sentences ([23] below and section 7).

I have now laid out the basic case for the Traditional Pursuit interpretation. There remains, so far as "Truth and Meaning" goes, the task of parsing the pages that immediately follow the introduction of the proposal, which have provided much of the grist for the opponent's mill. I take that up in the next section, and the following section I look at work contemporary with "Truth and Meaning." For an abbreviated route through the paper, a reader can skip to section 7 for the light that later work sheds on the project in "Truth and Meaning."

5. Problematic Aftermath?

The main textual sources of the Replacement Theory, at least so far as "Truth and Meaning" goes, lie in the six paragraphs that follow the one from which I have last quoted, and in particular in paragraphs 1, and 3-6, of those. The last three need to be read in the light of the treatment of the theory as an empirical theory. I quote the first of these paragraphs in full in [21], followed by commentary (bracketed letters added).

> [21] [a]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. [b] 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. [c] 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. [d] 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. [e] 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'. [f] So much to the good for what I call a theory of meaning, and so much, perhaps, against my so calling it. (p. 24)

[21a] asserts what we might expect at this point, that there is a connection between a Tarski-style definition of truth and the concept of meaning. [21b], however, is often cited as evidence that Davidson is not really interested in the concept of meaning after all, but is instead suggesting replacing it. This is on the face of it, however, a perverse reading, since Davidson has just said that there is a connection between the concept of meaning and a definition of truth of the kind that Tarski has shown how to construct! Are we not to take him seriously? The reason commentators have not taken him seriously, I think, is that they have supposed that Davidson could not seriously have meant that giving conditions necessary and sufficient for the truth of a sentence, that is, a condition that obtains just in case the sentence is true, would tell us much about the meaning of a sentence. And that is certainly true. But this is to ignore the context in which Davidson is making this remark. Specifically, Davidson is talking about giving truth conditions in the context of a Tarski-style truth definition, one that meets Convention T. If you give necessary and sufficient conditions for a sentence by way of a Tarskian truth definition for it, and so give its truth conditions *in that sense*, you have indeed expressed precisely what it means by giving those conditions *using a sentence that translates the sentence for which you are giving truth conditions*.

And that this is what Davidson has in mind is shown by the next sentence [21c] in which he says that knowing *the semantic concept of truth*, that is, the concept that Tarski defines, is to know what it is for any sentence to be true and amounts in one good sense we can give to the phrase to understanding the language. For knowing the semantic definition of truth, and that it is the semantic definition, which is what Davidson likewise assumes here (see his retrospective remarks in "Reply to Foster," p. 173, specifically [23] below), amounts to knowing what it is for any sentence to be true in a way that puts one in a position to interpret each into a language one understands, via the canonical theorems (more on this below) that give the truth conditions. One can put it this way: if someone asks what a certain sentence means, for example, what 'John palters' means, you can answer by saying 'John palters' is true iff John prevaricates in action or speech, where it is understood that you are not just giving a sentence that is the same in truth value as 'John palters' but which translates it. This uses a biconditional specifying truth conditions to give the meaning of a sentence, and in the context it is clear that that is what is intended. In just the same way, a Tarskian definition of the semantic concept of truth for a language gives the meaning of each sentence of the language. One need only know the metalanguage, the definition, and that the definition is a Tarskian definition and so meets Convention T, and have a way of effectively picking out the theorems in virtue of which the definition meets Convention T. This would put one in a position to interpret any sentence of the language, and in light of this it seems entirely reasonable to claim that "this amounts, in one good sense we can give to the phrase, to understanding the language." One can worry about various aspects of this. Perhaps we want to put some constraints on the axioms to ensure that the proofs reveal semantic structure, and perhaps the sense of understanding is theoretical rather than practical, but the point in the present interpretive context is that the connection between the concept of meaning and a Tarskian truth definition that Davidson is drawing attention to involves a concept of meaning that is connected with being able to interpret a sentence that someone uses, that is, to grasp what he means in using the sentence literally (or first meaning, as Davidson puts it in "A Nice Derangement of Epitaphs" (Davidson 2005)).

Ebbs fastens on the phrase "in one good sense we can give to the phrase" (p. 82) in [21c], but this is a thin reed on which to rest the Explication Interpretation, and a close reading of the paper up to this point, and a proper appreciation of the role of Convention T in constraining the semantic conception of truth, shows that there is *no reason* to read it as rejecting the concept of meaning and *a perfectly straightforward way* of understanding it that is compatible with Davidson doing exactly what he says he is setting out to do.

[21d] has been cited as well as a reason to think that Davidson is jettisoning the usual project, because he says his "freewheeling use of the word 'meaning'" is "apt to shock old hands." "Freewheeling" here has been interpreted to mean "nonstandard," and the nonstandard use is what is to be shocking to old hands. But what Davidson says immediately after this shows what he has in mind: what he calls a theory of meaning does not make use of *meanings* of sentences, or of words. And in light of the discussion in the first five pages of "Truth and Meaning," it is clear that what he means is that it does not assign meanings as entities to sentences or words. It is the rejection of the whole Fregean tradition that is apt to shock old hands. But the rejection of this philosophical approach to understanding meaning is not the rejection of the project of understanding meaning.

In [21e], Davidson says that what he calls a theory of meaning "falls comfortably within what Quine terms the 'theory of reference' as distinguished from what he terms the 'theory of meaning'." This has likewise been taken to be an abdication of the ambition to pursue the theory of meaning as opposed to the theory of reference, and so to indicate support for the Replacement Theory (or, in Ebbs's version, the Explication Interpretation). The contrast Quine has in mind is expressed in [22].

> [22] 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)

So what does Davidson mean when he says that what he calls a theory of meaning falls comfortably within what Quine terms the theory of reference? What Davidson has in mind is that the theory of truth *per se* draws only on the concepts that Quine lists as the concepts of the theory of reference. It is absolutely clear that this is so and that it is not only compatible with the account I have given of Davidson's project but entailed by it.

There is a mistake here on Davidson's part, which he acknowledges in retrospective remarks (see "Radical Interpretation" pp. 138-9, and "Reply to Foster," pp. 171-3). The mistake is to identify the truth theory with the meaning theory. For knowledge of the truth theory does not by itself put one in a position to use it to interpret another. One must also know certain things about it, such as that it meets Convention T, and one must have, as noted above, an effective method for picking out the theorems of it in virtue of which it satisfies Convention T (a canonical proof procedure). Davidson acknowledges the mistake in [23], which is drawn from "Reply to Foster."

> [23] 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 if from evidence, and such a person could not fail to realize that his theory satisfied the constraints. (p. 173)

We will return to "Reply to Foster" at greater length below. For now the point is that [21e] is exactly what one would expect Davidson to say on the interpretation I have given, given his identification of the meaning theory with the truth theory, which he later admits came from conflating the theory with the relevant body of knowledge he was supposing one would have about it in virtue of having constructed it from evidence. Against this background, the final remark in the paragraph calls for no special comment.

In sum, a careful reading of [21] shows that it supports the Traditional Pursuit interpretation, and that the alternative reading, whether the straight Replacement Theory or the Explication Interpretation, is forced to ignore everything in the paper up to this point and to take Davidson to be disingenuous in suggesting that he is providing a solution to the problem that he sets himself, and disingenuous in suggesting that there is a connection between the concept of meaning and a Tarskian definition of truth, and at the same time to ignore the fact that there is a perfectly reasonable story about what the connection is that makes what Davidson says a perfectly sensible continuation of the project he announces that he is pursuing in the paper.

I have yet to deal with paragraphs 3-6 of the aftermath of the introduction of the proposal. But an essential prelude is paragraph 2, in which Davidson notes that "A theory of meaning (in my mildly perverse sense [by which he means as noted that it does not appeal to meanings as entities]), is an empirical theory, and its ambition is to account for the workings of a natural language" (p. 24). The rest of the paragraph notes that for our own language it is easy to tell when a theory makes the correct predictions (this, by the way, contradicts Ebbs's claim that Davidson never suggests we can test the adequacy of a truth theory by appeal to our competence in our own languages (Ebbs, p. 85)). Here he is thinking that it is clear to us when a theorem counts as one in virtue of which the theory satisfies Convention T. The difficulty lies in developing a theory that has the right predictions.

> [24] Empirical power in such a theory depends on success in recovering the structure of a very complicated ability—the ability to speak and understand a language. We can tell easily enough when particular pronouncements of the theory comport with our understanding of the language; this is compatible with a feeble insight into the design of the machinery of our linguistic accomplishments. (p. 25)

There are two points I want to draw attention to in [24]. The first is just the idea that the theory is to be treated as an empirical theory. The second is the explicit aim of accounting for the ability to speak and understand a language. This is exactly the project announced at the outset in "Theories of Meaning and Learnable Languages," a project continuous with and of the same type as that pursued traditionally by linguists and philosophers.

The next four paragraphs (3-6) are among the most difficult in "Truth and Meaning," and Davidson himself remarks in a footnote to paragraph 6 that it is confused. At the risk of being tedious, I propose to work through them with some care, for the seeds of the Replacement Theory are frequently planted in these fields. The first of these is given in [25].

> [25] [a] The remarks of the last paragraph apply directly only to the special case where it is assumed that the language for which truth is being characterized is part of the language used and understood by the characterizer. Under this circumstance, the framer of a theory will as a matter of course avail himself when he can of the built-in convenience of a metalanguage with a sentence guaranteed equivalent to each sentence in the object language. [b] Still this fact ought not to con us into thinking a theory any more correct that entails '"Snow is white" is true if and only if snow is white' than one that entails instead:
>
> (S) 'Snow is white' is true if and only if grass is green.
>
> provided, of course, we are as sure of the truth of (S) as we are of that of its more celebrated predecessor [this echoes paragraph 6, and so Davidson's remarks about what is wrong with 6 apply here as well—more on this below]. Yet (S) may not encourage the same confidence

> that a theory that entails it deserves to be called a theory of meaning. (pp. 25-6)

It is [25b] that causes trouble. But [25a] is about what recourse we can have *if we are confirming a theory for our own language*. In [25b], the shift is obviously to a more general perspective when we ask what is to be said about the language from the point of view of anyone whether or not it is a language that that person already speaks. And so the question Davidson is raising is what counts for correctness of a truth theory, construed as an empirical theory, from that perspective.

The next paragraph, [26], helps to shed some light on this, together with a footnote, [27], added in 1982.

> [26] The threatened failure of nerve may be counteracted as follows. The grotesqueness of (S) is in itself nothing against a theory of which it is a consequence, 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 party to such an enterprise, 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. (p. 26)

> [27] Critics 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.)

Let us adopt the default interpretive assumption that the retrospective footnote expresses correctly Davidson's view of what he was aiming at in the original paragraph. Then it is clear that he does not think, after all, that (S) would be part of a correct truth theory for the language. He does not of course deny that it is true. So in [26] the point cannot be that any true biconditional of the form (T) is as good as any other in giving the meaning of an object language sentence. It cannot even be that it would be fine as long as we were as sure of its truth as of the truth of the canonical: 'Snow is white' is true iff snow is white. His idea, rather, is that in the general case, in which we test a theory for a language not our own, it is the totality of its predictions that are relevant. (Recall [20] in which this is made explicit; see also [29] below.) He makes clear that he does not think that (S) would survive this test. And in the footnote, if not in the original, he indicates his reason for thinking this. To get (S) as a theorem, the axioms for 'snow' and 'white' would have be roughly of the following form (I hedge because the full-blown theory would invoke sequences, or functions from variables to objects, as satisfiers, that is, what the predicates were true of):

> (t)(x)('snow' is true of x as used by s at t iff x is grass at t)
> (t)(x)('white' is true of x as used by s at t iff x is green at t)

These must get the right results not just for 'snow is white' but also for 'this is snow' and 'this is white'. But it would predict, for 'this is snow', for example, that it will be true on an occasion of utterance iff the speaker demonstrates something that is grass. And that is the wrong result. But, then, if he thinks that (S) could not be party to a successful theory for a natural language which has to get the right truth conditions for sentences containing demonstratives, what is the point of saying that

> [26c] If … (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 … be anything essential to the idea of meaning that remained to be captured.

What could it be except an expression of the conviction that a theory that was in fact true would *ipso facto* satisfy Convention T (again recall [20])?

And note here how strange this way of putting it would be *if* we took Davidson to be suggesting a replacement for, or Carnapian explication of, the notion of meaning, which is not bound to a "theory-neutral standard" of correctness. If that is what he is doing, why doesn't he say: "If (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 it would 'give the meaning' of 'Snow is white' in the sense of 'meaning' I am introducing/selecting." There would be no need to *counteract* a threatened failure of nerve if there is no substantive claim being made but only a stipulation.

We can confirm this interpretation by looking at the next paragraph, [28], and at a remark that Davidson makes, [29], in "Radical Interpretation" about his methodology (that he has radical interpretation in mind as the proper standpoint from which to confirm a theory even in "Truth and Meaning" is made clear in the middle paragraph on p. 37).

> [28] What appears to the right of the biconditional in sentences of the form '*s* is true if and only if *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 if and only if *s* is. (p. 36)

> [29] In philosophy we are used to definitions, analyses, reductions. Typically these are intended to carry us from concepts better understood, or clear, or more basic epistemologically or ontologically, to others we want to understand. The method I have suggested fits none of these categories. I have proposed a looser relation between concepts to be illuminated and the relatively more basic. At the centre stands a formal theory, a theory of truth, which imposes a complex structure on sentences containing the primitive notions of truth and satisfaction. These notions are given application by the form of the theory and the nature of the evidence. The result is a partially interpreted theory. The advantage of the method lies not in its free-style appeal to the notion of evidential support but in the idea of a powerful theory interpreted at the most advantageous point. This allows us to reconcile the need for a semantically articulated structure with a theory testable only at the sentential level. The more subtle gain is that very thin evidence in support of each of a potential infinity of points can yield rich results, even with respect to the points. By knowing only the conditions under which speakers hold sentences true, we can come out, given a satisfactory theory, *with an interpretation of each sentence*. It remains to make good on this last claim. The theory itself at best gives truth conditions. *What we need to show is that if such a theory satisfies the constraints we have specified, it may be used to yield interpretations*. (pp. 137-8; emphasis added)

[28] should be interpreted in light of Davidson's earlier remark in "Truth and Meaning" about holism (see the discussion at [12] above and the hyperbolic remarks in the first full paragraph of "Truth and Meaning" on p. 22): when we have put aside meanings as entities, the meaning of words and expressions is to be sought in their systematic contributions to the meanings of sentences. He is thinking that the truth theory is an empirical theory, and that a merely true truth theory for a natural language, in the light of the need to accommodate indexicals, will suffice for it to satisfy Convention T. So when a T-sentence for *s* is a consequence of a theory that gets it right for every sentence of the language, the sentence used to give its truth conditions contributes to determining its meaning by being party to a whole theory that gets it right for the totality of sentences of the language. As he thinks an empirically

confirmed theory for a natural language will satisfy Convention T, and reveal the role of each word in it in virtue of their roles in sentences in general, this tells us what there is to know of the meaning of *s*.

Why the remark, "not by pretending synonymy"? Might this not be interpreted as denying that '*p*' translates *s*? That would be a strained interpretation in the light of the rest of the paper! The point is just that what fixes it as giving the interpretation of the object language sentence is not that the biconditional says that it translates the object language sentence—manifestly it does not *say* that—but that it is a consequence of a theory that suffices for the language as a whole, and that, Davidson is suggesting, suffices for it to satisfy Convention T. (One might also take this remark to be a rejection of the very intelligibility of synonymy, though it is difficult to see why it should be so construed. But in any case, Davidson does not in this paper show any antipathy to the very idea of synonymy, or analyticity for that matter. As noted earlier, he remarks, later in the paper (p. 33), when noting that the truth theory leaves untouched the analysis of particular words, that "synonymy, as between expressions, goes generally untreated, [and] so also synonymy of sentences, and analyticity." This isn't to dismiss these notions as unintelligible, but only to say that having a truth theory in hand does not deal with them. See also, in this connection, his free use of "synonymy" in [51] in his analysis of 'entails that' in "Reply to Foster," discussed at the end of section 7.2 below).

Davidson notes that as he conceives of it, the theory of truth is an empirical theory. When he considers the confirmation of such a theory for speakers in general, he suggests that, in light of the fact that many natural language sentences are context sensitive, the mere task of developing an empirically adequate theory will suffice by itself to ensure that it meets Convention T. Thus, if (S) ''Snow is white' is true iff grass is green', were the issue of such a theory (and Davidson is clear that he does not think it could be), it would be interpretive. The choice of the example here, which Davidson himself denies could be party to the enterprise, is perhaps unfortunate. But the point being urged is just that empirical adequacy plausibly secures interpretiveness, that from thin evidence at a lot of points a rich result can be secured, namely, a theory that is not only true but satisfies Convention T, and, hence, which can be used for interpretation.

That this is the right interpretation of what is going on here is strongly supported by passage [29] from "Radical Interpretation." The suggestion is that by treating a truth theory modified to accommodate a natural language as an empirical theory, we impose a very strong constraint on it. The evidence, truth of various biconditionals, and, as he makes clear, their ability to track truth counterfactually as well, is thin evidence at each point, not evidence directly about interpretability, but since there are an infinity of points, the idea is to extract rich results: "knowing only the conditions under which speakers hold true sentences, we can come out, given a satisfactory theory, with an interpretation of each sentence." And he is clearly distinguishing between truth conditions in the sense of conditions under which a theory is true, and interpretations, and the goal is to show that a theory that satisfies the conditions can *in fact* be used for interpretations. While it is better worked out in "Radical Interpretation," one can see this as the idea that lies behind these remarks in "Truth and Meaning."

Ebbs places considerable weight on the conditional in [26c]. Ebbs writes:

> [E1] On the most natural and plausible reading, Davidson is asserting that if, contrary to what he takes to be so, (S) followed from a well-confirmed theory of truth for English that led to the invariable pairing of truths with truths and falsehoods with falsehoods, then there would not be anything central to the idea of meaning that remained to be captured. The point of passage [d] [this is [26] above], as I read it, is to express Davidson's commitment to his theory, even in cases where it would violate the s-means-that-p requirement [here Ebbs has in mind the requirement

> that the theory meet Convention T]. On the reading that Lepore and Ludwig favor, however, this cannot be the point of passage [d], since, according to their reading, Davidson is committed to the *s*-means-that-*p* requirement, which the biconditional (S) violates. … If we take [d] as a subjunctive conditional, it seems we must see Davidson as confused or mistaken about whether (S) satisfies the *s*-means-that-*p* requirement. [For:] … on the subjunctive reading we are now considering, it follows from Lepore and Ludwig 's interpretation of Davidson that despite his supposed commitment to the *s*-means-that-*p* requirement, he failed to see that by our ordinary, pretheoretic standards, [m] is false. (p. 87)

It seems that Ebbs is thinking that since we all recognize that [S] is not interpretive, Davidson could not have meant by [26c] to express the thesis that getting a truth theory that had true consequences for all sentences of the language would suffice for it to meet the appropriate analog of Convention T for natural languages. Suppose, however, that someone had said,

> If, contrary to fact, a parrot discoursed, reasoned and philosophized as well as a man, then there would not be anything central to the idea of a person that remained to be captured.

Would we *deny* that the philosopher who asserts this thinks that discoursing, reasoning and philosophizing like a man are sufficient to be a person *on the grounds that everyone knows that parrots are not persons*? Evidently not. Nor should we draw the parallel conclusion substituting (S) for the role of the parrot and substituting following from a truth theory that entailed true (counterfactual supporting—see [31] below) T-sentences for every object language sentence for the role of discoursing, reasoning and philosophizing as well as a man.

There is a final paragraph to deal with, and of this Davidson himself says in a retrospective footnote that it is confused. Here is the paragraph in [30] and the footnote in [31].

> [30] It may help to reflect that (S) is acceptable, if it is, because we are independently sure of the truth of 'Snow is white' and 'Grass is green'; but in cases where we are unsure of the truth of a sentence, we can have confidence in a characterization of the truth predicate only if it pairs that sentence with one we have good reason to believe equivalent. It would be ill advised for someone who had any doubts about the colour of snow or grass to accept a theory that yielded (S), even if his doubts were of equal degree, unless he thought the colour of the one was tied to the colour of the other. Omniscience can obviously afford more bizarre theories of meaning than ignorance; but then, omniscience has less need of communication. (pp. 26-7)
>
> [31] This paragraph is confused. What it should say is that sentences of the theory are empirical generalizations about speakers, and so must not only be true but also lawlike. (S) presumably is not a law, since it does not support appropriate counterfactuals. It's also important that the evidence for accepting the (time and speaker relativized) truth conditions for 'That is snow' is based on the causal connection between a speaker's assent to the sentence and the demonstrative presentation of snow. For further discussion see Essay 12 ["Reply to Foster"]. (p. 26, n. 11)

First, what Davidson appears to be saying in this passage is that the acceptability of a sentence like (S) depends on our assurance that it is true, and that if one doesn't independently know that each side is true, then we can be confident of the truth theory that gives this result only if we can be confident of the theory resting on facts about the language that don't depend on having independent reason for thinking each sentence to be true. The strange thing about this is the suggestion that nonetheless its merely being true as opposed to being the product of a theory that got it right for all sentences is somehow the touchstone for adequacy. The retrospective footnote says that the paragraph is confused. Davidson does not say exactly how he thinks it is confused. He says that it should say that the theory, being an empirical theory about speakers (*s* is true as used by *U* at *t* iff *p*), must be lawlike, and (S) is not lawlike. (I presume that he is thinking that in a counterfactual situation in which grass is blue but snow is white, the theory would predict that speakers would not accept 'Snow is white', and that this would be incorrect for speakers of English.) So again he disclaims the acceptability of (S), and the point of talking about it seems to be to draw attention to what he thinks makes for satisfaction of Convention T (the acceptability of the theory for interpretation). In light of this footnote, it can be seen that, at least in retrospect, he was thinking that part of what made for adequacy for the purposes of interpretation is not just the theory's getting it right as a matter of fact, but getting the predictions about truth conditions, as seen from the perspective of speakers of the language, right in counterfactual circumstances as well. Whether this is indeed adequate is another question, and plausibly later Davidson came to think it was not (see the discussion of "Radical Interpretation" in section 7.1).

To take stock, in section 5, we have looked carefully at the discussion in the immediate aftermath of the introduction of the proposal to use a theory of truth in pursuit of a meaning theory for natural languages. The first paragraph, on examination, is merely an elaboration of the basic idea. Knowledge of the semantic concept of truth for a natural language, specifically, a truth definition that satisfies Convention T, and, what Davidson assumes but does not make explicit, knowledge that it is such a definition, puts one in a position to understand the language, and the contribution of words to sentence meaning. It may be apt to shock old hands because it does not advert to assigning meanings as entities to words or sentences, and it (the truth theory specifically) does not invoke any concepts other than those of the theory of reference—which is not to say that knowledge that it is a semantic definition of truth does not involve knowledge that it satisfies Convention T. An axiomatic theory for the semantic concept of truth satisfies Convention T, but does not say that it does so. (Again, Davidson notes this explicitly in later work and acknowledges he was not as clear as he could have been about it in "Truth and Meaning," as noted above in connection with [23]; see also his remarks on p. 138 in "Radical Interpretation.")

I said earlier that one should not attribute to a philosopher a project at odds with the one that he announces he is engaged in if there is a reasonable way to construe him as engaged in it. In a close reading of "Truth and Meaning," I have given a way of construing Davidson as pursuing the project that he announces himself as engaged in that is not only natural and reasonable but the only one that makes good sense of everything that he says. The Replacement Theory, and its Explication variant, is strained, is never suggested by Davidson himself, and requires us to think that Davidson is repeatedly being disingenuous about what he is up to. Ebbs says that Davidson was trying to explicate "those aspects of our uses of ['translates', 'means that', and 'interprets'] that are most important to us" (2012, p. 103). Ebbs never says what the target is however, what aspects of our use of those terms Davidson was supposed to be trying to capture, or why. He admits, in a footnote, that "Davidson's application of the method [of explication] is not as explicit" as he would like (2012. p. 103, n.12), for Davidson "says little about why he thinks his constraints capture what matters to us in our unregimented uses" of semantic terms (2102, p. 103, n.12). In fact, Davidson says not just little but nothing that touches on this at all. I suggest that there is an obvious reason for this. He was not doing what Ebbs claims him to have been doing. All Davidson's explicit remarks on method (e.g., in [29] above) suggest he is engaging in a straightforward attempt to say what it is "for words to mean what they do" (p. xiii). On Ebbs's interpretation of Davidson, it is quite unclear what he is supposed to have been aiming at, quite unclear how to evaluate it, and quite unclear what its significance is supposed to be. The project I have attributed to Davidson is philosophically subtle, dialectically clever, motivated by a sophisticated and illuminating critique of the traditional appeal to meanings as entities, constrained by what is generally agreed to be a central requirement on any theory of language, namely, that its constitutive features be intersubjectively available, and if it is correct, is of fundamental importance for our understanding of language and meaning. The Replacement or Explication Theory, on the other hand, attributes to Davidson an esoteric doctrine which he never announces, an attitude toward the concept of meaning he never expresses or argues for, and a puzzling positive suggestion put forward under what would have to be, on his own view, a misleading label, for who knows what end. Which is to be preferred?

6. Other Relevant Passages From Near Contemporary Work

I have concentrated on a close reading of the relevant sections of "Truth and Meaning" because this is the epicenter of the interpretive dispute. Yet "Truth and Meaning" is not the only essay Davidson wrote on the topic. The interpretive line I have pursued is reinforced by looking at "Truth and Meaning" in the context of Davidson's later work. I drew on some of that work in the previous section, passages from "Radical Interpretation" and "Reply to Foster." I return to these in the next section. In this section, I consider some passages in "Semantics for Natural Languages" (1970), originally read at a conference in 1968, and in "In Defence of Convention T" (1973), originally read at a conference in 1970.

"Semantics for Natural Languages" begins, in [32], not surprisingly, with an announcement of the same project as that of "Theories of Meaning and Learnable Languages" and "Truth and Meaning":

> [32] A theory of the semantics of a natural language aims to give the meaning of every meaningful expression, but it is a question what form a theory should take if it is to accomplish this. Since there seems to be no clear limit to the number of meaningful expressions, a workable theory must account for the meaning of each expression on the basis of the patterned exhibition of a finite number of features. Even if there is a practical constraint on the length of the sentences a person can send and receive with understanding, a satisfactory semantics needs to explain the contribution of repeatable features to the meaning of sentences in which they occur. (p. 55)

He then immediately repeats the positive proposal of "Truth and Meaning" in [33].

> [33] I suggest that a theory of truth for a language does, in a minimal but important respect, do what we want, that is, give the meanings of all independently meaningful expressions on the basis of an analysis of their structure. (p. 55)

There is no shying away here from the project of giving the meaning of sentences of a natural language or of construing this as a matter of saying how it is that ordinary speakers understand sentences as uttered. The positive proposal is to accomplish the task by articulating a theory of truth. Again: if there is a way of seeing how articulating a theory of truth could do the job described, it would be perverse to suppose that Davidson, after announcing the project, immediately turns his back on it without saying that that is what he is doing.

There is the phrase "in a minimal but important respect", of course, but we have seen already why he would be cautious about having done all that a theory of meaning could be expected to do in the discussion at the end of section 2. But there is also a point to this in the context of this essay, located in his discussion of the role of general knowledge in a context in disambiguating utterances on p. 59, where he says that by "granting this … we accept a limitation on what a theory of truth can be expected to do," though "[w]ithin this limitation it may still be possible to give a theory that captures an important concept of meaning." What's left out here is not meaning, but those aspects of understanding that involve the application of general knowledge to resolving ambiguity in contexts of speech.

What properties do we look for in a truth theory? The answer is given in [34].

> [34] An acceptable theory should … account for the meaning (or conditions of truth) of every sentence by analyzing it as composed, in truth relevant ways, of elements drawn from a finite stock. A second natural demand is that the theory provide a method for deciding, given an arbitrary sentence, what its meaning is. … A third condition is that the statements of truth conditions for individual sentences entailed by the theory should, in some way yet to be made precise, draw upon the same concepts as the sentences whose truth conditions they state. (p. 56)

The first sentence here says that the truth theory should account for the meaning of every sentence in terms of its significant parts. The parenthetical 'conditions of truth' of course refers to the fact that it will do so by proving a translational T-sentence for the object language sentence which states conditions of truth using a sentence that translates it or interprets it relative to contextual parameters. The second sentence requires a mechanical procedure, a proof procedure, for generating each translational T-sentence. The third can be seen as a condition on meeting Convention T, for if we use concepts not expressed in the object language sentences in stating conditions under which they are true, we do not provide sentences that give their meaning in the metalanguage.

Might it be doubted that Davidson has in mind here a theory that meets Convention T (or an analog for natural languages)? This suggestion is refuted by his observation (on p. 58) that in order to "accommodate the indexical, or demonstrative, elements in a natural language ... Convention T must be revised to make truth sensitive to context." (See (Lepore and Ludwig 2005, ch. 5 sec. 2) for one way to do this: in a nutshell, what we want is that a relativized T-sentence such as '(S)(t)("I am hungry" understood as if uttered by S at t is true iff S is hungry at t)' to be a canonical theorem iff the corresponding sentence '(S)(t)("I am hungry" understood as if uttered by S at t means that S is hungry at t)' is true.) Davidson says further: "Convention T, suitably modified to apply to a natural language, provides a *criterion of success* in giving an account of meaning" (p. 61; emphasis added). And in light of the way of spelling this out just indicated, it is perfectly clear how this should be so.

Let's retreat a bit in the text to consider what support Davidson offers for his claim. On page 60, in [35], he has the following to say.

> [35] One relatively sharp demand on a theory for a language is that it give a recursive characterization of sentencehood. … In defining sentencehood what we capture, roughly, is the idea of an independently meaningful expression. But meaningfulness is only the shadow of meaning: a full-fledged theory should not merely ticket the meaningful expressions, but give their meanings. … now I should like to say a bit more in support of the claim that a theory of truth does 'give the meaning' of sentences.

Immediately following this, in [16], which we have quoted before, Davidson explains in the most straightforward way possible the connection between meeting Convention T and giving the meaning of sentences:

> [16] A theory of truth entails, for each sentence s, a statement of the form 's is true if and only if p' where in the simplest case 'p' is replaced by s. [And in the more complex cases we appeal to translation or interpretation relative to context.] Since the words 'is true if and only if' are invariant, we may interpret them if we please as meaning 'means that'. So construed, a sample might then read '"Socrates is wise" means that Socrates is wise'. (p. 60)

Here we know that in the general case Davidson has in mind a theory that meets Convention T or an analog for natural languages. Replacing 'p' by s guarantees the translation requirement is met, and this guarantees that we can replace 'is true if and only if' with 'means that' salva veritate. All the information is already available in the T-sentence, relative to the knowledge that it is a canonical theorem of a theory that meets Convention T.

But let us look at the next two paragraphs, [36] and [37], which have given at least some commentators pause.

> [36] This way of bringing out the relevance of a theory of truth to questions of meaning is illuminating, but we must beware lest it encourage certain errors. One such error is to think that all we can learn from a theory of truth about the meaning of a particular sentence is contained in the biconditional demanded by Convention T. What we can learn is brought out rather in the *proof* of such a biconditional, for the proof must demonstrate, step by step, how the truth value of the sentence depends upon a recursively given structure.

About this, for example, Ebbs says "This passage is superficially compatible with Ludwig and Lepore's reading, according to which the central goal of a Davidsonian truth theory is to display the recursive structure of a sentence in a way that meets the *s*-means-that-*p* requirement" (2012, p. 80). But, indeed, it is not just superficially *compatible* with that reading. It *entails* that reading, since the demand that the theory meet Convention T *just is the s-means-that-p requirement*, as we have noted. But Ebbs goes on to say that if we adopt this reading of Davidson, according to which the theory is to meet Convention T, as Davidson demands, we fall into a second error that Davidson warns against, in passage [37], which begins with the remark quoted earlier about Convention T being a criterion for success of a truth theory giving an account of meaning:

> [37] [a] Convention T, suitably modified to apply to a natural language, provides a criterion of success in giving an account of meaning. But how can we test such an account empirically? [b] Here is the second case in which we might be misled by the remark that the biconditionals required by Convention T could be read as giving meanings, for what this wrongly suggests is that testing a theory of truth calls for direct insight into what each sentence means. But in fact, all that is needed is the ability to recognize when the required biconditionals are true. This means that in principle it is no harder to test the empirical adequacy of a theory of truth than it is for a competent speaker of English to decide whether sentences like '"Snow is white" is true if and only if snow is white' are true. (pp. 61-2)

Davidson remarks further, in [38]:

> [38] [a] I have been imagining the situation where the metalanguage contains the object language, so that we may ask a native speaker to react to the familiar biconditonals that connect a sentence and its description. [b] A more radical case arises if we want to test a theory stated in our own language about the language of a foreign speaker. Here again a theory of truth can be tested, though not as easily or directly as before. The process will have to be something like that described by Quine in Chapter 2 of *Word and Object*. [c] We will notice conditions under which the alien speaker assents to or dissents from, a variety of his sentences. The relevant conditions will be what we take to be the truth conditions of his sentences. We will have to assume that in simple or obvious cases most of his assents are to true, and his dissents from false, sentences—an inevitable assumption since the alternative is unintelligible. (p. 62)

Ebbs has this to say (pp. 80-1):

> [E2] Davidson's point here is that even in the case in which the metalanguage contains the object language, we need not test a theory of meaning by asking whether [M]-sentences are true. [Sentences of the from 's means that p'.] The constraint that we must satisfy is that the biconditionals of the theory are true, not that they state the meanings of sentences. His criteria for constructing and testing theories of truth that are to serve as theories of meaning do not appeal to or presuppose notions of synonymy or translation. This is crucial to the project of explaining how to give empirical content to truth theories for languages that we do not yet know. … And in Davidson's view there is no better way to judge the adequacy of a truth theory that is to serve as a meaning theory for a language that we do not yet know than to employ the empirical tests and satisfy the constraints he proposes:

> [39] What no one can, in the nature of the case, figure out from the totality of the relevant evidence cannot be part of meaning. (p. 235)

So far so good! Yes, Davidson does hope to show how a theory meeting certain formal and empirical constraints (which don't already presuppose we know what theory is correct) can be shown *ipso facto* to satisfy Convention T, and, hence, to show how a truth theory meeting informative constraints can serve as a meaning theory by enabling us, if we like, to derive true statements of the form 's means that p' (or suitably relativized variants for natural languages). But this is not the conclusion that Ebbs draws. Instead:

> [E3] When one reads this passage [That is [39] at the end of [E2]] in the light of passage [b] [this is the combination of [37b] and [38b] above], it is natural to conclude that Davidson rejects the *s*-means-that-*p* requirement, according to which our pre-theoretical judgments about the truth values of [M]-sentences place substantive, independent constraints on the acceptability of a truth theory that is to serve as a meaning theory—constraints that are prior to and independent of Davidson's proposed method for figuring out from the totality of the relevant evidence which truth theories can serve as meaning theories. (Ebbs, p. 81)

In [E3], Ebbs says that Davidson rejects the *s*-means-that-*p* requirement. As we have noted, this is just the requirement that the theory satisfy Convention T if it is to count as giving the meanings of object language sentences. Does Davidson say that meeting Convention T is a criterion for success in giving an account of meaning? If he does, then Ebbs is simply wrong. Now recall that [37] begins with [37a] (which is not part of the material Ebbs quotes in his [b]).

> [37a] Convention T, suitably modified to apply to a natural language, provides a criterion of success in giving an account of meaning. But how can we test such an account empirically?

Here Davidson says that Convention T *is* a criterion for success in giving an account of meaning. Therefore, Ebbs is simply wrong.

Everything that follows in [37] is directed at the question how we can test such an account empirically, that is, how we can tell that a theory satisfies Convention T, and so is suitable for giving the meanings of object language sentences. The second error that Davidson refers to is not the error of thinking that Convention T is a criterion of success (how could it be given the way the passage begins!) but the error of thinking that confirmation *has to go* by way of direct insight into what a sentence means. This is not the same as saying we might not rely on this. When it is our own language, it is not difficult to identify target translational T-sentences with minimal explanatory insight into meaning, since disquotation or something close will often do. This doesn't work, of course, when the language is not our own. In that case, we employ the principle of charity—here "assume in simple or obvious cases that most of his assents are true" [38c]—and read off from the conditions that prompt the assents what they are about (a footnote to this paragraph cites Davidson's essays on radical interpretation where a more sophisticated version is given—see the discussion in the next section). The assumption in the case of translating others is that Charity provides a principled way of identifying conditions under which another's sentences are true that provide an interpretation of them in the context. The error Davidson has in mind is thinking that there is no substantive answer to the question when a truth theory satisfies Convention T in terms of constraints it has to meet that do not themselves presuppose anything about which theory is correct.

Why does Ebbs read any of this as a rejection of the requirement that a truth theory that can be used to interpret a language must meet Convention T (or a suitable analog for natural languages), given that Davidson explicitly requires this?

One possibility is that Ebbs takes the requirement that a truth theory satisfy Convention T if it is to be used for interpretation to entail that the only way to confirm such a theory is to rely on pretheoretic judgments about the truth values of [M]-sentences, and it is clear that Davidson does not think that is the only way to confirm such a theory. But the requirement does not entail that, and, of course, in the case of another's language, it is not an option.

But [E3] and [E2] with the reference to [39] suggest another line of argument. Ebbs's objection to the *s*-means-that-*p* requirement is that it would place substantive, independent constraints on whether a truth theory can serve as a meaning theory, constraints "prior to and independent of Davidson's proposed method." It seems that Ebbs is (i) taking Davidson's claim [39]

> [39] What no one can, in the nature of the case, figure out from the totality of the relevant evidence cannot be part of meaning. (p. 235)

to be a foundational principle for him, (ii) assuming that the results will obviously not conform to intuitive constraints on meaning, and (iii) deducing that Davidson couldn't have meant to be explaining, or giving an account that he intended to conform to, our ordinary understanding of what it is to understand what another says. This is reinforced in passage [E4]:

> [E4] For Davidson the relevant evidence for testing a truth theory for a natural language L that is to serve as a meaning theory for L does *not* include our pre-theoretical judgments of the truth values of [M]-sentences. The problem for Lepore and Ludwig's reading is that even when Davidson is articulating and defending the most radical and counterintuitive consequences of this theory of meaning, such as the inscrutability of reference and the indeterminacy of meaning …, he insists that we have no grip on meaning that is firmer than, or independent of, what one can figure out from the totality of the relevant evidence. (2012, p. 85)

It seems that the crucial thought is that Davidson's constraint in [39] has counterintuitive consequences, and that this is what shows that he can't be intending to require that a truth theory meet Convention T. Now, central to this line of reasoning is the assumption that it is obvious (and obvious to Davidson) that the constraint that Davidson lays down is not met by the notion of understanding involved in our ordinary conception of communicative utterances and that Davidson agreed with this. Then one would be forced to conclude that Davidson did not think that what he was doing was giving an account of what is involved in meaning in the sense of what we take ourselves to understand in ordinary communicative exchanges.

The inescapable fact, however, is that this make no sense of Davidson's appeal to Convention T as a criterion for a truth theory that is to serve in interpreting others. One would have to take him either to be confused, or to be intending Convention T be reinterpreted, without ever saying so (I believe this is in fact what Ebbs thinks—see the discussion of the third component of his argument in section 7), in terms of his new reconstructed notion of meaning, and then to be making an obscure joke in asserting that Convention T is a criterion for a successful theory. There is no hint of this in Davidson's writings. But neither is there any need to see him as being disingenuous.

What Davidson says is perfectly intelligible and defensible (which is not to say that it is ultimately correct), if we take him to be committed to the view that because language is by its nature a medium for communication, evidence for meaning must be intersubjectively available. The sentence that leads into the material quoted in [39], in fact, which Ebbs does not quote, is: "The semantic features of language are public features." That the semantic features of language are in a suitable sense public features is not an extraneous requirement he imposes on language for who knows what purposes of his own, but a requirement that is in fact quite widely accepted in one form or another. Davidson, to be sure, has his own take on it, which is expressed in the idea that ultimately the evidence can be described in purely behavioral and third person terms. But this is certainly not something that is obviously wrong, if it is wrong, and it is motivated by what anyone must recognize as a central function of language. And if it is, appealing to it as a constraint on meaning is not an abandonment of the ordinary notion but a recognition of a central feature of it. Of course, then, if one holds this, one will hold that meaning as we ordinarily think of it is constrained by its function, and so what we uncover by a systematic investigation of it under this constraint applies to meaning and understanding as it figures in our ordinary pre-reflective thinking. It is no objection to this that it might have what are at first glance surprising consequences, even some consequences that are in some ordinary sense counterintuitive (if not in the sense of being contrary to the requirements of that concept). In just this way articulation of conceptual structure often involves counterintuitive consequences, such as that the set of the even numbers and the set of the natural numbers has the same cardinality, or (though these are more controversial) that free will is compatible neither with determinism nor indeterminism, or that we know nothing about the external world, or that induction does not lead to justified belief, or that know-how is just propositional knowledge, and so on.

But doesn't this leave Davidson's account, given his full commitments, vulnerable to objection? Might we find that he gives too strong a reading of [39] in his construal of what counts as relevant evidence? Might we not judge that the constraints that Davidson lays down do not in fact suffice to confirm a theory that meets Convention T? Yes: but it cannot be an objection to an interpretation of a philosopher's project that it does not make it invulnerable to objection. Of what interest could the project be in the first place if that were so? It must make some substantive claims that are not obvious but for which reasons can be advanced for it to be a theory that has a hope of making an advance on our understanding.

Interestingly, Ebbs cites the doubts Lepore and I raised about the success of Davidson's project, on our construal of it, as evidence against our construal being correct. He says:

> [E5] In fact, by attributing the assumption that if from public cues an interpreter cannot recover a supposed semantic feature of a speaker's words, there cannot be any such feature to Davidson, while pointing out that this assumption is incompatible with the s-means-that-p requirement, Lepore and Ludwig themselves show, in effect, that Davidson is committed to rejecting the s-means-that-p requirement. It follows that on their reading … Davidson is committed to the s-means-that-p requirement and to rejecting it. (2012, p. 86)

Ebbs clearly takes this to be a reason to think that our interpretation, and, hence, criticism of Davidson, is incorrect. But this is not a plausible principle of interpretation. Consider the parallel objection to a criticism of a functionalist account of the mind:

> In fact, by attributing to functionalists the assumption that if from the functional organization of a system one cannot extract mental features

> there cannot be any such features, while pointing out that this assumption is incompatible with the requirement that phenomenal states (which—the objection might go—can be varied while retaining the same functional organization) be counted as mental, the critic herself shows, in effect, that the functionalist is committed to rejecting that phenomenal states be counted as mental. It follows that, on the critic's characterization of the functionalist, the functionalist is committed to the requirement that phenomenal states be counted as mental and to rejecting it.

No! Critic and theorist *agree* on what the theory has to accommodate, and the theorist *has a view about what suffices*, and *reasons* for thinking it does. When the critic charges that it is insufficient, the critic is not saying that the theorist both is and is not committed to accommodating what the theory aims to accommodate. The critic is saying that what is said to do so fails to. You could, if you liked, say that in one sense the functionalist is committed to saying phenomenal states are not mental because she is committed to functionalism and (contrary to what she thinks) there is no functionalist analysis of phenomenal states (we are supposing this for the sake of illustration). Hence, you could say: since she says that mental states are functional states, but phenomenal states are not, she is committed to the thesis that phenomenal states are not mental states. Then if she is independently committed to saying that phenomenal states are mental states, she both is and is not committed to phenomenal states being mental states. *But that is not how she sees it!* And it would be absurd to argue that this sense in which she is committed to phenomenal states not being mental states shows that she was not trying after all to give a functionalist reduction of phenomenal states. Aiming at a target and failing to hit it does not mean that you were not aiming at the target.

To be fair, Ebbs goes on to say that Traditional Pursuit Theory would be correct if there were "very good textual evidence that Davidson is independently committed to the *s*-means-that-*p* requirement, and that he … fails to see that his conjectures about the nature of meaning conflict with this supposedly independent commitment" (p. 86). But there is such evidence, even abundant evidence, for Davidson says again, and again, as we have seen, that a theory can be used for interpretation if it meets Convention T, and this is perfectly in line what how he describes his own aims in his project.

In the interest of thoroughness, it will be useful to look at one more passage, this one from "In Defence of Convention T." This paper is concerned to defend the philosophical significance of Convention T, both in regard to understanding truth and in regard to semantics more generally. In [40], Davidson transitions from remarks on the former subject to the latter.

> [40] A recursive theory of absolute truth, of the kind required by Convention T, provides an answer, *per accidens* it may at first seem, to quite another problem. This problem may be expressed as that of showing or explaining how the meaning of a sentence depends on the meanings of its parts. A theory of absolute truth gives an answer in the following sense. Since there is an infinity of T-sentences to be accounted for, the theory must work by selecting a finite number of truth-relevant expressions and a finite number of truth-affecting constructions from which all sentences are composed. The theory then gives outright the semantic properties of certain of the basic expressions, and tells how the constructions affect the semantic properties of the expressions on which they operate.
>
> In the previous paragraph, the notion of meaning to which appeal is made in the slogan 'The meaning of 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. 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. (pp. 70-1)

The remarks of particular interest lie in the second paragraph. Davidson says that the notion of meaning at issue is not one that "opposes meaning to reference, or a notion that assumes meanings are entities," and he says that "a theory of truth confers a clear content" on the slogan 'The meaning of the sentence depends on the meanings of its part'. Of course, the remark that it does not assume meanings as entities is straightforward. But a question arises about the idea that the relevant notion of meaning is not opposed to reference, for this might suggest that all he has in mind by meaning is reference, extension, and truth value. There are two bits of context that help to interpret this here. The first is that he is still clearly concerned with the semantic conception of truth on which the truth theory meets Convention T, and as noted earlier, at the end of section 4 above, later in the same paper in [20], he suggests, as he did in "Truth and Meaning," that a truth theory for a natural language which is true will satisfy Convention T because of the need to accommodate demonstratives and other context sensitive devices in the language. Thus, the output of the theory that he has in view are T-sentences which are interpretive, and the goal is a theory that will enable its possessor "to use that language in communication" (p. 74). This shows that he has more in mind that just a theory that assigns truth values to sentences. Second, it is helpful to recall Davidson's remarks in "Truth and Meaning" about his theory of meaning falling, "comfortably within what Quine terms the 'theory of reference' as distinguished from what he terms the 'theory of meaning,'" in [21e]. As we noted before, the point of this remark is that the only semantic vocabulary in the truth theory relates to reference, satisfaction and truth. In this first phase of his work, as he remarks later in "Reply to Foster," he was not distinguishing between the truth theory itself and the body of knowledge one had to have about it to use it for interpretation—knowledge which suffices to know that it satisfies Convention T and to pick out the right theorems, the ones in virtue of which it satisfies Convention T. The point of the remark that the notion of meaning is not one that opposes meaning to reference is that the theory that issues in interpretive T-sentences itself employs only concepts from the theory of reference. But we have seen that this is compatible with aiming at a theory of the language that puts one in a position to communicate with those who are native speakers of it.

7. Revisions to the Project in Later Work

In this section, I take a closer look at "Radical Interpretation," and "Reply to Foster," with some support from other sources, to provide further evidence for the reading of "Truth and Meaning" given above.

7.1 Radical Interpretation

"Radical Interpretation," a version of which was first read in May 1973, begins with two questions in [41].

> [41] Kurt utters the words 'Es regnet' and under the right conditions we know that he has said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on and to interpret his words: we can say what his words, on that occasion, meant. What could we know that would enable us to do this? How could we come to know it? (p. 125)

The point I want to draw attention to is simply that the project announced here is that of saying what we could know that would enable us to say what a speaker's words as used on an occasion *meant*. The point is made again in [42].

> [42] What knowledge would serve for interpretation? A short answer would be, knowledge of what each meaningful expression means. In German, those words Kurt spoke mean that it is raining and Kurt was speaking German. So in uttering the words 'Es regnet', Kurt said that it was raining (p. 126).

Davidson does not say that this answer is incorrect or confused or that it traffics in hopelessly muddled concepts. His complaint is rather that "it does not say what it is to know what an expression means" (p. 126).

He goes on to put aside the appeal to assigning "to each meaningful expression … an entity, its meaning" which at best, echoing the criticism of "Truth and Meaning," "hypostasizes the problems" (p. 126); and he puts two constraints on an answer. The first is that, since the "interpreter must be able to understand any of the infinity of sentences the speaker might utter" (p. 127), "we must put in finite form" what it is the interpreter might know that would enable him to do this. The theory should be such that "someone who knows the theory can interpret the utterances to which the theory applies" (p. 128). The second requirement is that the theory "can be supported or verified by evidence plausibly available to an interpreter" (p. 128). If one is working with one's own language, one can test the theory by appeal to "instances of particular interpretations recognized as correct" since we can tell "whether [the proposed theory] yields correct interpretations when applied to particular utterances" (p. 128). However, to deal "with the general case" we must appeal to "evidence that can be stated without essential use of such linguistic concepts as meaning, interpretation, synonymy, and the like" because in radical interpretation ("interpretation in one idiom of talk in another" (p. 126)) "the theory is supposed to supply an understanding of particular utterances that is not given in advance" (p. 128).

Thus, the project is to state a finite body of information knowledge of which would put anyone who had it in a position to interpret any potential utterance of a sentence of the language, in the sense of saying what it meant, and to show how one could come to possess this body of knowledge on the basis of evidence that can be described without using concepts from the theory of meaning. The point of the restriction on the evidence is to relate the concepts of the theory to more basic concepts by showing how the concepts of the theory are to be applied in the light of the application of the more basic concepts.

The statement of this project is completely independent of Davidson's proposal for a solution. I have gone over it in detail to emphasize that the goal of the project is clearly that of illuminating what is involved in meaning something by an utterance and in understanding it. There is nothing in the set up of the problem that suggests that the goal is instead to replace the concept of meaning with a more tractable concept. Quite the contrary.

Davidson goes on, of course, to suggest that an empirically confirmed Tarski-style truth theory can do the job, but the job is specifically the one just described. He says: "What follows is a defence of the claim that a theory of truth, modified to apply to a natural language, can be used as a theory of interpretation" (p. 131). And specifically, as I have said, the idea is to confirm a theory of truth that meets Convention T, or a suitable analog for natural languages, so that in light of that knowledge one can read off from the appropriate theorems what object language sentences (as used) mean or would mean. A bit later, he puts the strategy this way: "assuming translation, Tarski was able to define truth; the present idea is to take truth as basic and to extract an account of translation or interpretation" (p. 134). He does not say: the present idea is to replace the concept of meaning with the concept of truth, or to replace interpretation with assignment of truth conditions; nor does he suggest he is going to extract from the ordinary concept of meaning some extensional core. He is fully explicit about what he wants to do: "the *hope* is that by putting appropriate formal and empirical restrictions on the theory as a whole, *individual T-sentences will in fact serve to yield interpretations*" (p. 134; emphasis added).

That Davidson is after a theory that tells us what utterances mean (not just when they are true or "strongly" true) is reinforced by his description of the problem facing the radical interpreter. He supposes that we can help ourselves to evidence in the form of hold true attitudes toward sentences (beliefs that sentences are true). The problem is explained in [43].

> [43] … a speaker holds a sentence to be true because of what the sentence (in his language) means, and because of what he believes. Knowing that he holds the sentence to be true, and knowing the meaning, we can infer his belief; given enough information about his beliefs, we could perhaps infer the meaning. But radical interpretation should rest on evidence that does not assume knowledge of meanings or detailed knowledge of beliefs. (p. 135)

The problem then is explicitly to break into the circle of meaning and belief, given knowledge of hold true attitudes, and the method is to adopt the Principle of Charity as a regulative ideal governing interpretation. We must hold fixed one of the two factors, belief or meaning, that generate hold true attitudes, in order to solve for the other. The Principle of Charity holds fixed belief by treating the speaker's beliefs as by and large true, so that in light of the speaker's environment, we can get a grip on what he believes, and so solve for the meaning of the sentence he holds true. The justification for the Principle of Charity is that if the other speaks a language, he is of necessity interpretable (see [56] in the appendix), and Charity is a condition on this being so (for discussion, see Lepore and Ludwig 2005, chs. 13-15). The principle we rely on is [H] (though, as I have just said, this should be treated as in the nature of a regulative ideal).

> [H] S holds true ϕ if and only if S believes that p and ϕ means that p.

What is it that we are solving for again? *What ϕ means*. That is the target. That is what the project is in pursuit of, not what it puts in the trash bin. The project is to figure out how we could identify what utterances mean on the basis of evidence that does not presuppose that we already know it.

The truth theory enters the picture because it is the vehicle for the compositional meaning theory. Given [H], if we identify the content of a belief, and a hold true attitude based on it, we can identify truth conditions for the sentence held true which are interpretive, and this gives us a target theorem for a truth theory for the language that meets Convention T. More precisely, to confirm a truth theory that will enable us to interpret the speaker's language, we:

(1) identify correlations of hold true attitudes to sentences prompted by conditions in the environment with the conditions that prompt them (U holds true *s* iff *p*);
(2) to thereby identify a class of corresponding T-sentences (*s* is true iff *p*); and then
(3) to develop a truth theory that entails as many of these as we can
(4) compatibly with other reasonable constraints on interpretation, such as finding the person one is interpreting to be largely rational, epistemically and practically.

The Principle of Charity justifies (2), and this shows that the goal is in fact to develop a theory that satisfies Convention T (or its analog for natural languages).

Davidson explains the goal in [44] at the end of "Radical Interpretation":

> [44] If we knew that a T-sentence satisfied Tarski's Convention T, we would know that it was true, and we could use it to interpret a sentence because *we would know that the right branch of the biconditional translated the sentence to be interpreted*. Our present trouble springs from the fact that in radical interpretation we cannot assume that a T-sentence satisfies the translation criterion. What we have been overlooking, however, is that we have supplied an alternative criterion: this criterion is that the totality of T-sentences should (in the sense described above) optimally fit evidence about sentences held true by native speakers.

> *The present idea is that what Tarski assumed outright for each T-sentence can be indirectly elicited by a holistic constraint. If that constraint is adequate, each T-sentence will in fact yield an acceptable interpretation.* (p. 139; emphasis added)

Here again the goal is put independently of the proposal. What Tarski assumed outright was that the metalanguage sentence translates the object language sentence. It is that that is to be elicited indirectly (or, at any rate, the analog for natural languages once we accommodate context sensitivity). Ebbs's idea that Davidson doesn't really have translation in mind but some notion that is not theory independent makes no sense of the idea that we elicit indirectly *what Tarski assumed outright*.

Before leaving "Radical Interpretation," there are two more issues it will be useful to touch on. The first point is one mentioned at the end of section 4, namely, that in "Radical Interpretation," as Davidson says in "Reply to Foster" p. 171, he criticizes "his own earlier attempts to say exactly what the relation is between a theory of truth and a theory of meaning," and "tried to do better." He says that the "criticisms [he] there levelled against [his] earlier formulation are ... essentially those elaborated by Foster in the second part of his ... paper" (p. 172). And those criticisms were essentially that a truth theory that is merely true is not *ipso facto* adequate for interpretation. Davidson expresses it in this way in "Radical Interpretation":

> [45] ... on reflection it is clear that a T-sentence does not give the meaning of the sentence it concerns: the T-sentence does fix the truth value relative to certain conditions, but it does not say the object language sentence is true *because* the conditions hold. Yet if truth values were all that mattered, the T-sentence for 'Snow is white' could as well say that it is true if and only if grass is green or 2 + 2 = 4 as say that it is true if and only if snow is white. We may be confident, perhaps, that no satisfactory theory of truth will produce such anomalous T-sentences, but this confidence does not license us to make more of T-sentences. (p. 138)

This is a criticism of the proposal that Davidson makes in "Truth and Meaning" and in "In Defence of Convention T" (see [20] above)—the proposal that it is enough to require that a truth theory for a natural language to be true for it to satisfy the analog of Convention T for natural languages. As I said at the end of section 4, the fact that Davidson criticizes his own earlier proposal on these grounds shows that he has, contrary to Ebbs's claim, a theory independent target in mind.

Ebbs has a response to this. He claims that it is perfectly compatible with the Explication Interpretation that Davidson revise "his list of constraints on an adequate explication" (Ebbs 2012, p. 98). The revisions, he says, "may be viewed ... as an attempt to identify some new subset of the present motley of applications of 'meaning' and 'means that' that we now wish to preserve and clarify" (loc. cit.).

However, Ebbs's suggestion is not compatible with the what Davidson actually says. For if Davidson were merely changing his list of what aspects of the usage of the word "meaning" he wants to preserve (whatever these could be exactly), he would not be criticizing his "earlier attempts to say exactly what the relation is between a theory of truth and a theory of meaning." He would simply be changing from one project to another. Moreover, Davidson says that his criticisms of his own earlier proposals in "Radical Interpretation" are essentially those of Foster in part 2 of his paper (1976, p. 172). And Foster's criticism was that a true truth theory does not satisfy Convention T whether or not adapted for a context sensitive language. Foster says at the beginning of section 2 of his paper: "… we are seeking a method of constructing theories of meaning for particular languages which will yield the greatest philosophical insight into the nature of meaning and language in general. To yield this insight the theories must be genuinely interpretive: they facts they state must suffice for the mastery of the languages they characterize" (Foster 1976, p. 7). If Davidson had one usage in mind for the earlier work and a different one for later work, it would make no sense for him to say that he agrees with Foster about the problems with his earlier proposal. The only thing that makes sense of his saying that he agrees with Foster's criticism of his earlier proposal is that he, like Foster, requires the theory not satisfy Convention T, where this is not to be understood in terms of some ersatz notion of translation.

The second issue concerns an argument that Ebbs advances in footnote 10 of his paper, which is attached to the discussion just mentioned. Foster had charged that Davidson could not distinguish between (a) and (b). In the footnote, Ebbs says that Davidson's revised proposal was that the canonical theorems of a truth theory express laws that support counterfactuals. This is supposed to rule out theories that have such theorems as (a) as a consequence. Ebbs says, however, that even this revised criterion fails to suffice for a theory to satisfy Convention T on its ordinary reading (or its natural language analog) because it can't distinguish between (b) and (b').

(a) 'a is part of b' is true in English if and only if a is part of b and the Earth moves
(b) 'a is part of b' is true in English if and only if a is part of b
(b') 'a is part of b' is true in English if and only if a is part of b and 2 + 2 =4

The point is supposed to be that Davidson's revised criterion does not suffice for the theory to satisfy Convention T, and that "Davidson was surely aware of this kind of objection to his theory," but "as far as I know, Davidson never discusses this objection in print." Ebbs suggests that the reason is that "Davidson believes the reply is obvious: given his constraints on a satisfactory explication of meaning, if (b') is a canonical consequence of a well-confirmed theory of truth for English, then it 'gives the meaning' of 'a is a part of b' just as well as the more familiar theorem (b) does" (p. 99, n. 10).

Here Ebbs supposes that Davidson's constraint is simply that the theory express laws, for the claim is that since (b') is nomically necessary if (b) is, Davidson's constraints don't distinguish between them. But this is not the constraint that Davidson introduces in "Radical Interpretation," but rather an upshot of it. The constraint is expressed in a compressed form in [44]: "the totality of T-sentences should (in the sense described above) optimally fit evidence about sentences held true by native speaker." Thus, the requirement is that the theory be the best (or tie for the best) theory from the standpoint of the radical interpreter, who assembles his best account of a speaker as a rational agent responding to his environment and other speakers, and this is a considerably stronger constraint than merely laying down that the theory be true and counterfactual supporting. As Davidson makes clear elsewhere (see in [57]-[60], and [63]-[66] in the appendix), the theory of interpretation does not stand alone, but is a part of a total theory of the speaker as a rational agent. If that constraint is met, then the theory's theorems will express laws about the speakers of the language it treats, but the constraint itself requires more than that. In [45] above, Davidson points out that truth value doesn't distinguish between theorems that say that 'Snow is white' is true iff snow is white and theorems which say that it is true iff grass is green or 2 + 2 =4. Trouble with the latter is that they don't explain why 'Snow is white' is true. It is not that the theorems are not nomically necessary. Even if they were (and with a few small qualifiers that could hardly speak to the point that Davidson has in mind, they would be, e.g., speaking of pure snow and healthy (Bermuda) grass), they would clearly not be adequate. The conditions that explain why 'Snow is white' is true, when it is, are to be found in the uses to which speakers put the contained words, and this is to be identified from the standpoint an interpreter of those speakers as rational linguistic beings engaged in the practice of communication with one another.

7.2 Reply to Foster

Let us turn finally to "Reply to Foster." This paper was a response to a paper read by John Foster to the Oxford Philosophical Society in June 1974 (Foster 1976), in which Foster argues that a semantic theory should state knowledge sufficient for one to interpret the language for which it is a theory, but that a truth theory does not do so, and if we add what additional information about the truth theory would be needed, we violate a constraint that Davidson lays down on his project. Davidson's reply shows clearly that he is engaged in the traditional pursuit. He says, to begin with, in [46], that he agrees with Foster on what is required for a meaning theory.

> [46] I share [Foster's] bias in favour of extensional first-order languages; *I am glad to keep him company in the search for an explicitly semantical theory that recursively accounts for the meanings of sentences in terms of their structures;* and I am happy he concurs in holding that a theory may be judged adequate on the basis of holistic constraints. … *I think Foster is right in asking whether a proposed theory explicitly states something knowledge of which would suffice for interpreting utterances of speakers of the language to which it applies*. … I was slow to appreciate the importance of this way of formulating *a general aim of theories of meaning,* though elements of the idea appear in several early papers of mine [these are "Theories of Meaning and Learnable Languages," and "Truth and Meaning"]. (p. 171; emphasis added)

Thus, again, the aim is to state something knowledge of which suffices for interpreting speakers of a language, and note here that Davidson says he shares all these goals with Foster, and Foster is clearly engaged in the traditional enterprise, and is criticizing Davidson from that standpoint. *Davidson agrees he is engaged in the same enterprise, and aims to defend himself from the criticisms that Foster advances from that standpoint.*

In recounting his project, Davidson notes that if we knew a Tarskian truth theory for a language L, "and that it was such a theory, then we could produce a translation of each sentence of L, and would know that it was a translation" (p. 172). He continues in [47].

> [47] Since Tarski was interested in defining truth, and was working with artificial languages where stipulation can replace illumination, he could take the concept of translation for granted. But in *radical* interpretation, this is just what cannot be assumed. So I have proposed instead some empirical constraints on accepting a theory of truth that can be stated without appeal to such concepts as those of meaning, translation, or synonymy, though not without a certain understanding of the notion of truth. By a course of reasoning, I have tried to show that if the constraints are met by a theory, the T-sentences that flow from that theory will in fact have translations of *s* replacing '*p*'. (p. 172)

That is to say, the goal is to describe constraints on a truth theory, without invoking the concepts of the theory of meaning, that will suffice for the theory to satisfy Convention T. Knowing the theory meets the constraints, and knowing that meeting them suffices for the theory to meet Convention T, we are then in a position to provide a translation of each of the sentences (putting aside here context sensitivity), and if we know the language of the theory, of course, then we are in a position to interpret the object language sentences. This is not abandoning the theory of meaning, but pursuing it by a clever bit of indirection.

Of his position in "Truth and Meaning," he says this, which I have quoted before:

> [23] 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 if from evidence, and such a person could not fail to realize that his theory satisfied the constraints. (p. 173)

This shows, as I remarked before, that he was thinking, even in "Truth and Meaning," that the requirement that the theory be empirically confirmed for a natural language would suffice for it to meet Convention T. Further, in discussing Foster's objection, Davidson says that Foster grants that the constraints he [Davidson] offered are adequate to ensure that it satisfies Convention T—i.e. to ensure that in its T-sentence, the right branch of the biconditional really does translate the sentence whose truth value it is giving.

The objection of Foster's he focuses on is the charge that one could have a theory that met Convention T but not know that one did, and then not know something sufficient for interpretation, but that if one then states what one has to know about the theory, one has to use intensional notions, such as that a translational T-theory *states* that … , which, Foster claims, Davidson has officially barred himself from using.

Davidson's response to the first part of this charge is to agree, but to add that his view all along was that you had to know not just the theory but also that it met constraints sufficient for it to satisfy Convention T. Davidson makes the connection with knowledge of the meaning of sentences explicit in [48].

> [48] Someone who can interpret English knows … that an utterance of the sentence 'Snow is white' is true if and only if snow is white; he knows in addition that this fact is entailed by a translational theory—that it is not an accidental fact about that English sentence, but a fact that interprets the sentence. Once the point of putting things this way is clear, I see no harm in rephrasing what the interpreter knows in this case in a more familiar vein: he knows that 'Snow is white' in English *means that* snow is white. (p. 175; emphasis in the original)

Curiously, Ebbs takes this passage not only not to count against his Explication Interpretation of Davidson but to provide important support for it. Since it seems not to on the face of it, let us take a moment out the discussion of Davidson's response to Foster's criticisms to consider how Ebbs handles this endorsement of the connection between a translational truth theory and explicit statements of what object language sentences mean. Ebbs places the emphasis on "I see *no harm in rephrasing* what the interpreter knows in a more familiar vein." This may seem to be a matter of grasping after straws, but there is nothing else to fasten onto here. Ebbs says boldly: "this passage does not support the claim that Davidson is committed to the s-means-that-p requirement" (p. 95). But it simply does: for a translational theory is one that meets the *s*-means-that-*p* requirement, for it is one that satisfies Convention T, and these are equivalent. Ebbs goes on to say: "… there is no incompatibility between the explicational reading of Davidson that I have outlined and his willingness to use [M]-sentences, as long as we do not take Davidson to regard his uses of those sentences as providing an independent constraint on his theory of interpretation" (p. 95). And Ebbs says that [48] provides "ample reason" not to take Davidson as committed to the theory satisfying Convention T, for "[a]s the emphasized phrase [the no harm phrase] suggests, Davidson is saying that if we accept his explication of 'means that,' then *there is no harm in affirming*" that someone who knows a T-sentence for 'snow is white' and knows it is entailed by an translational truth theory, knows that it is a fact that interprets 'snow is white'," that is, knows that 'Snow is white' means that knows is white (presumably in Davidson's supposed special sense of 'means that').

Is there ample reason in [48] to think that Davidson was not committed to the requirement that a truth theory satisfy Convention T in order for it to be able to be used for interpretation? Does the 'no harm' passage provide ample reason to think this? Recall that a translation truth theory *just is one that meets Convention T*. Recall that Foster is himself clearly not working with some explicational reading of translation or interpretation, and that Davidson is responding to Foster's criticisms, and that he aims to show that Foster misses the mark. Recall for a moment the last sentence of [47]: "By a course of reasoning, *I have tried to show that if the constraints are met by a theory*, the T-sentences that flow from that theory will *in fact* have *translations* of *s* replacing '*p*'" (emphasis added). Recall the first sentence of [44]: "If we knew that a T-sentence satisfied Tarski's Convention T, we would know that it was true, and we could use it to interpret a sentence *because* we would know that the right branch of the biconditional translated the sentence to be interpreted" (emphasis added). [48] provides no reason, let alone ample reason, for thinking that Davidson did not think that it was a constraint on a truth theory being used for interpretation was that it meet Convention T. On the contrary, it shows decisively that Davidson did think that, and when the passage is taken in context, as shown, the explication reading looks not merely eccentric but a little bit daft.

What *did* Davidson mean by saying "Once the point of putting things this way is clear, *there is no harm* in rephrasing what the interpreter knows"? Why does he not say: We can rephrase what the interpreter knows by saying that he knows that 'Snow is white' means that snow is white? A first question here is what 'this way' is to refer to. The two options are that it refers to what follows or to what precedes. But there is no particular way he has put what he has just said that needs to have special attention drawn to it. This contrasts with what he goes on to say, where he can be seen as putting in other words (that might be misleading) something he has already said. We can take 'this way' then to refer to saying that the interpreter knows that 'Snow is white' in English means that snow is white. So what he is saying is that once we see the point of putting what the interpreter knows as a matter of knowing that 'Snow is white' in English means that snow is white, there is no harm in doing so. And what is the point? The point is simply that the interpreter has knowledge sufficient to interpret that sentence, that is, to understand utterances of it, and that, in one good sense we can give to it, this amounts to knowing what the sentence means. What then is the harm that missing the point might give rise to? It is just failure to grasp the form of the knowledge that the interpreter has and how it enables the interpreter to interpret sentences of the object language, something that is not conveyed by simply saying that the interpreter knows that 'Snow is white' in English means that snow is white.

Let us now turn to the final component in Ebbs's argument, which involves a claim about Tarski and a claim about Davidson, and is designed as a strike against what has been a constant invocation in this discussion of the fact that Davidson so plainly requires that if a truth theory is to serve for interpretation, it must meet a suitable analog of Convention T for natural languages. I have reserved this for the penultimate section because it is directed particularly against things that Davidson says in "Radical Interpretation" and "Reply to Foster."

The first claim involves Tarski's point in using 'Convention' in Convention T, the adequacy condition on a materially correct definition of a truth predicate for a formal language. The point of using 'Convention', according to Ebbs, is to express that there is something arbitrary about its use as a criterion of adequacy. The arbitrariness, accord to Ebbs, arises because Convention T does not aim at a standard of correctness "uniquely determined by our current or previous uses" of a term to be defined. Hence, the use of 'Convention' is to be explained by Tarski's aiming to provide an *explication* of the concept of truth. Convention T expresses, presumably, that aspect of the usage of 'true' that Tarski wants to capture. Ebbs cites as evidence Tarski's disdain for the question "What is the right conception of truth?," about which he say that it is "so vague that no definite solution is possible" in his 1944 paper "The Semantic Conception of Truth" (Tarski 1944, , p. 355).

The claim about Davidson is that he understood “Convention” in “Convention T” in this way as well. Then when Davidson says “our outlook inverts Tarski’s” (p. 150), Ebbs interprets this, on the basis just mentioned, in the following way: “to invert Tarski’s approach is to explicate “means” or “translates” by assuming a prior grasp of truth” (2012, p. 101).

> [E6] Hence, when Davidson says “I want a theory that satisfies Convention T” he is not endorsing Tarski’s uncritical reliance on our evaluations of [M]-sentences. On the contrary, what Davidson seeks is a set of constraints on an empirical truth theory that is analogous to the simple biconditionals that Tarski places as constraints on his explication of “true-in-L.” (loc. cit.)

Ebbs takes this to be a reason to treat every reference to the requirement that a truth theory satisfy Convention T for it be used for interpretation to involve Convention T reinterpreted so that the notion of translation or interpretation it invokes is itself an explicated notion, one whose content is given by the constraints that are supposed to suit the theory for use in interpretation, not one which tests the adequacy of the constraints. We have already remarked that this is not compatible with what Davidson actually says—see e.g. the paragraph immediately following [E4] and following [44] above. But let’s consider the argument in its own terms, taking each claim in turn.

First, with respect to Tarski, we do want to understand why he uses the term “Convention” in “Convention T.” But Ebbs’s explanation relies on just the bare use of the word ‘Convention’ itself and Tarski’s claim that he “does not understand what is at stake in ‘disputes about the right conception of truth’“ (1944, p. 355). When one looks further at what Tarski says, it becomes clear that Ebbs’s is not the best explanation.

Ebbs says that the point of the Convention is to avoid “controversy about what [a term] ‘really’ means” (p. 100), that it is not “right or wrong, but more or less useful to us,” and that it “in effect specifies the uses of [a term] that we find clear, unproblematic, and worth preserving” (loc. cit.). He says that Tarski rejects the view that “there is a uniquely correct definition of the term [true]” (loc. cit.). But by this he does not mean merely that Tarski thinks that ‘true’ is ambiguous in ordinary usage, for earlier he distinguishes between selecting one ordinary meaning of a term like ‘bank’ from offering an explication. The point is, as he puts it earlier, in connection with meaning, “to replace some ... ordinary concepts of meaning with a different concept or notion characterized by one’s philosophical theory” (op. cit. p. 92). The point applied to Tarski is that he aims to replace some ordinary concept of truth (at least one of the conceptions, if are there more than one expressed by ‘true’) with a different concept, one defined by some subset of ordinary usage associated with that conception.

Given this, however, the passage that Ebbs cites from Tarski’s 1994 paper on “The Semantic Concept of Truth” does not support his interpretation. When Tarski says, in the passage quoted by Ebbs, that he has no intention of entering into dispute about what the right conception of truth is, it is not because there is not a definite everyday conception of truth he wants his truth definitions to conform to, but because he does not think there is point to talking about which of *the different conceptions of truth expressed by “true”* in everyday language is the “right” one. That is, the term is ambiguous, and philosophers seem to engage in a dispute over which of the various different conceptions expressed by the word ‘true’ is the right conception. The problem is, as he says explicitly, that “the sense in which the phrase ‘the right conception’ is used has never been made clear” (1944, p. 355). The debate here is as pointless as would be a debate over which of the conceptions of bank expressed by ‘bank’ is the right conception.

In fact, Tarski had in mind a definite conception of truth that is to provide the standard of adequacy for an adequate definition of a truth predicate for a formal language. Tarski notes in “On the Concept of Truth in Formal Languages” (Tarski 1983), where he originally introduces Convention T, that there are different conceptions of truth. But he isolates one as the one that he is concerned with, namely, what he calls the correspondence conception: "I shall be concerned exclusively with grasping the intentions which are contained in the so-called *classical* conception of truth ('true--corresponding with reality') in contrast, for example, with the *utilitarian* conception (true--in a certain respect useful)" (1983, 153). This is his same concern in "The Semantic Concept of Truth," where he says "The main problem is that of giving a satisfactory definition of this notion, i.e., a definition which is materially adequate and formally correct" (Tarski 1944, p. 341). He goes on to say (loc. cit.; emphasis added): "The desired definition does not aim to specify the meaning of a familiar word used to denote a novel notion; on the contrary, *it aims to catch hold of the actual meaning of an old notion*. We must then characterize this notion precisely enough to enable anyone to determine whether the definition fulfills its task." Tarski goes on to say that the word 'true' is ambiguous in everyday language and says "we must indicate which conception will be the basis of our discussion" (1944, p. 342). And he says, unequivocally: "We should like our definition to do justice to the intuitions which adhere to the classical Aristotelian conception of truth—intuitions which find their expression in the well-known words of Aristotle's Metaphysics: *To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true*." (He cites this formulation in a footnote in "On the Concept of Truth" (1983, p. 155)). This he calls the classical conception, and identifies it as the one that is his target. Convention T is introduced to serve as "a more precise expression of" the intuitions expressed in this classical formulation of the correspondence theory of truth (1944, p. 343). Thus, Ebbs is mistaken in saying that Tarski did not intend Convention T to express a criterion of adequacy that is uniquely determined by a prior conception of truth.

Why call Convention T a 'convention'? An answer to this question that conforms to the points made here is suggested by Marian David (who brings additional passages to bear as well):

> ... judging from indications present in Tarski's own works, the conventionalist aspect of Convention T seems intended to reflect that a choice has been made by Tarski, that he has chosen to make precise [better to say in conformity with Tarski's way of putting it 'provide a precise expression of'] the classical conception of truth rather than some other conception. (David 2008, p. 155)

That is, the conventional element attaches to the use of the term 'adequate definition of truth' and reflects a decision about which conception of truth guides the project of characterizing a materially adequate definition of truth for a formal language. Indeed, in the statement of Convention T, Tarski says a formal definition of a symbol "will be called *an adequate definition of truth*" if it has all appropriate instances of the T-scheme as consequences (1983, 188). In the footnote on that page, he says the convention can be converted to a normal definition in the metalanguage. So it is a convention because it introduces stipulatively the use of the phrase 'adequate definition of truth' for a predicate for a formal language, and the point is to fix which of the ordinary notions its extension is to conform to for the language.

But doesn't this just support Ebbs's reading after all? For what Tarski calls an adequate definition of truth, after all, does not, and is not regarded by him, as expressing the ordinary concept of truth, not even the classical conception! For, first, he does not think a definition can be given for a universal truth predicate because (inter alia) it would have to apply to languages which contained their own truth predicate, which gives rise to the semantic paradoxes. And, second, his goal is explicitly to provide a method for constructing truth definitions that are materially adequate for formal work in logic, and not to capture the intension of the term.

About this we can make three points. First, we should not forget that what Ebbs is interested in is what the adequacy condition is intended to do. For what he wants to argue is that *just as Tarski's adequacy condition is not meant to capture the ordinary notion of correspondence truth*, neither is Davidson's condition of adequacy supposed to capture the ordinary notion of meaning. If we keep this in mind, we can see that the fact that the definitions that the condition of adequacy is a condition on don't express the ordinary conception is not to the point. It is enough to undermine Ebbs's argument that in fact the adequacy condition is intended to express in a precise way the classical conception of truth. Second, Tarski could not have intended that the truth definitions he shows how to construct be explications in Ebbs's sense of the ordinary concept of truth. He is not trying to isolate some subset of uses of 'is true' and define a universal truth predicate that captures that subset of uses. He provides a method of defining for each language in a certain class a predicate that we can know to have the right extension for its language. These are all distinct definitions, and none is to be regarded as the explication of the ordinary notion. This is why Ebbs's focus has to be on the criterion of adequacy itself. Third, Tarski did intend that the condition of adequacy ensure that definitions of truth predicates for particular formal languages have exactly the extension of the concept of truth as restricted to those languages. If we are to take seriously the idea that Davidson modeled his project on Tarski's, then we would expect minimally that Davidson would aim to get the extension of 'x translates y' correct by laying down his adequacy conditions, which Ebbs denies that Davidson aims to do.

Turning to Ebbs's second claim about Davidson, this is obviously weakened if the interpretation Ebbs offers of Tarski is implausible. In fact, once we see what Tarski aimed at, it reinforces rather than undermines our interpretation of what Davidson aims to do. Davidson did, after all, take Tarski's work on truth as a model, and what Tarski sought to do was to capture in a rigorous way an ordinary notion. So, too, for Davidson: he sought an understanding of that notion of meaning that underlies Convention T, translation, by putting constraints on a true theory that suffice to elicit theorems that meet Convention T. But even apart from this, interpreting Davidson's remarks about his inverting Tarski's approach as expressing commitment to a project of explication seems strained. It does not conform to the tenor Davidson's remarks, and if it is the word 'convention' that is supposed to signal that someone's criterion of adequacy is in some sense stipulative, Davidson does not use 'convention' in the place you would expect, that is, he does not characterize his own suggestion for what will suffice for a truth theory to be usable for interpretation as a convention. If Davidson were thinking as Ebbs suggests, it would be natural for him to say: Tarski offered a convention for giving a definition (explication) of 'true', Convention T. I offer in the same spirit a convention for giving a definition (explication) of 'interprets', etc. But, of course, there's no hint of anything like this in what Davidson says, and no indication that Davidson reads Tarski in the way that Ebbs suggests that Tarski (mistakenly, as I have argued) be read.

In short, there is no way to muscle reflections on Convention T into a defense of the Explication Interpretation.

Now let us return to Davidson's response to Foster. The response to the first half of the charge was to agree with it but note that he always thought you had to know that the truth theory met Convention T to be used for interpretation. To the second half of the charge, he says in [49] that it was not part of his project to eschew the use of intensional notions in this context.

> [49] My way of trying to give an account of language and meaning makes essential use of such concepts as those of belief and intention, and I do not believe it is possible to reduce these notions to anything more scientific or behavioristic. What I have tried to do is to give an account of meaning (interpretation) that makes no essential use of unexplained linguistic concepts. … It will ruin no plan of mine if in saying what an interpreter knows it is necessary to use a so-called intensional notion—one that consorts with belief and intention and the like. (p. 176)

And it is not just concepts that that consort with belief and intention and the like that are permissible. Davidson notes that 'entails that’ would be a slightly more appropriate notion to use than Foster’s ‘says that’. He suggests in [50] an analysis of ‘entails that’ which appeals explicitly to the concept of synonymy as between sentence and utterance.

> [50] If a theory T entails that ‘Snow is white’ is true in English if and only if snow is white, then T has as a logical consequence a sentence synonymous with my utterance of ‘“Snow is white” is true in English if and only if snow is white’.

The second component, he notes, brings in a specifically linguistic concept. But he denies in [51] that this is a problem.

> [51] This does not make the account circular, for those conditions [on a theory of truth that are to elicit translations of object language sentences into metalanguage sentences] were stated, we have been assuming, in a non-question-begging way, without appeal to linguistic notions of the kind we want to explain. So the concept of synonymy or translation that lies concealed in the notion of entailment can be used without circularity when we come to set out what an interpreter knows. Indeed, in attributing to an interpreter the concept of a translational theory we have already made this assumption. (p. 178)

At the risk of beating a dead horse: notice that the notion of synonymy is invoked in an analysis of the notion of entailment. This is the ordinary notion of entailment, and it is the ordinary notion of synonymy that is invoked. And it is precisely that notion generalized—translation—an account of which is sought by way of putting constraints on a truth theory sufficient for it to meet Convention T.

Why in the end does Davidson not take the step to giving an explicit meaning theory? There is a straightforward reason which is connected with his analysis of the role of (for Davidson, apparent) sentential complements of verbs that create an intensional context. This is explained in [52] in the very last paragraph of “Reply to Foster.”

> [52] On a point of some importance, I think Foster is right. Even if everything I have said in defence of my formulation of what suffices for interpretation is right, it remains the case that nothing strictly constitutes a theory of meaning. A theory of truth, no matter how well selected, is not a theory of meaning, while 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. This does not, however, make it impossible to say what it is that an interpreter knows, and thus to give a satisfactory answer to one of the central problems of the philosophy of language. (p. 179)

Davidson analyzes ‘Galileo said that the earth moves’ as involving in use semantically two distinct token sentences, ‘Galileo said that. The earth moves.’ In the first, ‘that’ refers to the second. ‘Galileo said that’ expresses (roughly) that some utterance of Galileo’s samesays that (Davidson 2001a). He analyzes ‘x entails that p’ along the same lines: “If a theory T entails that ‘Snow is white’ is true in English if and only if snow is white, then T has as a logical consequence a sentence synonymous with my utterance of ‘“Snow is white” is true if and only if snow is white’“ (p. 178). It is the demonstrative element in that analysis he has in mind in saying that a statement that a translational theory entails certain facts cannot be a formal theory because of irreducible indexical elements in the sentences that express it. And it is clear that his attitude toward ‘x means that p’ would be the same. The analysis of this would closely parallel that of ‘x entails that p’. To say that ‘snow is white’ means that snow is white is to say (roughly) that ‘snow is white’ is synonymous with my utterance of ‘snow is white’. We can note that this provides additional support, if any were needed, for the view that Davidson aimed to specify constraints the meeting of which sufficed for canonical theorems of a truth theory to provide context relative specifications of truth conditions that interpreted utterances in the same sense as that involved in his analysis of 'x entails that p'.

8. Conclusion

The Introduction to *Inquiries into Truth and Interpretation* begins in [53] with an account of Davidson's general project.

> [53] What is it for words to mean what they do? In the essays collected here I explore the idea that we would have an answer to this question if we knew how to construct a theory satisfying two demands: it would provide an interpretation of all utterances, actual and potential, of a speaker or group of speakers; and it would be verifiable without knowledge of the detailed propositional attitudes of the speaker. The first condition acknowledges the holistic nature of linguistic understanding. The second condition aims to prevent smuggling into the foundations of the theory concepts too closely allied to the concept of meaning. A theory that does not satisfy both conditions cannot be said to answer our opening question in a philosophically instructive way. (p. xiii)

Davidson does not write, "What is it for words to mean what they do in a sense of 'meaning' that will be given stipulatively in what follows?" He does not say "What is it for words to mean what they do when we abstract away from much of what we mean by "meaning" to preserve only such and such features?" Nor, if he had intended either of these things, would what he goes on to say after this be an intelligible continuation. He rather announces what is a central question in the philosophy of language, and makes a suggestion about what sort of theory would provide a satisfactory answer to it. The Replacement Theory, even Ebbs's sophisticated version of it, the Explication Interpretation, requires us to read Davidson as either confused about what would count as an answer to his question or to be disingenuous about what he is up to. The Replacement Theory, even clothed in the doctrine of Carnapian explication, should be an interpretation of last resort, something we accept because we cannot otherwise make sense of what Davidson says. But it turns out that all of the passages that have been cited in favor of it are better understood as a straightforward pursuit of the traditional project, and the misunderstandings are grounded in a failure to grasp just how ingenious Davidson's proposal for circumventing the problems of the tradition is. The Replacement Theory arises out of a partial, fragmentary reading of Davidson, and is aided by a misunderstanding of his relation to his historical context, which sees him as carried along by certain contemporary currents of thought rather than steering his own course. It misses the main point of Davidson's proposal. It makes him less rather than more interesting. It ignores his real and substantial contribution to the theory of meaning. It cannot be made compatible with the full range of texts it has to deal with. Its source lies mainly in a few pages in one paper, "Truth and Meaning," read out of context, and outside the context of the rest of Davidson's work, and in scattered passages interpreted in isolation from their contexts and in the light of the commitments already accrued in the misreading of those earlier pages. It attributes to Davidson an unannounced esoteric doctrine with obscure unarticulated goals. This simulacrum of Davidson has nothing to do with the real thing. The real Davidson straightforwardly does what he says he is doing. He pursues the traditional goal of illuminating what it is for words to mean what they do through a clever and illuminating bit of indirection, drawing on insights from both Tarski and Quine, but transforming each in integrating them into his own project. While the Replacement Theory has become entrenched in philosophical lore, it is a fundamental mistake about Davidson which obscures his most important insights. It is time to put the Replacement Theory to rest. May it Rest in Peace.

Appendix: Quine's relation to Davidson

In this appendix, I consider the relation of Quine and Davidson's projects in the theory of meaning.

I think that part of what underlies the interpretive dispute I have with Ebbs is traceable to a difference in view about what Davidson has in mind in urging that we must be able to interpret another, ultimately, on the basis of evidence that does not presuppose anything about meaning or related matters. I believe that Ebbs sees this as in the vein of Quine's insistence on a standard of clarity in the development of a theory of communication (in Quine's hands this takes the form of an a theory of translation—what is preserved in communication) that requires us to leave behind our folk theory or folk conception of it as confused, vague, and inadequate. The standard is a broadly empiricist one: show what the cash value in observation is for the claims of the theory, and then read this into the content of the theory.

Davidson did look to Quine for inspiration. But Davidson's standard of evidence is not due to behaviorism or to some general empiricist commitment or to a project of scientific revision of our ordinary conceptual scheme. It is due to the fact that interpretation is, as Quine put it in the first sentence of the preface to *Word and Object*, "a social art" (1960). If I want to be interpreted by you, I must make myself intelligible to you, and so what I convey to you must be something that I can expect you to be able to figure out on the basis of the evidence you have. You cannot read my mind, but must perforce go on behavior. We both know this. We must both then take ourselves to be engaged in a transaction in which what is exchanged is recoverable from what is equally accessible to both of us. This reduces ultimately (or so it seems) to a description of behavior that does not presuppose anything about meaning or the attitudes. This is a point Davidson took from Quine. As Quine puts it clearly in [54] from "Epistemology Naturalized."

> [54] The sort of meaning that is basic to translation, and to the learning of one's own language, is necessarily empirical meaning. … Language is socially inculcated and controlled; the inculcation and control turn strictly on keying of sentences to shared stimulation. … Surely one has no choice but to be an empiricist so far as one's theory of linguistic meaning is concerned. (1969, p. 81)

But in taking this point from Quine, Davidson also transposed it. In an interview conducted with Ernie Lepore in 1988, Davidson described how he came to his approach to the theory of meaning in [55].

> [55] [a] … what's not in 'Truth and Meaning' but what lies behind it is the years of teaching philosophy of language [at Stanford] without anyone to give me any guidance, really without any background in the subject. So I started out as many people did in those days, reading Ogden and Richards's *The Meaning of Meaning* and Charles Morris. Now what looked like the central problem to them was to define the concept of meaning: x means y, where x is a word or a phrase or a sentence and God knows what y was supposed to be—and you wanted: iff what? That is how a lot of people were thinking about philosophy of language. Really smart people sought analyses of particular locutions, but never said anything about how you could tell whether you had come up with a correct solution or on what grounds you criticize these things aside from just ad hoc arguments. So I think perhaps I felt more frustrated by this situation that I found the subject to be in than I think other people did. On the one hand, so many issues seemed rather sharp: What is meaning? How do you even think about it? Where do you start? [b] And somewhere along the line I discovered Tarski and I thought: you don't even want to ask the question what is meaning. It's the wrong question. It was a huge shift of perspective to get away from worrying about what it is to talk about the meaning of a predicate. Reading Tarski made me realize

> that there's a way to get around all that— [c] and somewhere along there Quine showed up at the Center for Behavioral Studies at Stanford. At that point they invited people who were at the center to bring up an associate, and I had a term off and I agreed to just come and read a manuscript version of what was to become his *Word and Object*. I really didn't do anything else that term except read it over and over again, trying to understand what was going on. And when I did, I thought it was terrific. And I saw again that it was a whole way of approaching problems in the philosophy of language that other people hadn't caught on to, hadn't even thought about, and it seemed much more promising, and so I sort of slowly put what I thought was good in Quine with what I had found in Tarski. And that's where my general approach to the subject came from. (Davidson 2004b, pp. 257-8)

There are several things to take away from these remarks.

First, in [55a], there is Davidson's dissatisfaction with the (then) tradition, its focus on the question 'What is meaning?" and its focus on trying to say *in other words* what 'means' means. Many people tried to give analyses of the meanings of particular expressions, and this of course presupposes some grasp on what 'means' means, but they had no clear criterion for success or standard for criticism.

Second, in [55b], Davidson notes the shift in perspective that comes with recognition of how Tarski's work could be applied in the theory of meaning. Instead of asking directly, 'What is meaning?', one could instead ask what it is to give a meaning theory for a particular language, recognizing the importance, in determining meaning, of saying what contribution each semantically primitive expression makes to all of the sentences in which it appears. This provides in turn a purchase on the adequacy of particular accounts of meaning: whatever proposal one makes for some range of discourse, e.g., belief sentences, the account must be compatible with the words making the same systematic contributions in different contexts and our understanding of complex expressions resting on our grasp of the primitive components and their mode of composition. This perspective is applied in "Theories of Meaning and Learnable Languages" to criticize prominent analyses of quotation, indirect discourse, and belief sentences. What Davidson saw in Tarski's work was a way of achieving a compositional meaning theory without appeal to meanings as entities by way of formulating a truth theory for a language that met Convention T. If we think about how we would do this for our own language, we would use axioms in giving satisfaction conditions that employ terms the same in meaning (or appropriately related for context sensitive expressions) as those for which we give satisfaction conditions. Then, relative to a certain notion of a minimal or canonical proof (see (Lepore and Ludwig 2005, chapter 7, sec. 3) for discussion and a simple example), the T-theorems we derive which have no semantic vocabulary on the right hand side will be translational. The proofs would show how the primitive expressions for which axioms are given contribute to fixing, in virtue of their meaning, the interpretive truth conditions of the sentences in which they appear. The totality of the axioms then express the role of each primitive expression in fixing interpretive truth conditions in any sentence in which it can appear.

Third, in [55c], we see that the final ingredient is provided by Quine's observation that meaning should be constrained by the basic function of language. This constraint, as Davidson understood it, pertained to the ultimate evidence that is the basis for correct attributions of meaning. Once he had the idea about how a Tarski-style theory could be exploited to specify, relative of course to the assumption that it met (in an appropriate way) the requirement that it satisfy Convention T, the meaning of any sentence in the language (by giving interpretive truth conditions), the final step was to treat it as an empirical theory to be confirmed from the standpoint of the interpreter of another who works with the fundamental data on which interpretation must be based. This is what he got from Quine. And this then provides a way of approaching the theory of meaning that seeks to illuminate what meaning is, not by trying to provide a definition in other words of the predicate 'means', but instead by showing how a theory that can be used to interpret another, which provides an account of the systematic contribution of each semantically primitive expression in the language to the interpretive truth conditions of each sentence in which it appears, can be confirmed on the basis of evidence that does not presuppose any of the facts which the theory is concerned with, or any facts access to which presupposes any of those facts (see again passage [29] in connection with this).

Davidson puts it this way in his 1990 Dewey Lectures, "The Structure and Content of Truth" in [56].

> [56] What we should demand … is that the evidence for the theory be in principle publicly accessible, and that it not assume in advance the concept to be illuminated. The requirement that the evidence be publicly accessible is not due to an atavistic yearning for behavioristic or verificationist foundations, but to the fact that what is to be explained is a social phenomenon. … the correct interpretation of one person's speech by another must in principle be possible. … what has to do with correct interpretation, meaning, and truth conditions is necessarily based on available evidence. … language is intrinsically social. … meaning is entirely determined by observable behavior, even readily observable behavior. That meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language. (1990, p. 314)

Davidson makes clear that he is concerned with the basic issues in the philosophy of language that had exercised Ogden and Richards and Morris, but was dissatisfied with traditional approaches, and, of course, not only, as he notes here, in their emphasis and methods, but also in the appeal to entities in the theory of meaning as a prop for theories that provide no insight into how meaning is related to the facts that ground it. His solution took from both Tarski and Quine. Tarski provided a recursive framework for articulating how words contribute to the interpretive truth conditions of sentences. Quine provide a crucial insight about how to cast the problem of treating it as an empirical theory by emphasizing the constraint on the ultimate data for interpretation and the centrality of the standpoint of the interpreter in understanding language. Davidson does not give any hint here that he follows Quine in every respect, and no hint that he thinks he is turning his back on the project (and more on this momentarily). He says, rather discretely, "I sort of slowly put what I thought was good in Quine with what I had found in Tarski." It is clear, for example, that Davidson did not follow Quine in rejecting the propositional attitudes as "creatures of darkness" or seek to give behaviorist reductions of these notions. Instead they play a central role in his own account of interpretation, as expressed in [57] from "Thought and Talk,"

> [57] … it should not be thought that a theory of interpretation will stand alone, for as we noticed, there is no chance of telling when a sentence is held true without being able to attribute desires and being able to describe actions as having complex intentions. This observation does not deprive the theory of interpretation of interest, but assigns it a place within a more comprehensive theory of action and thought. (p. 162).

and he denies, in [58], they can be reduced to other concepts,

> [58] Adverting to beliefs and desires to explain action is … a way of fitting an action into a pattern of behaviour made coherent by that theory. This does not mean, of course, that beliefs are nothing but patterns of behaviour, *or that the relevant patterns can be defined without using the concepts of belief and desire*. Nevertheless, there is a clear sense in which attributions of belief and desire, and hence teleological explanations of belief and desire, are supervenient on behaviour more broadly described. (op. cit., p. 159)

Furthermore, he is clear that it is the common concepts with which he is concerned, for in commenting on the status of the maxims of interpretation he advances he says in [59], from "A New Basis for Decision Theory":

> [59] It should be emphasized that these maxims of interpretation are not mere pieces of useful or friendly advice; rather they are intended to externalize and formulate (no doubt very crudely) essential aspects of the *common concepts of thought, affect, reasoning and action.* What could not be arrived at by these methods is not thought, talk, or actions. (1985, p. 92, emphasis added)

He extends the same point to linguistic concepts ("Belief and the Basis of Meaning") in [60].

> [60] *Everyday linguistic concepts* are part of an intuitive theory for organizing more primitive date, so only confusion can result from treating these concepts and their supposed objects as if they had a life of their own" (p. 143; emphasis added).

The intuitive theory of which they are a part is a unified theory of thought, meaning and action. This picture is reinforced in the next exchange in the interview. Lepore at this point remarks, "So Quine had very little influence on your philosophy of language until very late, until you were in your forties. This I think would be a great surprise to many readers of your work." Davidson says in response: "That's right. My philosophy of language didn't grow out of my relationship with Quine at all" (Davidson 2004a, p. 258). (Quine and Davidson had been friends since Davidson was a graduate student at Harvard, but Davidson's dissertation was in classical philosophy, on Plato's *Philebus*.) Lepore puts it explicitly to Davidson that his is not a revisionist program in [61].

> [61] … readers might leave with the not uncommon impression that Davidson's philosophy of language is really just modified Quine. That would be a mistake. … He starts off clearly from a revisionist point of view. As early as his paper 'The Problem of Meaning in Linguistics', he's telling us that only very few features of our ordinary concept of meaning are salvageable. You don't think that at all. I don't see a revisionist perspective in your writings. (loc. cit.)

Lepore then asks Davidson what the difference is between his and Carnap's program, and whether he [Davidson] influenced Michael Dummett's interpretation of Frege as "trying to devise a theory of meaning in your sense long ago" (op. cit., p. 259). Davidson responds in [62].

> [62] I think the idea that there was a way of thinking philosophically about meaning tied to the idea of getting a serious semantic theory for as much of natural language as you could—well, I was the first person to say that, and I say it in 'Truth and Meaning'. There I suggested that my dream was to try to do for the semantics for natural language what Noam Chomsky was doing for the syntax of natural language. But he didn't have quite the same concept of a theory as I did. He knew what it was like to give a recursive definition of a sentence, for example. But when I was writing that paper, I couldn't believe no one thought about it that way. So I looked about in Carnap, in Reichenbach, and in Quine, and none of them was even describing this as a project. Tarski discouraged everybody by saying, of course, you can't do this for natural language. Quine never thought of it in terms of a theory at all. Of course, his discussion of translation could, if you think of it now with a little twist, ... be redescribed or reexpressed in a Tarski-like way. But he certainly wasn't thinking about it this way at the time he was first writing about it in *Word and Object*. (op. cit., p. 259)

We should appreciate the fact that in responding to Lepore here Davidson does not dispute what he says. Lepore puts it to Davidson explicitly: "I don't see a revisionist perspective in your writings." Davidson does not say, "No, actually, I am a revisionist." He does not take the opportunity to explain that he is really engaged in a project of Carnapian explication. He says that in "Truth and Meaning" he had suggested that his "dream was to try to do for semantics for natural language what Noam Chomsky was doing for syntax of natural language" (see [11] above). Davidson had his own agenda in the theory of meaning, and it was not Quine's. It was informed by his work on and thinking about the problems in the philosophy of language in the 1950s at Stanford—before the major influence from Quine, which came in 1958-9 when he read the manuscript of *Word and Object* when Quine was at the Center for Advanced Study in the Behavioral Sciences—by his work on experimental decision theory with Suppes and McKenzie at Stanford, and by his exposure to Tarski's work in the mid 1950s. What Davidson did was to integrate insights from a number of different contemporary strands in the philosophy of language into a completely novel approach, one that took important insights from Quine, but worked them, along with others, into the interests he had in overcoming the limitations of contemporary approaches to the theory of meaning. He did not aim to turn his back on that project but to provide an enlightened version of it.

To return to Ebbs: Ebbs sees Davidson's insistence on explaining meaning ultimately in terms of the evidence available to the radical interpreter as an attempt, like Quine's, to set the project of understanding meaning on a firm footing, leaving behind the vague, imprecise, confused notions at play in ordinary understanding of thought and talk. But this is to foist onto Davidson a project of Quine's which he does not share. His insistence on explaining meaning in terms of the evidence available from the standpoint of radical interpretation rests on his conviction that this places a constraint on the nature of meaning that arises from the fact that its function in facilitating communication requires it to be available intersubjectively. Behavior described in intentional terms is excluded because this would require prior identification of attitude contents, which he thinks cannot be recovered independently of interpreting utterances. (One should recall here the central thesis of "Thought and Talk" (Davidson 1975) that only linguistic beings have propositional attitudes.) But what we want to understand from this standpoint, as noted above, are the everyday linguistic and semantic concepts (and the allied psychological attitude concepts) that are "part of an intuitive theory for organizing more primitive data" ("Belief and the Basis of Meaning," p. 143); "only confusion can result from treating these concepts and their supposed objects as if they had a life of their own" (loc. cit.). Davidson thought that the concepts we actually deploy are concepts whose application conditions are to be understood in terms of more primate data, as shown in [56], [58]-[59], and in [63]-[66].

[63] The interlocking of the theory of action with interpretation will emerge in another way if we ask how a method of interpretation is tested. In the end, the answer must be that it helps bring order into our understanding of behavior. (1974, p. 161)

[64] There are conceptual ties between the attitudes and behavior which are sufficient, given enough information about actual and potential behavior, to allow correct inference to the attitudes. (2001b, p. 100)

[65] I have been engaged in a conceptual exercise aimed at revealing the dependencies among our basic propositional attitudes at a level fundamental enough to avoid the assumption that we can come to grasp them—or intelligibly attribute them to others—one at a time. My way of performing this exercise has been to show how it is in principle possible to arrive at all of them at once.

What makes the task practicable at all is the structure the normative character of thought, desire, speech, and action imposes on correct attributions of attitudes to others, and hence interpretations of their speech and explanations of their actions. (2004c, p. 166)

> [66] Belief, like the other so-called propositional attitudes, is supervenient on facts of various sorts, behavioral, neurophysiological, biological, and physical. … The point is … understanding. We gain one kind of insight into the nature of the propositional attitudes when we related them systematically to one another and to phenomena on other levels. As interpreters, we work our way into the whole system, depending much on the pattern of interrelationships. (2001c, p. 147)

There is no need to see him as engaging in a hidden project of explication to understand why he takes the position he does on this.

Before leaving this topic, it will be useful to look at a passage in "Belief and the Basis of Meaning," in which Davidson makes some explicit remarks about the relation of his program to Quine's. This occurs at the end of a section of the paper in which he discusses the analogy between (i) the problem in empirical decision theory of identifying degrees of belief and preferences by appeal to choice behavior, and (ii) identifying what a speaker believes and what he means by his sentences by appeal to evidence about the conditions under which he holds sentences true. Davidson summarizes the central ideas in [67].

> [67] … behavioural or dispositional facts that can be described in ways that do not assume interpretations, but on which a theory of interpretation can be based, will necessarily be a vector of meaning and belief. One result is that to interpret a particular utterance it is necessary to construct a comprehensive theory for the interpretation of a potential infinity of utterances. The evidence for the interpretation of a particular utterance will therefore have to be evidence for the interpretation of all utterances of a speaker or community. Finally, if *entities* like meanings, propositions, and objects of belief have a legitimate place in explaining speech behavior, it is only because they can be shown to play a useful role in the construction of an adequate theory. (p. 149; emphasis added)

In the next two paragraphs Davidson expresses his debt to Quine and says, in [68], something about how he thinks of the relation of his project to Quine's.

> [68] The appreciation of these ideas, which we owe largely to Quine, represents one of the few real breakthroughs in the study of language. I have put things in my own way, but I think that the differences between us are more matters of emphasis than of substance. Much that Quine has written understandably concentrates on undermining misplaced confidence in the usefulness or intelligibility of concepts like those of analyticity, synonymy, and meaning. I have tried to accentuate the positive. Quine, like the rest of us, wants to provide a theory of interpretation. His animadversions on meanings are designed to discourage false starts; but the arguments in support of the strictures provide foundations for an acceptable theory. (p. 149)

It is clear that in this paragraph Davidson is minimizing his differences with Quine. And it would be understandable if it encouraged the view that Davidson is signaling that he is engaged in Quine's project. Does he not say: "I think that the differences are more matters of emphasis than of substance"? But let us first recall the local context, and then attend carefully to what Davidson says following this.

In the opening sentence in [68], 'these ideas' refers to the ideas expressed in [67], and of course everything here is simply an expression of the commitment, derived from the function of language, to understanding meaning and propositional attitudes by way of investigating how a theory deploying these concepts is to be confirmed on the basis of evidence that doesn't presuppose knowledge of the right theory. Davidson has indeed put this in his own way. In particular, he makes use of propositional attitude vocabulary that Quine eschews as insufficiently clear, he characterizes an intermediate state of evidence as being hold true attitudes toward sentences, he explicitly frames the theory in semantic terms, expresses the criterion of adequacy using the notion of translation, and he sets up the problem as that of breaking into the circle of meaning and belief. All of this is alien to Quine's own way of conceptualizing the problem of giving an scientifically respectable account of language in terms of the notion of stimulus meaning in *Word and Object* (1960, ch. 2). Are these mere matters of emphasis and not matters of substance? Be that as it may, Davidson here clearly seeks to find and emphasize the common ground between him and Quine, not to dwell on the differences. But this doesn't entail that there no differences, and the differences I have already noted show that another difference lies in their attitude toward ordinary linguistic, semantic, and propositional attitude concepts.

Note how Davidson goes on. He says Quine has understandably concentrated on "undermining *misplaced confidence* in the usefulness or intelligibility of" certain concepts (emphasis added). Davidson can agree that there is misplaced confidence in the usefulness and intelligibility of these concepts, but still seek to show how to understand them. He says that Quine "like the rest of us, wants to provide a theory of interpretation." But who does Davidson have in mind by "the rest of us"? Earlier in the paper, Davidson says that the "[t]heory of interpretation is the business jointly of the linguist, psychologist and philosopher" (p. 142). The rest of us comprises all those interested in the traditional project. Does Davidson think that Quine is engaged in the same project? The inclusion of the clause "like the rest of us" here suggests that Davidson is well aware that Quine is not regarded as doing what the rest of us are doing. But perhaps the case can be made at some level of abstraction at which we think the project is making what sense we can of linguistic communication. Then the point is that we can see Quine as contributing to the project. And what Davidson says about how he takes Quine's contributions suggests that he thinks that their importance lies in clearing away bad ideas and in laying down important constraints on an adequate account. This is indicated in the next two things Davidson says.

First, Davidson says that Quine's "animadversions on meanings are designed to discourage false starts," where the plural indicates Davidson has in mind meanings as entities, the utility of which he criticized in "Truth and Meaning." Second, Davidson says that "the arguments in support of the strictures provide foundations for an acceptable theory." An acceptable theory of what? Of interpretation, but in the terms Davidson has used to explain it. Here he signals that his project involves accepting the requirement that we do better than the tradition has on its project, incorporates Quine's insights and strictures, and then provides the outline of an acceptable theory.

When Davidson goes on from this point to describe how his proposal differs from Quine's, he introduces the proposal that the theory used to interpret others not take the form of a translation manual but of a theory of truth that meets Convention T, and he remarks, in a passage quoted above, that we aim to invert Tarski's approach "to achieve an understanding of meaning or translation by assuming a prior grasp of the concept of truth" (p. 150); and specifically, by describing a way of "judging the acceptability of T-sentences that is not syntactical, and makes no use of the concepts of translation, meaning, or synonymy, but is such that acceptable T-sentences will in fact yield interpretations" (p. 150). This provides a way of grounding those notions in the primitive data that the theory of which they are a part are designed to help us organize. Davidson does take this from Quine, and he felt indebted to him. But as I mentioned above, he transposed what he took from him. He clearly does not take Quine's dismissive attitude toward meaning or the propositional attitudes but rather seeks to integrate Quine's insights into the foundation of an adequate theory of them.

Kirk Ludwig
Philosophy Department, Indiana University
ludwig@indiana.edu

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. *Philosophical Review* 101 (1):3–51.

Chihara, Charles S. 1975. Davidson's Extensional Theory of Meaning. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 28 (1):1-15.

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.

David, Marian. 2008. Tarski's Convention T and the Concept of Truth. In *New Essays on Tarski and Philosophy*, edited by D. Patterson: Oxford University Press.

Davidson, Donald. 1965. 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. Reprinted in Davidson 2001d.

———. 1967. Truth and Meaning. *Synthese* 17:304–323. Reprinted in Davidson 2001d.

———. 1970. Semantics for Natural Languages. In *Linguaggi nella Societa e nella Tecnica*. Milan: Comunita. Reprinted in Davidson 2001d.

———. 1973a. In Defence of Convention T. In *Truth, Syntax and Modality*, edited by H. Leblanc. Dordretch: North-Holland Publishing Company. Reprinted in Davidson 2001d.

———. 1973b. Radical Interpretation. *Dialectica* 27:314–328. Reprinted in Davidson 2001d.

———. 1974. Belief and the Basis of Meaning. *Synthese* 27:309–323. Reprinted in Davidson 2001d.

———. 1975. Thought and Talk. In *Mind and Language*, edited by S. Guttenplan. Oxford: Oxford University Press. Reprinted in Davidson 2001d.

———. 1976. Reply to Foster. In *Truth and Meaning: Essays in Semantics*, edited by G. Evans and J. McDowell. Oxford: Oxford University Press. Reprinted in Davidson 2001d.

———. 1985. A New Basis for Decision Theory. *Theory and Decision* 18:87–98.

———. 1990. The Structure and Content of Truth. *The Journal of Philosophy* 87 (6):279–328.

———. 2001a. On Saying That. In *Inquiries into Truth and Interpretation*. New York: Clarendon Press. First published, 1968.

———. 2001b. Rational Animals. In *Subjective, Intersubjective, Objective*. New York: Clarendon Press. First published, 1982.

———. 2001c. A Coherence Theory of Truth and Knowledge. In *Subjective, Intersubjective, Objective*. New York: Clarendon Press. First published, 1983.

———. 2001d. *Inquiries into Truth and Interpretation*. 2nd ed. New York: Clarendon Press. First published, 1984.

———. 2001e. Three Varieties of Knowledge. In *Subjective, Intersubjective, Objective*. New York: Clarendon Press. First published, 1988.

———. 2004a. An Interview with Donald Davidson. In *Problems of Rationality*. Oxford: Oxford University Press.

———. 2004b. *Problems of Rationality*. Oxford: Oxford University Press.

———. 2004c. A Unified Theory of Thought, Meaning, and Action. In *Problems of Rationality*. Oxford: Oxford University Press. First published, 1980.

———. 2005. A Nice Derangement of Epitaphs. In *Truth, Language, and History*. Oxford: Oxford University Press.

Ebbs, Gary. 2012. Davidson's Explication of Meaning. In *Donald Davidson on Truth, Meaning and the Mental*, edited by G. Preyer. Oxford: Oxford University Press.

Foster, John A. 1976. Meaning and Truth Theory. In *Truth and Meaning: Essays in Semantics*, edited by G. Evans and J. McDowell. Oxford: Clarendon Press.

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 J. 1982. Common Sense in Semantics. *Notre Dame Journal of Formal Logic* 23:174–218.

Lepore, Ernest, and Kirk Ludwig. 2005. *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford: 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.

———. 1969. Epistemology Naturalized. In *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Soames, S. 2008. Truth and Meaning: In Perspective. *Truth and Its Deformities: Midwest Studies in Philosophy* 32:1–19.

Soames, Scott. 1992. Truth, Meaning, and Understanding. *Philosophical Studies* 65 (1-2):17–35.

Stich, Stephen. 1976. Davidson's Semantic Program. *Canadian Journal of Philosophy* 6:201–227.

Tarski, Alfred. 1944. The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research* 4:341-376.

———. 1983. The Concept of Truth in Formalized Languages. In *Logic, Semantics, Metamathematics*. Indianapolis: Hackett Publishing Company. First published, 1934.