

March 2008

March 2008

Truth and Meaning – In Perspective

by

Scott Soames

School of Philosophy

University of Southern California

To Appear In

Truth and Its Deformities

Edited by Peter French

Midwest Studies in Philosophy

Volume XXXII

2008

Truth and Meaning – In Perspective **Scott Soames**

My topic is the attempt by Donald Davidson, and those inspired by him, to explain knowledge of meaning in terms of knowledge of truth conditions. For Davidsonians, these attempts take the form of rationales for treating theories of truth, constructed along Tarskian lines, as empirical theories of meaning. In earlier work¹, I argued that Davidson's two main rationales – one presented in "Truth and Meaning"² and "Radical Interpretation,"³ and the other in his "Reply to Foster"⁴ – were unsuccessful. Here, I extend my critique to cover an ingenious recent attempt by James Higginbotham to establish Davidson's desired result. I will argue that it, too, fails, and that the trajectory of Davidsonian failures indicates that linguistic understanding, and knowledge of meaning, require more than knowledge of that which a Davidsonian truth theory provides. I begin with a look at the historical record.

The Evolution of an Idea: A Historical Summary

When Davidson enunciated his idea, in the 1960s, that theories of meaning can be taken to be nothing more than theories of truth, it met with a warm reception. For devotees of Ordinary Language, its attraction lay in its promise of providing a theoretically respectable way of grounding claims about meaning, and distinguishing them from claims about use, that those who still placed meaning at the center of philosophy had come to recognize the need for.⁵ For

¹ Scott Soames, "Truth, Meaning, and Understanding," *Philosophical Studies*, 65, 1992; 17-35, and chapter 12 of *Philosophical Analysis in the Twentieth Century*, Volume 2 (Princeton and Oxford: Princeton University Press), 2003.

² Donald Davidson, "Truth and Meaning," *Synthese*, 17, 1967, 304-23; reprinted in *Inquiries into Truth and Interpretation*, (Oxford: Clarendon Press), 2001. Citations will be to the latter.

³ Donald Davidson, "Radical Interpretation," *Dialectica*, 27, 1973, 313-28; reprinted in *Inquiries into Truth and Meaning*. Citations will be to the latter.

⁴ Donald Davidson, "Reply to Foster," in Gareth Evans and John McDowell, eds. *Truth and Meaning*, (Oxford: Oxford University Press), 1976; reprinted in *Inquiries into Truth and Meaning*. Citations will be to the latter.

⁵ See chapters Parts 2-4 of *Philosophical Analysis in the Twentieth Century*, Vol. 2.

those laboring under the Quinean legacy of skepticism about analyticity, synonymy, and meaning, the idea afoot was that extensional notions from the theory of truth and reference were respectable, whereas intensional ones from the theory of meaning were not. This was an audience to which the Davidsonian program was bound to appeal.

It was one thing to claim that meaning has no special role to play in philosophy. As discomfiting as this was to Ordinary Language philosophers, it was something that Quineans could live with. Much more troublesome was the idea that meaning had no place in science. It certainly didn't seem that way to soldiers in the Chomskian revolution, who were busy transforming linguistics. The central work of the period, *Aspects of a Theory of Syntax*,⁶ enshrined the distinction between deep and surface structure, while championing the thought that a semantic theory of a natural language would interpret the deep structures of its sentences. To many, this brought to mind the Russellian distinction between logical and grammatical form. But what, it was wondered, is logical form, and what would it be to interpret it? Davidson laid the groundwork for answering these questions in a way that made sense to philosophers in the tradition of Russell, Tarski, Carnap, and Quine. For Davidsonians like Gilbert Harman, the logical forms of natural language sentences were their Chomskian deep structures, to interpret them was to give a truth theory for the language, and to see this as a theory of meaning was to see it as explicating what it is to understand the language.⁷

Though audacious, these ideas can be seen as the application of a familiar idea from philosophical logic. Since Tarski's seminal work on truth in the 1930s, it has been commonplace to view an interpreted formal language as the result of adding a model, plus a

⁶ Noam Chomsky, *Aspects of a Theory of Syntax*, (Cambridge: MIT Press), 1965.

⁷ Gilbert Harman, "Deep Structure as Logical Form," *Synthese*, 21, 1970, 275-97.

definition of truth-in-a model, to an uninterpreted formal system, thereby arriving at an assignment of truth conditions to every sentence.⁸ But if truth theories can be used in this way to *endow* sentences with meaning, then, surely, it seemed, they can also be used to *describe* the meanings of already meaningful sentences – provided, in the case of natural language, that we are clever enough to find the requisite logical forms to which to apply them. This was the technical task of the Davidsonian program. The philosophical challenge was to *justify* the claim that completing this task would yield a theory of meaning.

Coming up with this justification proved to be easier said than done. Consider again the use of a truth theory to *endow* sentences with meaning. Our announcement that we are using the theory to introduce an interpreted language contains a crucial piece of information not contained in the theory itself – namely, that certain of its theorems *are to be viewed as providing paraphrases* of the sentences the truth conditions of which they state. This suggests that if descriptive theories of meaning are to be put in the form of Tarskian truth theories, something beyond what they state must play a crucial role. Also, when we introduce interpreted formal languages, we typically don't have to choose which of the many theorems stating truth conditions of a single sentence provide *acceptable* paraphrases of it. Since potential paraphrases can often be proved to be extensionally equivalent, each is acceptable for the purposes of philosophical logic, or metamathematics. This is not true when our purpose is to give a descriptive theory of meaning. Thus, if a Tarskian truth theory is to fill the bill, it must be combined with something else that not only provides the information that meaning-

⁸ Alfred Tarski, "The Concept of Truth in Formalized Languages," and "On the Concept of Logical Consequence," in *Logic, Semantics, and Metamathematics*, 2nd. Edition, John Corcoran (Indianapolis: Hackett), 1983.

giving paraphrases are sought, but also specifies which of the many potential candidates are the genuine articles. This is the heart of the justificatory problem Davidson faced.

Initially, there was widespread optimism about its solution, together with widespread unclarity about what such a solution would require. The optimism was fueled by the attractiveness of the overall picture – which was seen as applying the proven advances of philosophical logic to the interpretation of natural language, without backsliding on Quine’s *au courant* skepticism about meaning. Davidson thought that systematic knowledge of truth and reference could do all legitimate work for which we need a notion of meaning. His strategy was to embrace Quine’s rejection of analyticity, synonymy, and our ordinary notion of meaning, substituting knowledge of truth and reference for it -- whenever there was something genuine to be captured. Since truth and reference are scientifically legitimate, such a theory was deemed respectable. Since it could be used to explain what it is to understand a language, it fit the emerging paradigm in linguistics. In short, was that one can have Quine, and Chomsky too.

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

*distinguished from what he terms the 'theory of meaning'. So much the good for what I call a theory of meaning, and so much, perhaps, against my so calling it.*⁹

There were, however, some conceptual flies in the ointment. The grand Davidsonian-cum-Quinean theme presupposed that truth and reference can be retained, while meaning is rejected. But it is not clear that our ordinary notions of truth and reference can be separated from our ordinary notion of meaning. Having rejected meaning as unscientific in *Word and Object*, and implicitly called ordinary reference into question by implicating it in his indeterminacy theses, Quine finished it off in "Ontological Relativity," calling for what was, in effect, its elimination.¹⁰ Surely, if the ordinary notion of an expression *referring* to a rabbit is to be eliminated, then the related notion of a predicate *being true of* a rabbit must also go. But this brings *truth* itself into play. How can one hold onto it, once one has abandoned its sister, *being true of*? For Quine, the question is moot, since he was willing to trade our ordinary notions of truth and reference for Tarski's disquotational replacements.¹¹

In the beginning, Davidson was too. Initially, he wrongly equated the notion of truth needed in his theories of meaning with Tarski-truth.¹² He was also a revisionist about reference.¹³ He didn't believe that his referential axioms stated facts about the world from which the truth conditions of sentences follow. On the contrary, these axioms had no independent content, and reflected no independent reality. Rather, they were seen as aspects of

⁹ Donald Davidson, "Truth and Meaning," p. 24, my emphasis.

¹⁰ W.V. Quine, *Word and Object* (Cambridge: MIT Press), 1960; "Ontological Relativity," in *Ontological Relativity and Other Essays* (New York: Columbia University Press), 1969.

¹¹ For discussion, see chapter 11 of *Philosophical Analysis in the Twentieth Century*, plus the final section of my paper, "The Indeterminacy of Translation and the Inscrutability of Reference," *Canadian Journal of Philosophy*, 29, 1999, 321-70.

¹² See my "Tarski's Theory of Truth," *Journal of Philosophy* 81, 1984, 411-29; and pp. 102-07, and 238-44 of my *Understanding Truth* (New York: Oxford University Press), 1999.

¹³ See, Donald Davidson, "'Reality Without Reference,'" *Dialectica*, 31, 1977, 247-53.

the total theory that derive their content entirely from their role in connecting theorems about the truth-conditions of sentences with one another. On this picture, one derives a statement of the truth conditions of S from statements about the reference of S's parts. The contents of these referential statements are abstracted from their role in deriving statements about the truth conditions of other sentences containing those parts. But since those other sentences contain additional words not in S, further referential axioms are required to derive theorems stating their truth conditions, thereby linking their interpretations to that of S. And so it goes, until the contents of every sentence and word are intertwined with, and dependent upon, the contents of every other sentence and word. *In the end, Davidson thought, our understanding any word or sentence is conceptually dependent on our understanding of every other word and sentence* – a radical version of meaning holism, akin to Quine's own.

We decided a while back not to assume that parts of sentences have meanings except in the ontologically neutral sense of making a systematic contribution to the meaning of the sentences in which they occur. Since postulating meanings has netted nothing, let us return to that insight. One direction in which it points is a certain holistic view of meaning. *If sentences depend for their meaning on their structure, and we understand the meaning of each item in the structure only as an abstraction from the totality of sentences in which it features, then we can give the meaning of any sentence (or word) only by giving the meaning of every sentence (and word) in the language.* Frege said that only in the context of a sentence does a word have meaning; in the same vein he might have added that only in the context of a language does a sentence (and therefore a word) have meaning.¹⁴

That was the grand philosophical canvass on which Davidson painted. Some of its main elements, like the idea that Tarskian truth predicates can be used for Davidson's purposes

¹⁴ Davidson, "Truth and Meaning," p.22, my emphasis.

were just errors to be recanted.¹⁵ Other parts – like Quine’s critique of analyticity and synonymy, and his indeterminacy theses – are vulnerable to powerful objections.¹⁶ Although the Davidsonian program continues to this day, much of the original philosophical background for it has fallen away. I will not, therefore, presuppose it in what follows. Instead, I will freely substitute the ordinary notion of truth for Tarski’s, I will not assume that intensional semantic notions are illegitimate, and I will not rely on any kind of semantic holism. My question is, *How, if at all, can one justify Davidson’s claim that theories of truth qualify as theories of meaning?* I will approach this question with as little philosophical baggage as possible.

The Problem of Justification

Davidson originally held that a truth theory for L qualifies as a theory of meaning, if knowledge of what it states is sufficient for understanding L. The problem was in showing that his theories satisfied the condition. How can knowledge of a truth theory be sufficient for understanding meaning, when its theorems give truth conditions of sentences only in the weak sense of pairing them with materially equivalent claims? If all I know about S is expressed by the theorem [‘S’ is true iff P], I can readily draw the conclusions expressed by [S doesn’t mean that \sim P] and [S doesn’t mean that Q], where the claim made by Q is obviously incompatible with that made by P. But how does one move from these modest negative results to interesting positive conclusions about what S does mean? Initially, Davidson thought that compositionality gave the answer. In compositional theories, theorems stating the truth conditions of sentences are derived from axioms interpreting their parts. Thus, he reasoned, “accidentally true”

¹⁵ See pp. 422-24 of my “Tarski’s Theory of Truth,” and pp. 102-07 of *Understanding Truth*. Also, D. Davidson, “The Structure and Content of Truth,” (The Dewey Lectures 1989), *Journal of Philosophy*, 87, 1990, 279-328.

¹⁶ See *Philosophical Analysis in the Twentieth Century*, chapters 16 and 17 of Vol. 1, chapters 11 and 12 of Vol. 2.

statements of truth conditions, like [‘Snow is white’ is true iff grass is green] will not be generated without simultaneously generating falsehoods, like [‘Snow is grass’ is true iff grass is grass] and [‘Trees are green’ is true iff trees are white]. Since truth theories must be *true* in order to qualify as theories of meaning, Davidson believed that the problem of “accidentally true” statements of truth conditions wouldn’t arise. Instead, he thought, truth theories that are both true and compositional will end up deriving only those statements ‘*S* is true iff *P*’ in which the sentence replacing ‘*P*’ is a close enough paraphrase of the sentence replacing ‘*S*’ that “nothing essential to the idea of meaning ... [would remain] to be captured.”¹⁷

John Foster showed him to be wrong.¹⁸ Let L_S be an extensional fragment of Spanish. Suppose one has a true, compositional truth theory T_1 of L_S that delivers a *translational T-theorem* -- [‘*S*’ is true in L_S iff *P*] in which *P* means the same as *S* – for each sentence of L_S . We now construct a new theory T_2 by replacing all axioms of T_1 interpreting a word, phrase, or sentence-forming construction with new axioms stating different, but extensionally equivalent, interpretations. Since T_1 is both true and compositional, so is T_2 -- despite the fact that *all* T-theorems of T_2 may, like (1), be nontranslational.

1. ‘Mis pantalones son verdes’ is true in L_S iff my pants are green and first-order arithmetic is incomplete.¹⁹

Knowledge of these theorems is not sufficient to understand L_S . So, T_2 can’t be a theory of meaning, even though it satisfies Davidson’s constraints. The problem remains, even if we assume constraints strong enough to rule out all but *translational truth theories* – defined as

¹⁷ Page 26, “Truth and Meaning.”

¹⁸ J.A. Foster, “Meaning and Truth Theory,” in *Truth and Meaning*.

¹⁹ Here, and throughout, I put aside complications caused by indexicals.

those that entail a translational T-theorem for each sentence of the language. Knowing what is stated by the translational truth theory T1 is no more helpful in coming to understand L_S than knowing what is stated by the nontranslational T2 -- *unless one also knows, of that which is stated by T1, that it is expressed by a translational theory*. If one wrongly thinks that T1 is nontranslational, then knowledge of the truth conditions it states won't yield knowledge of meaning. Thus, knowledge of what is stated by even the best truth theories is insufficient for understanding meaning.

Davidson's response was make the obvious minimal revision of his justificatory proposal. On the revised view, *what makes a translational truth theory TT a correct theory of meaning is that knowledge of the claim made by the conjunction of its axioms, plus knowledge, of that claim, that it is made by a translational theory are, together, sufficient for understanding L.*²⁰ The idea is this: (i) Knowledge of the conjunction of axioms of TT allows one to derive a translational T-theorem for each S, and thereby to pair S with *the claim expressed by a translation of S*. (ii) Knowledge, of this truth-conditional knowledge, that it is expressed by a *translational* theory, allows one to identify the claim paired with S as the claim *expressed by S*. In this way, one comes to understand the language.

However, this reasoning fails. For each sentence S, a truth theory will have infinitely many T-theorems [*'S' is true iff P*] among its consequences. If the theory is translational, at least one will be translational. But nothing in the theory specifies which one. Knowledge of a theory known to be translational doesn't allow one to separate translational from nontranslational theorems. Thus, it *doesn't* suffice for understanding the language.²¹

²⁰ Davidson, "Reply to Foster."

²¹ See my "Truth, Understanding, and Meaning."

The natural response is to add a definition of *canonical theorem* to truth theories, picking out, for each S, a unique T-theorem as translational. However, it is doubtful that this would provide the needed justification. Once information about canonicity is added, the only role played by knowledge of the canonical truth theorem CTT is that of allowing one to identify a claim in which S is paired with a certain content, which is then stipulated to be the content expressed by a translation. *Neither the truth of CTT, nor the fact that it states the truth conditions of S, plays any role in interpreting S.* All it does is supply a translational pairing, which could be supplied just as well in other ways. One could get the same interpretive results by replacing the truth predicate in a translational truth theory with *any arbitrary predicate F whatsoever*. Whether or not the resulting theory is true makes no difference. To interpret S, all one needs to know, of the canonical F-theorem, is that it links S with *the content expressed by a translation of S*. No one would conclude from this that translational F-theories count as theories of meaning. Why, then, suppose that translational truth theories do? So far, we have been given no answer.²²

The collapse of this justificatory attempt should not be lamented. Prompted by Foster's objection into invoking a notion paraphrase beyond anything in the truth theories themselves, we face the dilemma of appealing to a notion strong enough to overcome the objection, at the cost of robbing truth and reference of their cherished roles in explicating meaning, or of not overcoming the objection at all. If the justificatory enterprise is both to succeed, and be worth the candle, some way out of this dilemma must be found. Some believe that the answer lies in

²² This point should not be obscured by the fact that Davidson did not, in his reply to Foster, explicitly involve the notion of a *translational* truth theory. Instead, he invoked the idea of a truth theory that is known to satisfy a set of controversial constraints about what meaning and interpretation are, and what it is to verify theories of such. Unless satisfaction of these constraints guarantees translationality (which it would seem it does not), knowing that they are satisfied won't be sufficient to interpret sentences, and Davidson won't have an answer to Foster.

psycholinguistic speculation.²³ According to them, a translational T-theory counts as a correct theory of meaning because speakers unconsciously use it to interpret sentences. On this view, canonical T-theorems are statements [*'S'* is true iff *P*] that, as a matter of psycholinguistic fact, terminate the interpretive derivations of ordinary speakers.

There are several reasons to doubt this. First, it requires a robust “language of thought” distinct from the natural languages, used to state the truth conditions of natural language sentences. Presumably this system must include Mentalese counterparts of all natural-language predicates. But it is implausible to think that I have any way of linguistically representing the concept of a microwave oven other than by using the English expression. Second, such a theory simply passes the buck from explaining what it is to understand natural language to explaining what it is to understand Mentalese. Third, if natural language sentences are understood by interpreting them in antecedently understood Mentalese, then all that is required is translation from the former to the latter. Although a canonical truth theory, stated in Mentalese, could, in principle, provide this, it could not do so efficiently. The derivations of T-theorems are long and cumbersome. If all one wanted was a translation, no one would dream of using them over other, more efficient methods. Why then suppose that Nature foisted them on us? Fourth, there is no compelling psycholinguistic evidence I am aware of that supports this psychologizing of the Davidsonian project. Finally, even if, by the purest luck, it were to turn out that actual English speakers in fact used internalized Davidsonian truth theories as imagined, this wouldn't be a semantic fact about English, but a psychological quirk about us. If a new speaker came along, who assigned all English expressions precisely the interpretations we do, but used a different method for translating into Mentalese, he would still be an English

²³ Richard Larson and Gabriel Segal, *Knowledge of Meaning* (Cambridge: MIT Press), 1995.

speaker. Psychologizing the Davidsonian semantic program isn't a way of saving it; it's a way of killing it.

Higginbotham's Justificatory Idea: A First Approximation

That program remains one of the active approaches in semantics today – despite the crisis created by the so-far unsuccessful attempts to solve its justificatory problem. By far the most promising suggested solution that I know of derives from James Higginbotham's "Truth and Understanding."²⁴ Semantic theories, he says, should tell us what sentences, and the expressions that make them up, mean. He says:

The kind of meaning that a *sentence* has, however, is determined by what it may be used to *say*, and the kind of meaning that words and phrases have is determined by their contributions to the meanings of the sentences in which it occurs. It could therefore be proposed that semantic theory is charged with establishing, formally, all of the facts to the effect that so-and-so means such-and-such, or at least all such facts as come readily to the lips of native speakers, hoping in this way to clarify the nature and extent of the human capacity for language. I think that this simple answer is correct.²⁵

He warns us, however, not to expect theorems that mention meaning. Rather, he proposes to explicate meaning in terms of truth, reference, knowledge, and understanding. It is facts involving these notions that semantic theories are expected to account for.

The facts that semantics must account for comprise the context-independent features of the meaning of expressions that persons must know if they are to be competent speakers of the languages to which those features are assigned. What they must know, I suggest, consists of: facts about the reference of expressions, about what other people know and are expected to know about the reference of

²⁴ *Philosophical Studies*, 65, 1992, 3-16; reprinted in Mark Richard, ed. *Meaning*, (Oxford: Blackwell), 2003; citations are to the reprinted version.

²⁵ Pp. 255-6.

expressions, about what they know about what one is expected to know about the reference of expressions, and so on up. From this point of view, meaning does not reduce to reference, but knowledge of meaning reduces to norms of knowledge of reference. Such norms are *iterated*, because knowledge of meaning requires knowledge of what others know, including what they know about one's own knowledge.²⁶

He adds:

As a speaker of English, you are expected, for example, to know that 'snow is white' is true if and only if snow is white; to know that 'snow' refers to snow, and that 'is white' is true of just the white things...²⁷

Applying his point about iteration, we may add that, as a speaker of English you are also expected to know *that speakers of English are expected to know*: that 'snow is white' is true iff snow is white, that 'snow' refers to snow, and that 'is white' is true of just the white things. If, and only if, you know all such things, the idea goes, you understand English.

The plausibility of this idea is illustrated by what it tells us about an earlier example. Higginbotham would say: if you know not only (i) *that 'mis pantelones son verdes' is true in Spanish iff my pants are green*, but also (ii) *that necessarily, one who understands that sentence knows (i)*, and so on for further iterations, then you know that the sentence means that my pants are green, *and not that my pants are green and arithmetic is incomplete*. Repeating this for every sentence should, he suggests, be sufficient for understanding the language.

In order to test this idea, as a defense of Davidson, we must make it precise. The idea is that a theory the theorems of which give the contents of linguistic norms, iterated knowledge of which is necessary and sufficient for understanding L, qualifies as a theory of meaning of L.

²⁶ P. 257.

²⁷ P. 257.

Let C1 be a first approximation of this claim (understanding a theorem to be any logical consequence of T).

- C1 Any theory T satisfying (a) and (b) is an acceptable theory of meaning for L.
- (a) For each theorem T_T of T, (i) knowledge of T_T is necessary for understanding L, (ii) knowledge that *knowledge of T_T is necessary for understanding L* is necessary for understanding L, and so on for further iterations.²⁸
- (b) The knowledge specified in (a) is sufficient for understanding L.

The justificatory argument is completed by C2.

- C2. Translational, Davidsonian truth theories satisfy C1.

Why This First Approximation Won't Do.

However, C2 won't do. Many theorems of translational truth theories – both axioms and their logical consequences -- fail to satisfy C1(a). Let L be a fragment of English corresponding to the first-order predicate calculus. The best candidates for theorems satisfying C1(a) are the referential axioms for names and predicates, plus those for the truth-functional connectives. However, even here the case isn't clear. Suppose L doesn't contain any *reference* predicate relating words and things, any *knowledge* predicate relating agents and propositions, any devices for referring to expressions, any expression designating L, any *understanding* predicate holding between agents and languages, or any *necessity* operator. Such a language, though primitive, surely could be spoken – presumably as a first language. It is debatable whether an individual's mastery of L would be sufficient to credit him or her with implicit knowledge of what its terms refer to, and what its predicates are true of. But even if it is sufficient, what of the iterations? Must a native speaker of L know *that in order for anyone to*

²⁸ (i) and (ii) may equivalently be put: (i) it is required, and hence necessary, that one who understands the sentences and other expressions of L (with the meanings they actually have) knows T_T , (ii) it is required, and hence necessary, that one who understands the sentences and other expressions of L (with the meanings they actually have) knows (i).

understand L it is required, and hence, necessary that one know that 'Plato' refers to Plato and that 'is human' is true of o iff o is human – even though he or she may have no way of expressing the concepts that make up this alleged knowledge? It is not obvious that this is necessary. It is one thing to credit a speaker with implicit knowledge guiding his or her use of language. It is quite another to credit the speaker with implicit knowledge attributing this implicit knowledge to others. Thus, it is doubtful that the referential axioms satisfy C1(a).²⁹

Matters get worse for Tarski's quantificational axioms, and for his definition of truth in terms of *truth relative to an assignment*. Do all L-speakers who understand its quantificational sentences know these things? *Do they also know that in order to speak L one must know them, and hence that every L-speaker does? Do they, still further, know that every speaker knows the previous iterated claim?* It hardly seems likely. I understand L, yet I don't know that every speaker of L knows the Tarski's quantificational axioms -- let alone that they know that I know the very thing I take myself not to know. These consequences of C1 and C2 are incredible.

Next consider theorems that aren't axioms. Let T/R be a translational theorem that specifies the truth (or reference) conditions of a sentence S (or expression E) in terms of a strict paraphrase of S (or E). Putting aside the question of whether translational theorems like this satisfy condition C1(a), we have two nontranslational cases to consider. In case 1, we let T_{Early} be any theorem that appears on a line earlier in the derivation of T/R than the line on which TR appears. It's not plausible that T_{Early} must satisfy C1(a). The only grounds for thinking that it must are grounds for thinking (i) that speakers actually employ this derivation to generate T/R, which they use to interpret S, or E, (ii) that every speaker knows that other speakers do the same, and would refuse to count anyone as knowing L whose understanding of S, or E, came

²⁹ Thanks to William Dunaway for discussion of points in the previous paragraph.

about in some other way, and (iii) that every speaker knows that every speaker knows this. As before, such results are incredible. Although I understand L, I don't think that I have the knowledge of other speakers, or that they have the knowledge of me, needed for the condition to be satisfied. In case 2, we let T_{Extra} be a theorem of the truth theory that is neither translational, nor one needed to derive any translational theorem. Surely there is no reason to think that it satisfies C1(a). Note, it doesn't help to claim that T_{Extra} won't count as "canonical." Since C1 and C2 are stated in terms of simple theoremhood, their conjunction is refuted.

Reformulating the Idea

Perhaps we can do better by reformulating the justificatory claim as RC1.

RC1 A theory T which identifies some canonical subset SubT of its theorems, and correctly says that it satisfies (a) and (b), is an acceptable theory of meaning for L.

- (a) For each member T_T of SubT, (i) knowledge of T_T is necessary for understanding L, (ii) knowledge that *knowledge of T_T is necessary for understanding L* is necessary for understanding L, and so on for further iterations.
- (b) The knowledge specified in (a) is sufficient for understanding L.

RC2. Properly augmented translational theories of truth and reference satisfy RC1.

This revision avoids our earlier problems by excluding some of the problematic theorems.

Doing this requires adding new theoretical machinery to the truth theory defining canonical theorems, and specifying, via the empirical claim T_C , the work they are supposed to do.

T_C The class of *canonical* theorems of T satisfies (a) and (b) of RC1.

This expansion T_+ of the truth theory T is what is claimed to be the theory of meaning for L.

Davidsonian theories of truth and reference alone are *not* enough. Instead, the theories that genuinely explain meaning contain Davidsonian theories as parts, while making further claims about knowledge and understanding. Since questions about what counts as understanding sentences and other expressions are closely tied to questions about their meanings -- over and

above their truth and reference conditions – the exciting initial thought that Davidsonian theories would explicate meaning in wholly extensional terms has gone by the boards. This need not be an objection. But it is a fact.

Evaluating the Expanded Idea: Why We Still Don't Have a Justification

There is, however, an objection in the wings. Suppose that both (2) and (3) are theorems of an expanded, translational truth theory T_+ of a fragment L of English -- as well they might.

2. '2 = 2 and if $n = 2$, there are numbers x, y, z such that $x^n + y^n = z^n$, but if $n > 2$, there are no such numbers' is true in L iff *2 = 2 and if $n = 2$, there are numbers x, y, z such that $x^n + y^n = z^n$, but if $n > 2$, there are no such numbers.*
3. '2 = 2 and if $n = 2$, there are numbers x, y, z such that $x^n + y^n = z^n$, but if $n > 2$, there are no such numbers' is true in L iff *if $n = 2$, there are numbers x, y, z such that $x^n + y^n = z^n$, but if $n > 2$, there are no such numbers.*

If (2) and (3) are *both* canonical, then it will be left unclear whether the quoted sentence means that which is expressed by the right-hand side of (2), or that which is expressed by the right side of (3). If, following Higginbotham, we take *what is said* by an utterance of S to be a good guide to the meaning of S , then the right-hand sides of (2) and (3) *will*, it seems, differ in meaning -- since an assertive utterance of the former will result in an assertion of the claim expressed by the first conjunct (that the number 2 = the number 2), whereas an assertive utterance of the latter will not. Given that the quoted sentence presumably doesn't mean both, we conclude that T_+ fails to satisfy clause (b) of RC1, and, so, is false (assuming that one who doesn't know the meaning of the quoted sentence doesn't understand it).

The same failure occurs, even if *only* (2) is counted as canonical, so long as *both* (2) and (3) satisfy (a) of RC1. If they do, then appeal to this condition won't determine whether *what the quoted sentence means* is given by the right side of (2), or the right side of (3). Thus, RC1(b) will be in the same jeopardy as before. Before, T_+ claimed that (2) and (3) both satisfy RC1(a). Here, T_+ says this about (2), while remaining silent about (3) – which, we are

presently assuming, does satisfy the condition. Since this silence doesn't entail anything about the meaning of the quoted sentence, T+ is thrown into the same doubt as before.

The difficulty presented by these sentences can be avoided only if we can show that although (2) satisfies RC1(a), (3) doesn't. But this is doubtful. To understand the quoted sentence, one must be familiar with the number 2, and know what it is to identify it with a number n – in which case the first conjunct will be superfluous, since, in that case, one can't help knowing that $2 = 2$. Thus, knowledge of (2) and (3) *will* go hand in hand -- in which case knowledge of (3) will be necessary for understanding L, if knowledge of (2) is. Still, it might be objected, even if knowledge of (2) and (3) *do* necessarily go hand in hand for competent speakers, *knowledge that* knowledge of (2) and (3) necessarily go hand in hand might *not* be required for competence. All it takes for this iterated knowledge claim to fail is for there to be a misguided philosopher who – though himself a competent speaker of L who knows both (2) and (3) – doubts, and so does *not* know, that this *must* be true of all competent speakers. Given the nearly unbounded reach of such possible doubt, we can't rule this out – which means that we *can't*, in the end, be sure that (3) does satisfy condition RC1(a).

Although this may seem all to the good, it will vindicate T+ as a theory of meaning only if *we can be sure* that (2) does satisfy the condition. Can we be sure of that? Consider mathematical nonfactualists who accept Fermat's last theorem while denying that any mathematical statements are true, and so reject (2). Though I take such philosophers to be mistaken about the scope of the truth predicate, I don't doubt that they understand the sentence quoted in (2), even though they don't believe, and hence don't know, (2).³⁰ As a result, *I don't*

³⁰ See Mark Richard, "Deflating Truth," *Philosophical Issues* 8, 1997, 57-78 for a defense of the philosophical coherence (if not correctness) of such nonfactualists. My response is found on pp. 88-93 of that volume. For further discussion, see my "Understanding Deflationism," *Philosophical Perspectives* 17, 2003, 369-383.

believe, and so I don't know, that knowledge of (2) is necessary for understanding L. Surely this doesn't mean that *I'm* not a competent speaker, or that *I* don't understand the quoted sentence. Hence, our final attempt to justify Davidsonianism is unconvincing. The justificatory problem thus remains unsolved.

The Disconnect Between Theory and Practice

Earlier, I noted that in "Truth and Meaning," Davidson responded to the worry that, since the theorems of his truth theories are only material biconditionals, his theories might issue in true but grotesquely nontranslational theorems like

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

and thereby fail to count as theories of meaning. As I indicated, his response was heroic.

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.

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 us 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.*³¹

³¹ P. 26, "Truth and Meaning," my emphasis.

This refusal to appeal to an antecedently understood notion of meaning, synonymy, or translation to constrain acceptable theories of truth -- or to justify taking them to be theories of meaning -- was, in effect, quietly abandoned nine years later, in the wake of Foster's objection. By that time, however, the original methodological refusal had become securely embedded in a procession of empirical analyses of linguistic phenomena advancing the Davidsonian program.

For example, in 1968, just one year after "Truth and Meaning," Davidson analyzed

4a. Galileo said that the earth moves

as having the structure

4b. Galileo said that: The earth moves

in which 'that' is a demonstrative used by the speaker to refer not to *what Galileo said*, nor, of course, to Galileo's utterance, but to *the speaker's utterance* of 'The earth moves.' On this picture, the theory of truth, cum theory of meaning, issues in a theorem along the lines of (4c).³²

4c. An utterance, by x, of 'Galileo said that the earth moves', containing as a subpart x's utterance u of 'the earth moves', is true iff some utterance of Galileo and x's utterance u make Galileo and x samesayers.

Over the years, much has been said for and against this analysis. Although it inspired a progression of increasingly sophisticated successors, it is, I think, fair to say that no one today stands by it in its original form. The point, however, is not to rehearse its shortcomings, but to note how it fits Davidson's admonition in "Truth and Meaning" *not* to construe the right-hand sides of the T-theorems of his interpretive theories as "pretending" to capture the meanings of that which appear on the left. This point is illustrated by (4c), since one can know that which is said by my utterance of (4a) without knowing anything about me, or any utterance of mine, and

³² Donald Davidson, "On Saying That," *Synthese*, 19, 1968-9, pp. 130-146.

since what is said by my utterance *u* could have been true, even if *u* had had a different content, or if neither it, nor I, had existed. Given these intensional and hyperintensional differences, one cannot regard (4c) as translational in any interesting sense. Although this may have seemed acceptable according to the justificatory picture sketched in “Truth and Meaning,” it is not acceptable according to post-Foster attempts at justification – all of which rely, in one way or another, on translational theorems. In changing the justification of the program, Davidson, in effect, narrowed the class of empirical analyses capable of advancing it.

This lesson has yet to be learned. Even though the justificatory story has changed -- imposing strong constraints on *canonical* theorems -- empirical analyses offered to advance the program often don't take these constraints seriously. A recent analysis of propositional attitudes by Richard Larson and Peter Ludlow is a case in point.³³ The proposal, which is arguably the most sophisticated Davidsonian successor of the one in “On Saying That,” uses the rules for deriving T-theorems to assign annotated phrase structure trees, called “interpreted logical forms,” to sentences. For example, (5b) is the interpreted logical form assigned to (5a).

5a. John speaks Spanish.

5b. <S, truth>

<NP, John>

<VP, John>

<V, John, Spanish>

<NP, Spanish>

<'John', John>

<'speaks', <John, Spanish> >

<'Spanish' Spanish>

These interpreted logical forms are taken to be the objects of attitude verbs. Thus, (6) is said to be true iff Mary asserts / believes (5b).

6. Mary asserts / believes that John speaks Spanish.

³³ “Interpreted Logical Forms,” *Synthese* 95, 1993, 305-356.

Spelled out in words, the T-theorem interpreting (6) amounts to (7).

7. 'Mary asserts/believes that John speaks Spanish' is true iff Mary asserts / believes the interpreted phrase marker whose root node is the pair <'S', truth>, which dominates a pair of nodes <'NP', John> and <'VP', John>, where <'NP', John> dominates <'John', John>, and <'VP', John> dominates a pair of nodes <'V' <John, Spanish>>, and <'NP', Spanish>, with the first of these nodes dominating <'speaks', <John, Spanish>> and the second dominating <'Spanish', Spanish>.

Is (7) true? It is hard to say. Since its right-hand side is a theoretical claim we have no pre-theoretic grasp of, it's not easy to judge the truth or falsity of (7), without already having accepted the theory. This is awkward, since it is by evaluating such theorems that we are supposed to test the theory itself.³⁴ Suppose, for the sake of argument, that (7) is true. Although this would support the analysis, *as a theory of truth*, it would not vindicate it *as a theory of meaning*. Theorems like (7) don't supply paraphrases for the sentences they purport to interpret, and so don't conform to post-Foster attempts to justify the Davidsonian program. Nor, in my opinion, do they conform to any other viable justificatory approach.³⁵ All too often, truth-theoretic analyses are offered, even though no attempts to justify their claim to be theories of meaning have been successful. Even worse, the most sophisticated empirical analyses are sometimes inconsistent with the most sophisticated of the justificatory attempts. This is not a healthy state for what purports to be an empirically viable theory of meaning to be in.

What is the Alternative?

Given the problems of both the justification and execution of the Davidsonian program, we would do well to consider an alternative. The justificatory problem arose directly from

³⁴Elsewhere, I have made suggestions about how to extract empirically testable claims from such theorems, and hence come to a judgment about their truth or falsity. See pp. 147-159 of *Beyond Rigidity*, (New York: Oxford University Press), 2002; and "Truth and Meaning: The Role of Truth in the Semantics of Propositional Attitude Ascriptions," in *Proceedings of the 7th International Colloquium on Cognitive Science*, edited by Kępa Korta and Jesus M. Larrazabal, (Dordrecht: Kluwer), 2003, 21-44.

³⁵ Larson and Ludlow appear to adopt a version of the psycholinguistic justification discussed earlier.

Davidson's initial conviction that theories of meaning mustn't talk about meaning. Since his truth theories make no statements about what sentences mean, justifying their use as theories of meaning has always been a challenge. Theories that do state informative truths about what the sentences of a language mean are a natural alternative. By this, I don't mean theories that derive an instance of (8) for each sentence S of a language, from axioms about the parts of S .

8. 'S' is true in L iff S.

No one knows how to construct theories that do that in an illuminating way. Rather, I mean theories that recursively assign certain entities to sentences – identified as their meanings -- on the basis of their semantically significant structure.

The idea, which goes back to Frege and Russell, was rejected by Davidson 41 years ago in the first few pages of "Truth and Meaning."

"Up to here we have been following in Frege's footsteps...But now, I would like to suggest, we have reached an impasse: the switch from reference to meaning leads to no useful account of how the meanings of sentences depend on the meanings of the words ... that compose them. Ask, for example, for the meaning of 'Theaetetus flies'. A Fregean answer might go something like this: given the meaning of the name 'Theaetetus' as argument, the meaning of 'flies' yields the meaning of 'Theaetetus flies' as value. That vacuity of the answer is obvious. We wanted to know what the meaning of 'Theaetetus flies' is; it is no progress to be told that it is the meaning of 'Theaetetus flies'."³⁶

"The contrast here between a real and pretended account will be plainer still if we ask for 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

³⁶ "Truth and Meaning," pp. 20-21.

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. [This is what he sees no prospect of. Thus he concludes]... Paradoxically, the one thing meanings do not seem to do is oil the wheels of a theory of meaning – at least as long as we require of such a theory that it non-trivially give the meaning of every sentence in the language. 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.’³⁷

Davidson’s objection to meanings as entities is that they can’t be used to *non-trivially give the meanings of sentences*, or to play any useful role in theories of meaning. The objection, though not entirely without force, is not true. Combining the ideas of Russell and Tarski, we can recursively assign structured Russellian propositions – the constituents of which are objects and properties – to all sentences of a language. Adding a theory of truth for propositions, gives us a theory that specifies the truth conditions of sentences, identifies which are synonymous with which, provides a natural account of attitude ascriptions, and lays the foundation for a theory of the assertions made, and the beliefs expressed, by sincere assertive utterances. Like all empirical theories, this one is underdetermined by the data for it. However, since its claims are testable, the theory is capable of empirical confirmation or disconfirmation.

But does it, one may ask, really *give us the meanings of sentences*? Yes and no. To take the simplest sort of example, we may suppose that it tells us the meaning of (9a) is (9b).

- 9a. A is larger than B.
- b. < the relation of being larger than, <a, b>>

If the theory is correct, then the theorem pairing it with (9b) is a true theoretical description of the meaning of (9a). However, *this way* of giving us the meaning of (9a) is not one that would

³⁷ Ibid., pp. 21-22.

allow us to come understand it, if we didn't already. One might dismiss this as unimportant, since as long as the meanings of sentences are correctly identified, the fact that they aren't presented in a way suitable for language learning means only that the theory is no replacement for the language lab. Since our interest in semantic theories is theoretical, not pedagogical, this is no loss. However, there is a difficulty here that goes much deeper. What is (9b), after all, but a simple set-theoretic structure, the standard set-theoretic expansion of which is (9c)?

9c. {the relation of being larger than, {the relation of being larger than, {a, {a,b}}}}

How could this structure be the meaning of anything, let alone (9a)? There is nothing in it to indicate that the relation *larger than* is being predicated of anything, or, if it is, what exactly it is predicated of. Does (9c) represent a as being larger than b? Does it represent b as being larger than a? Does it represent a as being larger than a? Or is (5c) not representational at all? Surely, there is nothing in this set-theoretic structure by virtue of which it represents anything as being one way rather than another. But if it's not representational, then it doesn't have truth conditions, in which case, it can't be the meaning of any sentence.

This, I suspect, is what lies behind Davidson's worry about propositions, and his dismissive remarks about theories of meaning that invoke them. Since, any other abstract structure that we can identify and make precise, will be similarly non-representational, the problem can't be solved by selecting any such structure as the meaning of (9a). One could, of course, take propositions to be inherently and intrinsically representational, and so *sui generis*. However, this is a council of despair. Davidson wouldn't accept such obscuritism, and we shouldn't either. If we posit structured propositions as meanings of sentences, we ought to explain what they are, and how they are able to play the roles we assign to them.

The only way to do this is, I believe, to acknowledge that propositions aren't intrinsically representational. Here's the idea. We retain the conception of propositions as structured complexes, the constituents of which are objects and properties. To say that certain constituents make up a complex is to say that, in the complex, the constituents stand in certain relations to one another. The complex is, in effect, *the standing of the constituents in those relations*. What these relations are depend on the specific abstract structures we take propositions to be. Which structures these are doesn't matter. Suffice it to say, for our purposes, that the proposition expressed by (9a) is a complex in which a, b and the larger-than relation stand in a certain relation R. How does it come about that this entity – a's and b's standing in R to *larger than* -- represents a as being larger than b? The answer rests not on anything intrinsic to R, but on the *interpretation* placed on R by the way that we use it.

Though abstractly expressed, the idea is commonplace. Take maps, for example. On my map, the dot labeled 'Los Angeles' is (roughly) two inches below and $\frac{1}{2}$ an inch to the right of the dot labeled 'San Francisco'. The standing of these dots in this spatial relation on the map *represents* the city Los Angeles as being (roughly) 320 miles south and 80 miles east of the city San Francisco. It does so, in part, because of the interpretation we give to the relation *being two inches below and half an inch to the right of* on the map. This is the kind of interpretation we give the propositional relation R, in interpreting the complex in which a and b, stand in R to *larger than*. In both cases -- the map and the proposition -- our interpretation of a relation that the constituents of a structure stand in is what endows the structure with representational properties, and hence, truth conditions. A proposition, like a map, is something we interpret.

This idea comes from *The Tractatus*.³⁸ There, propositions are taken to be sentences that we use a certain way. Sentences are complexes in which words and phrases stand in certain structural relations. For example, sentence (9a) is the standing of the name ‘A’ to the name ‘B’ in a certain grammatical relation involving the predicate ‘is larger than’. Call this relation R_G . According to Wittgenstein, we interpret R_G as standing for the predication of the *larger than* relation of the objects represented by its arguments, thereby bringing it about that (9a) represents a as being larger than b.

At this point, it may be objected that propositions have dropped out of the picture. But they need not. We still need something to play the roles of the meaning of synonymous sentences, of what we assert by uttering such sentences, and of what those who accept such sentences believe. The sentences in question may differ in vocabulary, and in some aspects of superficial syntactic structure. However, if they express the same proposition, then utterances of them “say the same thing,” and express the same belief. To assert a proposition is to assertively utter, inscribe, or produce some representation that expresses it. A similar point holds for beliefs and other (non-perceptual) propositional attitudes. To bear such an attitude toward a proposition is to bear a more basic relation to a propositional vehicle that expresses it.

What, on this account, is it for an abstract structure – like (9b/c) – to count as the proposition that a is larger than b? It is for us to use the structure to predicate *larger than* of a and b. What is it for us to use the structure in that way? It is, very roughly, for us to use the grammatical structure of some sentence or other representation, the semantic contents of the

³⁸ Ludwig, Wittgenstein, *Tractatus Logico Philosophicus* (London: Routledge and Kegan Paul), 1922. For a brief discussion of Wittgenstein’s theory of propositions (interpreted sentences), see pp. 215-16 of volume 1 of my *Philosophical Analysis in the Twentieth Century* (Princeton and Oxford: Princeton University Press), 2003.

constituents of which are a, b, and *larger-than*, to predicate the latter of the former. In these cases, the representational properties of propositions are grounded in, and explained by, the representational properties of sentences, *not the other way around*.

The picture is complicated by the fact that there is, I think, one kind of propositional attitude we bear to propositions that isn't mediated by representations that expresses them. The attitude involves perception. When I see an object *o as being red*, I see both *o* and the color, which is a kind of property. Since perception is a form of cognition, my perceptual experience involves my predicating the color of the object. I don't, by virtue of this cognitive activity, thereby *see* the proposition that *o* is red. However, since the proposition is part of the content of my perceptual state, I do come to bear a propositional attitude toward it. What counts as bearing this attitude toward a proposition in which redness is predicated of *o* is simply that my perceptual experience involves this predication. In both the perceptual and the linguistic case, the explanation of what is predicated of what in the proposition bottoms out in predication as a cognitive activity of agents – in the one case, in the way agents interpret different *perceived* propositional constituents, in the other case in the way they interpret linguistic representatives of those constituents.

Many details of this story remain to be filled in. The task of doing so has a constructive part and a foundational part. In the constructive part we use propositions as theoretical constructs in linguistic and cognitive theories, and subject those theories to empirical test. In the foundational part, we explain what propositions are, how they acquire their representational properties, and how we are related to them. Since the tasks run in tandem, advances in one need not wait on progress in the other. This, it seems to me, is the most promising alternative to the Davidsonian approach to semantics.