

## REFERENCE: SOME RECENT PHILOSOPHICAL ISSUES

Jaegwon Kim

A *semantical theory* for a language is intended to represent what one must *know* to be able to interpret utterances of that language. What is the form such a theory must take, and what information must it provide about the language and its speakers? These are among the basic problems debated in the current Anglo-American philosophy of language. The aim of the present paper is to sketch the outlines of two influential approaches to the nature of semantical theory, namely the "building block" theory and the "holistic theory," and clarify the philosophical and methodological motivation for the holistic approach. The paper is basically an introductory survey of the current debate in this area, with a special focus on Donald Davidson's work.

When you know how to interpret, or understand, the sentences of a language, you have a certain *cognitive* capacity; there is something that you know that someone who does not know how to interpret that language does not know. Not all of this knowledge may be capturable in propositional form, that is, in a *theory*. However, an important core of this knowledge may be so capturable; in this paper we assume that this is so. Let a *theory of meaning* or *semantical theory* about a language be a theory such that knowing it suffices for the interpretation and understanding of the utterances of that language. That is, anyone who knows this theory can interpret the meaning of each and every sentence of this language. We also assume that the speaker of this language "knows" this theory, at least implicitly—in the following rough sense: we can explain his speech by positing that he knows this theory.<sup>1</sup>

These points place heavy demands on theories of meaning. This is because the cognitive ability alluded to that is connected with "knowing a language" involves not only the ability to recognize grammatical sentences but knowing how to use that language to describe facts and communicate with other speakers of the language.<sup>2</sup> Constructing such theories for particular languages is of course an empirical-theoretical task of the special sciences. As philosophers, we are interested in the *form* and *methodology* of such theories, for what counts as *understanding* a

<sup>1</sup> This is essentially Donald Davidson's notion of "semantical theory." The present paper is intended, largely, as an introduction to his philosophy of language, which is currently very influential in the United States and Great Britain. Davidson's principal papers are available in a useful recent collection, *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984). See especially "Truth and Meaning," "Semantics for Natural Languages," "Radical Interpretation," "Thought and Talk," and "Reality without Reference." In preparing this paper I am indebted to Bradford Petrie.

<sup>2</sup> Perhaps, there can be interpreters who are not speakers of a language, namely interpreters who are not full-fledged communicators, although of course all communicators must be interpreters. The discussion in this paper is not materially affected if we explicitly confine ourselves to interpreters.

language, or what counts as getting a semantical theory *right*, is a conceptual and methodological issue, not an empirical one.

It is perhaps useful to stress that a semantical theory in this sense for a given language must not presuppose the speaker's ability to interpret another language. That would only prompt the further question of what his ability to interpret this other language consists in, and this is precisely our original question, namely, what it is to understand, or be able to interpret, a language. The question, therefore, concerns *radical interpretation*—that is, interpretation of a language that does not presuppose any other language to have already been interpreted.

As I said, semantical theories about particular languages are empirical theories, and should therefore be verifiable on the basis of empirical evidence. But there is a further condition. This evidence must be *describable without the use of semantic concept*, for the epistemic situation in which our would-be interpreter is placed is one of radical interpretation, and we cannot assume that any semantical facts about the language are already available to him. Furthermore, the empirical evidence that confirms a semantical theory of a language must be *the kind of evidence that we can assume to be available to the interpreter and, therefore, to the typical speakers of the language*, because it's in virtue of learning and internalizing this theory that they are able to use the language, to interpret and communicate. The learning process requires that they be able to select a correct interpretive theory for the language on evidence that is available to them; and using the language requires that they be able to apply it to particular interpretive situations and be able to tell whether or not their interpretations are successful. The kind of evidence involved here is precisely the evidence that confirms, or disconfirms, semantical theories in the present sense.

To summarize, there are two preliminary conditions characterizing a semantical theory T for language L: (1) anyone who knows T knows how to interpret every sentence of L; (2) T must be verifiable on empirical evidence describable without the use of semantic concepts and which, moreover, is available to typical speakers of L.

It is generally, though by no means universally, agreed that T must generate a *truth condition* for every sentence of language L<sup>3</sup>. (We are supposing for simplicity that L consists only of assertoric sentences, not interrogatives or imperatives; but obviously assertoric sentences play a fundamental role in any language. It is clear that imperatives and interrogatives are intimately tied to certain assertive sentences, imperatives with their satisfaction conditions and interrogatives with their presuppositions.) What is a truth condition? The truth condition of a sentence S is a condition c such that c obtains iff S (as a sentence of L) is true. We can complicate this to take into account tenses, demonstratives, etc., but we can ignore these complications here. It is thought that knowing the truth condition of an utterance is essential to understanding what the utterance means, and how the utterance succeeds in communicating certain information.

<sup>3</sup> See Davidson's "Truth and Meaning."

We need not here discuss whether the generation of a truth condition for each and every sentence of L is *sufficient* for a semantical theory of L. Donald Davidson has forcefully argued that it is sufficient, provided that the truth condition generation meets a further set of constraints. On the other hand, many philosophers have argued against this. Among the most influential critics of the Davidsonian semantical program is Michael Dummett.<sup>4</sup> It could be argued, for example, that a truth-condition semantics of the sort Davidson has in mind can at best give an account of “referential meaning,” and that it must be supplemented by two things: (1) a theory of Fregean “sense,” and (2) a theory of “pragmatic force.” These complications again do not concern us, as long as we agree, provisionally, that truth-condition semantics is a necessary ingredient in any complete semantical theory. It is the theory that connects language with the world, and hence explains how by the use of language we are able to communicate information about the world.

But how do we generate truth conditions for sentences of L? Listing for each sentence its truth condition will not do, for there is a potential infinitude of sentences for any language of the sort we are interested in, and this makes any such task an impossible one to complete. And even if it could be completed, it is not something that can be learned by any speaker, and this means that no such theory could give an account of how in virtue of having internalized such a theory a speaker can use the language. But there is another, perhaps more important, reason why such a list will not do as a semantical theory. By listing truth conditions for whole distinct sentences, it will not be able to explain how components of a sentence contribute to the meaning, or truth condition, of the whole sentence. Thus, such a theory will have nothing to say about what it is that “snow” contributes to the truth conditions of both “Snow is white” and “Snow is cold,” or how “Snow is white and cold” entails “Snow is cold.”

This means that a semantical theory must generate truth conditions for an infinitude of sentences from a finite base by iterated applications of a finite set of rules. That is, it must be a *finite* theory. Further, the finite base of the theory must begin with interpretations of sub-sentential components—“words.” Truth conditions of “simple” sentences must be generated from the semantical interpretations of these sub-sentential elements, and those for ever more complex sentences must be generated by iterated applications of “composition rules.” Only such a theory can throw light on how “meanings” or “interpretations” for sentences are constructed out of simpler semantical units, and thus explain how an interpreter is able to assign meaning to every sentence of this language.

Let us call a sentence of the following form a “T-sentence”:

S is true if and only if p,

where “p” is a place-holder for a sentence (of the metalanguage) and S is a name (“structural description”) of a sentence in L, our target language. A T-sentence

<sup>4</sup> See e.g. Dummett’s “What is a Theory of Meaning?” in Samuel Guttenplan, ed., *Mind and Language* (Oxford: Oxford University Press, 1975).

is plausibly viewed as giving the truth condition of sentence S. Thus, the program of generating a truth condition for every sentence S of language L becomes that of appropriately deriving a T-sentence for S from the axioms of the semantical theory for L.

Obviously, the place to begin is the atomic sentences of L, that is, sentence of the form “a is F” or “a is R to b” which contain no sentences as proper parts. A natural suggestion is to introduce the relation of “refers to,” and formulate a general truth condition for atomic sentences as follows:

“a is F” is true iff “a” refers to object A, “F” refers to property f, and A has f (or the object “a” refers to has the property “F” refers to).

And similarly for atomic sentences involving relational predicates. If L is an extensional first-order language we know how to proceed from here in the usual way. If modalities and demonstratives are involved, things become more complicated, but in familiar ways; there are other recalcitrant constructions to worry about, such as mass terms, tenses, reference to events, etc., and especially indirect discourse involving psychological verbs, such as “believe,” “want,” and “hope.” Much work has been done to overcome these obstacles; but whether or not, and to what extent, this Davidsonian program of truth-conditional semantics will meet with total success is not our present concern.

Our concern centers on the unanalyzed notion of “reference” used in the foregoing statement of truth conditions for atomic sentences. Presumably, a semantical theory incorporating it must begin with a series of axioms of the following form:

“a” refers to object A

“F” refers to property f,...

It is this relation of reference (“naming,” “designation,” etc.), it may seem to us, that connects our basic terms with extralinguistic things—physical objects, properties, functions, etc. It is here, it seems, that our language makes contact with the world, and ultimately it is in virtue of this relation that by the use of language we manage to say something *about* the world. Thus, reference seems to be our fundamental semantical relation, one that anchors language to reality. And for truth-conditional semantics, the relation seems essential to a proper generation of truth conditions.

But it also seems that we cannot leave this relation unanalyzed, because if our semantic theory is to capture the cognitive capacity of an interpreter, we may not simply assume that these basic semantical facts involving reference are available to him. The theory is a radical interpretive theory, and we want it to explain what it is to know that a term refers to an object, and how the interpreter can know this on the basis of evidence available to him. In order to meet our second constraint on semantical theories, therefore, it appears that reference must be analyzed or explicated in such a way that the evidential basis of our semantical theory is free of the threat of circularity.

Hartry Field has argued for the need for an analysis of the concept of reference.<sup>5</sup>

<sup>5</sup> In his “Tarski’s Theory of Truth,” *Journal of Philosophy* 69 (1972): 347-75. This is an influential paper advocating the “building block” approach.

He argues that there are reasons for accepting the following two assumptions; (1) semantical facts are not among the *ultimate* facts of the world; in fact, there are no semantical facts over and beyond the fundamental physical facts of the world; and (2) semantical relations are "real" relations, and entities posited by (correct) semantical theory are among the real entities of the world. The first is the assumption of "Physicalism"; the second is the assumption of "Semantical Realism." From these two assumptions, it seems to follow that referential relations cannot be facts above and beyond the physical facts of the world, and that they must be reducible to a physical base.

Thus, Field believes that the referential relation must be physically analyzed. If we have qualms about his commitment to physicalism, we can state this reductive position in the following way: the referential relation cannot be taken as an unanalyzed primitive—it must be given an analysis that does not make use of other semantic concepts. The idea is therefore this: our semantical theory begins with a (nonsemantical) analysis of the referential relation, and then develops essentially along the lines of Tarski's celebrated construction of a truth definition, supplemented with special constructions to handle demonstratives, tenses, indirect discourse, etc.

This approach is now often referred to as the "building block" theory for obvious reasons. It is at least one initially plausible way to look at semantical theory, at least truth-conditional semantics.

But what are the prospects for an acceptable analysis of the referential relation? They do not seem very bright. We can look at the situation in the following way. Either the sought-after analysis is purely physicalistic—that is, it makes use only of physical terms—or it is mentalistic, making essential use of psychological concepts. Can there be an acceptable mentalistic analysis of reference? It is plausible to think that any such analysis would have to make use of "intentional mental concepts," that is, concepts of those psychological states that have "propositional content" or can be said, to use Brentano's well-known and well-worn phrase, to be "directed upon" objects. Examples of such concepts are, of course, those of believing, intending, thinking, inferring, etc. In fact, some philosophers (e.g. Paul Grice, R. M. Chisholm) have tried to explain meaning and linguistic reference in terms of these intentional psychological states.

However, there are some powerful arguments, advanced notably by Davidson, that in constructing a semantical theory for a language, we may not presuppose a prior and independent theory of our subjects' intentional psychological attitudes, a theory that enables us to attribute such states to the speakers of the language. According to Davidson, a theory that interprets language L, that is, a theory that assigns meanings to utterances of L, and a theory that interprets the contents of the psychological states of the speakers of L, are so intimately interconnected and interdependent that to assume one is available while constructing the other is in fact to beg the question. I cannot present the details of Davidson's arguments here; however, their gist will be evident in a later section of this paper. Evidently, we will not have made much progress with understanding reference if we analyze

“semantical reference” (“Term *a* refers to object *x*”) on the basis of such intentional psychological terms as the speaker’s *intention* to refer to *x* by the use of term *a* or his *thought of* or *about* *x*. Semantical reference and intentionality (in Brentano’s sense) are conceptually too close to each other for one to serve as an explicative basis for the other, although of course noting the precise nature of their interconnection is important. In any event, I believe there is a good deal to be said for Davidson’s thesis that a *unified theory* must be developed that simultaneously interprets our subjects’ psychological states and interprets their utterances. Semantical theory, on this view, must be regarded as part of a general theory of human behavior.

Can there then be a purely physical analysis of reference? Again, the prospects appear dim. It seems evident that the referential relation is causal and conventional, and involves very diverse physical mechanisms in different cases. Moreover, if there is to be a physical reduction of reference, this would have to be via a reduction to the psychological; that is, a psychological reduction of reference must precede the physicalization, for it seems undeniable that psychological elements, such as perceptual identification and intention to refer or communicate, are essentially involved in many referential relations. It is therefore exceedingly difficult to imagine that some unitary physical process underlies the relation between a sign and the object it refers to in every actual and possible case of reference in every actual and possible language. This is because psychological states seem capable of realization in vastly different physical structures; for example, we cannot expect the physical state realizing pain in one biological species to be the same as that realizing pain in another species with a very different physical structure.<sup>6</sup> Although we may grant that semantical facts are ultimately supervenient upon physical facts, this does not imply that the former are “reducible” or “analyzable” or “definable” in terms of the latter.<sup>7</sup> Supervenient dependence is only a necessary condition *in res* for the possibility of reduction. Whether a reduction is in fact possible depends on the question whether the particular dependencies between the supervenient properties or facts and those upon which they supervene are “simple” or “perspicuous” enough to make the reduction possible.

Furthermore, even if a physical reduction were possible, a “building block” theory based on such a reduction is unlikely to meet the second of our constraints on semantical theories stated earlier. As you will recall, this condition states that such a theory must be verifiable and learnable on the basis of evidence available to the speakers of the language. This was to ensure that such a theory can explain how language can be learned and used in communication. Now the kind of physical reduction that the “building block” theory contemplates would have to involve, at one stage or another, the complex neurophysiological mechanisms of human and other

<sup>6</sup> This is the famous “multiple realization argument” familiar in philosophy of mind and psychology. See Hilary Putnam, “The Nature of Mental States” in his *Mind, Language and Reality: Philosophical Papers II* (London: Cambridge University Press, 1975).

<sup>7</sup> See my “Concepts of Supervenience” in *Philosophy and Phenomenological Research*, 45.2.

language users. That is, the physical reductive basis of the fact that, in language L, the referential relation holds for a given sign and a given object, must involve the neurophysiological processes of the speakers of this language. One obvious reason is that, as noted above, this relation must involve psychological facts about the speakers, and these facts, if physically reducible, would presumably reduce to neurophysiology. But then the referential relations would be inaccessible to normal speakers or interpreters. They have no sophisticated knowledge of the brain processes underlying linguistic behavior; our ancestors lacked even a crude and approximate understanding of the physiological basis of language. But we somehow manage to learn and interpret languages. This seems to show that a physical reduction would be totally irrelevant to the main aim of semantical theory, that is, to capture and systematize what we as interpreters and speakers know.

There is a more direct argument against the "building block" theory, called "the permutation argument." As it captures the basic insight of the "holistic theory," the main rival of the "building block" theory, I shall describe it in some detail.

The permutation argument runs essentially as follows. In developing a semantical theory for a language, the controlling linguistic evidence consists of *sentences*—in particular, what sentences the speakers of the language "hold true" and what sentences they do not so hold. Let us call this a "truth distribution" over the utterances of language L. When a semantical theory generates truth conditions for the utterances of L, it also generates a truth distribution over them. According to the advocates of the holistic approach, the truth distribution generated by a semantical theory must match the truth distribution of the speakers of L as closely as possible. (This is a form of the much discussed "principle of charity" in interpretation or translation.) That is to say, if a sentence of L is generally held true by the speakers of L, our semantical theory for L must, *ceteris paribus*, generate a truth condition for that sentence under which that sentence is in fact true (in practice, this means being true by our light, that is, by the interpreter's light). Further, it is only at the sentential level that we can correlate the speakers' behavior, especially intentional actions, with their speech. Interpretively relevant behavior is determined by, and reflects, intentional psychological states like belief and desire, and these are states involving *propositions*, that is, contents that are describable only by full sentences. It is my belief that *it will rain today* and my desire that *I not get wet* that lead to my behavior of taking the umbrella with me. And these beliefs and desires are in important ways ascertained through the subjects' linguistic behavior involving full sentences, such as assent and dissent. This is why truth distribution is of fundamental importance to interpretation.

Davidson says:

"We don't need the concept of reference; neither do we need reference itself, whatever that may be. For if there is one way of assigning entities to expressions (a way of characterizing 'satisfaction') that yields acceptable results with respect to the truth conditions of sentences, there will be endless ways that do as well. There is no reason, then, to call any one of these relations 'reference'

or 'satisfaction'.'<sup>8</sup>

The core of the permutation argument then is this: If one way of assigning referents to our expressions works right in generating a truth distribution that matches the speakers' truth distribution, there are indefinitely many ways of assigning referents to expressions that will do the same. That is to say, any system of individual referential relations that connect signs to their referents can be permuted in many ways without affecting the truth distribution generated. This means that referential relations are essentially fluid, and give us no fixed point in reference to which a semantical theory is to be developed. In constructing a Tarski-style truth theory, we would normally fix referential relations between our basic symbols and their "referents"; but there is no fixed way, no right or wrong way, of doing this. The only thing that needs to be preserved is the truth distribution; reference may fall where it may. Semantical theory is not based on the analysis of the referential relation as its foundation.

We can consider how the permutation argument works by considering a simple example: Consider a sentence whose logical form is "a is F." Assume that this sentence is generally held true by our subjects, and we want to generate a truth condition that makes it in fact true. We can do this by interpreting "a" to refer to Socrates and interpret "F" to refer to the property of being wise. But, trivially, the same effect can be produced by interpreting "a" to refer to Alcibiades and "F" to refer to the property of being handsome. John Wallace has shown how this can be done systematically,<sup>9</sup> but we have said enough to give you an idea of how this might be possible.

The permutation argument seems to show this: we do not need a determinate system of referential relations to get our semantics right—at least, our truth-conditional semantics right. To put it another way: two semantical theories may be equally correct and explain everything that needs to be explained equally well, but differ widely in the referential relations they posit. And as far as semantical considerations go, there is, and can be, no evidence to decide which system of referential relations is the "correct" one. There is no fact of the matter here about what is "correct" or "incorrect."

Davidson's own theory is "holistic" in the following sense. First, as noted above, Davidson views semantical theory in the context of developing a global theory of the speakers' behavior. It is only in such a context of constructing an overall explanatory theory of behavior that we can develop a semantical theory that explains speech. This global theory is an "intentional theory" in that one of its fundamental goals is to attribute to the subjects contentful psychological states, "intentional states"—beliefs, desires, and the rest. And in doing this we are under the constraint of rationality that requires that the total system of such states attributed to a subject be maximally rational and coherent in relation to available evidence, and must

<sup>8</sup> "Reality without Reference" in *Inquiries into Truth and Interpretation*, p. 224.

<sup>9</sup> In "Only in the Context of a Sentence do Words Have Any Meaning," *Midwest Studies in Philosophy* 2 (1977).

also meet a certain minimum standard of absolute rationality. Unless this constraint is met, there would be no reason to think, Davidson argues, that we are here dealing with a subject who has a system of beliefs or who has speech (as opposed to mechanical, mindless production of phonetic noise).

The interdependence of speech and thought, that is, interdependence between semantical theory and psychological theory, can be seen briefly as follows: the speaker's utterance "This rose is red" is determined by at least the following two conditions, his belief that this rose is red, and the semantical fact (which he knows qua speaker of the language) that "This rose is red" means that this rose is red (and also his desire to speak the truth). Given the utterance, if we know the speaker believes that this rose is red (and wanted to express this belief), then we can know that "This rose is red" means that this rose is red. Conversely, given the utterance, if we know "This rose is red" means that this rose is red, we can know that the speaker believes that this rose is red. Thus, given the utterance and belief, we can infer meaning; given the utterance and meaning, we can infer belief. This is all pretty trivial. What makes the job of the interpreter nontrivial, that is, what makes a semantical theory nontrivial, is this: we are given the utterances, and must solve for both belief and meaning at the same time. This, according to Davidson, is what makes semantical theory inescapably holistic, and why we need to view semantical interpretation in the general context of interpreting human behavior.<sup>10</sup>

In this brief paper I have reviewed two influential contrasting approaches to the question of what "meaning" is, and of what "understanding meaning" is, the "building block" theory and the "holistic" theory. Each has its ardent advocates and equally ardent critics. The debate has been intense and philosophically fruitful, and is likely to continue to be so.

## REFERENCES

- Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford.
- Dummett, M. (1975) 'What is a Theory of Meaning,' in S. Guttenplan, ed., *Mind and Language*, Oxford University Press, Oxford.
- Field, H. (1972) 'Tarski's Theory of Truth,' *Journal of Philosophy* 69, 347-75.
- Kim, Jaegwon (1984) 'Concepts of Supervenience,' *Philosophy and Phenomenological Research*, 45. 2.
- Putnam, H. (1975) *Mind, Language and Reality: Philosophical Papers II*, Cambridge University Press, London.
- Wallace, J. (1977) 'Only in the Context of a Sentence do Words Have Any Meaning,' *Midwest Studies in Philosophy* 2.

<sup>10</sup> See "Thought and Talk" and "Belief as the Basis of Meaning" in *Inquiries into Truth and Interpretation*.

Department of Philosophy  
The University of Michigan  
Ann Arbor, MI 48109  
U.S.A.

Visiting Professor (Fall Semester, 1984)  
Department of Philosophy  
Seoul National University