

## Interpretation: Keeping in Touch with Reality

Gilead Bar-Elli

Davidson upheld the following central theses:

1. In a narrow sense a theory of meaning (for a language) is basically a Tarski-like theory of truth for that language. In a wider sense it includes also an account of how someone could come to know that such a theory was true.<sup>1</sup>
2. Such a theory is extensional – in particular, it does not appeal to Fregean meanings (senses) or to ways in which things are conceived and apprehended.<sup>2</sup>
3. Meaning and belief are inter-dependent: ascribing beliefs to persons and determining the content of their utterances are two inter-dependent tasks of the same interpretative theory. Meaning and belief are two “variables” that are determined simultaneously by the same holistic theory.<sup>3</sup>

I doubt whether these three theses, which characterize basic features of Davidson's conception of meaning, are coherent. The narrow sense of (1) has (2) as an obvious corollary. The wider sense of (1), or some aspects of it, is partly explicated by (3). But this doesn't seem to cohere with (2). When theory of meaning is taken in the wider sense to include an account of how one could come to know that such a theory of truth was true (for a given language), and when the full force of the marriage between meaning and belief, as in (3), is appreciated, (2) seems to me doubtful.

There are two main reasons for my doubts. The one focuses on the fact that the “correctness” of T-sentences, when a truth theory is combined with a theory of belief ascription, depends on the particular ways the truth conditions are conceived and

described, so that even logically equivalent descriptions are set apart: some may form a correct T-sentence, while others do not.

The other reason has to do with the fact that belief ascriptions do not relate a person to situations and events *simpliciter* (simple objects of cognition) but to situations and events conceived in terms of objects, their properties and relations, as given to us in particular ways. In belief-ascriptions, these, unlike what Davidson thinks of reference and satisfaction in truth-theory, are not theoretical constructs that can, after having served their purpose, be disregarded, but are rather essential components of belief-ascription, and hence, on Davidson's program, of understanding.

Evidently, there is a close connection between these two reasons. Construing the truth conditions as "intensional" (or non-extensional), and their "correctness" as sensitive to the ways they are conceived (in an interpretation) is evidently related to the interpreter's beliefs, and their intensionality. Many issues concerning these topics have been widely discussed in the literature<sup>4</sup>. I do not discuss this literature here, and try to shall confine my following remarks to distilling what seems to me to have direct bearing on the reasons for doubting the coherence of Davidson's three theses stated above.

### Meaning and the Correctness of a Truth Theory

Ascribing beliefs to others is an essential element in understanding them and interpreting what they say. How are such ascriptions justified, and on what basis do we make them? First and foremost, of course, on the basis of what they say. But this can be of any help only when we know what they say, that is, understand it. However, as Davidson repeatedly pointed out, knowledge of what people believe is necessary for understanding their idiolect and utterances, not less than the other way around. We seem to be trapped in a vicious circle, in which knowing what people believe is

necessary for understanding their utterances, and this understanding is necessary for knowing what they believe. Admitting the circle, Davidson dispels its viciousness by his theory of interpretation, in which belief and meaning are like two unknown variables determined simultaneously. A central claim here is that the theory is thoroughly extensional. This is the main point I want to question.

A long and respectable tradition tells us that interpretation consists in assigning meanings to words, and that understanding consists in grasping the meanings of words. These meanings comprise the meanings (senses) of sentences, which determine their "truth conditions." Davidson (as so many others) ascribes this conception to Frege. I shall not expand here on the accuracy of the ascription. Let me just remark that Davidson (as so many others) seems to undermine some essential elements of the Fregean conception, which set it apart from many of its followers. In particular he seems to undermine the significance of Frege's "context principle," which gives priority to the meaning of sentences, and construes the meaning of words as their systematic contribution to the meaning of sentences.<sup>5</sup>

Davidson rejects this tradition. Although he holds fast to the Fregean idea that truth conditions are constitutive of the meaning of sentences, he claims that ascribing meanings (senses) to words is neither necessary, nor useful, in explaining this. Understanding and interpretation consist in forming a theory of meaning for a language (or an idiolect). And meanings of words are superfluous in such a theory: meanings, by his well-known metaphor, don't oil the wheels of a theory of meaning ("Truth and Meaning," 1967, 20).

A theory of meaning for a language consists, in Davidson's lights, in a Tarski-like theory of truth for that language: an axiomatic theory that entails all the "correct" T-

sentences for the language. A T-sentence is a sentence of the form “ ‘p’ is true iff q,” where ‘p’ is a structural description of a sentence, and ‘q’ is an expression or description of its truth conditions. In Tarski’s theory this is done in terms of reference (or satisfaction), but Davidson’s idea is that these notions are theoretical constructs that, having done their job, can, like Wittgenstein’s ladder, be thrown away, after fulfilling their job: they are not essential features of understanding, for which only a partial understanding of truth is assumed.

In dealing with actual languages, this general scheme must be filled-in with substantial and empirically constrained content. We must filter out theories that entail distorted T-sentences in which the truth conditions of, e.g., “snow is white” may be proved to be that grass is green. For example, we want a truth-theory for German, phrased in English, in which the T-sentence ‘ “Schnee ist weiss” is true (in German) iff snow is white’ can be proved. But, since the theory is wholly extensional, there seems to be no inner constraint guarantying that “distorted” truth theories don’t emerge, in which unwanted T-sentences such as ‘ “Schnee ist weiss” is true (in German) iff grass is green’ are provable. We would also want to eliminate entailments like ‘ “Schnee ist weiss” is true (in German) iff snow is white or grass is black’, from having the status of T-sentences. I guess that even sentences like “ “Schnee ist weiss” is true (in German) iff XYZ’ (where XYZ is a sophisticated physical description of snow being white) shouldn’t have this status.<sup>6</sup> In short, we want the truth conditions to be not only extensionally correct, but to be conceived and described in the “right” way. Intuitively, we want the right-hand side to have the same meaning as, or to be synonymous with, the sentence mentioned on the left. In Tarski’s original idea this was guaranteed either by assuming translation or by the fact that the meta-language contained the object

language, and the right-hand side of a T-sentence was in fact a disquotation of the left. But Davidson, aiming at a general theory of meaning, assumes neither.

Davidson considers various aspects of truth-theories that reduce the chances of getting such distorted theories. Since 'Schnee', 'weiss', 'Gras', 'grün', etc., appear in so many German sentences the theory will have to deal with, the complexity of the truth theory may put such restrictive conditions on the axioms governing these words that the chances of getting distorted T-sentences become very small. He also claims that T-sentences should be regarded to be laws that support counterfactuals. These claims raise notorious problems into which I shall not enter here.

There is however another (and related) problem, which is perhaps more basic: What does it mean to say that a certain T-sentence is distorted or incorrect? What is the nature of the evidence for that? Davidson's basic idea was that a truth theory yields as theorems all the "correct" T-sentences for the (interpreted) language. A T-sentence is supposed to be the evidence and the test for a truth-theory; its "correctness" must therefore be established independently of the theory.<sup>7</sup> A correct T-sentence is supposed to be one on whose right-hand side we have the "right" truth conditions of the (interpreted) sentence described on the left-hand side, and this seems to be in some way 'intensional': it must be sensitive, e.g., to the difference between true sentences such as "snow is white" and "grass is green." It must be sensitive even to the difference between "snow is white" and "snow is white or grass is black," or "XYZ," etc. This does not contradict Davidson's claim that the theory itself is extensional, but it raises a doubt about its significance. The truth-theory may indeed be extensional, but it is intensionally constrained in a way that must be part of a theory of meaning in the wide sense: what is known in understanding the language. Leaving that aside for the

moment, let us look at how we know the “correct” truth conditions, and what their being “correct” consists in.

There are, it seems to me, two strands in Davidson’s writings for answering these questions. The one we can call the local, direct approach; the other - the global or holistic approach.

### The Local Direct Strand - Triangulation

Some crucial steps here can be sketched as follows. Imagine the following typical situation of “radical translation”. The interpreter notices that

- a. X utters ‘p’ in a situation S (with an assertoric force or some other “pro-attitude”).

The interpreter then concludes that

- b. X holds-true the (utterance of the) sentence ‘p’ in the situation S.

Gathering such information about other speakers in similar situations he may generalize

- c. Speakers of L hold-true utterances of ‘p’ in situations of kind S.

On this basis he may hypothesize that the following T-sentence is correct:

- d. T: ‘p’ is true (in L, say, German) iff q (‘q’ is the interpreter’s description of S).

Having reached this stage, and taking into account many other factors of what we would regard as standard speech conditions, the interpreter can now add intensional idioms like saying and believing, and conclude, for instance, that the speaker have said that q, or believed that q. She may even conclude, quite reasonably, that S caused X to

believe that q (just as it caused the interpreter to believe that q). This expresses an aspect of the interpreter's attitude towards the speaker as having a language, as belonging to a speech community, and as a rational human being, basically like the interpreter.

One point to note is that all steps in this imaginary field-work are in some way hypothetical. The interpreter may be wrong in each of them. She can, for instance, be wrong in the very first stage of identifying the utterance, or in recognizing an assertoric force or pro-attitude towards the utterance on behalf of the speaker. And she can, of course, be wrong in the way she describes the situation S, in what she takes to be the salient features she assumes the speaker to be attending to, etc. But she cannot, according to Davidson, be too grossly wrong, for then there would be nothing to be wrong about. But this is of no moment to our concerns here.

Another point is the importance of the real situation S. The situation we imagine is an example of what Davidson calls triangulation (or one type of it).<sup>8</sup> We have three factors here which must be coordinated: the speaker, X, the interpreter, Y, and the situation, S. What is the role of S here? It seems to be necessary for the above transitions: our interpreter can advance in the stages sketched above, because she assumes (believes) that the situation S has caused X to utter 'p' and to hold it true (in that situation). If she didn't think there was a real causal link between the situation S and the utterance of 'p', she wouldn't have any reason for the holding-true ascription. This then is the real, or the supposed, touchstone of reality: the real situation S is a real cause of the utterance and of the speaker's holding his utterance true.

It should be stressed that no appeal to a theory of meaning or to a truth theory for the language concerned is involved here. The whole process seems to be local and

“bottom-up,” proceeding from holding an utterance true in a situation to believing  $q$  in the situation and to the appropriate T-sentence. We should not be misled by calling step (d) above “a T-sentence,” alluding thereby to its being a potential test for a truth-theory for that language. It can (and should) be such a test: the theory should have it as one of its (infinitely many) theorems. But it is established as such a test independently of the theory. It is only after having so many such correct T-sentences that we can embark on the task of formulating such a truth theory and claiming it to be the heart of a theory of meaning for that language. Such a theory may have a corrective effect: in working it out the interpreter may be led to correct or change some of her initial T-sentences, so that the overall theory should accord with the maxims of rationality, and their ascription to the speaker. But this doesn’t change the basic, bottom-up, character of what being a “correct” T-sentence is here.

Various questions are raised by this description. Let me mention one. On the above local approach the situation  $S$  seems really to be essential. But the general tenor of Davidson’s theory may cast doubt on it. What is really needed for the interpretation is that the interpreter should **believe** that the situation  $S$ , or what the interpreter **believes** to be the situation  $S$ , caused the utterance. If the situation didn’t in fact cause the utterance yet the interpreter believed it did, that should be enough (for her ascribing to the speaker the holding-true attitude towards his utterance). And if, on the other hand, the situation did in fact cause the utterance but the interpreter didn’t believe it, the real causal link wouldn’t be relevant for her interpretation. Interpreters, in ascribing beliefs to others, seem thus to be trapped in their “veil of beliefs,” as they are, according to Davidson, in justifying beliefs in general.<sup>9</sup>

### The Global, Holistic Strand



Here then we come to the other, holistic strand in Davidson theory of meaning and belief. Ascribing beliefs and meanings are two inter-dependant factors in the interpretative enterprise, which must be “solved” simultaneously: the interpreter’s task is to ascribe meanings to the speaker’s utterances so that a “maximal” match results between her own and the speaker’s beliefs. "Maximal match" here is not a simple additive notion: different beliefs may have different weights, so that agreement on whether Egypt is a democracy does not have the same significance as on whether snow is white or that the person in front of you is a human being. Various aspects of humanity and rationality shape this notion in a non-monotonic way. Principles of rationality and the principle of charity are reigning over this global, holistic enterprise. The main dictate of these principles is that the meaning theory for the speaker’s language must be such that "most" (qualified as above) of the resulting beliefs ascribed to the speaker should be true according to the interpreter’s lights: “Interpret your fellow-man as you would like to be interpreted yourself.” This may be put as the Davidsonian semantic twist to the Jewish command: “Thou shall love your fellow-man as yourself.”

It is this maximization, which not only rules out distorted truth theories, but also gives content to the very idea of a “correct” theory and a “correct” T-sentence. This is what “correct” means here: being part of a maximal match between the speaker’s beliefs and the interpreter’s. This is the main difference between the two strands: on the local one the theory may indeed have a corrective effect, in checking and correcting the initial hypotheses forming the T-sentences. But these are revisions of a process that is basically local and bottom-up. On the global one, the theory is not only corrective, but definitional: the correctness of a T-sentence is defined by its belonging to the overall theory. Establishing the correct T-sentences is here a top-down process.

But if maximal match of beliefs, as qualified above, is the reigning principle, the role of real situations in the world may seem quite mysterious, and the emerging conception of truth seems to be coherentistic, Davidson's late reservations about his earlier use of this term notwithstanding. In fact it is not clear why we should appeal to truth and truth conditions. We should be satisfied, on this strand, with believing-true conditions, whose analogue of a T sentence is:

B-T: 'p' is believed-true (by speakers of L) iff it is believed that q. ('Schnee ist weiss' is believed-true (in German) iff it is believed that snow is white.)

This can be arrived at as a generalization of many instances of the form:

'p is believed-true by speaker s iff it is believed, by both speaker and interpreter, that q.

It is of no help to retort here, in the standard Davidsonian manner mentioned above, that real situations and episodes of the world are operative in forming and determining the contents of beliefs by being causally connected to them. For they are thus operative only when the interpretive theory is construed as truth-theory, when the right-hand sides of T-sentences give the truth conditions of their lefts, as we did in the direct, bottom-up strand sketched above. But there is no reason to see the theory in this way, if all we need and all we get is a maximal match of beliefs (not of true beliefs). On this view, all that a B-T-sentence gives us is a belief-match between speaker and interpreter – their being right or wrong doesn't matter.

Davidson may be persuasive in claiming that questioning the truth of all or most of our beliefs is incoherent.<sup>10</sup> But this claim carries persuasion only on an independent adoption of a realistic standpoint on truth, and on the causal relation between assenting to an utterance and the real situation that prompts it; it is not implied by the holistic conception itself and by the "top-down" construal of the correctness of theory.

Davidson was right to emphasize that a T-sentence is not translational: it relates a sentence to real situations or episodes. But he seems to have undermined the significance of the fact that in triangulation, and in establishing the correctness of T-sentences, these situations and episodes are conceived of and described in terms of objects, their properties and relations given to us in particular ways.<sup>11</sup> This is true of the interpreter herself and of the beliefs she ascribes to the speaker. Since beliefs and meanings are determined simultaneously, these particular ways of conceiving the elements of the situations must be deemed essential to the interpretative enterprise. If that is so, the fact that the theory itself is extensional seems to be of a rather limited significance. It seems to turn out then that on Davidson's own theory, understanding a language essentially involves senses, ways of conceiving or of being given, much in the spirit of what Frege and Wittgenstein taught, or at least much more so than Davidson has suggested.

*The Hebrew University of Jerusalem*

---

## Notes

- <sup>1</sup> The narrow sense is repeatedly argued for in Davidson's writings. For the wider sense, see for instance the following formulation: "What we want is an **account of what it is to understand a language** [...] what it would suffice someone to know if he were to understand an arbitrary utterance of a speaker (my answer: a theory of truth for that speaker's language), plus an account of how someone could come to know that such a theory was true" ("Epistemology and Truth," 1988, 190).
- <sup>2</sup> This is a central and recurring thesis of Davidson's. See, for instance, "Truth and Meaning" (1967, 23-24). The issue is however rather intricate: Davidson's basic idea is that a theory of

---

meaning should impose structure on sentences so as to explain how the meanings of sentences depend on the meanings of their words. This in itself is not a direct argument against non-extensional theories. But, for Davidson, extensional theories, and primarily Tarski's definition of truth (or its axiomatic variant), are the best models we have for such explanations.

<sup>3</sup> See "Belief and the Basis of Meaning" (1974). This again is a central and recurring theme. Cf. for instance the following formulation: "What we hold true, what we believe, determines what we mean, and thus, indirectly, when our sentences are true" ("Epistemology and Truth," 1988, 189-190.)

<sup>4</sup> Many of these are collected in *Truth and Interpretation* (Ernie Lepore, ed.), Blackwell, 1986; *The philosophy of Donald Davidson* (Lewis Hahn, ed.), open Court 1999. Particularly pertinent is also Bilgrami A. *Belief & Meaning*, Blackwell, 1992.

<sup>5</sup> See G. Bar-Elli, *The Sense of Reference – Intentionality in Frege* (De Gruyter, 1996), ch. 5.

<sup>6</sup> This is often referred to in the literature as the first Foster problem. See J. Foster: "Meaning and Truth Theory", in G. Evans and J. McDowell (eds.), *Truth and Meaning* (Oxford, 1976).

<sup>7</sup> "The **evidence** for the correctness of a theory of truth must come at the level of T-sentences; these also constitute the testable predictions of the theory [...] The evidence that the utterances of a certain sentence by a speaker are true under the conditions provided by a T-sentence consists in facts about the conditions that cause the speaker to hold the sentence true and to take that sentence to express a belief of his" ("Epistemology and Truth," 1988, 189).

<sup>8</sup> Triangulation is central, of course, not only to Davidson's account of belief, but to thought and perception in general: "Central to my account of empirical content is the process of **triangulation**, which narrows down the relevant [distal??] causes of perceptual beliefs, and makes possible grasp of the concept of objectivity. The type of cause repeatedly singled out as the cause of assent to a given perceptual sentence then constitutes the content of that sentence and of the belief that sentence can be used to express" ("Reply to McDowell," 1999, 107. See also, for instance, "The Second Person," 1992). Also: "What makes communication possible is the fact that many of the same objects, events, and aspects of the world are salient for all humankind, frequently eliciting observably similar responses" ("Reply to Quine," 1999, 82).

- 
- <sup>9</sup> See for instance the following formulation: “Neurath was right in rejecting the intelligibility of comparing sentences or beliefs with reality... Nor can such events be considered in themselves to be evidence, unless, of course, they cause us to believe something” (“Empirical Content,” 1982, 173).
- <sup>10</sup> See for instance: “By paying close attention to what gives content to our beliefs and meaning to our sentences, we see that it is **impossible for most of our perceptual beliefs to be false**. What we hold true, what we believe, determines what we mean, and thus, indirectly, when our sentences are true. (“Epistemology and Truth,” 1988, 189-190).
- <sup>11</sup> In light of this, another Davidsonian thesis seems to be of a rather restricted validity. Davidson held that “The perspective on language and truth that we have gained is this: what is most open to observation is sentences and their uses, and truth is the semantic concept we understand best. **Reference** and related semantic notions like satisfaction are by comparison **theoretical** (as are the notions of word, predicate, sentential connectives and the rest); there is no question about their correctness beyond the question whether they yield a satisfactory account of the truth conditions of sentences and the relations among sentences” (ibid., p. 181). This may be convincing with regard to truth and meaning, but much less so with regard to belief-ascription and triangulation. Here the situation is conceived, as Davidson himself has repeatedly said, in terms of its constituents: objects, their properties and relations. We don’t have a notion of situation prior to these, and we don’t conceive of these constituents as theoretical constructs determined by their functional role in a theory of situation.