The Folly of Trying to Define Truth

Donald Davidson

The Journal of Philosophy, Volume 93, Issue 6 (Jun., 1996), 263-278.

Stable URL:
http://links.jstor.org/sici?sici=0022-362X%28199606%2993%3A6%3C263%3ATFOTTD%3E2.0.CO%3B2-Q

THE JOURNAL OF PHILOSOPHY

VOLUME XCIII, NO. 6, JUNE 1996

THE FOLLY OF TRYING TO DEFINE TRUTH*

In the *Euthyphro*, Socrates asks what holiness is, what "makes" holy things holy. It is clear that he seeks a definition, a definition with special properties. He spurns the mere provision of examples or lists, asking in each case what makes the examples examples, or puts an item on the list. He rejects merely coextensive concepts ("something is holy if and only if it is dear to the gods"): what makes something dear to the gods is that it is holy, but not vice versa. The dialogue ends when Socrates begs Euthyphro to enlighten him by coming up with a satisfactory answer; Euthyphro decides he has another appointment.

The pattern of attempted definition, counterexample, amended definition, further counterexample, ending with a whimper of failure, is repeated with variations throughout the Socratic and middle Platonic dialogues. Beauty, courage, virtue, friendship, love, temperance are put under the microscope, but no convincing definitions emerge. The only definitions Plato seems happy with are tendentious characterizations of what it is to be a sophist. He also gives a few trivial samples of correct definitions: of a triangle; of mud (earth and water).

In the *Theaetetus*, Plato attempts to define empirical knowledge. Like many philosophers since, he takes knowledge to be true belief plus something more—an account that justifies or warrants the belief. It is the last feature which stumps him (again foreshadowing the subsequent history of the subject). It seems no more to occur to Plato than it has to most others that the combination of causal and rational elements that must enter into an analysis of justified belief

* A modified version of this paper was written for, and will appear in, a festschrift for Professor Pranab Sen.

0022-362X/96/9306/263–78

(as it must into accounts of memory, perception, and intentional action) may in the nature of the case not be amenable to sharp formulation in a clearer, more basic, vocabulary.

What is important in the present context, however, is the fact that in attempting to define knowledge, it is only with the concept of warrant that Plato concedes defeat. He does not worry much about the equal involvement of knowledge with truth and belief.

Again, though, Plato was simply blazing a trail that other philosophers over the ages have followed: you follow his lead if you worry about the concept of truth when it is the focus of your attention, but you pretend you understand it when trying to cope with knowledge (or belief, memory, perception, and the like). We come across the same puzzling strategy in David Hume and others, who forget their skepticism about the external world when they formulate their doubts concerning knowledge of other minds. When a philosopher is troubled by the idea of an intentional action, he would be happy if he could analyze it correctly in terms of the concepts of belief, desire, and causality, and he does not for the moment worry too much about those (at least equally difficult) concepts. If memory is up for analysis, the connections with belief, truth, causality, and perhaps perception, constitute the problem, but these further concepts are pro tem taken to be clear enough to be used to clarify memory, if only the connections could be got right. It is all right to assume you have an adequate handle on intention and convention if your target is meaning. I could easily go on.

There is a lesson to be learned from these familiar, though odd, shifts in the focus of philosophical puzzlement. The lesson I take to heart is this: however feeble or faulty our attempts to relate these various basic concepts to each other, these attempts fare better, and teach us more, than our efforts to produce correct and revealing definitions of basic concepts in terms of clearer or even more fundamental concepts.

This is, after all, what we should expect. For the most part, the concepts philosophers single out for attention, like truth, knowledge, belief, action, cause, the good and the right, are the most elementary concepts we have, concepts without which (I am inclined to say) we would have no concepts at all. Why then should we expect to be able to reduce these concepts definitionally to other concepts that are simpler, clearer, and more basic? We should accept the fact that what makes these concepts so important must also foreclose on the possibility of finding a foundation for them which reaches deeper into bedrock.

We should apply this obvious observation to the concept of truth: we cannot hope to underpin it with something more transparent or easier to grasp. Truth is, as G. E. Moore, Bertrand Russell, and Gottlob Frege maintained, and Alfred Tarski proved, an indefinable concept. This does not mean we can say nothing revealing about it: we can, by relating it to other concepts like belief, desire, cause, and action. Nor does the indefinability of truth imply that the concept is mysterious, ambiguous, or untrustworthy.

Even if we are persuaded that the concept of truth cannot be defined, the intuition or hope remains that we can characterize truth using some fairly simple formula. What distinguishes much of the contemporary philosophical discussion of truth is that though there are many such formulas on the market, none of them seems to keep clear of fairly obvious counterexamples. One result has been the increasing popularity of minimalist or deflationary theories of truth—theories that hold that truth is a relatively trivial concept with no "important connections with other concepts such as meaning and reality."[1]

I sympathize with the deflationists; the attempts to pump more content into the concept of truth are not, for the most part, appealing. But I think the deflationists are wrong in their conclusion, even if mostly right in what they reject. I shall not pause here to give my reasons for refusing to accept correspondence theories, coherence theories, pragmatic theories, theories that limit truth to what could be ascertained under ideal conditions or justifiably asserted, and so on.[2] But since I am with the deflationists in being dissatisfied with all such characterizations of truth, I shall say why deflationism seems to me equally unacceptable.

Aristotle, as we all know, contended that

(1) To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.

When Tarski[3] mentions this formulation in 1944, he complains that it is "not sufficiently precise and clear," though he prefers it to two others:

(2) The truth of a sentence consists in its agreement with (or correspondence to) reality.

[1] These words are quoted from Michael Dummett's jacket blurb for Paul Horwich's *Truth* (Cambridge: MIT, 1991). This is not, of course, Dummett's view.

[2] I spell out my reasons for rejecting such views in "The Structure and Content of Truth," this JOURNAL, LXXXVII, 6 (June 1990): 279-328.

[3] "The Semantic Conception of Truth," *Philosophy and Phenomenological Research*, IV (1944): 342-60.

(3) A sentence is true if it designates an existing state of affairs (*ibid.*, p. 343).

In 1969, Tarski[4] again quotes (1), and adds,

> [T]he formulation leaves much to be desired from the point of view of precision and formal correctness. For one thing, it is not general enough; it refers only to sentences that "say" about something "that it is" or "that it is not"; in most cases it would hardly be possible to cast a sentence in this mold without slanting the sense of the sentence and forcing the spirit of the language (*ibid.*, p. 63).

He adds that this may be the reason for such "modern substitutes" for Aristotle's formulations as (2) and (3).

In the *Wahrheitsbegriff*, however, Tarski[5] prefers the following informal statement:

(4) A true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so (*ibid.*, p. 155).

It seems to me that Aristotle's formulation is clearly superior to (2), (3), and (4); it is more in accord with Tarski's own work on truth; and Tarski's comment that (1) is "not general enough" is strangely out of keeping with the spirit of his own truth definitions.

(1) is superior to (2)-(4) for three reasons. First, (3) and (4) mention states of affairs, thus suggesting that postulating entities to correspond to sentences might be a useful way of characterizing truth. ("A true sentence is one that corresponds to the facts," or "If a sentence is true, there is a state of affairs to which it corresponds.") But facts or states of affairs have never been shown to play a useful role in semantics, and one of the strongest arguments for Tarski's definitions is that in them nothing plays the role of facts or states of affairs. This is not surprising, since there is a persuasive argument, usually traced to Frege (in one form) or Kurt Gödel (in another), to the effect that there can be at most one fact or state of affairs. (This is why Frege said all true sentences name the True.) Tarski's truth definitions make no use of the idea that a sentence "corresponds" to anything at all. We should not take seriously the mention of "states of affairs" in such remarks of Tarski's[6] as this: "[S]emantical concepts

[4] "Truth and Proof," *The Scientific American*, CCXX (1969): 63-77.

[5] "The Concept of Truth in Formalized Languages," in *Logic, Semantics, Metamathematics* (New York: Oxford, 1956), pp. 152-278 (originally published in German in 1936).

[6] "The Establishment of Scientific Semantics," in *Logic, Semantics, Metamathematics*, pp. 401-08.

express certain relations between objects (and states of affairs) referred to in the language discussed and expressions of the language referring to those objects" (*ibid.*, p. 403).

A second reason for preferring Aristotle's characterization of truth is that it avoids the awkward blanks marked by the words 'so and so' in Tarski's version (4); one is hard pressed to see how the blanks are to be filled in. Aristotle's formula, on the other hand, sounds much like a generalization of Tarski's convention-T.

The third reason for preferring Aristotle's characterization is that it makes clear, what the other formulations do not, that the truth of a sentence depends on the inner structure of the sentence, that is, on the semantic features of the parts. In this it is once again closer to Tarski's approach to the concept of truth.

Tarski's convention-T, which he understandably substitutes for the rough formulas I have been discussing, stipulates that a satisfactory definition of a truth predicate 'is true' for a language L must be such as to entail as theorems all sentences of the form

s is true-in-L if and only if p

where 's' is replaced by the description of a sentence, and 'p' is replaced by that sentence, or a translation of the sentence into the metalanguage. Since it is assumed that there is an infinity of sentences in L, it is obvious that, if the definition of the truth predicate is to be finite (Tarski insisted on this), the definition must take advantage of the fact that sentences, though potentially infinite in number, are constructed from a finite vocabulary. For the languages Tarski considered, and for which he showed how to define truth, all sentences can be put into the form of an existential quantification, or the negation of an existential quantification, or a truth-functional compound of such sentences. So how "incomplete," from Tarski's point of view, is Aristotle's formulation (1)? It deals with four cases. There are the sentences that "say of what is that it is not": in modern terms it is a false sentence that begins 'It is not the case that there exists an x such that...'. An example might be: 'There does not exist an x such that $x = 4$'. Then there are sentences that "say of what is not that it is"; for example: 'There exists an x such that $x = 4 \,\&\, x = 5$'. There are sentences that "say of what is that it is"; for example: 'There exists an x such that $x = 4$'. And, finally, there are sentences that "say of what is not that it is not"; for example, 'It is not the case that there exists an x such that $x \neq x$'. According to the classical formulation, sentences of the first two kinds are false and of the second two kinds are true. Tarski is so far in agreement. What would Tarski

add? Just the truth-functional compounds (beyond those involving negation) of the types of sentences already mentioned; these are true or false on the basis of the truth or falsity of the kinds of sentences already provided for. Of course, Tarski also showed in detail how the truth or falsity of the first four types of sentences depended in turn on their structure.

Thus, the classical formulation regarded as an informal characterization is "incomplete" in only a minimal way compared to Tarski's own work, and is better than Tarski's informal attempts to state the intuitive idea. Needless to say, someone might question the extent to which natural languages can be adequately characterized using such limited resources; but this is a comment equally applicable to Tarski.

Despite his nod in the direction of a correspondence theory, in which sentences are said to correspond to facts, Tarski ought not to be considered as giving comfort to serious partisans of correspondence theories, nor should Aristotle. For neither Aristotles's formula nor Tarski's truth definitions introduce entities like facts or states of affairs for sentences to correspond to. Tarski does define truth on the basis of the concept of satisfaction, which relates expressions to objects, but the sequences that satisfy sentences are nothing like the "facts" or "states of affairs" of correspondence theorists, since if one of Tarski's sequences satisfies a closed sentence, thus making it true, then that same sequence also satisfies every other true sentence, and thus also makes it true, and if any sequence satisfies a closed sentence, every sequence does.[7]

If Tarski is not a correspondence theorist (and he certainly does not hold a coherence theory or a pragmatic theory or a theory that bases truth on warranted assertability), is he a deflationist? Here opinions differ widely: W. V. Quine thinks he is, and so does Scott Soames. John Etchemendy thinks Tarski simply says nothing about truth as a semantic concept, and Hilary Putnam, though for somewhat different reasons, agrees.[8]

If Tarski has said "all there is to say" about truth, as Stephen Leeds, Paul Horwich, and Soames all contend, and Quine has

[7] At one time I suggested calling Tarski's concept of truth a correspondence theory on the strength of the role of sequences in satisfying closed sentences, but I subsequently withdrew the suggestion as misleading. For the suggestion, see "True to the Facts," in *Inquiries into Truth and Interpretation* (New York: Oxford, 1984). For the retraction, see "Afterthoughts, 1987," in A. Malichowski, ed., *Reading Rorty* (Cambridge: Blackwell, 1990), pp. 120-38.

[8] For references, and further discussion, see my "The Structure and Content of Truth."

strongly hinted, then a sort of deflationary attitude is justified; this is not quite the same as the "redundancy" view, but close to it. The redundancy view, taken literally, is the same as the disquotational view taken literally: we can always substitute without loss a sentence for that same sentence quoted, and followed by the words 'is true'. What Tarski added, as Michael Williams and others have pointed out, is a way of predicating truth of whole classes of sentences, or of sentences to which we do not know how to refer; you may think of this as an elaboration of the redundancy theory in that it allows the elimination of the truth predicate when applied to sentences of a language for which that predicate has been defined.

At the same time that we credit Tarski with having shown how to make sense of remarks like 'The English sentence Joan uttered about Abbot was true' or 'Everything Aristotle said (in Greek) was false' or 'The usual truth table for the conditional makes any conditional true that has a false antecedent', we have to recognize that this accomplishment was accompanied by a proof that truth cannot (given various plausible assumptions) be defined in general; there can be no definition of 'For all languages *L*, and all sentences *s* in *L*, *s* is true in *L* if and only if ... *s* ... *L* ...'. In other words, Tarski justified the application of a truth predicate to the sentences of a particular language only by restricting its application to the sentences of that language. (It is ironic that in much recent writing on deflationary theories, Tarski has been taken to have lent support to the idea that there is a single, simple, even trivial, concept of truth.)

A deflationary attitude to the concept of truth is not, then, encouraged by reflection on Tarski's work. One can adopt the line advanced by Putnam and Etchemendy that Tarski was not even doing semantics, despite his insistence that he was; but this construal of Tarski does not support a deflationary theory: it simply denies the relevance of Tarski's results to the ordinary concept of truth. If, on the other hand, one takes Tarski's truth definitions to say something about the relations of specific languages to the world, one cannot at the same time claim that he has told us all there is to know about the concept of truth, since he has not told us what the concept is that his truth definitions for particular languages have in common.

I think that Tarski was not trying to define *the* concept of truth—so much is obvious—but that he was *employing* that concept to characterize the semantic structures of specific languages. But Tarski did not indicate how we can in general reduce the concept of truth to other more basic concepts, nor how to eliminate the English predicate 'is true' from all contexts in which it is intelligibly applied to sentences.

Convention-T is not a rough substitute for a general definition: it is part of a successful attempt to persuade us that his formal definitions apply our single pretheoretical concept of truth to certain languages. Deflationists cannot, then, appeal to Tarski simply because he demonstrated how to handle the semantics of quantification for individual languages. Leeds, Horwich, Williams, and others who have contended that all Tarski did was reveal the usefulness of an otherwise dispensable concept are wrong. They are right that we need a truth predicate for the purposes they, along with Tarski, mention; but they fail to note the obvious fact that at the same time Tarski solved one problem he emphasized another: that he had not, and could not, given the constraints he accepted, define or fully characterize truth.

Over the years, Quine has said a number of things about truth, but there has been, from early days until the most recent, what seems a consistent embrace of a deflationary attitude. Thus, Quine has made much of the "disquotational" aspect of the truth predicate, the fact that we can get rid of the predicate 'is true' after the quotation of an English sentence simply by removing the quotation marks as we erase the truth predicate. As Quine put it in *From a Logical Point of View*,[9] we have a general paradigm, namely,

(T) '——' is true-in-*L* if and only if ——

which, though not a definition of truth, serves to endow 'true-in-*L*' with

> every bit as much clarity, in any particular application, as is enjoyed by the particular expressions of *L* to which we apply [it]. Attribution of truth in particular to 'Snow is white'... is every bit as clear to us as attribution of whiteness to snow (*ibid.*, p. 138).

In *Word and Object*, Quine[10] remarks that "To say that the statement 'Brutus killed Caesar' is true, or that 'The atomic weight of sodium is 23' is true, is in effect simply to say that Brutus killed Caesar, or that the atomic weight of sodium is 23" (*ibid.*, p. 24). The theme is repeated thirty years later in *Pursuit of Truth*[11]:

> ...there is surely no impugning the disquotation account; no disputing that "Snow is white" is true if and only if snow is white. Moreover, it is a full account; it explicates clearly the truth or falsity of every clear sentence (*ibid.*, p. 93).

[9] Cambridge: Harvard, 1961.
[10] Cambridge: MIT, 1960.
[11] Cambridge: Harvard, 1990.

"Truth," he summarizes, "is disquotation" (*ibid.*, p. 80). On this matter, Quine has not changed his mind.

It is the disquotational feature of truth, in Quine's opinion, which makes truth so much clearer a concept than meaning. Comparing theory of meaning and theory of reference, Quine says that they constitute "two provinces so fundamentally distinct as not to deserve a joint appellation at all."[12] The former deals with such tainted topics as synonymy, meaning, and analyticity. The concepts treated by the latter, which include truth, are by contrast "very much less foggy and mysterious...." For although 'true-in-*L*' for variable '*L*' is not definable, "what we do have suffices to endow 'true-in-*L*', even for variable '*L*', with a high enough degree of intelligibility so that we are not likely to be averse to using the idiom" (*ibid.*, pp. 137-38). "What we do have" is, of course, the paradigm (T) and the "expedient general routine" due to Tarski for defining 'true-in-*L*' for particular languages.

The disquotational feature of truth, wedded to the thought that this may exhaust the content of the concept of truth, encourages the idea that truth and meaning can be kept quite separate. But can they in general? Scattered remarks in Quine's work suggest otherwise. In 1936, Quine published the brilliant and prescient "Truth by Convention."[13] In it he remarks that "in point of meaning...a word may be said to be determined to whatever extent the truth or falsehood of its contexts is determined" (*ibid.*, p. 89). It is hard to see how truth could have this power of determining meaning if the disquotational account were all there were to say about truth. Other passages in Quine suggest the same idea: "First and last, in learning language, we are learning how to distribute truth values. I am with Davidson here; we are learning truth conditions."[14] Or again, "Tarski's theory of truth [is] the very structure of a theory of meaning."[15]

Up to a point it may seem easy to keep questions of truth and questions of meaning segregated. Truth we may think of as disquotational (in the extended Tarski sense) and therefore trivial; meaning is then another matter, to be taken care of in terms of warranted assertability, function, or the criteria for translation. This is the line followed, for example, by Horwich in his recent book *Truth* (*op. cit.*),

[12] *From a Logical Point of View*, p. 130.
[13] Reprinted in *The Ways of Paradox* (Cambridge: Harvard, 1976).
[14] *The Roots of Reference* (La Salle, IL: Open Court, 1974), p. 65.
[15] "On the Very Idea of a Third Dogma," in *Theories and Things* (Cambridge: Harvard, 1981), p. 38.

by Soames,[16] and by Lewis.[17] It may, at least at one time, have been Quine's view. In *Word and Object*, in a passage that immediately precedes the remark that to say that the sentence 'Brutus killed Caesar' is true is in effect simply to say that Brutus killed Caesar, Quine despairs of a substantive concept of truth, and concludes that we make sense of a truth predicate only when we apply it to a sentence "in the terms of a given theory, and seen from within the theory" (*op. cit.*, p. 24). This is, I think, what Quine means when he says that truth is "immanent." The point is not merely that the truth of a sentence is relative to a language; it is that there is no transcendent, single concept to be relativized.[18]

Most recently, however, Quine muses that truth "is felt to harbor something of the sublime. Its pursuit is a noble pursuit, and unending"; he seems to agree: "Science is seen as pursuing and discovering truth rather than as decreeing it. Such is the idiom of realism, and it is integral to the semantics of the predicate 'true'."[19]

I turn now to Horwich's version of deflationism, for he seems to me to have accepted the challenge other deflationists have evaded, that of saying something more about an unrelativized concept of truth than we can learn from Tarski's definitions. Horwich's brave and striking move is to make the primary bearers of truth propositions—not exactly a new idea in itself, but new in the context of a serious attempt to defend deflationism. He is clear that he cannot provide an explicit definition of a truth predicate applying to propositions, but he urges that we really have said all there is to know about such a predicate (and hence the predicate it expresses) when we grasp the fact that the "uncontroversial instances" of the schema:

The proposition that *p* is true if and only if *p*

exhaust its content. (The limitation to "uncontroversial instances" is to exclude whatever leads to paradox.) The schema is taken as an axiom schema: the totality of its instances constitute the axioms of his theory.

This theory is, of course, incomplete until the controversial instances are specified in a non-question-begging way; and since the

[16] "What Is a Theory of Truth?" this JOURNAL, LXXXI, 8 (August 1984): 411-29.

[17] "Languages and Language," *Minnesota Studies in the Philosophy of Science*, volume VII (Minneapolis: Minnesota UP, 1975).

[18] The preceding paragraphs on Quine are partly quoted and partly adapted from a longer and more detailed study of Quine on truth: "Pursuit of the Concept of Truth," in P. Leondardi and M. Santambrogio, eds., *On Quine: New Essays* (New York: Cambridge, 1995). The relevant pages are 7-10.

[19] *From Stimulus to Science* (Cambridge: Harvard, 1995), p. 67.

set of axioms is infinite, it does not meet one of Tarski's requirements for a satisfactory theory of truth. But perhaps the first difficulty can be overcome, and the second may be viewed as the price of having an unrelativized concept of truth. There are, further, the doubts many of us have about the existence of propositions, or at least of the principles for individuating them.

All these considerations give me pause, but I plan to ignore them here. I want to give deflationism its best chance, since it seems to me to be the only alternative to a more substantive view of truth, and most substantive views are in my opinion, as in Horwich's, clear failures. But although I enthusiastically endorse his arguments against correspondence, coherence, pragmatic, and epistemic theories, I cannot bring myself to accept Horwich's "minimal" theory.

I have two fundamental problems with Horwich's theory, either of which alone is reason to reject it if it cannot be resolved; and I do not myself see how to resolve them.

The first problem is easy to state: I do not understand the basic axiom schema or its instances. It will help me formulate my difficulty to compare Horwich's axiom schema with Tarski's informal (and ultimately supplanted) schema:

'____' is true if and only if ____

Tarski's objection (among others) is that you cannot turn this into a definition except by quantifying into a position inside quotation marks. The complaint ends up with a question about the clarity of quotations: How does what they refer to depend on the semantic properties of their constituents? It has sometimes been proposed to appeal to substitutional quantification, and one may wonder why Horwich cannot generalize his schema:

(p) (the proposition that p is true if and only if p)

by employing substitutional quantification. But here Horwich quite rightly explains that he cannot appeal to substitutional quantification to explain truth, since substitutional quantification must be explained by appeal to truth.

Why, though, does Horwich not try generalizing his schema by quantifying over propositions? The answer should be: because then we would have to view ordinary sentences as singular terms *referring* to propositions, not as *expressing* propositions. This brings me to the crux: How are we to understand phrases like 'the proposition that Socrates is wise'? In giving a standard account of the semantics of the sentence 'Socrates is wise', we make use of what the name

'Socrates' names, and of the entities of which the predicate 'is wise' is true. But how can we use these semantic features of the sentence 'Socrates is wise' to yield the reference of 'the proposition that Socrates is wise'? Horwich does not give us any guidance here. Could we say that expressions like 'the proposition that Socrates is wise' are semantically unstructured, or at least that after the words 'the proposition that' (taken as a functional expression) a sentence becomes a semantically unstructured name of the proposition it expresses? Taking this course would leave us with an infinite primitive vocabulary, and the appearance of the words 'Socrates is wise' in two places in the schema would be of no help in understanding the schema or its instances. A further proposal might be to modify our instance of the schema to read:

> The proposition expressed by the sentence 'Socrates is wise' is true if and only if Socrates is wise.

But following this idea would require relativizing the quoted sentence to a language, a need that Horwich must circumvent.

So let me put my objection briefly as follows: the same sentence appears twice in instances of Horwich's schema, once after the words 'the proposition that', in a context that requires the result to be a singular term, the subject of a predicate, and once as an ordinary sentence. We cannot eliminate this iteration of the same sentence without destroying all appearance of a theory. But we cannot *understand* the result of the iteration unless we can see how to make use of the same semantic features of the repeated sentence in both of its appearances—make use of them in giving the semantics of the schema instances. I do not see how this can be done.

My second difficulty with Horwich's theory is more dependent on my own further convictions and commitments. Horwich recognizes that to maintain that truth has, as he says, "a certain purity," he must show that we can understand it fully in isolation from other ideas, and we can understand other ideas in isolation from it. He does not say there are no relations between the concept of truth and other concepts; only that we can understand these concepts independently. There are several crucial cases so far as I am concerned, since I do not think we can understand meaning or any of the propositional attitudes without the concept of truth. Let me pick one of these: meaning.

Since Horwich thinks of truth as primarily attributable to propositions, he must explain how we can also predicate it of sentences and utterances, and he sees that to explain this without compromising

the independence of truth, we must understand meaning without direct appeal to the concept of truth. On this critical matter, Horwich is brief, even laconic. Understanding a sentence, he says, does not *consist* in knowing its truth conditions, though if we understand a sentence we usually *know* its truth conditions. Understanding a sentence, he maintains, consists in knowing its "assertability conditions" (or "proper use"). He grants that these conditions may include that the sentence (or utterance) be true. I confess I do not see how, if truth is an assertability condition, and knowing the assertability conditions *is* understanding, we can understand a sentence without having the concept of truth.

I realize, however, that this is disputed territory, and that heavy thinkers like Michael Dummett, Putnam, and Soames, following various leads suggested by Ludwig Wittgenstein and H. P. Grice, believe that an account of meaning can be made to depend on a notion of assertability or use which does not in turn appeal to the concept of truth.

My hopes lie in the opposite direction: I think the sort of assertion that is linked to understanding already incorporates the concept of truth: we are *justified* in asserting a sentence in the required sense only if we believe the sentence we use to make the assertion is true; and what ultimately ties language to the world is that the conditions that typically cause us to hold sentences true *constitute* the truth conditions, and hence the meanings, of our sentences. This is not the place to argue this. For now I must simply remark that it would be a shame if we had to develop a theory of meaning for a speaker or a language independently of a theory of truth for that speaker or language, since we have at least *some* idea how to formulate a theory of truth, but no serious idea how to formulate a theory of meaning based on a concept of assertability or use.

I conclude that the prospects for a deflationary theory of truth are dim. Its attractions seem to me entirely negative: it avoids, or at least tries to avoid, well-marked dead ends and recognizable pitfalls.

Let me suggest a diagnosis of our aporia about truth. We are still under the spell of the Socratic idea that we must keep asking for the *essence* of an idea, a significant *analysis* in other terms, an answer to the question what *makes* this an act of piety, what *makes* this, or any, utterance, sentence, belief, or proposition true. We still fall for the freshman fallacy that demands that we *define* our terms as a prelude to saying anything further with or about them.

It may seem pointless to make so much of the drive to define truth when it is unclear who is trying to do it: not Tarski, who proves it

cannot be done; not Horwich, who disclaims the attempt. Who, then, *admits* to wanting to define the concept of truth? Well, that is right. But. But the same ugly urge to define shows up in the guise of trying to provide a brief criterion, schema, partial but leading hint, in place of a strict definition. Since Tarski, we are leery of the word 'definition' when we are thinking of a concept of truth not relativized to a language, but we have not given up the definitional urge. Thus, I see Horwich's schema on a par *in this regard* with Dummett's notion of justified assertability, Putnam's ideally justified assertability, and the various formulations of correspondence and coherence theories. I see all of them as, if not attempts at definitions in the strict sense, attempts at *substitutes* for definitions. In the case of truth, there is no short substitute.

Now I want to describe what I take to be a fairly radical alternative to the theories I have been discussing and (with unseemly haste) dismissing. What I stress here is the *methodology* I think is required rather than the more detailed account I have given elsewhere. The methodology can be characterized on the negative side by saying it offers no definition of the concept of truth, nor any quasi-definitional clause, axiom schema, or other brief substitute for a definition. The positive proposal is to attempt to trace the connections between the concept of truth and the human attitudes and acts that give it body.

My methodological inspiration comes from finitely axiomatized theories of measurement, or of various sciences, theories that put clear constraints on one or more undefined concepts, and then prove that any model of such a theory has intuitively desired properties—that it is adequate to its designed purpose. Since among the models will be all sorts of configurations of abstract entities, and endless unwanted patterns of empirical events and objects, the theory can be applied to, or tested against, such specific phenomena as mass or temperature only by indicating how the theory is to be applied to the appropriate objects or events. We cannot demand a precise indication of how to do this; finding a useful method for applying the theory is an enterprise that goes along with tampering with the formal theory, and testing its correctness as interpreted.

We are interested in the concept of truth only because there are actual objects and states of the world to which to apply it: utterances, states of belief, inscriptions. If we did not understand what it was for such entities to be true, we would not be able to characterize the contents of these states, objects, and events. So in addition to the

formal theory of truth, we must indicate how truth is to be predicated of these empirical phenomena.

Tarski's definitions make no mention of empirical matters, but we are free to ask of such a definition whether it fits the actual practice of some speaker or group of speakers—we may ask whether they speak the language for which truth has been defined. There is nothing about Tarski's definitions that prevents us from treating them in this way except the prejudice that, if something is called a definition, the question of its "correctness" is moot. To put this prejudice to rest, I suggest that we omit the final step in Tarski's definitions, the step that turns his axiomatizations into explicit definitions. We can then in good conscience call the emasculated definition a theory, and accept the truth predicate as undefined. This undefined predicate expresses the *general*, intuitive, concept, applicable to any language, the concept against which we have always surreptitiously tested Tarski's definitions (as he invited us to do, of course).

We know a great deal about how this concept applies to the speech and beliefs and actions of human agents. We use it to interpret their utterances and beliefs by assigning truth conditions to them, and we judge those actions and attitudes by evaluating the likelihood of their truth. The empirical question is how to determine, by observation and induction, what the truth conditions of empirical truth vehicles are. It bears emphasizing: absent this empirical connection, the concept of truth has no application to, or interest for, our mundane concerns, nor, so far as I can see, does it have any content at all.

Consider this analogy: I think of truth as Frank Ramsey thought of probability. He convinced himself, not irrationally, that the concept of probability applies in the first instance to propositional attitudes; it is a measure of degree of belief. He went on to ask himself: How can we make sense of the concept of degree of belief (subjective probability)? Subjective probability is not observable, either by the agent who entertains some proposition with less than total conviction and more than total disbelief, or by others who see and question him. So Ramsey axiomatized the pattern of preferences of an idealized agent who, more or less like the rest of us, adjusts his preferences for the truth of propositions (or states of affairs or events) to accord with his values and beliefs. He stated the conditions on which a pattern of such preferences would be "rational," and in effect proved that, if these conditions were satisfied, one could reconstruct from the agent's preferences the relative strengths of that agent's desires and subjective probabilities. Ramsey did not suppose

everyone is perfectly rational in the postulated sense, but he did assume that people are nearly enough so, in the long run, for his theory to give a content to the concept of subjective probability—or probability, as he thought of it.

A brilliant *strategy*! (Whether or not it gives a correct analysis of probability.) The concept of probability—or at least degree of belief—unobservable by the agent who has it and by his watchers, linked to an equally theoretical concept of cardinal utility, or subjective evaluation, and both tied to simple preference by the axiomatic structure. Simple preference in turn provides the crucial empirical basis through its manifestations in actual choice behavior.

We should think of a theory of truth for a speaker in the same way we think of a theory of rational decision: both describe structures we can find, with an allowable degree of fitting and fudging, in the behavior of more or less rational creatures gifted with speech. It is in the fitting and fudging that we give content to the undefined concepts of subjective probability and subjective values—belief and desire, as we briefly call them; and, by way of theories like Tarski's, to the undefined concept of truth.

A final remark. I have deliberately made the problem of giving empirical content to the concept of truth seem simpler than it is. It would be *relatively* simple if we could directly observe—take as basic evidence—what people *mean* by what they say. But meaning not only is a more obscure concept than that of truth; it clearly involves it: if you know what an utterance means, you know its truth conditions. The problem is to give *any* propositional attitude a propositional content: belief, desire, intention, meaning.

I therefore see the problem of connecting truth with observable human behavior as inseparable from the problem of assigning contents to all the attitudes, and this seems to me to require a theory that embeds a theory of truth in a larger theory that includes decision theory itself. The result will incorporate the major norms of rationality whose partial realization in the thought and behavior of agents makes those agents intelligible, more or less, to others. If this normative structure is formidably complex, we should take comfort in the fact that the more complex it is, the better our chance of interpreting its manifestations as thought and meaningful speech and intentional action, given only scattered bits of weakly interpreted evidence.

DONALD DAVIDSON

University of California/Berkeley