

Symposium

on

Understanding Truth

By

Scott Soames

Précis

Philosophy and Phenomenological Research
Volume LXV, No. 2, 2002

Precis of Understanding Truth

Scott Soames

Understanding Truth aims to illuminate the notion of truth, and the role it plays in our ordinary thought, as well as in our logical, philosophical, and scientific theories. Part one is concerned with substantive background issues: the identification of the bearers of truth, the basis for distinguishing truth from other notions, like certainty, with which it is often confused, and the formulation of positive responses to well-known forms of philosophical skepticism about truth. Part two explicates the formal theories of Alfred Tarski and Saul Kripke, including their treatments of the Liar paradox, and evaluates the philosophical significance of their work. Part three extends important lessons drawn from Tarski and Kripke to new domains: vague predicates, the Sorites paradox, and the development of a larger, deflationary perspective on truth.

Part one attempts to diffuse five different forms of truth skepticism, broadly conceived: the view that truth is indefinable, that it is unknowable, that it is inextricably metaphysical, that there is no such thing as truth, and the view that truth is inherently paradoxical, and so must either be abandoned, or revised. An intriguing formulation of the last of these views is due to Alfred Tarski, who argued that the Liar paradox shows natural languages to be inconsistent because they contain defective, and ultimately incoherent, truth predicates. I argue in response that on a plausible interpretation of his puzzling notion of an inconsistent language, Tarski's argument turns out to be logically valid, but almost certainly unsound, since one of its premises can be seen to be indefensible. Similar results are achieved for other forms of truth skepticism.

Once the grounds for truth skepticism are cleared away, I turn to Tarski's definition of truth. I indicate how he conceived the problem, how his criteria of adequacy guarantee that any definition satisfying them introduces a predicate that applies to all and only object-language truths, and how he approached the technical problem of formulating a definition that would allow him to derive what he regarded as a "partial definition" of truth for each sentence of the object language. Next I explain the formal techniques employed in his inductive definitions, the method of turning those definitions into explicit definitions (where possible), and the way in which his definitions can be shown to be materially adequate. The explication concludes with a discussion of truth and proof in the language of arithmetic, and the outlines of Tarski's theorem of the arithmetic indefinability of arithmetical truth. I also touch on his conceptions of quotation and quantification, and discuss his fallacious argument against using substitutional quantification into quotes to define truth.

I next turn to the question of whether Tarski's definition can be taken to be an analysis of truth. Since his formally defined truth predicates do not mean the same as our ordinary predicate, *is true*, it is clear that the presupposed conception of analysis cannot be one that requires synonymy of *analysandum* with *analysans*. I claim that what is required is that one's formally defined truth predicates be capable of playing the role of truth in all theoretical contexts in which that notion is needed. I argue that although Tarski's truth predicates can play many of the roles demanded of truth, they cannot play the role of truth in theories of meaning and interpretation. Crucial to the argument is my contention that there is a conceptual connection between our ordinary notions of truth and meaning, in virtue of which statements of the truth conditions of sentences provide some information about their meanings. Since there is no corresponding connection between Tarski's notion of truth and the notion of meaning, statements of the Tarski-truth conditions of sentences provide no information whatsoever about meaning. The explanation of this difference between ordinary and Tarskian truth predicates is, I

suggest, that whereas a sentence is understood to be true in the ordinary sense iff it expresses a true proposition, no such connection to propositions is built into Tarski's truth predicates.

I then turn to the Liar. Earlier, I explained how Tarski proved that the language of arithmetic does not contain its own truth predicate by showing that the claim that a language both satisfies certain minimal conditions and contains its own truth predicate leads to a contradiction. I now confront the puzzle that it seems obvious that does English satisfy the relevant conditions and while containing its own truth predicate. Of course this cannot be. In light of this I explore the well-known view that English is really an infinite hierarchy of languages defined by a hierarchy of Tarski-style truth predicates. I explain the construction and show how, on this view, different versions of the paradox are blocked. I then turn to problems with the approach, the most serious of which is the irresistible urge to violate the hierarchy's restrictions on intelligibility in the very process of setting it up. We tend to forget this because we imagine ourselves taking a position outside the hierarchy from which it can be described. However, when we realize that the hierarchy is supposed to apply to the language we are using to describe it, the paradox returns with a vengeance, threatening to destroy the very thing that is meant to avoid it. This prompts an investigation of a different approach that blocks the paradoxes by abandoning bivalence.

To this end I construct a philosophical model of partially defined predicates that would vindicate this idea. The crucial contention is that a predicate may be introduced into a language by rules that provide sufficient conditions for it to apply to an object, and sufficient conditions for it to fail to apply, but no conditions that are both individually sufficient and jointly necessary for it to apply, or for it to fail to apply. Because the conditions are mutually exclusive but not jointly exhaustive, there will be objects for which there are no possible grounds for accepting either the claim that the predicate applies to them, or the claim that it doesn't. I illustrate how a language could come to

contain such predicates, and I argue that there is a natural way of understanding the truth predicate in which it conforms to this model. On this view it turns out that there are sentences, including Liar sentences like *This sentence is not true* and "Truth Tellers" like *This sentence is true*, about which the rules determining whether or not a sentence is true provide no result. Because of this both the claim that such sentences are true, and the claim that they are not true, must be rejected, thereby blocking the usual paradoxical results.

As promising as this may seem, I argue that it does not constitute a general solution to the Liar paradox. The reason it doesn't has to do with the dynamic character of the paradox; the very activity of solving it in a particular case provides material for recreating it in a new and strengthened form. I show that even if the truth predicate in a particular language is partially defined in the manner envisioned, once speakers are in a position to describe it as such, and to explain how the original version of the Liar is thereby blocked, they will be in possession of conceptual material that can be used to restate the paradox in a form resistant to the original solution. While this does not, in my view, undermine the rationale for taking the ordinary truth predicate to be partially defined, it does limit what can be achieved, philosophically, by such an analysis.

All of this is put forward as the proper philosophical model for understanding Saul Kripke's formal theory of truth. In my view, the model incorporates and makes explicit essential philosophical insights guiding his formal construction. However, there are two complicating factors. First, Kripke's construction is compatible with different philosophical interpretations. Second, some of his own remarks invoke a philosophical model different from the one I suggest, with a markedly different interpretation of the gaps that characterize liars and truth tellers. I argue that in these remarks Kripke uncharacteristically misdescribes the real philosophical import of his own formal construction.

I put the model of partially defined predicates to further use in developing a theory according to which vague predicates are both partially defined and context sensitive. Given such a predicate, one begins with a set of things which the rules of the language clearly determine that it applies to -- its default determinate-extension -- and a set of things which the rules of the language clearly determine that it does not apply to -- its default determinate-antiextension. Since these sets don't exhaust all cases, speakers have the discretion of adjusting the extension and antiextension so as to include initially undefined cases. Typically this discretion is exercised by explicitly characterizing an object o for which the predicate is initially undefined as being F , or as being not F . When a speaker does this, and other conversational participants accept the speaker's characterization, the extension (or antiextension) of the predicate is adjusted so as to include o , plus all objects that bear a certain similarity relation to o . Often the relevant similarity relation involves perceptual indistinguishability, or something akin to it. Thus, when o is explicitly characterized as F , the extension of the predicate is adjusted so as to include o plus all objects indistinguishable, or virtually indistinguishable, from o .

These are the features of vague predicates upon which the Sorites paradox is built. Given the right kind of case, one can construct a chain $x_1 \dots x_n$, such that x_1 is definitely F , x_n is definitely not F , and each x_{i+1} bears the relevant similarity relation (of virtual indistinguishability) to x_i . In this sort of situation one can often create a sense of puzzlement by walking a subject through the sequence and asking whether successive items are F . One starts with the an initial characterization of x_1 as F . The rule governing application of the predicate to new cases will then determine that, by current conversational standards, x_2 , which bears the relevant similarity relation to x_1 , counts as F as well. When shown x_2 and asked to confirm this, the subject is merely being asked to explicitly characterize something as F that already counts as being F by the previously accepted conversational standards. Once the subject has done this, however, x_3 comes

implicitly to be characterized as F, and the process can be repeated all the way to x_n , which is not F.

This is the dynamic version of the Sorites paradox. I argue that there is nothing inherently contradictory or incoherent about it. What it shows is that the boundary lines fixing the extensions and antiextensions of certain predicates are inherently unstable. Although this is not a practical problem for speakers in ordinary situations, and although it does not represent any theoretical incoherence in the semantics of vague terms, it does explain the discomfort one feels when presented with the dynamic version of the Sorites. One is uncomfortable because one feels pressured into establishing stable boundary lines that cannot be maintained. The solution is to recognize that there is no requirement that the boundary lines for the application of vague predicates be stable in this way. There are, of course, other versions of the Sorites. I argue that the proposed theory provides plausible and instructive solutions to them as well.

The book closes with an attempt to articulate a broad philosophical perspective on truth incorporating important insights of Tarski and Kripke, without their artificial restrictions on the bearers of truth. Among these insights none is more important than their essentially deflationist conception of truth. For Tarski and Kripke truth is not a contentious metaphysical or epistemological notion, and a successful analysis of it should not be laden with controversial philosophical consequences. Rather, the content of the claim that a putative truth bearer is true is equivalent to that of the truth bearer itself, a fact that endows the truth predicate with an important practical and theoretical utility. As powerful and compelling as this idea is, I argue that there is no hope of maintaining it as long as the bearers of truth are taken to be sentences. The claim that *Snow is white* is a true sentence of English is simply not equivalent, in a sufficiently strong and interesting sense, to the claim that snow is white. Thus, if deflationism about truth is to be maintained, it must be restricted to propositions rather than sentences. Once this is recognized the relevant equivalence claim is extremely plausible: the proposition that

snow is white is equivalent, in an appropriately strong sense, to the proposition that it is true that snow is white.

Still, a number of important questions about the truth remain. Although plausible answers have been given to some, in my view no existing theory adequately answers them all. To illustrate this, I close by examining a range of deflationary theories of truth -- including the classical redundancy theory, Peter Strawson's performative theory, and Paul Horwich's minimal theory -- and attempt to separate what is correct, and worth preserving, in them from what is not.