Soames on the Logical Empiricists on Truth, Meaning, Convention, and Logical Truth

Mario Gómez-Torrente
Instituto de Investigaciones Filosóficas, UNAM, Mexico City

Scott Soames's new installment of his series *The Analytic Tradition in Philosophy*, its volume 2, subtitled *A New Vision*, is one more *tour de force* in the author's historiographic work. The volume devotes well over half of its pages to superb expositions and discussions of doctrines and ideas in philosophy of language, logic and science by adherents or fellow travelers of logical empiricism. In the very limited space I have it is of course impossible to cover every topic that seems to me of special interest in this part of the book. But when reading Soames, I much more often than not discover myself agreeing with him, so any attempt on my part to cover many of Soames's topics would result in rather unexciting reading anyway. Instead I will focus on a relatively small, but—I think—important set of historical and philosophical issues where I found myself disagreeing to some extent with Soames. These are issues surrounding Tarski's theory of truth and its connections with the notion of meaning on the one hand, and Quine's Carrollian objection to the conventionalist linguistic doctrine of logical truth, on the other.

1. Truth and Meaning

A well-known criticism of Tarski's theory of truth is that Tarskian defined truth predicates don't stand in the appropriate relations to the notion of meaning. For example, as Soames reminds the reader,

> instances of schema (33a), which contain our ordinary truth predicate, are obvious a priori truths, whereas instances of (33b), which contain a Tarskian truth predicate for L, are neither obvious nor knowable a priori.
>
> 33a. If s means in L that P, then s is true in L iff P.
>
> b. If s means in L that P, then s is T in L iff P.
>
> It is the obviousness and availability of (33a) that allows claims of the form *s is true in L iff P* to provide information about meaning... The unavailability of (33b) prevents similar conclusions from being drawn from claims of the form *s is T in L iff P*. (Soames 2018, 278)

(33b) is not necessarily available because a Tarskian truth predicate for L is just a defined predicate, introduced by stipulation as an abbreviation of a more or less complicated condition stated in terms of logic, set theory, and concepts of the metalanguage corresponding to concepts of L; so there is no necessary connection between facts of meaning about L and what stipulations some stipulator makes about a Tarskian truth predicate for L. Another manifestation of the stipulative nature of Tarskian truth predicates is that Tarskian T-biconditionals, appropriate claims of the form *s is T in L iff P*, are *a priori* and necessary (even when the corresponding claim of the form *s means in L that P* is true), being logico-mathematical consequences of stipulations, while corresponding "truth-biconditionals", appropriate claims of the form *s is true in L iff P*, are neither *a priori* nor necessary. A consequence of this is in turn that T-biconditionals cannot be said to "give the meaning" of sentences of languages like English or its fragments, and knowledge of those biconditionals cannot provide knowledge of meaning; if meaning is given by truth conditions, and knowledge of meaning by knowledge of truth conditions, these conditions have at best to be provided by truth-biconditionals, if only because the truth conditions provided by truth-biconditionals (unlike the conditions provided by T-biconditionals) are conditions that don't derive from mere stipulations, and that are appropriately contingently attached to the corresponding sentences.

Soames criticizes both Tarski and Carnap for not seeing any of this, and for thinking that Tarski's explication of the concept of truth can capture more than it really can as regards the connections between truth and meaning. While the criticism may be fair to some extent, I wonder exactly to what extent it is fair, and I suspect that it is not fair to a large extent; I will explain the grounds for my suspicions in what follows.

In the case of Tarski, the textual evidence Soames offers is scarce and is not always adequately presented in context. For example, Soames (2018, 248) recalls that in his 1944 expository paper on truth, Tarski says that his definition "does not aim to specify the meaning of a familiar word used to denote a novel notion; on the contrary it aims to catch hold of the actual meaning of an old notion" (Tarski 1944, 341), and takes this as evidence that Tarski thinks that a truth predicate defined via his method means almost the same as the ordinary truth predicate (as suitably restricted to a language fragment). But Tarski says this right after saying very clearly that he will consider a definition of truth satisfactory if it is *materially adequate* and formally correct (and nothing more), and the sentence quoted by Soames is offered by Tarski in anticipation of a possible objection that since there is no unique meaning of 'true', there need not be a unique extension of 'true' nor a unique way of satisfying the condition of material adequacy: "In order to avoid any ambiguity, we must first specify the conditions under which the definition of truth will be considered adequate from the material point of view" (Tarski 1944, *ibid.*, right before the sentence quoted by Soames). Thus, when Tarski speaks of "catching hold" of the meaning of 'true', it is appropriate to think, he is just thinking of catching hold of the extension, or the material adequacy conditions, *determined by the intended meaning* of 'true' (which, as he famously goes on to explain, is the meaning embedded in the "Aristotelian" conception of truth).

Soames also notes that Tarski "says in section 13 [of Tarski (1944)] that his notion of truth can be used to define semantic notions including *consequence*, *synonymy*, and *meaning*", and he adds: "He would not have said this had he not believed that his defined truth predicate for L comes very close to capturing the ordinary notion *being a true sentence of L*" (Soames 2018, 249). One problem with this inference is that, as is well known, Tarski was always very clear that the notions defined by his methods were only meant to recover a certain limited preferred core of the corresponding ordinary notions. The notions of consequence, synonymy and meaning that he has in mind may have been conceptually far away from the ordinary notions, and he may have been perfectly aware of this. In the case of consequence, it is fairly clear that he was so aware.[1] In the case of synonymy and meaning, he refers us to definitions given in Carnap (1942), as Soames notes. But the definition of meaning given here is just a definition of 'designation' (denotation) (Carnap 1942, 31-2, 50-1), and the definition of synonymy (*ibid.*, 55) is just a definition of codesignation. Both Tarski and Carnap (see e.g. Carnap's discussion *ibid.*, 54-5) were evidently aware that the ordinary notions of meaning and synonymy are non-extensional notions, and yet they use 'meaning' and 'synonymy' in these contexts for purely extensional notions, apparently because they have in mind the thesis or hypothesis that the meaningful or at any rate preferred core of those notions is extensional. This is a thesis that Tarski always adhered to, while Carnap was always more sympathetic to the idea of explicating non-extensional aspects of

[1] In his well-known paper on the concept of logical consequence, Tarski says:

> With respect to the clarity of its content the common concept of consequence is in no way superior to other concepts of everyday language. Its extension is not sharply bounded and its usage fluctuates. Any attempt to bring into harmony all possible vague, sometimes contradictory, tendencies which are connected with the use of this concept, is certainly doomed to failure. We must reconcile ourselves from the start to the fact that every precise definition of this concept will show arbitrary features to a greater or less degree. (Tarski 1936, 409)

pre-theoretical notions, and he sought to pursue the idea, though only in later work culminating in *Meaning and Necessity*.

Carnap is certainly more expansive than Tarski about the alleged properties of his explications, and Soames takes some Carnapian claims, especially in Carnap (1942), as evidence for his critical assessment. Key here is this passage quoted by Soames:

> The rules [of a semantical system] determine a truth-condition for every sentence of the object language, i.e. a sufficient and necessary condition for its truth. In this way the sentences are interpreted by the rules, i.e. made understandable, because to understand a sentence, to know what is asserted by it, is the same as to know under what conditions it would be true. To formulate it in still another way: the rules determine the meaning or sense of the sentences. (Carnap 1942, 22)

Soames says that "this paragraph makes sense only if the truth conditions are stated using the ordinary truth predicate of sentences. If Tarski's defined truth predicate is intended, the remarks are absurd" (Soames 2018, 285); but since it is obvious from other claims made by Carnap in the surrounding text that he is making his claims about Tarskian truth predicates, Soames concludes that Carnap was badly confused about the ability of such predicates to capture the relations between the ordinary notion of truth and the notion of meaning. By contrast with Soames, however, I suspect that Carnap could not have been so badly confused here, and that attention to context again makes this idea reasonable.

Just a few pages before the passage quoted by Soames, Carnap distinguishes descriptive semantics ("the description and analysis of the semantical features either of some particular historically given language, e g French, or of all historically given languages in general" (Carnap 1942, 11) from *pure semantics*. Pure semantics is "the construction and analysis of semantical

systems" (*ibid.*, 12), and Carnap is very clear that his treatise "is devoted to *pure semantics*" (*ibid.*, 14), stressing that although "there will occasionally also occur examples referring to semantical features of historical languages", "these examples are in fact meant as referring to semantical systems which either are actually constructed or could be constructed in close connection with those languages" (*ibid.*, 14). Crucially, he is also absolutely clear that

> The rules of a semantical system *S* constitute, as we shall see, nothing else than a definition of certain semantical concepts with respect to *S*, e g. 'designation in *S*' or 'true in *S*'. Pure semantics consists of definitions of this kind and their consequences; therefore, in contradistinction to descriptive semantics, it is entirely analytic and without factual content. (*ibid.*, 12)

As he also puts it a bit later, "[in pure semantics] we lay down definitions for certain concepts, usually in the form of rules, and study the analytic consequences of these definitions" (*ibid.*, 13).

Carnap is thus perfectly aware that T-biconditionals, which only arise in his constructed or stipulated "semantical systems", are analytic (as he understands 'analytic'), and thus necessary and *a priori*. Given such a perfect awareness, it's hard to see how he could have thought that T-biconditionals could interpret or give the meaning of, say, English or a fragment of it, as these would be non-constructed languages, for which the matter of the truth conditions of their sentences would be a contingent issue. But as we have seen, it's also independently clear that that cannot really be what he thought. What he is saying in the passage quoted by Soames, given that his book deals exclusively with pure semantics, must of course be that T-biconditionals determine an interpretation for his *stipulated* semantical systems. Is that so absurd? *Prima facie* it doesn't sound so implausible as to say that T-biconditionals interpret (fragments of) natural

languages; semantic stipulations in a natural or "historical" language *are* analytic and somehow determine an interpretation for the expressions which are the subject of the stipulation: if I say (in English) *(I stipulate that) the (new) sentence 'ret' is true iff rain is wet* (where 'true' is the ordinary truth predicate), intuitively the sentence *'Ret' is true iff rain is wet* becomes analytic for me, even though by its means I do in a way come to provide an interpretation of 'Ret' for my own use and understanding. Now what Carnap is claiming is that (for example) a stipulation such as *(I stipulate that) 'Ret' is T in S iff rain is wet* (where 'T in S' is a Tarskian truth predicate) determines an interpretation for and provides the stipulator with an understanding of 'Ret'. And I don't think that's evidently absurd, either.

First, the analyticity of *'Ret' is T in S iff rain is wet* does not make the claim absurd, as we have illustrated. And second, the fact that a defined Tarskian truth predicate instead of the ordinary truth predicate is involved in the sentence *'Ret' is T in S iff rain is wet* does not make the claim absurd by itself. The stipulator who introduces a Tarskian truth predicate in English as a metalanguage works under many assumptions formulatable (and formulated) in English, among which are, in particular, things such as Tarski's "Convention T", which directs him to define a Tarskian truth predicate in such a way that in T-biconditionals there is a correspondence in meaning between the two sides (compare Carnap's extended discussion of Convention T as a part of the metatheory of a semantical system, on pp. 26ff. of Carnap 1942). For example, he introduces his stipulations along with intentions such as the intention that *'Ret' is T in S* mean the same as (and that 'Ret' be a translation of) *Rain is wet*. In this case, such an intention accompanies the stipulation *'Ret' is T in S iff rain is wet* if 'T in S' is a Tarskian truth predicate, and thus provides an interpretation for 'Ret' concurrently with the stipulation. (Note that sentences of Soames's form (33b), such as *If 'Ret'*

means in S that rain is wet, then 'Ret' is T in S iff rain is wet, *are* analytic and *a priori for the stipulator*.) It's then not implausible at all to suppose that Carnap, in the passage quoted by Soames, is saying that T-biconditionals provide the stipulator with an interpretation for the relevant sentences, given that they arise in the context of semantical systems subject to a number of assumptions essential to the development of such systems, such as the adequacy condition that we know as "Convention T". Surely someone other than the stipulator, aware only of the stipulation *'Ret' is T in S iff rain is wet* by itself, will not know what the meaning or truth conditions of 'Ret' are in S, but we don't need to read Carnap as thinking otherwise. Only if Soames could show that Carnap meant to claim that the stipulative sentence *by itself* interprets 'Ret' would we have proof that he was badly confused; but as far as I can tell, Soames has not shown that.

A different matter is whether other authors, such as Davidson, may have been confused in the way described by Soames, i.e. in thinking that T-biconditionals can by themselves interpret object language sentences, including sentences of natural language fragments. On this question I agree with Soames that everything points to Davidson having been indeed confused, at least in his early publications on truth and meaning. But I think we don't need to see a Davidsonianism *avant la lettre* in Carnap's relatively unambitious claims about Tarskian truth predicates in stipulated semantical systems.

2. Convention and Logical Truth

Soames is often quite critical of the errors or insufficiencies that one can see with hindsight in the work of the authors he studies. But there are some exceptions. One of them is provided by his exposition of Quine's (1936) celebrated Carrollian argument against the conventionalist doctrine of logical truth, which he, along with the vast majority of commentators in the analytic tradition, takes

to constitute a "powerful critique of the program of grounding a priori knowledge in knowledge of meaning" (Soames 2018, 306). Quine's argument may be powerful, and I myself don't have any sympathies with the conventionalist doctrine, but I have the impression that there have only been a few half-hearted attempts to defend conventionalism against the argument, and that we could only benefit if more attention were given to this issue. Recently, as Soames notes, several authors have defended the view that, against the common impression, Carnap was not the target of Quine's argument and/or Carnap did not hold the doctrine criticized by Quine. But, regardles of the merits of these exegetical claims, they do not question the power of the argument against the conventionalist doctrine as specified.

Perhaps the conventionalist doctrine that Quine specifies is indeed refuted by his argument. Quine considers a conventionalist view according to which the set of logical truths is actually infinite, and according to which the conventions that have made it the case and that can justify (for the conventionalist) that each one of that infinity of logical truths are true are explicit general conventions that have made this the case and have justified it by *implying* that infinity of truths or having them as logical consequences ("Logical truths, being infinite in number, must be given by general conventions rather than singly; and logic is needed then to begin with, in the metatheory, in order to apply the general conventions to individual cases" (Quine 1960, 108)). But that creates the problem for the conventionalist that he is appealing to facts of implication whose metaphysical or epistemic basis is still unaccounted for; if he then proposes that these facts are no more than additional conventions, a regress of Carrollian reminiscences begins:

> In the adoption of the very conventions ... whereby logic itself is set up, ..., a difficulty remains to be faced. Each of these conventions is general, announcing the truth of every

> one of an infinity of statements conforming to a certain description; derivation of the truth of any specific statement from the general convention thus requires a logical inference, and this involves us in an infinite regress. (Quine 1936, 96)

The objection does seem definitive against conventionalism as specified. Quine also considers a variant, one in which the thesis that the general logical conventions are explicitly "adopted" or "set up" is abandoned, and it is proposed that they are instead simply "observed" implicitly. To this Quine objects that

> [i]n dropping the attributes of deliberateness and explicitness from the notion of linguistic convention we risk depriving the latter of any explanatory force and reducing it to an idle label. We may wonder what one adds to the bare statement that the truths of logic and mathematics are a priori, or to the still barer behavioristic statement that they are firmly accepted, when he characterizes them as true by convention in such a sense. (Quine 1936, 99)

The objection, in other words, is that, unlike in the case of "explicit" conventionalism, here we would not have a distinction between general logical conventions and other firmly accepted, "implicitly observed" general propositions that need not be true or may even be false.

Quine's objection here doesn't have the air of "definitiveness" of the objection against "explicit" conventionalism, for it does nothing to suppress the possibility that some characteristic other than "deliberateness and explicitness" may distinguish the implicit general logical conventions from other firmly observed general propositions; that such a characteristic exists is not obviously impossible, and in fact some authors have pursued this line of response to Quine on behalf of conventionalism. However, I think a more effective Quinean objection would be as follows. Any sensible kind of "explicit" conventionalism

(including the doctrines that Quine's argument is directed against), is not a historical thesis, but a thesis about the "context of justification". A sensible conventionalist doesn't hold that anyone at any moment has stipulated the conventions she talks about, but merely that, were one to do so, an appropriate set of sentences would become (logical) truths and the ground for an (*a priori*) justification of each one of those truths would be laid down. The Quinean problem with "explicit" conventionalism is that it fails even as a thesis about the context of justification. Now the "implicit" conventionalist doesn't seek to avail herself of the distinction between context of justification and historical fact. She holds that as matter of historical fact some general logical convention is "observed" from the start. That this "observing" obtains must then imply that the appropriate set of sentences become (logical) truths and that the ground for an (*a priori*) justification of each one of those truths exists as a matter of fact. But if the fact about conventions being observed consists in that the infinity of logical truths (of some general form or forms) are at some historical point asserted as true by someone, "implicit" conventionalism is evidently false, for no such fact exists. If it consists in that some finite subset of that infinity of logical truths are at some historical point asserted as true by someone, this fact is evidently insufficient to determine that the infinity of logical truths are true. If it consists in that some finite fact of this kind yields, together with other facts, the truth of the infinity of logical truths, we need to know what other facts these are; but these cannot be merely facts that some finite subset of logical truths are at some historical point asserted as true by someone, and they cannot be merely facts that some general logical conventions have been at some point asserted by someone, for the reasons made clear in Quine's argument against "explicit" conventionalism. But it appears that no other, "richer" facts are available to the conventionalist.

Be this as it may, I think yet a different variant of conventionalism, not even mentioned by Quine, needs to be explicitly considered, namely a variant where what is abandoned is the assumption that the set of logical truths is infinite. We might call this "finitary" conventionalism. The finitary conventionalist might argue as follows. Undoubtedly no infinite set of sentences are (logically) true by convention at any particular moment. But one could adopt conventions about what is (logically) true as one thinks of new particular logical truths. Even if no infinite set of truths are in fact true by convention at any particular moment, there is no obstacle to a *potentially* infinite set of logical truths. This is all that is required by the conventionalist intuition that the logical character of certain sentences can be explained as if it originated in the introduction of certain conventions, and that the peculiar epistemic nature of those sentences can also be explained in this way.

The finitary conventionalist insists, then, that Quine's argument presupposes that the criticized conventionalist position requires it to be the case that the relevant conventions determine the existence of an actually infinite set of logical truths. Quine observes, correctly, that this could only happen in virtue of non-conventional facts of some kind, for example infinitary mathematical facts about the derivability of the logical truths from basic axioms via the application of basic rules, or simply about the truth of an infinite set of basic axioms. But the finitary conventionalist rejects any appeal to such facts, and quite understandably. Surely it is part of the spirit behind the conventionalist position to reject appeals to logical or mathematical facts that are not grounded in human decisions or constructions. And how could an infinitary fact be grounded in human decisions or constructions? The finitary conventionalist points out that, in denying the existence of such facts, he is doing greater justice

to the conventionalist spirit than the conventionalists of the kind Quine seeks to refute.

It's not difficult to see that such a position would share in a certain spirit that fed many historical conventionalist and constructivist views of the relevant period, even if Carnap's conventionalism was not finitary. Finitary conventionalism is in line with the ontological finitism of some logical empiricists and fellow travelers, such as Neurath and Tarski. And it is also in line with the radical conventionalism and constructivism about logical and mathematical truth of (at least some readings of) the Wittgenstein of the 1930's, according to which every use or acknowledgment of a new sentence as a logical truth is ultimately no more than a basic (unjustified) new convention—similarly, for the finitary conventionalist, logical truths come to be known as they get to be individually stipulated as true by the person, or the community, that sets up the corresponding conventions.[2]

These are just a few indications that finitary conventionalism needs to be given more attention than it has been given. I hope that if more attention is given to it, we will get a better assessment of the scope and limits of Quine's critique of conventionalism about logical truth than we have had so far.

[2] There also analogies between the finitary conventionalist position and classical mathematical constructivism. Just as the constructivist holds that there is no actual infinite set of mathematical truths (or any other actually infinite set, of course), the finitary conventionalist holds that there is no actually infinite set of logical truths. And just as the Brouwerian constructivist holds that mathematical truths get to be known as they get to be constructed by the ideal mathematician's mind, the finitary conventionalist holds that logical truths come to be known as they get to be individually stipulated as true by the person, or the community, that sets up the corresponding conventions. (There are other aspects where the analogy breaks down. For example, it is surely part of the finitary conventionalist position that there is a good degree of freedom the stipulator has when choosing the sentences stipulated as true, while it is surely a part of the Brouwerian position that what the ideal mathematician's mind can construct is not similarly unconstrained.)

References

Carnap, R. (1942), *Introduction to Semantics*, Harvard University Press, Cambridge (Mass.).

Quine, W. V. (1936), "Truth by Convention", in O. H. Lee (ed.), *Philosophical Essays for A. N. Whitehead*, Longmans, Green and Co., New York. Reprinted in Quine, *The Ways of Paradox and Other Essays*, Harvard University Press, Cambridge (Mass.), 1966, 70-99. References to the reprint.

Quine, W. V. (1960), "Carnap and Logical Truth", *Synthese* 12. Reprinted in Quine, *The Ways of Paradox and Other Essays*, Harvard University Press, Cambridge (Mass.), 1966, 100-125. References to the reprint.

Soames, S. (2018), *The Analytic Tradition in Philosophy. Volume 2. A New Vision*, Princeton University Press, Princeton (N.J.).

Tarski, A. (1936), "On the Concept of Logical Consequence", in Tarski, *Logic, Semantics, Metamathematics*, 2nd edn., Hackett, Indianapolis, 1983, 409-420.

Tarski, A. (1944), "The Semantic Conception of Truth. And the Foundations of Semantics", *Philosophy and Phenomenological Research* 4, 341-376.