Tarski’s Convention T continues to have a pervasive impact on philosophical thought about truth. Here is how Tarski states it:

A formally correct definition of the symbol ‘Tr’, formulated in the metalanguage, will be called an adequate definition of truth if it has the following consequences:

(α) all sentences which are obtained from the expression ‘x ∈ Tr if and only if p’ by substituting for the symbol ‘x’ a structural-descriptive name of any sentence of the language in question and for the symbol ‘p’ the expression which forms the translation of this sentence into the metalanguage;

(β) the sentence ‘for any x, if x ∈ Tr then x ∈ S {x is a sentence}’ (in other words ‘Tr ⊆ S’).

Its influence is surely deserved, but I believe nevertheless that there is a great deal of confusion about the importance of Convention T. It is commonly assumed that a definition of truth—or, more generally, any good theory of it—must imply, for the sentences of a given language, a set of biconditionals of the form delimited in clause (α), in which the sentences embedded on the right-hand side are equivalent in meaning to those mentioned on the left-hand side.2 In §1 I will draw and justify some distinctions, and in particular stress the importance of distinguishing the claim that it is sufficient for adequacy in a theory or definition of truth that it imply biconditionals of the sort mentioned in clause (α) from the claim that it is necessary for adequacy in a theory or definition of truth that it do so. I will then discuss the role of

---

1Tarski (1956), pp. 187-188.
2That is, as Tarski’s has it, where “p” is a translation of x—not, it should be noted, of “x ∈ Tr”, which would invite an excessively strong reading of the biconditional.

Douglas Patterson is Assistant Professor of Philosophy at Kansas State University.
the sufficiency claim in currently popular deflationary theories of truth. In §2 I will explain that inflationary theories of truth, by contrast, should not be required to imply such biconditionals, taking a typical if somewhat schematic correspondence theory of truth as my example. I will then comment in §3 on why the claims of §2 are so often missed, and in §4 on what the results established there tell us about truth and the debate between deflationary and inflationary or substantivalist theories thereof.

Two notes on Tarski are in order. First, many will have noticed that Convention T as written says “if” and not “if and only if”: it apparently makes only the sufficiency claim. However in my experience logicians read it as making both claims, citing the definitional-sounding “will be called” and also footnote 1 to page 188 of Tarski (1956), which claims that were only the metalanguage formalized, Convention T would be a definition of “materially adequate definition of truth”. I address Tarski’s views elsewhere; here I will accept Convention T as it is commonly read. My own view is that although Tarski was favorably inclined toward the necessity clause, his important results do not depend on it. However, this is a paper on how Convention T has been received and not, for the most part, one on Tarski himself.

Second, we should wonder what notion of implication is at issue in Convention T; Gupta points out that:⁶

Convention T, as stated above, is not fully precise; it does not delineate the notion of implication that it invokes. The notion plainly cannot be that of logical implication. For that makes Convention T much too strict: even Tarski’s own definition (in CTFL) fails to be materially adequate. The relevant notion of implication is, therefore, one that allows the use of some non-logical resources. But what resources are allowed—analytical truths, mathematical truths, or something else? Tarski himself does not provide, as far as I know, an explicit answer to the question. In practice he allows himself the use of limited syntactic information about the object language and some simple principles of higher-order logic (equivalently, set theory).

The point is quite well taken. Gupta himself addresses it by allowing the implication of the appropriate instances of the schema to be logical implication by a proposed definition plus an additional set of sentences (subject to constraints that bar circularity). Fully applying this strategy requires a view about what logical as opposed to non-logical implication is, which would introduce a host of complications into our present discussion. In what follows I will assume roughly the sorts of principles Tarski does without explicitly considering the point further.⁴ We should also note explicitly that the implication at issue in Convention T is not material, since if it were, the necessity claim of Convention T would be trivial: every definition or theory of truth is such that either it is not correct, or the instances of the disquotation schema are correct.⁵

§1. Deflationary Theories of Truth and Convention T

The basic idea behind Convention T is that, given that certain instances of the T-Schema

(T-Schema) s is true in L if and only if p

namely those where the sentence substituted for “p” translates s, are definitions of truth for particular sentences, a definition that implies all of them is a definition of truth for all of the sentences of a language. Tarski writes:⁶

⁵Belnap (1993), p. 132, makes a related point.
Douglas Patterson

In order to make clear the content of the concept of truth in connexion with some one concrete sentence of the language with which we are dealing we can apply the same method as was used in §1 in formulating the sentences (3) and (4). We take the scheme (2) \( \{ x \text{ is a true sentence if and only if } p \} \) and replace the symbol ‘\( x \)’ in it by a name of the given sentence, and ‘\( p \)’ by its translation into the metalanguage. All sentences obtained in this way... naturally belong to the metalanguage and explain in a precise way, in accordance with linguistic usage, the meaning of phrases of the form ‘\( x \) is a true sentence’ which occur in them.

He continues:  

Not much more in principle is to be demanded of a general definition of true sentence than that it should satisfy the usual conditions of methodological correctness and include all partial definitions of this type as special cases; that it should be so to speak, their logical product.

Convention T is then introduced as a restatement of “this postulate”, as Tarski calls it. The idea is that since 

\[ s \text{ is true in German if and only if snow is white} \]

defines “is true” for German when \( s \) is “Schnee ist weiss”, in that it allows “is true” to be eliminated relative to the contexts and prior theory in which Tarski is interested, then something that implied such a biconditional for each sentence of German would be an adequate definition of truth for German: it would allow “is true” to be eliminated from any context in which it appears, and would, if properly constructed, be conservative. That, then, is the attraction of Convention T—it states a criterion of adequacy on the definition of truth: just construct something that implies the instances of the T-Schema mentioned in clause \( (\alpha) \), and truth has been defined.

Though Tarski explicit formulates Convention T to apply to definitions of truth according to his conception of definition (one that stresses eliminativity and relativization to a prior theory and set of contexts) it is commonly taken to apply more generally to accounts of truth of various sorts, and for our purposes here we may accept this generalization.  

Whatever one takes as one’s theory or account of truth—a tidy universally quantified biconditional like “\( \forall s, s \text{ is true if and only if } s \text{ corresponds to a fact} \)” a list of instances of the Disquotation Schema, an ineliminative circular definition, or what have you—Convention T so generalized says that certain of its consequences determine whether or not it is a good theory or account of truth.

There are, however, two sorts of difference that implication of the instances delimited in clause \( (\alpha) \) could make. Implication of such instances could be sufficient or could be necessary for a theory’s being an adequate theory of truth. In order to facilitate discussion, let us restate Convention T, dropping explicit mention of clause \( (\beta) \), since it isn’t at issue, and separating the necessity and sufficiency claims for clarity. We may also add a bit of generality by replacing the notion of translation with the notion of equivalence in meaning.  

**Convention T**  

\[ \text{(sufficiency)} \text{ A theory of truth for a language } L \]  
\[ \text{is adequate if it implies for each sentence } s \text{ of } L \]  
\[ \text{an instance of the T-Schema such that the used} \]

---


8That is, the theories of truth considered in the current literature on truth generally aren’t explicitly relativized to a prior theory and set of contexts, but Convention T is still taken to concern them. Belnap (1993) provides an excellent discussion of what is specific to definitions as Tarski construes them.

9Tarski himself speaks of equivalence in meaning in place of translation on occasion; see, e.g., Tarski (1956), pp. 168-169.
Douglas Patterson

sentence \([p]\) is a metalanguage sentence equivalent in meaning to \(s\).

(necessity) A theory of truth for a language \(L\) is adequate only if it implies for each sentence \(s\) of \(L\) an instance of the T-Schema such that the used sentence \([p]\) is a metalanguage sentence equivalent in meaning to \(s\).

The difference between the sufficiency and the necessity clauses is often obscured in discussions of Convention T. Soames, for instance, writes: ¹⁰

The importance of schema T for Tarski is further illustrated by his characterization of individual instances like ‘Snow is white’ is true if and only if snow is white) as “partial definitions” of truth... If such instances are thought of as partial definitions, then the task of defining truth for an entire language may be seen as that of finding a way of generalizing the “partial definitions” so as to cover every sentence of the language.

Tarski expressed this idea by requiring that a definition of truth must logically entail an instance of schema T for each sentence of \(L\).

The “must” here is crucially obscure. Tarski’s idea is that since an instance of the T-Schema where \(s\) is equivalent in meaning to the sentence substituted for \([p]\) is an adequate definition of truth as applied to a single sentence, a definition that implies such biconditionals for all the sentences of a language is an adequate definition of truth for that language. All this establishes, however, if accepted, is that it is sufficient for adequacy in a definition of truth that it imply such T-

biconditionals; whereas Soames’s “must” suggests necessity.¹¹ Likewise, Marian David writes: ¹²

Tarski goes on to suggest that an adequate definition of truth for sentences should have the biconditional

\[(1) ‘Snow is white’ is a true sentence if and only if snow is white\]

as one of its consequences.

This is ambiguous as well: does it mean “should, in order to be adequate” or “should, in order to avoid inadequacy”? We get our answer a page later when David writes: ¹³

Impressed by Tarski’s observation, one might propose that for a definition of sentence-truth to be adequate it is necessary that it implies all the instances of (T) (the T-Schema).

and then argues that deflationism consists in accepting the sufficiency claim as well.

This lack of clarity is unfortunate, because the necessity and sufficiency clauses stand in very different relations to the current state of thought about truth. In general deflationary theorists of truth endorse the sufficiency clause of Convention T. Leeds, for instance, writes: ¹⁴

It is not surprising that we should have use for a predicate \(P\) with the property that “___” is \(P\) and “___” are always interdeducible. For we frequently find ourselves in a position to assert each sentence in a certain infinite set \(z\) (e.g. when all the members of \(z\)

¹⁰Soames (1999), pp. 68-69. As we have seen, Gupta points out that it is a mistake to claim that Tarski thought that a definition should logically entail the biconditionals (Gupta [2002], p. 3).

¹¹On the preceding page Soames makes clear that he takes Tarski to be committed both to sufficiency and to necessity. The point here is that the “must” obscures the fact that the claims about partial definition can support only the claim that implication of the biconditionals is sufficient.


Douglas Patterson

share a common form); lacking the means to formulate infinite conjunctions, we find it convenient to have a single sentence which is warranted precisely when each member of \( z \) is warranted. A predicate \( \text{P} \) with the property described allows us to construct such a sentence: \( (x) (x \in z \rightarrow \text{P}(x)) \). Truth is thus a notion that we might reasonably want to have on hand, for expressing semantic ascent and descent, infinite conjunction and disjunction. And given that we want such a notion, it is not difficult to explain how it is that we have been able to invent one: the Tarski sentences, which axiomatize the notion of truth, are by no means a complicated or recondite axiomatization; the possibility of moving from this axiomatization to the explicit truth definition was always latent in the logical structure of language, though it took a Tarski to discover it.

Truth is useful, we may say, as a device of (what Quine calls) disquotation... To explain the utility of disquotation we need say nothing about the relations between language and the world.

Here the view is that a good theory of truth just is one that “axiomatizes” instances of the Disquotation Schema for a language: we have what we want in an account of truth when we have something that implies certain instances of the T-Schema where the used sentence is equivalent in meaning to the mentioned one. This just is the sufficiency clause. Likewise, Horwich’s account of truth for utterances is as follows:\(^{13}\)

We have an inclination to accept any instance of

\[ (\text{D}) (u \in {}^{*}p^{*}) \rightarrow (u \text{ is true if and only if } p) \]

which, together with the disquotation schema (D), yields the more general schema

\[ (\text{DT}) (\text{Int}(u) \in {}^{*}p^{*}) \rightarrow (u \text{ is true if and only if } p) \]

Our grasp of truth for utterances is constituted by our inclination to accept instances of this schema. Moreover, these instances are the axioms of the minimal theory of that property.

Here an adequate account of truth for utterances is one that takes as axioms instances of the T-Schema, this time with the requirement of equivalence in meaning explicit in the antecedent of a conditional rather than implicit in the restriction to the disquotation schema. Horwich’s account does little more than imply instances of the T-Schema where the sentence used to state an utterance \( u \)’s truth-condition is equivalent in meaning to (is the interpretation of, on Hor-
Douglas Patterson

which’s account of “interpretation”) \( u \) itself. It is an assumption of the claim that this is a sufficient account of truth that a theory that is more or less the “logical product”, in Tarski’s phrase, of such instances is an adequate account. This, again, just is the sufficiency clause of Convention T.

Examples could be multiplied. In general, deflationists are happy to take a set of instances of the T-Schema or the more restrictive Disquotation Schema as a sufficient account of truth for a language. It is this commitment of deflationary theories on which I will focus in what follows. Though I won’t press the point here, I am of the opinion that endorsement of the sufficiency clause is in fact the clearest criterion of deflationism available—one clearer and more tractable than the usual talk about whether or not truth is a “real” property. For present purposes, however, it will suffice to understand that the sufficiency clause is of contemporary relevance because it is an assumption of so many presentations of the very popular deflationary view. The reader who remains unconvinced that her favorite form of “deflationism” is committed to the sufficiency clause may read the rest of this paper as a commentary on the independent issue of the status of Convention T and those forms of “deflationism” that are committed to it.

On a conception of truth that incorporates the sufficiency clause, truth is to be explained in terms of certain instances of the T-Schema, those in which the sentence substituted for “p” is equivalent in meaning to \( s \). The appeal to equivalence in meaning is crucial to the position, since not all instances of the T-Schema would make for an adequate account of truth. The instances of the “disquotation-negation schema”

\[[p] \text{ is true in } L \text{ if and only if } \sim p\]

such as “snow is white” is true in English if and only if snow is not white” are, after all, instances of the T-Schema, though nobody would suggest that an expression defined in terms of them would be a truth-predicate.

Instances of the T-Schema, however, where the sentence substituted for “p” is equivalent in meaning to \( s \) are suited to do the job the deflationist demands of them: when sentence \( s \) in language \( L \) really is equivalent in meaning to the English [p], then it will be the case that \( s \) is true in \( L \) if and only if \( p \), a point summed up in the meaning-truth schema

\[(MT) \text{ If } s \text{ means in } L \text{ that } p \text{ then } s \text{ is true in } L \text{ if and only if } p.\]

Disquotationalism singles out such instances by placing a syntactic restriction on them that guarantees equivalence in meaning between the sentence substituted for “p” and \( s \) (subject to a ban on heterophony and various context-dependent semantic features), while other forms of deflationism get the equivalence in other ways: the presentential theory, for instance, guarantees that an instance of the T-Schema where “s is true” inherits its content from a tokening of the sentence substituted for “p” is fit for inclusion in an account of truth, since for one sentence-tokening to be anaphorically dependent on another is for it to inherit its content from that other.\(^{16}\)

On all of these views, the reason for singling out the instances just so is to produce equivalence in meaning between the sentence substituted for “p” and \( s \). If quotations referred to the negations of the sentences they embed, disquotationalists would insist that their theories consisted of the instances of the disquotation-negation schema. Hence though the notion of meaning isn’t explicitly in play in every deflationary theory’s identification of the relevant instances, it does in every case play a role in the justification of the preferred method of selection.\(^{17}\)

\(^{16}\) In order for this to hold good if “p” is context-sensitive, the tokening of it on which “s is true” depends must occur in the same context as the instance of the T-Schema.

\(^{17}\) The point is sometimes obscured by the practice of talking about “the” instances of the T-Schema as being all and only those where \( s \) is equivalent in
Thus, in all these cases, Convention T relies upon an antecedent notion of meaning. It follows from this that if we accept a theory of truth because we accept Convention T and accept that the theory satisfies it, we cannot explain meaning in terms of truth-conditions on pain of undermining the interest Tarski and others take Convention T to have. This can be established as follows. Convention T tells us what it is sufficient to do in order to introduce, into a metalinguistic theory for some object language, a truth-predicate for the sentences of that object language. Convention T says, using a meta-metalinguistic predicate “object language sentence $s$ is equivalent in meaning to metalanguage sentence $[p]$”, that it is sufficient for a predicate introduced into the metalanguage to be a truth-predicate for the object language that the definition of that predicate imply instances of the T-Schema where the metalanguage sentence substituted for “$p$” is equivalent in meaning to the object language sentence $s$.

Now the question is: under what conditions could we introduce such an equivalence-in-meaning predicate itself? There are no doubt many candidates, and Tarski says nothing on the topic. However, we can note this: if introducing the equivalence-in-meaning predicate requires us to have a meaning to $[p]$. Davidson, for instance, writes “if we characterize the T-Sentences by their form alone, as Tarski did, it is possible, using Tarski’s methods, to define truth using no semantical concepts” (Davidson [1984], p. 74). “The T-Sentences” for a language are clearly supposed to be those delimited by clause $(a)$ of Convention T. Tarski does not, however, “characterize them by their form alone”: he characterizes them by their form and the (presumably semantic) fact that the used sentence translates the mentioned sentence. Soames likewise understands “Schema T” as having only instances where $[p]$ is a “paraphrase” of $s$ (Soames [1999], p. 246, and footnote 15 thereto).

18 Some sort of enumerative, dictionary-like mapping seems most in the spirit of Tarski’s approach. See Mou (2001) for a development of the idea. Such an approach raises questions about the meta-meta-metalinguistic treatment of the conditions under which a meta-metalinguistic translation mapping of object language sentences onto metalanguage sentences is itself adequate. I will spare the reader a discussion of this.

19 Convention T might also assume that a truth-predicate for the object language is available as a primitive in the meta-metalanguage; this modification would not affect the point in the text that this would undermine the interest of Convention T.

20I here neither assert nor deny that this clause would be easy to formulate, or that it would be needed.
object language in order to introduce the notion of equivalence in meaning that is used by Convention T in stating a condition of adequacy on a definition of truth for the object language.

Showing what it takes to define truth for a language in one metalanguage by assuming that the task is already accomplished in another is hardly what actual philosophers have wanted out of a condition of adequacy on a definition of truth. Tarski and others like him have assumed that a condition of adequacy on a definition of truth for a language is of interest only if it doesn’t assume that truth is defined for that language in some other way, since they have been interested, not merely in introducing a truth-predicate for a language into some language regardless of whether one already exists in others, but rather in what it takes to introduce such a predicate on the assumption that one is not antecedently available at all. Tarski surely would not have been satisfied if he had found that he had to explain translation in terms of truth. Likewise, the modern day deflationist would be chagrined to learn that in fact the notion of meaning appealed to in his or her theory (e.g., for instance, the proponentialist’s notion of the content that is inherited anaphorically by “s is true”) had to be explained in terms of truth and truth conditions.

Technically there might be some interest in defining truth so as to imply the biconditionals while assuming the availability of a truth-predicate in some meta-metalanguage, but the definition of truth in question would be bled of the philosophical import that definitions implying the biconditionals are generally taken to have. Hence, in order to have this import, views that take implication of instances of the T-Schema, where s is equivalent in meaning to the sentence substituted for “p”, to be sufficient for adequacy in a theory of truth must refrain from explaining meaning in terms of truth-conditions. Conceptions of truth that include the sufficiency clause that are intended to have such import, including many if not all deflationary conceptions, must therefore be packaged with some non-truth-conditional account of meaning. Fortunately, many such alternative conceptions of meaning are on offer: Field (1994) surveys and combines a range of possibilities, including theories that treat meaning in terms of verification conditions, indication relations, and conceptual roles and Brandom (1994) works out in great detail an account of meaning in terms of inference, to mention just two of many examples. The proponent of the sufficiency clause must pin her hopes on at least one such theory.

§2. Inflationary Theories of Truth and Convention T

The sufficiency clause of Convention T, supplemented with an adequate account of meaning in terms of something other than truth-conditions, offers an attractive conception of what a theory of truth ought to do, one that deflationary theories admirably (indeed, often trivially) meet. The idea that implication of the instances of the T-Schema in question is the central goal of a theory of truth captures the intuition that such biconditionals “axiomatize” the notion of truth, that the instances of the disquotation schema codify the “transparency” of the truth-predicate, and so on. These are important and common views about truth, and the sufficiency clause does them justice.

What should we make, by contrast, of the necessity clause, the claim that any good theory of truth must imply instances of the T-Schema where [p] is equivalent in meaning to s? It is important to pause for a moment and reflect
on the view expressed in this clause as to what a theory of truth should do. What it requires is that a theory or definition of truth encode particular semantic information about the sentences of a particular language, since the instances of the T-Schema in question express not merely the truth-conditions but also the meanings of the sentences they mention. They may, of course, do this even if the metalinguage itself lacks an “is equivalent in meaning to” predicate; the point is that were the metalinguage enriched to include such a predicate, the truth of the theorems would be preserved when we substituted “is equivalent in meaning to” for “is true if and only if”. This is a rather counterintuitive view: why should saying what truth is require implications sufficient, given the introduction of an “is equivalent in meaning to” predicate, for saying to which metalanguage sentences various object language sentences are equivalent in meaning? It is one thing to say that a theory that encodes a certain kind of information about meaning establishes what truth is as well, and quite another to say that any theory that shows what truth is will also encode information about the meanings of the sentences of some particular language.

Counterintuitive when examined or not, the necessity clause of Convention T is commonly but tacitly assumed in discussions of truth, as we saw above in passages from David. Another example can be seen in this quotation from Horwich:

> The basic thesis of deflationism, as I see it, is that the disquotation schema
>
> “p” is true if and only if p
>
> is conceptually fundamental. By this I mean that we accept its instances in the absence of supporting argu-

---


---

Theories of Truth and Convention T

... more specifically, without deriving them from any reductive premise of the form

\[ x \text{ is true} = x \text{ is } F \]

which characterizes any of the traditional (‘inflationary’) accounts of truth, such as the correspondence theory, the coherence theory, the verificationist theory, and the pragmatist theory.

Here it is presupposed that an inflationist does try to derive instances of the disquotation schema from her theory of truth.24 This is tantamount to the idea that an inflationary theory of truth is intended to encode information about the meanings of the sentences of some particular language. But there is no reason to think that an inflationary theory will be like this: a theory of what truth is will tell one that truth for each sentence of a language consists in the instantiation, for the sentence in question, of whatever property the theory identifies with truth, but this of itself need not give one enough information to specify a truth condition for the sentence in question in a way that could also figure as a specification of its meaning.

In any case, inflationary theorists of truth must reject the necessity clause, since their theories simply don’t imply the biconditionals. Suppose we did endorse the necessity clause of Convention T, maintaining that an inflationary theory of truth must imply the biconditionals, and assume for the sake of argument that our inflationary theory is a simple correspondence theory:

\[ \forall x, x \text{ is true if and only if } x \text{ corresponds to a fact.} \]

The idea, then, is that we have to show, assuming our account of correspondence, for each sentence s of a given lan-

---

24I assume here that Horwich assumes that roughly the resources Tarski assumes are available: limited syntactic information and set theory.
guage, that $s$ corresponds to a fact if and only if $p$, where $[p]$ is equivalent in meaning to $s$—for from this, by the theory, it will follow that $s$ is true if and only if $p$. How will we proceed?

Consider the sentence $s$. By universal instantiation of our correspondence theory we have:

(1) $s$ is true if and only if $s$ corresponds to a fact.

Now the question is: how are we going to argue from (1) to $[s$ is true if and only if $p]$ where $[p]$ is equivalent in meaning to $s$? We need to fill the gap in:

(1) $s$ is true if and only if $s$ corresponds to a fact.
(2) ...

\[ \text{s is true if and only if } p. \]

However, unless we have some information about meaning, we can draw no conclusions from the hypothesis that $s$ corresponds to a fact, since we don’t know to what fact $s$ would correspond.

Some ways of supplying this information will result in implication of the biconditional without any help from the correspondence theory. For instance, suppose that we are supplied with some information about meaning in the form of a “means that” ascription. The argument would then be

(1) $s$ is true if and only if $s$ corresponds to a fact.
(2) $s$ means that $p$.

\[ \text{s is true if and only if } p. \text{ (2, BY MT)} \]

In this case, of course, the conclusion is an appropriate instance of the T-Schema, but the correspondence theory plays no role in deriving it: (1) is irrelevant. On a conception of meaning on which MT holds, to state the meaning of a sentence always implies of itself what its truth condition is. We therefore cannot non-trivially “derive” the instance from the correspondence theory in this manner. The assumption that meaning is truth-conditional does all of the work, while the substantial claim about what truth is does none. Likewise we might go quotational, and try taking instances of the disquotation schema as additional premises:

(1) $[p]$ is true if and only if $[p]$ corresponds to a fact.
(2) $[p]$ is true if and only if $p$. (DISQUOTATION SCHEMA)

\[ \text{[p] is true if and only if } p. \text{ (2, REITERATION)} \]

But this just makes the disquotationist’s point: if one is going to assume the instances of the disquotation schema, then one doesn’t need a correspondence theory of truth in order to get instances of the T-Schema where $s$ is equivalent in meaning to the sentence substituted for “$p$”. Why not, then, just get rid of the inflationary theory of truth? If one’s focus is entirely on the T-Schema, it can very easily look as though the inflationary account of truth is completely useless.

In other cases the correspondence theory is at least required to imply the relevant instance of the T-Schema for $s$. In one simple case we might proceed as follows:

(1) $s$ is true if and only if $s$ corresponds to a fact.
(2) $s$ corresponds to a fact if and only if $p$.

\[ \text{s is true if and only if } p. \text{ (2, BY MT)} \]

This gets the job done as long as $[p]$ is equivalent in meaning to $s$. The point, however, is that the correspondence theory has been augmented with claims about the cond-
tions under which particular sentences correspond to particular facts. Such claims aren’t implied by the account of truth as correspondence itself—e.g. by the account of truth as consisting in a certain kind of causal relation to objects, or in certain structural isomorphisms to states of affairs—and the unsupplemented correspondence theory therefore isn’t adequate if the necessity clause of Convention T is correct. If one defends a theory of truth based on a “reductive premise” of the sort illustrated by the simple correspondence theory, one had better not endorse the necessity clause.25

Another kind of case provides a good illustration of the interplay between supplying information about meaning and being recognizably deflationary. Suppose for simplicity that $L$ consists only of atomic predications, and let $s$ be “Aristotle is a philosopher”.26 Now consider a familiar sort of theory based roughly on Tarski’s own way of defining truth.27 This theory will include a clause like “an atomic predication $F_n$ is true if and only if what the name $n$ refers to falls within the extension of the predicate $F$”. This theory won’t imply an instance of the T-Schema where the sentence substituted for “$p$” is equivalent in meaning to “Aristotle is a philosopher”, however, unless it is augmented with information about to whom “Aristotle” refers and about who falls within the extension of “is a philosopher”. So if we take the “theory” of truth here to consist simply of the general clause that determines the truth-conditions of sentences based on name reference and predicate extension, it doesn’t imply the biconditionals and hence isn’t adequate if the necessity clause is correct.

We might, if we endorse the necessity clause, rectify this by accounting assignments of objects to names and extensions to predicates as also part of our theory of truth. The biconditionals would then be implied, and the theory would be adequate by the standards of the necessity clause. This, of course, is exactly what Tarski himself did. However, the theory would then be open to all of the criticisms that are usually directed against Tarski’s definitions—e.g. it would apply only to the sentences of one language, it would need to be supplemented when the language was extended, and so on.28 One buys implication of the biconditionals at the price of endowing a theory with deflationary characteristics, since in doing so one wed’s one’s account of what truth is to facts about the meanings of the sentences of a particular language. One could say the same of the manner of deriving the biconditionals that in each case takes as an additional premise [$s$ corresponds to a fact if and only if $p$] where [$p$] is equivalent in meaning to $s$. The conditions under which a sentence corresponds to a fact are determined in part by its meaning. Thus, if an alleged “correspondence theory” includes claims about the conditions under which particular sentences correspond to facts, it offers an account of truth that depends on facts about the meanings of particular sentences, and such a theory of truth needs to be modified whenever new sentences are added to the language or the meanings of old sentences change. This dependence is a hallmark of deflationary theories.

The inflationary theorist of truth must reject the necessity clause of Convention T. From the above it might look as though everyone ought to reject it, but in fact given these points an attractive argument for deflationism and against inflationism can be mounted. One need simply maintain that instances of the T-Schema where $s$ is equivalent in meaning to the sentence substituted for “$p$” express such

25I do not claim originality for the point that inflationary theories don’t of themselves imply the biconditionals. Gupta and Belnap make the point at (1993), pp. 26-27; and David is clear, at (1994), p. 104, that the inflationary theory doesn’t imply them. My purpose here is to draw the consequences of this point more clearly and forcefully than they have been drawn before.

26I borrow the example from an anonymous referee.

27The seminal treatment here is Field (1972).

28Field (1972). Note that Field’s own revision of Tarski’s definition based on replacing the enumerative definitions with causal theories doesn’t imply the biconditionals unless supplied with information about what refers to what.
important features of truth, or of the concept of truth, or of the meaning of “true”, that any good theory of truth ought to include or imply these instances. This is the import of Leeds’s claim that the instances of the disquotation schema “axiomatize” the notion of truth, as well as of Williams’s oft-repeated dictum that when we have endorsed such instances we have said “just about everything there is to be said about truth”.29 If the instances say just about everything, surely any theory that fails to imply them says almost nothing about truth and therefore cannot be a good theory of it. In general the idea that the biconditionals express central and important things about truth, things of a sort to which a theory of truth ought to be held responsible, is widely endorsed. Anyone who participates in this consensus is committed to the necessity clause and thus to the falsehood of inflationism. I will return to the possibility of arguing against inflationism on this basis in §4; here the point is that it is a substantial, non-trivial commitment of inflationism that such arguments must be misguided.

§3. Convention T and the Replacement Thesis
What makes these points so easy to miss? I think they are easily obscured because the claim that an inflationary theory must imply instances of the T-Schema where the sentence substituted for “p” is equivalent in meaning to s appears to be a consequence of a plausible general claim about the reduction of one property to another. Consider a property F, and suppose that we want to reduce it in the sense that we want to establish that being F consists in being F’, for some F’. Then, I take it, we have to show at least that F’ can “do all of the important theoretical work” that F does, and we show this by showing that claims about F’ that we endorse independently of the reduction imply, for all of the important sentences that concern F, sentences that differ from them only in that they concern F’ where the originals concern F. That is, when considering the claim that F reduces to F’, the Replacement Thesis holds:

(RT) A property F reduces to a property F’ only if claims about F’ that are independent of this reduction imply substitution-variants of all important sentences concerning F where explicit reference to F is replaced with explicit reference to F’.

For instance, if it is claimed that temperature reduces to mean molecular kinetic energy,30 then claims about mean molecular kinetic energy that are independent of the claim that temperature reduces to it (claims from statistical mechanics) have to imply substitution-variants of important claims about temperature; e.g. it has to follow from the account of mean molecular kinetic energy and other claims independent of the reductive thesis about temperature that increasing mean molecular kinetic energy at fixed volume increases pressure, since increasing temperature at fixed volume increases pressure.

I include the qualification “important” to allow that good reductions need not imply all of the claims that involve the reduced term.31 For instance, suppose “Steve doesn’t go outside when the temperature is below zero” is true. Does this mean that statistical mechanics must imply the corresponding substitution variant of this claim? No, because temperature’s role in explaining Steve’s behavior, though it might be a key explanatory role of temperature to those in-

30I am aware that there is a good deal of commonly unacknowledged complexity to this familiar example. I choose it only for its familiarity, and take no position on the reductive thesis itself.
31“Important” is intended to allow readings weaker than “essential”, so that one might maintain that a good reduction should imply substitution variants of lawlike generalizations involving the reduced term even if one’s views about essential properties take some of these not to be essence-constituting. My intent is simply to leave these issues open.
interested in Steve, is not something that goes into the account of what temperature is: if temperature didn’t play this explanatory role, temperature would still be what it in fact is. So the standards that govern such reductions, whatever they are in detail, allow that not every claim involving a reduced term be implied by the reduction plus the theory of the reductive base; only those claims involved in determining what the reduced property is need be involved.\footnote{Note that I do not claim that the proponent of a reductive thesis denies the claims that aren’t implied by it plus the theory of the reductive base. The proponent of the reduction of temperature needn’t deny our claim about Steve, and the inflationist about truth, I will be at pains to emphasize, does not deny the biconditionals at issue in Convention T.}

RT captures the idea loosely formulated in the claim that a good reductive base can “do all of the important theoretical work” of a reduced property, because it requires that all of the important theoretical relationships concerning the reduced property be relations concerning the reductive base, with reference to the reduced property supplied merely as a kind of veneer, so to speak, over the more basic phenomena. Though reductive relations among properties are more complicated than was perhaps once assumed, something close to RT looks right, and I will not question RT here. For our purposes the point is that an inflationary theory of truth such as a correspondence theory makes a reductive claim about truth, one that is thereby subject to RT.

Now, when we consider an inflationary theory such as the correspondence theory, the widely endorsed idea that instances of the T-Schema where the sentence substituted for “p” is equivalent in meaning to s tell us something important about truth lulls us into thinking that, by RT, the correspondence theory will somehow imply them. As we saw in the last section, however, this is not the case. An account of correspondence itself, one spelled out in terms of causal relations or structural isomorphisms, might be rich and informative yet not imply any of the relevant instances of the T-Schema. Since the inflationist is offering a reductive theory of truth, and since RT is sound, the only option for the inflationist is to deny that these instances of the T-Schema are important sentences about truth in the sense at work in RT.

Thus, whereas for the deflationist a set of instances of the T-Schema for a language in each of which s is equivalent in meaning to [p] says “just about everything that there is to be said about truth”; for an inflationist, not only is there more to be said about truth, but such biconditionals say nothing important about what truth is at all. This is not to deny that a set of them for a language settles the intension of the truth-predicate for that language; the point is rather that they tell one nothing about what truth is. The difference between the deflationist and the inflationist is not, therefore, that the latter adds something to the former’s theory of truth. Features of the truth-predicate such as its “transparency”, the fact that certain instances of the Disquotation Schema hold good, cannot be part of what an inflationary theory of truth is intended to explain.

Even inflationists can be misled here. In “The Metaphysics of Truth”, Michael Devitt acquiesces quite readily in the demand that the correspondence theory explain why people who understand the concept of truth accept the instances of the “equivalence schema” for the language they speak and then concedes defeat at explaining how to meet it.\footnote{Devitt (2001), pp. 601-602.}

I start my case by acknowledging something that may seem to count against the correspondence theory and in favor of the deflationary theory. It is a striking fact that people competent with the truth-term tend to believe instances of the equivalence schema, for example that ‘Snow is white’ is true if and only if snow is white. This fact is easy for a deflationary theory to explain because, according to that theory, such beliefs are obvious once one has mastered the truth term: what is be-
Douglas Patterson

One who, with Tarski, holds that truth is defined by instances of the T-Schema where s is equivalent in meaning to the sentence substituted for “p” can explain why those who have beliefs involving the concept of truth believe these instances: if the instances determine what “is true” means, then one who knows what it means knows what these biconditionals say. If explaining why people accept the biconditionals is something that a theory of truth must do, deflationary theories are adequate and inflationary theories are not. If, by contrast, such biconditionals do not define truth, what explains why people so reliably believe such instances of the T-Schema for the sentences of their language? Why, as Devitt asks, are the instances so obvious, when the correspondence theory—which Devitt assumes is supposed to explain why they hold—is so complicated and unobvious?

For the inflationist, however, the explanation should come, not from the account of truth, but the account of meaning: people accept such biconditionals for languages they understand because meaning is truth conditional and understanding a language is knowing what its sentences mean. The accounts of meaning and the understanding of it work together to imply that what speakers know when they understand their language is in part what the truth-conditions of its sentences are. These are expressed by instances of the T-Schema where the sentence substituted for “p” really does state the condition under which s is true, at least some of which are also instances where the sentence substituted for “p” is equivalent in meaning to s. Hence these knowledgeable speakers acknowledge the truth of these biconditionals when presented with them. To say this, however, is not to say that an explanation of what truth is has to explain why meaning is truth conditional, or why grasping meaning requires knowledge of truth-conditions. These are, rather, tasks for the truth-conditional theory of meaning. On an inflationary theory properly understood, it is simply not the case, as Devitt assumes, that that “the nature of truth is such that the equivalences hold” between [“p” is true] and [p]. It is rather the nature of meaning that explains this. This is why the instances can be obvious to speakers even though the correct account of the nature of truth is not obvious.

§4. Conclusion: Prospects for the Debate

As I noted at the end of §2, anyone who thinks that instances of the T-Schema of the sort at issue in Convention T—for example, instances of the Disquotation Schema—tell one something important about truth is committed to the falsehood of inflationism. As I also noted there, it might be thought that inflationism’s disconnection from the instances gives the deflationist material for an argument against the inflationist: surely the idea that the instances somehow spell out what truth is, is a powerful and attractive one—it was good enough for Tarski, and it is good enough for current deflationists. If we’ve learned that an inflationary theory of truth can’t respect this intuition, and doesn’t allow that the notion of truth is somehow spelled out in instances of the T-Schema where s is equivalent in meaning to the sentence substituted for “p”, isn’t this just a demonstration of the fail-

---

ure of inflationism?

I think the inflationist has persuasive responses available here. The first is to reiterate that though such instances are not an important part of the theory of truth, they are the foundation of the theory of meaning. They are not unimportant tout court. The fact that truth plays a role in them, though it doesn’t tell one anything about what truth is, does tell one something about the importance of truth in our lives: our understanding of the meanings of our words rests upon, and is spelled out in terms of, our grasp on truth and on conditions for being true. This division of labor between the theories of truth and meaning is analogous to such divisions found elsewhere. Reliabilism about knowledge does not commit one to the claim that the theory of reliability must issue in particular ascriptions of knowledge; and hedonism about moral value does not require that the account of pleasure must issue in particular ascriptions of moral value.35 This is not to deny the importance of the link between reliability and knowledge, or pleasure and moral value—quite the contrary, as any proponent of these views could recognize. However, not every explanatory role that something plays goes into constituting what it is. In the same way, truth-conditionality about meaning does not require the theory of truth to issue in particular ascriptions of meaning. The instances of the Disquotation Schema hold because when two sentences are equivalent in meaning, one may always be used to state the other’s truth condition, precisely because meaning just is truth conditional and hence having the same meaning is having the same truth-condition. These facts, and not the nature or definition of truth, explain the disquotational role of the truth-predicate. If the inflationist can muster a good argument in favor of inflationism, I believe that these points suffice to respond to the advocate of the necessity clause.

The central difference between inflationary and deflationary theories of truth concerns the relative explanatory priority of truth and meaning. The inflationist may make use of truth in the explanation of meaning, while an adherent of the sufficiency clause, such as a deflationist, must explain meaning in other terms if her theory is to have the interest it is usually taken to have. The primary task for the proponent of the sufficiency clause, therefore, is to establish the adequacy of some account of meaning that does not explain meaning in terms of truth, while the primary task for the inflationist is to argue that the best account of meaning for the sentences of an interesting language such as English must of necessity give an important explanatory role to truth and truth conditions.36 It is at the level of the account of meaning, rather than the account of truth itself, that the dispute between these rival accounts must be settled.37

References

David, Marian (1994), Correspondence and Disquotation (New York: Oxford University Press).


36Bar-On, Horisk, and Lycan (1999) presents a good discussion of these points, and argues that in fact the deflationist is unwittingly committed to the truth-conditional theory.

37I thank Robert Brandom, Nuel Belnap, John MacFarlane, Lionel Shapiro, Mark McCullagh, Mark Lance, Gila Sher, anonymous referees, and several audiences for helpful comments on previous drafts.