# Tarski on Truth

John Peloquin

In “The Concept of Truth in Formalized Languages”,[1] Tarski provides a definition of truth for formalized languages which satisfies the following adequacy condition:

(T) Let $L$ be an interpreted object language, $M$ a suitable meta language for $L$, and $\Delta$ a monadic predicate in $M$ defining truth in $L$. Then $\Delta$ is *adequate* if and only if, for all sentences $\phi$ in $L$, if $\ulcorner\phi\urcorner$ is a (structurally descriptive) name of $\phi$ in $M$ and $\psi$ is a translation of $\phi$ in $M$, then the sentence

$$\Delta(\ulcorner\phi\urcorner) \leftrightarrow \psi$$

is derivable in $M$.[2]

By using the machinery of object and meta languages, Tarski is able to avoid certain difficulties associated with the ordinary conception of truth, like the liar paradox.

Intuitively, if $\Delta$ is an adequate truth predicate for $L$, then $\Delta$ applies to precisely the (names of) true sentences in $L$, where “true” has its ordinary meaning. Indeed, if $\phi$ is a sentence in $L$ and $\psi$ is a translation of $\phi$ in $M$, then $\phi$ and $\psi$ have the same meaning, so are either both true or both false. But then by adequacy, $\Delta$ applies to (the name of) $\phi$ if and only if $\psi$ is true, which holds if and only if $\phi$ is true. So $\Delta$ applies to all and only (names of) true sentences in $L$.

Now let $\Delta$ be Tarski’s truth predicate for $L$. Then since $\Delta$ is adequate, $\Delta$ is coextensive with our ordinary predicate “true in $L$”. But does it have the same meaning? Soames provides several reasons to think not. First, $\Delta$ applies directly to *sentences*, and more specifically to *sentences in the language L*, whereas our ordinary truth predicate applies directly to *propositions* (the

[1] [2]

[2] [2], p. 187–8, Convention T.

things expressed by sentences), and derivatively applies to sentences in any language capable of expressing those propositions. So $\Delta$ and our ordinary predicate have direct application to different classes of things.[3] Second, $\Delta$ tells us nothing apriori about the *meanings* of sentences in $L$, whereas our ordinary truth predicate (when applied to sentences) does so. To illustrate, let $L$ be a first-order language whose nonlogical vocabulary includes a constant symbol $s$ and a monadic predicate symbol $W$. Interpret $s$ as snow and $W$ as the property of being white, so that a translation of the atomic sentence $\ulcorner Ws \urcorner$ into English is "Snow is white". Then we obtain

(1) $\Delta(\ulcorner Ws \urcorner) \leftrightarrow$ Snow is white

The left side of (1) can be fully expanded using Tarski's definition of $\Delta$, which after simplification yields the logically equivalent

(2) There is an object $x$ such that $x =$ snow and $x$ is white if and only if snow is white.

But (2) tells us nothing about the meaning of the original sentence $\ulcorner Ws \urcorner$, so (1) does not either.[4] The problem is that in order to know that $\Delta$ is adequate and therefore applies to precisely the (names of) true sentences in $L$ (i.e. is a genuine truth predicate), we need to already know the meanings of the sentences in $L$. But just knowing that $\Delta$ gives us theorems like (1) does not tell us anything about the meanings of sentences in $L$.[5] The situation is different with our ordinary truth predicate, for which the principle

(3) If $\phi$ means in $L$ that $p$, then $\phi$ is true in $L$ if and only if $p$.

seems to be an analytic, apriori, and necessary truth.[6] There is no such apriori connection in the case of Tarski's truth predicate, so his analysis does not capture our ordinary notion of truth. Of course, this is not necessarily a bad thing. After all, our ordinary notion of truth runs into difficulties like the liar paradox, whereas Tarski's does not.

[3] [1], p. 98–9.
[4] [1], p. 103–4.
[5] [1], p. 105.
[6] [1], p. 105–6.

When discussing object languages of infinite order for which an adequate definition of truth cannot be given in a meta language,[7] Tarski observes that an alternate approach can be taken to characterize the true sentences in such an object language:

(T*) If $L$ is an interpreted object language and $M$ is a meta language for $L$ with a consistent theory, then if we extend $M$ by adding a new primitive monadic predicate symbol $T$ and axioms of the form

$$T(\ulcorner\phi\urcorner) \leftrightarrow \phi$$

for all sentences $\phi$ in $L$, then the resulting extension $M^*$ also has a consistent theory.[8]

This approach differs from the previous in that truth in the object language is not *defined* in the meta language, but rather a new truth primitive is added to the meta language and new axioms are added to the meta theory to characterize it. Consistency of the meta theory is maintained in (T*) because, roughly, if the theory of $M^*$ is inconsistent, then there must be some *finite* subset of the theory which is inconsistent (because proofs are finite). But this finite subset is a subset of the theory of a meta language $M^{**}$ for a subset of $L$ of finite order, where $M^{**}$ is just an extension of an appropriate subset of $M$ for which $T$ can be defined as before and the axioms proved. But since the finite subset is inconsistent, the theory of $M^{**}$ is inconsistent, so the theory of $M$ is inconsistent, contradicting the assumption in (T*).[9]

Since consistency of the meta theory is maintained using this approach, the liar paradox must not arise. But why exactly? The key is that, by adding $T$ to $M$ and characterizing $T$ with axioms involving sentences in $L$, we did not thereby make $M^*$ sufficiently universal to facilitate application of $T$ to (names of) the sentences in $M^*$ and therefore facilitate construction of the paradox. In other words, the approach provides only a *partial definition* of truth in $M^*$ which is not sufficient to generate the paradox in $M^*$.[10]

This approach has advantages over the definitional approach in that it is simpler while still

[7] [2], p. 247, Theorem I.
[8] [2], p. 256, Theorem III.
[9] [2], p. 256–7.
[10] [2], p. 262.

remaining perfectly adequate in the sense of (T), but its primary disadvantage is that it severely limits deductive power, yielding a highly incomplete meta theory. For example, we are unable to derive the general quantified law of noncontradiction, among other basic results.[11] We can always supplement the resulting meta theory with additional axioms, or additional deductive rules, but this will make it more difficult to prove that the resulting meta theory is consistent.[12] Another disadvantage is just the fact that it does fail to provide a definition of truth, which is a fundamental notion in both formal and informal investigations. If we do not wish to take this notion as a primitive, we must therefore use the other approach if possible.

## References


[1] Soames, S. *Understanding Truth.*

[2] Tarski, A. "The Concept of Truth in Formalized Languages." *Logic, Semantics, Metamathematics.*


[11] [2], p. 256.
[12] [2], p. 257–261.