

Class 18 - Tarski's Theory of Truth

I. From intensions to extensions

'Semantics' is a broad term, covering both sense and reference.

We have looked a little bit at the problems for constructing theories of meaning.

Frege's theory, which we might call naive semantics, posited meanings.

Naive semantics solved Frege's three puzzles.

It also explains the synonymies of 'snow is white' and 'la nieve es blanca'; and 'bachelor' and 'unmarried male'.

But naive semantics is saddled with an inflated ontology of propositions and senses.

The positivists tried to explain meaning in terms of observation and verification.

But, without an analytic/synthetic distinction, their claim that individual sentences are confirmed and disconfirmed independently, and that they thus have meanings independently of a broader background theory, seems indefensible.

Quine urged that we could maintain meaningfulness without meanings.

But, Quine's semantic holism seems incompatible with the basic tenets of compositionality.

Compositionality says that the meaning of a complex sentence is built from the meanings of its parts.

Semantic holism says that the basic unit of meaning is an entire theory; the parts lack meaning on their own.

From another direction, the IBS program, which promised to base meaning on speaker intentions, ran into the bloated ontology of propositions that saddled naive semantics.

Schiffer and Johnston led us to a despair about semantic theory in any guise.

One direction we might take from Quine's work is to abandon all hope of an intensional semantic theory. We can try to construct theories of reference and truth, instead, and see if there are any semantic facts left over.

The pressure to construct satisfying theories of reference and truth is greater if they are going to explain all explicable semantic properties, in the absence of a theory of meaning.

So, let's put aside the thorny questions about theories of meaning and look at theories of truth and reference.

II. Formal and ordinary conceptions of truth

Tarski presents his theory of truth as a modest step in the direction of rehabilitating the easier, extensionalist side of semantics.

It is perhaps worthwhile saying that semantics...is a sober and modest discipline which has no pretensions of being a universal patent-medicine for all the ills and diseases of mankind, whether imaginary or real (88).

While the essay we are reading is fairly non-technical, Tarski's great achievement is in the technical

foundations of semantics and model theory.

He constructs a formal definition of truth, for artificial languages like those of formal logic and mathematics.

Frege and Russell dreamed of logically perfect languages, and decried the inadequacies of natural language.

Tarski is less doctrinaire about his formal work.

See the end of §6, where he mentions, but does not express, the hope of replacing natural languages in science with formal languages.

But, Tarski holds that the formal construction is not merely formal.

The desired 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 (85).

The relation between theoretical and descriptive semantics is analogous to that between pure and applied mathematics, or perhaps to that between theoretical and empirical physics; the role of formalized languages in semantics can be roughly compared to that of isolated systems in physics (101).

Tarski's claims are that he provides a precise definition of truth within a formal language, for a formal theory, and that this formal definition may help us understand our ordinary conception of truth.

Tarski models his procedure and hope after the success of formalization for contentious concepts like those of negative and imaginary numbers.

Imaginary numbers were once called impossible.

Formal theories of complex numbers assuaged worries about their legitimacy.

Similarly, Tarski hoped, formal truth theories can legitimate uses of truth in empirical science, mathematics, and even the humanities, by giving a formal expression of our ordinary conception.

III. Inflationary and deflationary conceptions of truth

It is not necessarily clear what our ordinary conception of truth is.

I take our ordinary conception to be a correspondence theory, as it is traditional to interpret Aristotle.

To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true (*Metaphysics*, 1011b25).

According to the correspondence theory of truth, truth is a relation between words and worlds.

The correspondence theory is an inflationary theory of truth.

Tarski discusses other inflationary theories of truth in §3 of his essay.

The truth of a sentence consists in its agreement with (or correspondence to) reality (86).

And

A sentence is true if it designates an existing state of affairs (ibid).

All three of these inflationary characterizations of truth involve a words-worlds connection.

The inflationary theory is distinguished from a deflationary theory of truth.

The deflationary theory of truth has many proponents, all of whom have different ways of understanding and explaining deflationism.

An emphasis on Tarski's Convention T (see below) is common to all deflationists.

Like deflationists about meaning, they believe that there is no essence to truth, no single reduction of truth to a specific property.

Some deflationists claim that truth is just a device for simplifying long conjunctions.

If you said a lot of smart things at the party, I could list them all.

Or, I could just say:

1. Everything you said last night was true.

In 1, 'true' is eliminable, by a longer, clunky sentence, or set of sentences.

Such eliminations are essential to the deflationary conception of truth.

Sometimes, the deflationary theory is called a redundancy theory: to say that snow is white is true is just to say, redundantly, that snow is white.

So, we do not need 'true'; it just comes in handy.

Tarski argues that we need a concept of truth by considering

2. The first sentence written by Plato is true.

In 2, 'true' appears ineliminable.

Similarly, Tarski considers the claim:

3. All consequences of true sentences are true.

Tarski uses 2 and 3 to show that 'truth' plays an essential role in a theory.

Tarski's claim that 'truth' is essential may not have inflationary implications.

If 'true' is a device used to refer to other sentences, as in 2 and 3, it depends on what we think of those other sentences, the ones without 'true' and with content.

If we need a words-worlds relation in order to ascribe 'true' to a sentence, then truth will not be merely deflationary, or redundant.

If all there is to truth is eliminable, then perhaps there is no essence to truth.

Even Tarski's claim that the semantic conception of truth captures Aristotle's intent (§17) may not have inflationary implications, since Aristotle's original claim could be given a deflationary interpretation itself!

Both inflationists and deflationists have used Tarski's work for their own purposes, and Tarski himself does not clearly distinguish between the inflationary and deflationary interpretations.

He believes that the technical merits of his semantic conception will lead us to accept it.

It seems to me obvious that the only rational approach to [questions about the correct notion of truth] would be the following: We should reconcile ourselves with the fact that we are confronted, not with one concept, but with several different concepts which are denoted by one word; we should try to make these concepts as clear as possible (by means of definition, or of an axiomatic procedure, or in some other way); to avoid further confusions, we should agree to use different terms for different concepts; and then we may proceed to a quiet and systematic study

of all concepts involved, which will exhibit their main properties and mutual relations (94)

Furthermore, Tarski believes that the semantic conception is agnostic among any deeper philosophical debates.

We may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians - whatever we were before. The semantic conception is completely neutral toward all these issues (98)

IV. Why all the fuss about truth?

In the early twentieth century, truth had taken a pretty bad beating. The central problem is that regimenting truth within formal theories creates paradoxes. A paradox is a set of sentences that we have good reasons both to believe and to believe are unacceptable, or incompatible. A sentence that appears, impossibly, both true and false is a paradox. Various semantic notions in addition to truth create paradoxes as well.

The most important semantic paradox is the liar.

4. This sentence is false.

4, like any sentence, could be either true or false.
If 4 is true, then its content, which says that 4 is false, is true.
So 4 is false.
Since 4 is false, the denial of its content must be true.
So, 4 must be not false, or true.
Thus we have good reasons to believe that 4 is both true and false, which is a paradox.

There are grounds to question the paradoxicality of 4.
'This sentence' refers to itself, and so seems grammatically deviant.
We can refine 4 a little bit, removing 'this sentence', but it will not help.

4'. 4' is false.

We might reject the meaningfulness of 4 and of 4', due to its ungrammaticality, and dismiss the paradox. In contrast, the grammaticality of the semantic-paradoxical 5, due to Quine, is not in question.

5. 'Yields falsehood when appended to its own quotation' yields falsehood when appended to its own quotation.

Note that 5 is perfectly grammatical.

Another semantic paradox is attributed to Kurt Grelling.
Some predicates apply to themselves, whereas others do not.
'Polysyllabic' is polysyllabic; 'monosyllabic' is not monosyllabic.

Call a predicate heterological if it does not apply to itself.

'Monosyllabic' is heterological; 'polysyllabic' is autological, or homological.

Now, consider whether 'heterological' applies to itself.

If it does, then 'heterological' is not heterological.

But, if 'heterological' is not heterological, then it does not apply to itself, which means that it is heterological.

We can construct a statement involving 'heterological' whose truth value is puzzling.

6. 'Heterological' is heterological.

Tarski also discusses Richard's paradox of definability, at the end of §5.

One response to the semantic paradoxes is to banish semantic terms which cause them.

Another option is to attempt to create a consistent theory of truth.

Tarski's theory of truth avoids the paradoxes and rehabilitates the old Aristotelian view of truth by proscribing self-reference.

The core of Tarski's solution is his distinction between object languages and metalanguages.

V. Constructing a truth predicate

Imagine a language that does not contain the word 'true'.

Call that language the object language.

Technically, the object language contains only atomic sentences and the first-order logical structure.

Sentential functions build up all of the infinite sentences of the language.

We can not devise any of the semantic paradoxes in the object language because it is not semantically closed, i.e. it does not contain names of the terms of the object language, and it contains no truth predicate; see §8.

To examine and assess the object language we ascend to a metalanguage.

We translate the object language into the metalanguage, introducing names for all of the sentences of the object language.

Then, we add a predicate, 'True', which applies to some sentences of the object language.

We partition the sentences of the object language into two classes: the true and the false.

We need names of each of the object-language sentences in the metalanguage; we can just take their ordinary names.

The extension of a predicate is the set (or class) of objects of which that predicate holds.

The anti-extension of a predicate is the set (or class) of objects of which the predicate does not hold.

So, the extension of 'woodchuck' is the set of all woodchucks.

The anti-extension of 'woodchuck' is the class of all things that are not woodchucks.

(The anti-extension of 'woodchuck' is a class because it is too big to be a proper set!)

The extension of the truth predicate is the set of all the true sentences of the object language.

All the false sentences are in the anti-extension of the truth predicate.

Technically, we use the notion of satisfaction of an open sentence by a sequence to generate, recursively, the true sentences of the object language within the metalanguage; see §11.

There will be many more sentences of the meta-language than of the object language, but the truth predicate will only apply to the sentences of the object language.

That way, we will not be able to generate the semantic paradoxes within the metalanguage.

We can not write 4 in the object language, because the object language does not contain the term 'false'. And, we can not write it in the metalanguage, because while the metalanguage contains 'false', 'true' and 'false' only apply to sentences of the object language.

If we tried to construct 4 in the object language, 'this sentence' would refer to a sentence of the metalanguage, to which 'false' does not apply.

If you are bilingual, it helps to think of the object language and metalanguages as two different languages.

Take, for example, a fragment of Spanish which does not contain 'verdad' or any other semantic term, as your object language.

Take English as your metalanguage.

In Tarski's construction, 'true' and 'false' are English terms which apply only to Spanish sentences, and only in the metalanguage.

So, we can not write:

4". Este frase es false.

because 4" is Spanglish.

And we can not write 4 itself, because 'false' does not apply to any English sentence.

If we wanted to do so, we could construct a truth predicate for the metalanguage.

But, the truth predicate for the metalanguage can not be written within the metalanguage, on pain of paradox.

So, we have to go up one step to a meta-metalanguage.

In our example, the meta-metalanguage might be Norwegian.

We could introduce a term in the meta-metalanguage, 'sann', to apply only to true English sentences.

And, we could introduce a term, 'uetke', to apply only to false English sentences.

Thus, Tarski's construction entails an infinite series of metalanguages, each one containing truth predicates only for languages lower in the hierarchy.

VI. Tarski's theory of truth

The minimal condition for the truth predicate, for getting truth right, is what we call the T-schema, or Convention T:

7. p is true iff x

In 7, 'p' is the name of any object-language sentence, and x are the truth conditions of that sentence.

Instances of 7 are written in a metalanguage into which we have introduced names of all the sentences of the object language.

The truth conditions are written in the metalanguage.

Here are some instances of the T-schema:

8. 'the cat is on the mat' is true iff the cat is on the mat.
9. '2+2=4' iff 2+2=4
10. 'Barack Obama is president' is true iff the husband of Michelle Obama and father of Sasha Obama and Malia Obama is head of the executive branch of the United States of America.

Note that the truth conditions do not have to be expressed using the same terms as the sentence on the left.

Further, we could just name the sentence 'Barack Obama is president' X, in the metalanguage, so that we get:

- 10'. X is true iff the husband of Michelle Obama and father of Sasha Obama and Malia Obama is head of the executive branch of the United States of America.

To determine which sentences are true and which are false, we have to examine the truth conditions as given on the right hand side of instances of the T-schema.

If using a different language for the object language and the metalanguage helped, consider:

- 8'. 'El gato está en el alfombrilla' is true iff the cat is on the mat.

Notice that you could understand the truth conditions of 8' without understanding the meaning of the Spanish sentence on the left side.

Convention T, despite being a simple schema, is the essence of Tarski's much more complicated technical construction.

It is the centerpiece of Tarski's semantic conception of truth.

We shall call a definition of truth "adequate" if all these equivalences [i.e. instances of the truth schema] follow from it (87).

VII. The semantic conception of truth is strictly extensional

It may be slightly misleading to call Tarski's theory a semantic conception of truth.

For, it is a use of 'semantic' which is independent of the intensionalist meaning theories we have examined.

Any interpretation of Tarski as concerned with intensions is probably a mis-interpretation.

Consider:

Semantic notions are undoubtedly involved to a larger or smaller degree, in psychology, sociology, and in practically all the humanities. Thus, a psychologist defines the so-called intelligence quotient in terms of the numbers of *true* (right) and *false* (wrong) answers given by a person to certain questions... [A] student of literature may be strongly interested in the problem of whether a given author always uses two given words with the same *meaning* (100).

The latter example seems to extend Tarski's theory of truth to intensional questions.

I think it best to understand his use of 'meaning' here to involve only extensional concepts like reference.

VIII. Is Tarski a deflationist?

Tarski defends calling his theory of truth a semantic conception.

Semantics is a discipline which, speaking loosely, *deals with certain relations between expressions of a language and the objects* (or “states of affairs”) “referred to” by those expressions. As typical examples of semantic concepts we may mention the concepts of *designation, satisfaction, and definition...* (88).

Hartry Field, in his famous paper on Tarski’s theory, argues that Tarski is not the deflationist that the deflationists want him to be.

In order to use Convention T, we need to supplement it with some kind of account of why we chose certain sentences to be true and not others.

To see the problem, remember that we could understand the truth conditions in 8’ without understanding the Spanish sentence on the left.

It is not enough just to construct the extension and anti-extension of ‘True’.

We would want to understand the component parts of the Spanish expressions, and how they interact to form true or false sentences.

Convention T, by itself, does not provide that kind of explanation.

Thus, Tarski’s construction only reduces ‘truth’ to other semantic notions, like satisfaction.

It doesn’t get rid of semantic notions.

Note that Field’s criticism is not the criticism that Tarski considers in §15, that his concept of truth presumes, in circular fashion, logical connectives that are defined by the truth tables.

Against that (potentially devastating) objection, Tarski rightly argues that we can define all the sentential connectives together, implicitly, by their functions.

We need not appeal to the truth tables.

We can even introduce the connectives as undefined terms.

Such implicit definitions are acceptable to avoid circularity in the case of logical connectives, though they are not acceptable for terms with greater content, as, say, a way of avoiding the circularity in Quine’s hocus-pocus argument.

[An explanation of that last sentence could be a good term paper.

Tarski’s claim in §15 about the role of the connectives in proof theory, as opposed to semantic (or model) theory is a hint.]

If we are merely concerned with constructing a metalinguistic truth predicate, Convention T and its supplements (designation, satisfaction, definition) might suffice.

We might, in contrast, wish to take Tarski’s claim to constructing a semantic notion of truth seriously.

In that case, it would be useful to be able not merely to explain truth in terms of other semantic notions, but to reduce them to physical ones.

Recall the IBS attempt to reduce the semantic to the psychological, with an aim to physicalistic theories of mind.

Similarly, we would need to supplement Tarski’s theory with something like causal accounts of reference.

Tarski himself probably wanted a physical notion, but he explained truth in terms of satisfaction and primitive denotation.

We need, in addition to Convention T, an explanation of why the terms are true of the things of which they are true, in a way that is consistent with physics.

It is not that we could not add such an account to complete Tarski's theory.
But, once we do, the theory does not appear deflationary at all.
The concepts of satisfaction and denotation are referential, involving inflationary concepts of a words-worlds connection.

IX. Kripke and the truth hierarchy

Field's worry about Tarski's theory concerns its use by deflationists in a broader semantic project.
But, there are technical concerns about his constructions, as well.
Tarski's construction produces a hierarchy of languages.
Each 'true' occurs at a different level in the hierarchy.
There are cases in which we do not know which level our 'true' belongs to.
Consider:

11. Everything George W. Bush said was false.

11 must be made in a metalanguage one step higher than anything that Bush ever said.
So, to know what level my 'false' belongs to, I need to know about all the levels of Bush's uses of 'true' and 'false'.
If Bush once claimed:

12. Everything Bill Clinton said was false.

then we also need to know all the levels of what Clinton said.
Furthermore, if Clinton were the speaker of 11, then Bush and Clinton become embroiled in an ugly semantic circle.

Kripke, in a paper called "Outline of a Theory of Truth," showed that we can construct a truth predicate for a language embedded within the object language itself, without creating paradox.
Here is a far-too-quick sketch of Kripke's approach.
Start with your base language, containing no logical connectives, quantifiers, or truth predicate.
Then, add a truth predicate.
Put into the extension all the standardly true sentences of the language, and into the anti-extension all the false sentences.
Choose an operator to map into the next level.
Standard operators value a conjunction as true iff... etc.,
Non-standard operator might produce a supervaluational truth predicate, or a three-valued logic truth predicate.
At each level, we bring along all the earlier sentences, and apply the truth predicate to them.
But, the truth predicate does not apply to sentences at its own level.
At a fixed point, some sentences, well-formed sentences will have no truth value.
In the minimal fixed point, this includes ungrounded sentences (the truth-teller, e.g.) as well as paradoxical sentences.
There are many different fixed points, and lots of technical work can be done with them.
An object-language truth predicate allows us to value many sentences that include 'true'.
For Kripke, we'll have all those sentences, and one truth predicate for all of them.
It's like encompassing the entire Tarskian hierarchy in one fell swoop.

X. Summary

There are at least two ways to look at Tarski's semantic theory of truth.

The first way is minimalist, and it focuses on the condition of adequacy, the T-schemas.

The second way is inflationist, and it focuses on the extent to which Tarski legitimizes our ordinary, correspondence notion of truth.

Deflationists often seem to be arguing against a more substantial notion of truth than correspondence.

They deny that there is an essence to truth.

But, correspondence theorists are not ordinarily committed to any kind of spooky essence.

Tarski himself makes that pretty clear.

In no interpretation of the term 'metaphysical' which is familiar and more or less intelligible to me does semantics involve any metaphysical elements peculiar to itself (100).

I wonder if the deflationist is arguing against a straw inflationist.

Still, the question of whether philosophers need any notion of truth (or Truth) is much discussed.

Tarski makes a compelling argument that science aims at truth.

The main reason we want consistent theories is because we know that an inconsistent theory contains a falsehood.

As soon as we succeed in showing that an empirical theory contains (or implies) false sentences, it cannot be any longer considered acceptable (102).

Some philosophers try to replace truth with warranted assertability, or coherence, or some other conditions.

I'm a big fan of truth, and Truth.

But, there are obvious epistemic worries about our access to it.

The old problem of whether we can assess a words-worlds connection, being embedded in one side of it, still resonates.

The importance of Tarski's theory of truth resides, in large part, in Davidson's hopes for an extensionalist semantic theory, which we will examine next.