Lecture 1: Tarski on Truth

Philosophy of Logic and Language — HT 2016-17

Jonny McIntosh
jonathon.mcintosh@balliol.ox.ac.uk

Alfred Tarski (1901-1983) was a Polish (and later, American) mathematician, logician, and philosopher. In the 1930s, he published two classic papers: 'The Concept of Truth in Formalized Languages' (1933) and 'On the Concept of Logical Consequence' (1936). He gives a definition of truth for formal languages of logic and mathematics in the first paper, and the essentials of the model-theoretic definition of logical consequence in the second. Over the course of the next few lectures, we'll look at each of these in turn.

1 Background

The Liar Paradox

Let sentence (1) = 'sentence (1) is not true'. Then:

1. 'sentence (1) is not true' is true IFF sentence (1) is not true
2. sentence (1) = 'sentence (1) is not true'
3. So, sentence (1) is true IFF sentence (1) is not true
4. So, sentence (1) is true and sentence (1) is not true

Premise 2. is clearly true, and premise 1. is an instance of:

> 'S' is true IFF S,

According to Tarski, the problem is that natural languages are *semantically closed*: for each sentence S, they contain another sentence S' that attributes truth (untruth) to S.

Tarski thinks that, to talk unproblematically about the true (untrue) sentences of one language, the *object* language, we need to use another language, a *metalanguage*.

Thus Tarski's task: show how to define a predicate $True_L$ in a metalanguage that applies to all and only the true sentences of the object language.

2 The Shape of the Problem

2.1 Formal Correctness

The definition must be *formally correct*, i.e. be (or be provably equivalent to) a sentence of the form,

$\forall x, True_L(x)$ if and only if $\phi(x)$,

where ϕ doesn't contain the predicate $True_L$ or any other predicates expressing otherwise obscure notions. (Definitions taking this form are called *explicit* definitions.)

This requirement ensures that the definition is not circular. However, it's obviously not enough to ensure that what is defined is a *truth* predicate. Consider, for example:

$\forall x, True_L(x)$ if and only if $x = x$.

This is formally correct, but obviously inadequate: it counts *every* sentence of L as true.

2.2 Material Adequacy

The definition must also be *materially adequate*. This is the requirement that the definition capture the core of the meaning of the word 'true'. To do this, it must at the very least be *extensionally adequate*. That is to say, unlike the definition above, it should yield a predicate $True_L$ that applies to all and *only* the true sentences of the object language

Tarski suggests that a (formally correct) definition of $True_L$ for a given object language L is materially adequate IFF it entails, for each sentence of L, an instance of the schema:

(T) 'X' is $True_L$ IFF p,

where 'X' is replaced by a name of the sentence and 'p' by a translation of that sentence in M. (If M is just an extension of L, 'p' will be replaced by the sentence itself.)

This criterion of material adequacy is Tarski's famous *Convention T*.

2.3 Tarski's Hierarchy

Various constraints on the relationship between L and M fall out of Convention T:

- M has to contain resources for *referring* to each of the sentences of L.
- M has to contain the sentences of L (or their *translations*).

Also:

- M has to be *distinct from* L. Why? To avoid the Paradox, Tarski thinks.

3 How Tarski Solves It

3.1 Finite Languages

Tarski also shows how to construct definitions along these lines for various sorts of languages. It's very easy in some simple cases. Consider L_1, which contains just two sentences: '1 + 1 = 2' and '1 + 1 = 3'. Assuming these have their ordinary meanings, a definition of a truth predicate for L_1, using English as our metalanguage, might be:

> $\forall s$(s is true$_{L_1}$ IFF
> ((s is '1 + 1 = 2' and one plus one is two) OR
> (s is '1 + 1 = 3' and one plus one is three)))

- Is this a formally correct definition? Yes.
- Is it also materially adequate? Yes.

3.2 Recursive Definitions

Suppose, for example, we consider L_2, which extends L_1 with the sentential connectives, '$\neg$' and '$\wedge$' — where these have the meanings that you're familiar with from 1st year logic. These operators can be iterated, so L_2 contains infinitely many sentences. In this case, Tarski suggests that we give a *recursive* definition: we first define the truth predicate for a certain basic set of sentences of the language, the *atomic* sentences, and then extend the definition to all the other sentences in terms of what we have said about the basic set. For example, a recursive definition of a truth predicate for L_2 might be:

> $\forall s$(s is true$_{L_2}$ IFF
> ((s is '1 + 1 = 2' and one plus one is two) OR
> (s is '1 + 1 = 3' and one plus one is three) OR
> (s is formed by prefixing a sentence s_1 with '$\neg$' and s_1 is not true$_{L_2}$) OR
> (s is formed by placing '$\wedge$' between sentences s_1 and s_2 and both s_1 and s_2 are true$_{L_2}$)))

- Is this a formally correct definition? Not obviously, but can be turned into one.
- Is it materially adequate? We can't check all the instances of **(T)** individually.

3.3 Quantification

Quantifiers pose a further problem. We cannot specify the truth conditions of quantified sentences, such as '$\exists x(x + x = 2)$', in terms of the truth values of their parts, as their parts don't generally *have* truth values. Tarski offers two methods for dealing with this. The first involves replacing variables with names of the objects in the domain of quantification. But this won't always work, as in general some objects in the domain of quantification won't have names. So for the general case, Tarski developed the method that will be familiar from the semantics of predicate logic. This involves the notion of a *variable assignment*, and treats the variables themselves as a kind of *temporary* name.

4 Philosophical Significance

At the very least, Tarski shows us how to give explicit definitions of predicates that are co-extensive with various substitution instances of the predicate 'is a true sentence of L'. This in itself is an impressive achievement. But does he do more?

4.1 Defining the Concept of Truth?

Does Tarski provide the means for defining the concept of truth, i.e. for defining predicates that have the same meaning as, say, the English predicate 'is true'?

Problems:

- The predicate 'is true' applies to propositions, not sentences.
- The English predicate 'is true' (or 'expresses a truth') applies to sentences of a *range* of languages, including English!

4.2 Explicating the Concept of Truth?

Does Tarski provide the means for *explicating* the concept of truth, i.e. for defining predicates that can replace the predicate 'is true' (or 'expresses a truth') in all legitimate theoretical contexts, but that lack its defects — e.g. don't give rise to contradiction?

Problems:

- A definition of truth cannot both (1) tell us what the defined predicate means *and* (2) tell us what the sentences of the object language mean, as in *truth conditional semantics*. For (2), we'd have to *already* know what the defined predicate means.
- Look at the definition of 'true$_{L_1}$'. It defines a property that '1 + 1 = 2' has in all possible worlds in which one plus one is two. But is '1 + 1 = 2' true in all such worlds? Isn't it *false* in (some) possible worlds where '2' refers to three?
- Tarski's definitions are not *projectable*: they don't tell us under what conditions other truth predicates, for other languages, hold of the sentences of those languages.
- Knowing the *truth* conditions of, say, '1 + 1 = 2', whether or not it's sufficient for knowing what that sentence means, at least gives *negative* information about its meaning. But knowing the conditions under which it is, say, true$_{L_1}$ does not.

Selected Bibliography

Starred items (*) are more introductory, and good places to start. Tarski (1933) is his first proper treatment of the topic, but his (1944) is a more accessible presentation. Even more accessible still is his (1969). Of the rest, Field (1972), Etchemendy (1988), Soames (1999), and Künne (2003) will be particularly useful in starting out on essays.

*Alexis P. Burgess and John P. Burgess (2010) *Truth* (Princeton UP), Ch. 2

Donald Davidson (1990) 'The Structure and Content of Truth' in *Journal of Philosophy* 87(6), pp. 279-328.

Michael Dummett (1959) 'Truth' in *Proceedings of the Aristotelian Society* 59, pp. 141-162.

John Etchemendy (1988) 'Tarski on Truth and Logical Consequence' in *The Journal of Symbolic Logic*, 53(1), pp. 51-79, §1.

Hartry Field (1972) 'Tarski's Theory of Truth' in *Journal of Philosophy*, 69(13), pp. 347-375.

*Mario Gómez-Torrente (2006) 'Alfred Tarski' in E. Zalta, ed. *Stanford Encyclopedia of Philosophy*: https://plato.stanford.edu/entries/tarski/

Richard Heck, Jr. (1997) 'Tarski, Truth, and Semantics' *The Philosophical Review* 106(4), 533-554.

*Wilfrid Hodges (2001) 'Tarski's Truth Definitions' in E. Zalta, ed. *Stanford Encyclopedia of Philosophy*: http://plato.stanford.edu/entries/tarski-truth/

*Richard Kirkham (1992) *Theories of Truth* (MIT Press), Ch. 5 and 6.

Wolfgang Künne (2003) *Conceptions of Truth* (OUP), §4.2.

Douglas Patterson (2011) *Alfred Tarski: Philosophy of Language and Logic* (Palgrave Macmillan), Ch. 4.

Hilary Putnam (1985) 'A Comparison of Something with Something Else' in *New Literary History* 17(1), pp. 61-79.

Gila Sher (1999) 'What is Tarski's Theory of Truth?' in *Topoi* 18(2), pp. 149-166.

Scott Soames (1999) *Understanding Truth* (OUP), Ch. 3 and 4.

Alfred Tarski (1933) 'The Concept of Truth in Formalized Languages' in his (1983) *Logic, Semantics, and Metamathematics*, 2nd revised edition (Hackett).

— (1944) 'The Semantic Conception of Truth' in *Philosophy and Phenomenological Research* 4(3), pp. 341-376.

— (1969) 'Truth and Proof' in *Scientific American* 220, pp. 63-77.