PHILOSOPHY OF LOGIC AND LANGUAGE

WEEK 4: TARSKI ON LOGICAL CONSEQUENCE

JONNY MCINTOSH

OVERVIEW

We've been looking at the significance of Tarski's theory of truth (and Kripke's alternative) for the Liar Paradox.

This week, we will look at its significance for thinking about the concept of **LOGICAL CONSEQUENCE**.

First, I'll explain the connection between Tarski's definitions of truth and contemporary **MODEL THEORY**.

Then, I'll explain how this suggests an appealing account of the concepts of **LOGICAL CONSEQUENCE** and **LOGICAL TRUTH**.

Next week, I'll look in a bit of detail at two problems for this account, the problem of **LOGICAL CONSTANTS** and some influential objections raised by John Etchemendy.

I'll also look at alternative account. Where Tarski's account is **MODEL-THEORETIC**, the alternative is **PROOF-THEORETIC**.

MODEL THEORY

The Tarski-style definitions of truth that we sketched were definitions of truth for **INTERPRETED** languages, languages whose sentences have meanings that make them either true or false.

These include languages such as the **LANGUAGE OF ARITHMETIC**. In these, **NON-LOGICAL** expressions such as '0', 'S' ('the successor of'), '+', and 'x' have fixed meanings.

In later work, Tarski showed how we can provide definitions of truth in a **MODEL** for **UNINTERPRETED** languages, languages whose sentences *don't* have meanings that make them either true or false.

These include languages such as L_1, L_2, and $L_=$ of first year. In these, non-logical expressions such as 'P', 'Q', 'a', and 'b' do *not* have fixed meanings.

Roughly, a model for a language specifies just enough information about its nonlogical vocabulary for assigning truth values to each of the sentences of the language.

A bit more precisely, a model for a language **L** is a nonempty domain **D** plus an appropriate assignment of denotations from **D** to the basic non-logical expressions of **L**.

For example, constants (names) might be assigned objects in **D** and *n*-place predicates might be assigned sets of *n*-tuples of objects in **D**.

We can then define truth in a model for an uninterpreted language by abstracting from definitions of truth (*simpliciter*) that we give for interpreted languages with the same vocabulary.

In the case of the uninterpreted language of predicate logic, the result is the definition of truth in a model (or **STRUCTURE**) that you're familiar with from 1st year.

Using this, we can define the notion of **LOGICAL CONSEQUENCE** as follows:

A sentence ϕ is a **LOGICAL CONSEQUENCE** of a set Γ of sentences — that $\Gamma \vDash \phi$ — IFF ϕ is true in every model in which every member of Γ is true.

And we can then go on to define the notions of **LOGICAL VALIDITY** and **LOGICAL TRUTH** as follows:

An argument whose premises are the members of a set Γ and whose conclusion is a sentence ϕ is **LOGICALLY VALID** IFF ϕ is a logical consequence of Γ

A sentence ϕ is a **LOGICAL TRUTH** ($\vDash \phi$) IFF ϕ is a logical consequence of $\varnothing$ — i.e. IFF ϕ is true in every model.

Given a proof system for a language, e.g. a set of rules of inference, we can also go on to investigate such metatheoretical questions as:

Whether the proof system is **SOUND**, i.e. whether $\Gamma \vdash \phi$ only if $\Gamma \vDash \phi$.

Whether the proof system is **COMPLETE**, i.e. whether $\Gamma \vdash \phi$ if $\Gamma \vDash \phi$.

These sorts of questions had been raised (and in some cases settled) before Tarski. Tarski's achievement, by showing us how to construct precise definitions of truth in a model, was to bring them *inside* mathematics.

Let's have a closer look at the sorts of issues that arise in providing an account of logical consequence, so as to better appreciate the philosophical merits of Tarski's account.

LOGICAL CONSEQUENCE

TRUTH PRESERVATION

ARGUMENT 1

1. Everyone smokes and everyone drinks
2. Everyone smokes and drinks

The premises of this argument might not both be true, but one thing we seem to be sure of is that, *if* they are both true, the conclusion is also true.

Otherwise put: it is not the case that the premises are all true and the conclusion is false. We'll say that such an argument is **TRUTH PRESERVING**.

In order for the conclusion of an argument to be a logical consequence of the premises, it is *necessary* that the argument be truth preserving. But it is obviously not *sufficient*.

ARGUMENT 2

1. London is the capital of the U.K.
2. So Paris is the capital of France

So what else is needed? There are broadly speaking two ideas. One appeals to the notion of **NECESSITY**. The other appeals to the notion of **FORMALITY**.

NECESSITY

The first thought: when the conclusion of an argument is a logical consequence of its premises, the argument is, in some sense, **NECESSARILY** truth preserving.

That is to say, is in some sense not **POSSIBLE** for the premises to be true and the conclusion false.

The source of this thought: logic is in some sense **INDEPENDENT** of how things actually are.

But what sense of 'necessity' is at issue here? Three different ideas are usually suggested.

The first is that it is **METAPHYSICAL** necessity that is at issue.

In other words: when the conclusion of an argument is a logical consequence of its premises, there is no **POSSIBLE WORLD** in which the premises are true and the conclusion is false.

This marks a difference between **ARGUMENT 1** and **ARGUMENT 2**. Although both are truth preserving, only **ARGUMENT 1** is, in this sense, necessarily truth preserving.

But it does not mark a difference between **ARGUMENT 1** and other arguments where, intuitively, the conclusion is *not* a logical consequence of the premises.

ARGUMENT 3

1. This cup contains water.
2. This cup contains H_2O.

We might think that the problem here is that, while the words 'water' and 'H_2O' necessarily refer to the same substance, this fact was an empirical discovery; it is not part of their *meanings* that they refer to the same substance.

This that it is **CONCEPTUAL** or **ANALYTIC** necessity that is at issue.

In other words: when the conclusion of an argument is a logical consequence of its premises, it is not conceptually possible for the premises to be true and the conclusion false.

This seems to mark a difference between **ARGUMENT 1**, on the one hand, and both **ARGUMENT 2** and **ARGUMENT 3**, on the other.

But the distinction between **ANALYTIC** and **SYNTHETIC** truths is unclear. (And famously attacked by Quine.)

And it is still not sufficient for marking a difference between **ARGUMENT 1** and *every* argument where, intuitively, the conclusion is not a logical consequence of the premises.

ARGUMENT 4

1. John is a bachelor.
2. So John is not married.

The final suggestion is that is it is **A PRIORI KNOWABILITY** that is at issue.

In other words: when the conclusion of an argument is a logical consequence of its premises, it is knowable *a priori* that it is not the case that the premises are true and the conclusion is false.

(This raises a host of issues. What is *a priori* knowledge? Do we have any? And if so, how is it even possible for us to have it?)

(Insofar as an account of how it is possible to have to *a priori* knowledge depends on the analytic/synthetic distinction, Quine's criticisms of the latter will have to be addressed.)

But there's a more immediate problem: the appeal to *a priori* knowability doesn't seem to distinguish **ARGUMENT 1** from **ARGUMENT 4** either!

In summary, while any of these three notions of necessity may help us to articulate *necessary* conditions on a conclusion's being a logical consequence of a set of premises, none of them seem to yield a *sufficient* condition.

FORMALITY

What more is needed? One influential idea appeals to the notion of **FORMALITY**.

The idea is that, while arguments like **ARGUMENT 4** are truth-preserving, they are not truth-preserving in virtue of their *form*, but rather in virtue of their *matter*.

One way to try to bring this out is to point out that **ARGUMENT 4** is an instance of a certain *pattern* of argument, obtained by replacing its non-logical expressions with schematic letters:

1. **a** is an **F**
2. So, **a** is not a **G**

And other instances of the same pattern are *not* truth-preserving:

ARGUMENT 5

1. Theresa is an MP
2. So, Theresa is not a Conservative

By contrast, **ARGUMENT 1** is an instance of a different pattern of argument:

the **LOGICAL FORM** of an argument (or sentence) is the pattern of argument (or sentence) obtained by replacing its non-logical expressions with schematic letters.

When an argument (or sentence) is an instance of a certain logical form, we may say that it is a **SUBSTITUTION INSTANCE** of that form.

1. Every **F** is a **G** and every **F** is an **H**
2. So, every **F** is a **G** and an **H**

There are three slightly different sources for this idea:

- Logic is **TOPIC-NEUTRAL**, applying to any subject matter whatsoever
- Logic is **ABSTRACT**, concerning structure rather than context
- Logic is **NORMATIVE**, dictating laws that apply to thinking *as such*

In these terms, it seems that the conclusion of an argument is a logical consequence of its premises *only if* every substitution instance of that argument is truth-preserving.

The idea, then, is that if the conclusion of an argument is a logical consequence of its premises, the argument is truth-preserving in virtue of its **LOGICAL FORM**, where...

Let's try to state the idea a little more precisely.

That is to say, it is a *necessary* condition on the conclusion's being a logical consequence of the premises that every substitution instance be truth-preserving.

(This will be accepted by anyone who accepts that there is such a thing as the logical form of a sentence, and so of an argument.)

Can we say something stronger? Can we say that it is also a *sufficient* condition on the conclusion's being a logical consequence of the premises?

This is the **SUBSTITUTIONAL** conception of logical consequence:

The conclusion of an argument is a logical consequence of its premises *if and only if* every substitution instance of that argument is truth-preserving.

(The substitutional conception is often associated with the Czech philosopher, logician, and mathematician, Bernard Bolzano (1781-1848), but it is perhaps more accurately associated with certain mediaeval philosophers, such as Buridan.)

One worry is that it may be that every substitution instance of an argument is truth-preserving not because the conclusion is a logical consequence of the premises, but because of the expressive limitations of the language.

For example, in a language that contains just one name, *a*, which denotes the number 2, and one predicate, *F*, which denotes even numbers, the sentence *Fa* will be a logical truth.

Another worry is that it may be that every substitution instance of an argument is truth-preserving because of contingent facts about the cardinality of the universe.

For example, since there are more than two objects, the sentence '$\exists x \exists y\ x \neq y$', which contains no non-logical expressions, turns out to be a logical truth too.

Tarski's account of logical consequence can be understood to belong to the same tradition as the substitutional conception, but it is slightly different.

Both can be thought of as explaining logical consequence in terms of **THE ABSENCE OF COUNTER-EXAMPLES**. But they offer different accounts of the range of potential counter-examples.

For proponents of the substitutional conception, a counter-example is, as we have seen, a substitution instance of an argument's logical form whose premises are all true and whose conclusion is false.

For Tarski, a counter-example is rather a **MODEL** in which the premises of the argument are all true and the conclusion is false.

Since a model pairs non-logical expressions not with other expressions in the language, but rather with appropriate denotations from the domain, this addresses the first worry.

The translation of 'Two is even', for example, will turn out to be false in some models that pair the translation of 'two' with the number 3.

And since different models have different domains of quantification, with different cardinalities, it also addresses the second worry.

Since there are domains with just one object, there are models in which the sentence '$\exists x \exists y\ x \neq y$' comes out as false.

SUMMARY

This week, I've sketched Tarski's **MODEL-THEORETIC** account of logical consequence and other notions.

And we've looked at the place of Tarski's account within the more general context of thinking about these notions.

We've looked at attempts to spell out logical consequence in terms of **NECESSARY** preservation, where the relevant notion of necessity is understood as:

- metaphysical necessity
- conceptual or analytic necessity
- *a priori* knowability

And we've looked at attempts to spell it out in terms of **FORMALITY**, finding that Tarski's model-theoretic account belongs to the same tradition as the **SUBSTITUTIONAL** conception.

Next week, we'll begin by looking at some problems for Tarski's account:

- the problem of **LOGICAL CONSTANTS**
- objections raised by John Etchemendy