14 Tarski's Theorem

Here is the Diagonalization Lemma again:

Theorem 47. *If T is a p.r. axiomatized theory which contains* Q*, and φ is a one-place open sentence of T's language, then there is sentence δ such that (i) $\delta \leftrightarrow \varphi(\overline{\ulcorner \delta \urcorner})$ is true, and moreover (ii) $T \vdash \delta \leftrightarrow \varphi(\overline{\ulcorner \delta \urcorner})$.*

This chapter applies each part of the Lemma, to give a pair of results about truth that are usually packaged together as *Tarski's Theorem*. We arrive at the deep contrast between the notion of truth and the notion of provability which Gödel saw as underlying the incompleteness phenomenon.

14.1 Truth-predicates and truth-definitions

Recall a familiar thought: 'snow is white' is true iff snow *is* white. Likewise for all other sensible replacements for 'snow is white'. In sum, we can endorse every sensible instance of *'φ' is true iff φ*. And that's because of the very meaning of the informal truth-predicate 'true'.

How can we add a truth-predicate to an interpreted formal language L which contains the language of basic arithmetic (as in Defn. 11)? Such a language will in general not have quotation marks or the like available; however, we can arithmetize syntax and use code numbers to refer to wffs. Assume that we have fixed on some normal scheme for Gödel numbering L-wffs. Then we can define a corresponding numerical property *True* as follows:

> *True*(n) is true iff n is the g.n. of a true sentence of L.

Now suppose, just suppose, we introduce some expression $\mathsf{T}(\mathsf{x})$ with one free variable which is so defined as to *express* this numerical property *True*. And – allowing for the possibility that we've had to extend L in introducing such an expression – let L^* be the result of adding a new wff $\mathsf{T}(\mathsf{x})$ to our initial language L if necessary. (So for the moment, we leave it open whether L^* is just L, which it would be if a suitable $\mathsf{T}(\mathsf{x})$ is in fact already definable from L's resources.)

Then, by the definition of *True* and of T, we have the following for any sentence φ of the original language L:

> φ is true iff *True*$(\ulcorner \varphi \urcorner)$ iff $\mathsf{T}(\overline{\ulcorner \varphi \urcorner})$ is true.

Hence, for any L-sentence φ, every corresponding 'T-biconditional'

$$\mathsf{T}(\overline{\ulcorner\varphi\urcorner}) \leftrightarrow \varphi$$

is true. Which motivates our first main definition:

Defn. 54. *An open L^*-wff $\mathsf{T}(\mathsf{x})$ (where L^* includes L) is a* truth-predicate *for L iff for every L-sentence φ, $\mathsf{T}(\overline{\ulcorner\varphi\urcorner}) \leftrightarrow \varphi$ is true.*

And here's a companion definition:

Defn. 55. *A theory Θ (with language L^* which includes L) is a formal* truth-theory *for L iff it provides an L^*-wff $\mathsf{T}(\mathsf{x})$ such that $\Theta \vdash \mathsf{T}(\overline{\ulcorner\varphi\urcorner}) \leftrightarrow \varphi$ for every L-sentence φ.*

Equally often, a truth-theory for L is called a 'definition of truth for L'.

In sum, a truth-predicate T for L is a predicate that applies to (the Gödel numbers for) exactly the true L-sentences, and so *expresses* truth; and a truth-theory for L is a theory built in a perhaps extended language which *proves* all the T-biconditionals for L sentences.

So far, that's all just a sequence of (natural enough) definitions. Now for our first big result.

14.2 The undefinability of truth

Suppose T is a nice arithmetical theory with language L. An obvious question arises: could T be competent to define truth *for its own language* (i.e., can T already encompass a truth-theory for L)? And the answer is immediate:

Theorem 54. *No consistent p.r. axiomatized theory T which contains Q can define truth for its own language.*

Proof. Assume T defines truth for L using an open sentence $\mathsf{T}(\mathsf{x})$. Since T has the right properties, part (ii) of the Diagonalization Lemma applies. Therefore we can apply the Lemma in particular to $\neg\mathsf{T}(\mathsf{x})$, so there must be some sentence L such that

1. $T \vdash \mathsf{L} \leftrightarrow \neg\mathsf{T}(\overline{\ulcorner\mathsf{L}\urcorner})$.

According to T then, L is a Liar sentence, which (as it were) says that it is false! But, by our initial assumption that T is a truth-theory for L using T as its formal truth-predicate, we also have

2. $T \vdash \mathsf{T}(\overline{\ulcorner\mathsf{L}\urcorner}) \leftrightarrow \mathsf{L}$.

But (1) and (2) together entail that T is inconsistent, contrary to hypothesis. So our assumption must be wrong: T can't define truth for its own language. ⊠

14.3 The inexpressibility of truth

That first theorem puts limits on what a nice theory can *prove* about truth. But we can go further: there are limits on what a theory's language can even *express* about truth.

Consider our old friend L_A for the moment, and suppose that there is an L_A truth-predicate T_A that expresses the corresponding truth property *True*$_A$. Then part (i) of the Diagonalization Lemma, applies (for the proof of this part only depended on the fact that we were dealing with a language which contains the language of basic arithmetic). So in particular, there will be some L_A sentence L such that

1. $\mathsf{L} \leftrightarrow \neg \mathsf{T}_A(\overline{\ulcorner \mathsf{L} \urcorner})$.

is true. But, by the assumption that T_A is a truth-predicate for L_A,

3. $\mathsf{T}_A(\overline{\ulcorner \mathsf{L} \urcorner}) \leftrightarrow \mathsf{L}$

must be true too. (2) and (3) immediately lead to contradiction again. Therefore our supposition that T_A expresses the property of being a true L_A sentence has to be rejected.

The argument evidently generalizes. Take any language L which includes the language of basic arithmetic, so that the first part of the Diagonalization Lemma is provable. Call that an arithmetically adequate language. Then by the same argument,

Theorem 55. *No predicate of an arithmetically adequate language L can express the numerical property* True_L *(i.e. the property of numbering a truth of L).*

14.4 The Master Argument for incompleteness?

Our second Tarskian result tells us that while you can express *syntactic* properties of a sufficiently rich formal theory of arithmetic (like provability) inside a theory via Gödel numbering, you can't express some key *semantic* properties (like truth) inside the same theory. And this points to a particularly illuminating take on the argument for incompleteness.

For example: truth in L_A isn't provability in PA, because while PA-provability *is* expressible in L_A (by a provability predicate Prov), truth-in-L_A *isn't* expressible. So assuming that PA is sound so that everything provable in it is true, this means that there must be truths of L_A which it can't prove. Similarly, of course, for other nice theories.

And in a way, we might well take this to be *the* Master Argument for incompleteness, revealing the true roots of the phenomenon. Gödel himself wrote (in response to a query)

> I think the theorem of mine that von Neumann refers to is ... that a complete epistemological description of a language A cannot be given

> in the same language A, because the concept of truth of sentences in A cannot be defined in A. *It is this theorem which is the true reason for the existence of undecidable propositions in the formal systems containing arithmetic.* I did not, however, formulate it explicitly in my paper of 1931 but only in my Princeton lectures of 1934. The same theorem was proved by Tarski in his paper on the concept of truth.

In sum, as we emphasized before, arithmetical truth and provability in this or that formal system must peel apart.

How does that statement of Gödel's square with the importance that he placed on the syntactic version of the First Theorem? Well, Gödel was a realist about mathematics: *he* believed in the real existence of mathematical entities, and believed that our theories (at least aim to) deliver truths about them. But that wasn't the dominant belief among those around him concerned with foundational matters. As I put it before, for various reasons (for a start, think logical positivism), the very idea of truth-in-mathematics was under some suspicion at the time. So even though semantic notions were at the root of Gödel's insight, it was extremely important for him – given the intended audience – to show that you don't need to deploy those semantic notions to prove incompleteness.