# Löb's theorem

Rasmus Blanck

March 28, 2019

## 1 Diagonalisation

- Recall the diagonal lemma: For each formula $\varphi(x)$, there is a sentence $\delta$ such that $\mathrm{T} \vdash \delta \leftrightarrow \varphi(\delta)$.
- Gödel famously proved the diagonal lemma for the formula $\neg\mathrm{Pr}_\mathrm{T}(x)$ to obtain a sentence that is true, but not provable.
- The idea was to formalise the expression "this sentence isn't provable".
- What about "this sentence is provable"?
- Let's assume T to be a consistent, recursively axiomatised extension of PA and fix a $\Sigma_1$ provability predicate $\mathrm{Pr}_\mathrm{T}(x)$ numerating "$x$ is provable from the axioms of T".

## 2 Proof of Gödel's first

- We have $\mathrm{T} \vdash \gamma \leftrightarrow \neg\mathrm{Pr}_\mathrm{T}(\gamma)$
- Suppose $\mathrm{T} \vdash \gamma$.
- Then there is a proof of $\gamma$ from the axioms of T, so $\mathbb{N} \models \mathrm{Pr}_\mathrm{T}(\gamma)$ and (by $\Sigma_1$-completeness of T) $\mathrm{T} \vdash \mathrm{Pr}_\mathrm{T}(\gamma)$.
- By construction of $\gamma$, $\mathrm{T} \vdash \neg\gamma$, a contradiction.
- So $\gamma$ isn't provable in T: $\mathrm{T} \nvdash \gamma$.
- This means that $\mathbb{N} \models \neg\mathrm{Pr}_\mathrm{T}(\gamma)$.
- But we have $\mathbb{N} \models \gamma \leftrightarrow \neg\mathrm{Pr}_\mathrm{T}(\gamma)$, so
- $\mathbb{N} \models \gamma$.
- We have ascertained that $\gamma$ is true but not provable in T.

## 3 Henkin's question

- Consider an $\eta$ such that $\mathrm{T} \vdash \eta \leftrightarrow \mathrm{Pr_T}(\eta)$. Is such an $\eta$ true, provable, etc?
- Kreisel had some things to say about this, see below.
- Let's try a proof similar to the one above:
- Suppose $\mathrm{T} \vdash \eta$. Then $\mathrm{Pr_T}(\eta)$ is provable, so $\eta$ is provable, so ...
- Suppose $\mathrm{T} \vdash \neg\eta$. Then $\neg\mathrm{Pr_T}(\eta)$ is provable, so $\eta$ isn't provable, so ...
- Clearly, we must do something else.

## 4 Löb's answer

- Löb identified certain conditions that "the ordinary" provability predicate satisfies:

  L1) If $\mathrm{T} \vdash \phi$, then $\mathrm{T} \vdash \mathrm{Pr_T}(\phi)$.

  L2) $\mathrm{T} \vdash \mathrm{Pr_T}(\phi) \wedge \mathrm{Pr_T}(\phi \to \psi) \to \mathrm{Pr_T}(\psi)$

  L3) $\mathrm{T} \vdash \mathrm{Pr_T}(\phi) \to \mathrm{Pr_T}(\mathrm{Pr_T}(\phi))$.

- For provability predicates satisfying these, the Henkin sentence is provable:

**Löb's Theorem.** *Suppose that* $\mathrm{Pr_T}(x)$ *satisfies L1-L3. If* $\mathrm{T} \vdash \mathrm{Pr_T}(\phi) \to \phi$, *then* $\mathrm{T} \vdash \phi$.

*Proof.* Let $\phi$ be any sentence such that $\mathrm{T} \vdash \mathrm{Pr_T}(\phi) \to \phi$. Let $\lambda$ be such that $\mathrm{T} \vdash \lambda \leftrightarrow (\mathrm{Pr_T}(\lambda) \to \phi)$. By construction of $\lambda$ and L1 we have that :

$$\mathrm{T} \vdash \mathrm{Pr_T}(\lambda \leftrightarrow (\mathrm{Pr_T}(\lambda) \to \phi)) \tag{1}$$

By L2:

$$\mathrm{T} \vdash \mathrm{Pr_T}(\lambda) \wedge \mathrm{Pr_T}(\lambda \to (\mathrm{Pr_T}(\lambda) \to \phi)) \to \mathrm{Pr_T}(\mathrm{Pr_T}(\lambda) \to \phi) \tag{2}$$

By (1) and (2):

$$\mathrm{T} \vdash \mathrm{Pr_T}(\lambda) \to \mathrm{Pr_T}(\mathrm{Pr_T}(\lambda) \to \phi) \tag{3}$$

By L2 again:

$$\mathrm{T} \vdash \mathrm{Pr_T}(\mathrm{Pr_T}(\lambda)) \wedge \mathrm{Pr_T}(\mathrm{Pr_T}(\lambda) \to \phi) \to \mathrm{Pr_T}(\phi) \tag{4}$$

By (3) and (4):

$$\mathrm{T} \vdash \mathrm{Pr_T}(\lambda) \wedge \mathrm{Pr_T}(\mathrm{Pr_T}(\lambda)) \to \mathrm{Pr_T}(\phi) \tag{5}$$

By L3:

$$\mathrm{T} \vdash \mathrm{Pr_T}(\lambda) \to \mathrm{Pr_T}(\phi) \tag{6}$$

By assumption on $\phi$

$$\mathrm{T} \vdash \mathrm{Pr_T}(\phi) \to \phi \tag{7}$$

So by (6) and (7):

$$\mathrm{T} \vdash \mathrm{Pr_T}(\lambda) \to \phi \tag{8}$$

Then, by construction of $\lambda$:

$$\mathrm{T} \vdash \lambda \tag{9}$$

By L1:

$$\mathrm{T} \vdash \mathrm{Pr}_{\mathrm{T}}(\lambda) \tag{10}$$

But then, by (8):

$$\mathrm{T} \vdash \phi \tag{11}$$

□

# 5 Gödel's 2nd

If T is consistent, r.e., and sufficiently strong, then $\mathrm{T} \nvdash \mathrm{Con}_{\mathrm{T}}$.

Kripke's proof of Löb's theorem from G2.

- Suppose $\mathrm{T} \vdash \mathrm{Pr}_{\mathrm{T}}(\phi) \to \phi$.
- Then $\mathrm{T} + \neg\phi \vdash \neg\mathrm{Pr}_{\mathrm{T}}(\phi)$, by contraposition and the deduction theorem.
- For each $\phi$ we have $\mathrm{T} \vdash \neg\mathrm{Pr}_{\mathrm{T}}(\phi) \leftrightarrow \mathrm{Con}_{\mathrm{T}+\neg\phi}$, meaning
- $\mathrm{T} + \neg\phi \vdash \mathrm{Con}_{\mathrm{T}+\neg\phi}$.
- This contradicts Gödel's 2nd.
- Therefore $\mathrm{T} + \neg\phi$ must be inconsistent.
- Hence $\mathrm{T} \vdash \phi$.

The other way around:

- Suppose $\mathrm{T} \vdash \mathrm{Con}_{\mathrm{T}}$, that is
- $\mathrm{T} \vdash \neg\mathrm{Pr}_{\mathrm{T}}(\bot)$.
- By definition, $\mathrm{T} \vdash \mathrm{Pr}_{\mathrm{T}}(\bot) \to \bot$.
- By Löb's theorem, $\mathrm{T} \vdash \bot$, so T is inconsistent.

# 6 Kreisel on Henkin's problem

**Theorem 1.** *There is a formula* $\mathrm{Pr}_1(x)$ *and a term* $t_1$ *such that*

1. $\mathrm{Pr}_1(x)$ *represents provability in* T*,*
2. $\mathrm{T} \vdash t_1 = \ulcorner \mathrm{Pr}_1(t_1) \urcorner$
3. $\mathrm{T} \vdash \mathrm{Pr}_1(t_1)$*.*

**Theorem 2.** *There is a formula* $\mathrm{Pr}_2(x)$ *and a term* $t_2$ *such that*

1. $\mathrm{Pr}_2(x)$ *represents provability in* T*,*
2. $\mathrm{T} \vdash t_2 = \ulcorner \mathrm{Pr}_2(t_2) \urcorner$
3. $\mathrm{T} \vdash \neg\mathrm{Pr}_2(t_2)$*.*

Let $S(x, y)$ be the substitution function: $S(x, y)$ is the Gödel number of the formula that results from replacing the free variable of the formula with Gödel number $x$ by the numeral for $y$. We use $\ulcorner \phi \urcorner$ to denote the numeral for the Gödel number of the formula $\phi$. Example: let $\phi(x) := x < 7$. Then $S(\ulcorner \phi \urcorner, 5) = \ulcorner 5 < 7 \urcorner$.

We're essentially going to have to show that for every formula $\phi(x)$, we can effectively find a term $t$ such that

$$t = \ulcorner \phi(t) \urcorner.$$

This is sometimes called *strong diagonalisation*.

*Proof of Theorem 1.* Consider the formula $\mathrm{Pr_T}(\ulcorner S(x, x) \urcorner) \vee S(x, x) = S(x, x)$. This formula has a Gödel number, say, $k$. Let $t_1 = S(k, k)$. What is $t_1$? It is the (Gödel number of the) result of replacing the free variable in the formula with GN $k$ with the numeral for $k$. That is,

$$t_1 = \ulcorner \mathrm{Pr_T}(S(k, k)) \vee S(k, k) = S(k, k) \urcorner$$

Since $t_1 = S(k, k)$, we can substitute further:

$$t_1 = \ulcorner \mathrm{Pr_T}(t_1) \vee t_1 = t_1 \urcorner$$

We can now define the formula $\mathrm{Pr_1}(x) := \mathrm{Pr_T}(x) \vee x = t_1$.[1]

Let's show that $\mathrm{Pr_1}(x)$ satisfies 1)-3) of Theorem 1. Since $\mathrm{Pr_T}(x)$ represents provability in T, $\mathrm{Pr_1}(x)$ also does so, and therefore 1) holds. 2) was just shown above, since $t_1 = \ulcorner \mathrm{Pr_T}(t_1) \vee t_1 = t_1 \urcorner = \ulcorner \mathrm{Pr_1}(t_1) \urcorner$. For 3), it suffices to note that $\mathrm{Pr_1}(t_1) := \mathrm{Pr_T}(t_1) \vee t_1 = t_1$, and since $t_1 = t_1$ is provable in T, so is $\mathrm{Pr_1}(t_1)$. $\square$

*Proof of Theorem 2.* Use the same method to obtain a term $t_2$ such that

$$t_2 = \ulcorner \mathrm{Pr_T}(t_2) \wedge t_2 \neq t_2 \urcorner$$

Let $\mathrm{Pr_2}(x) = \mathrm{Pr_T}(x) \wedge x \neq t_2$. Then $\mathrm{Pr_2}(x)$ satisfies 1)-3) of Theorem 2. $\square$

Why isn't this accepted as a solution to Henkin's problem? It is thought that neither $\mathrm{Pr_1}(x)$ nor $\mathrm{Pr_2}(x)$ actually expresses the property "$x$ is provable in T". That is, the formulas are extensionally correct, but intentionally incorrect. It seems that even Kreisel agreed ([2], p. 681).

# Bibliography


[1] Boolos, G. (1993). The Logic of Provability, Cambridge University Press.

[2] Halbach V., and A. Visser (2014). Self-reference in Arithmetic I, RSL 7, pp. 671–691.

[3] Henkin, L. (1952). A problem concerning provability, JSL 17, p. 160.

[4] Kreisel, G. (1953). On a problem of Henkin's, Indagationes Mathematicae 15, p. 405–406.

[5] Löb, M. H. (1955). Solution of a Problem of Leon Henkin, JSL 20(4), pp. 115–118.


[1] This construction is attributed to Henkin in a footnote to [4]. Kreisel's own construction is slightly more complicated.