Löb's Theorem and Provability Predicates in Rocq

TYPES 2025, Glasgow

Janis Bailitis Dominik Kirst Yannick Forster

June 9, 2025

SAARLAND UNIVERSITY
COMPUTER SCIENCE

Inria

Introduction

- Sufficiently strong formal systems S have **provability predicates** $\mathrm{Pr}(x) : \mathbb{F}$
  - $S \vdash \varphi$ iff $S \vdash \mathrm{Pr}(\overline{\varphi})$
  - Many different of various strengths, even for same formal system

Theorem (Gödel, 1931)

If $\mathrm{Pr}(x)$ *and* S *are sufficiently strong, and* $S \vdash \varphi \leftrightarrow \neg\mathrm{Pr}(\overline{\varphi})$, *then* φ *is independent.*

Problem (Henkin, 1952)

What if $S \vdash \varphi \leftrightarrow \mathrm{Pr}(\overline{\varphi})$?

Theorem (Löb, 1955)

If $\mathrm{Pr}(x)$ *and* S *are sufficiently strong, and* $S \vdash \varphi \leftrightarrow \mathrm{Pr}(\overline{\varphi})$, *then* $S \vdash \varphi$.

Löb's Theorem and Motivation

Theorem (Löb's theorem, 1955)

Let $\mathrm{Pr}(x)$ *and* S *be sufficiently strong. For all sentences* φ,

$$S \vdash \mathrm{Pr}(\overline{\varphi}) \to \varphi \textit{ implies } S \vdash \varphi.$$

- Implies **Gödel's second incompleteness theorem** (If $S \vdash \neg\mathrm{Pr}(\overline{\bot})$, then $S \vdash \bot$)
  - ➤ Mechanised only once: Paulson (2015, Isabelle). Tedious details.
  - ➤ We extend Paulson's proof to Löb's theorem
- **Gödel's first incompleteness theorem** mechanised often[1]
- Kirst and Peters: Computational proof of first theorem, synthetic
  - ➤ Based on Beklemishev (2011) and textbooks by Kleene
  - ➤ Leave second theorem as future work

[1] Shankar (1986); O'Connor (2005); Harrison (2009); Paulson (2015); Popescu and Traytel (2019); Kirst and Peters (2023)

Our Work

Is there a less tedious proof of Löb's theorem?

- Gross, Gallagher, Fallenstein (2016): Löb's theorem in Agda
- Historically known to have intricate proof
- Many proof techniques known to fail
- Can a synthetic perspective simplify arguments?
  - $\rightarrow$ Usually, technically intricate details vanish, up to 90% shorter proofs

‘Sufficiently Strong’ in View of Löb’s Theorem

‘Sufficiently strong’ provability predicates:

Hilbert-Bernays-Löb (HBL) Conditions (Hilbert-Bernays (1939), Löb (1955))

$\mathrm{Pr}(x) : \mathbb{F}$ satisfies

- **necessitation** if $S \vdash \varphi$ implies $S \vdash \mathrm{Pr}(\overline{\varphi})$
- **the distributivity law** if $S \vdash \mathrm{Pr}(\overline{\varphi \to \psi}) \to \mathrm{Pr}(\overline{\varphi}) \to \mathrm{Pr}(\overline{\psi})$
- **internal necessitation** if $S \vdash \mathrm{Pr}(\overline{\varphi}) \to \mathrm{Pr}(\overline{\mathrm{Pr}(\overline{\varphi})})$

‘Sufficiently strong’ theories:

Diagonalisation Property (Carnap (1934))

S has **diagonalisation property** if for all $\varphi(x)$ there is sentence G s.t.

$$S \vdash G \leftrightarrow \varphi(\overline{G}).$$

HBL + Diagonalisation property = Löb’s theorem (abstract argument)

Church's Thesis (CT)

- CT: 'Every function is computable in a concrete model of computation.'[1]
- Results based on a variant of CT for arithmetic ($\mathsf{CT}_{\mathsf{PA}}$ / CT_{Q}):[2]

Axiom ($\mathsf{CT}_{\mathsf{PA}}$, Hermes and Kirst (2022))

For all $f : \mathbb{N} \to \mathbb{N}$ there is $\varphi_f(x_1, x_2) : \mathbb{F}$ such that for all $n : \mathbb{N}$,

$$\mathsf{PA} \vdash \forall y.\, \varphi_f(\overline{n}, y) \leftrightarrow y = \overline{f\,n}.$$

- Consistent for CIC[1]

[1] Kreisel (1965) as well as Troelstra and van Dalen (1988).
[2] We use EPF_μ (Richman (1983), Forster (2021)) which implies $\mathsf{CT}_{\mathsf{PA}}$ (Kirst and Peters '23).
[3] See also Pédrot (2024), Swan and Uemura (2019)

Exploiting Church's Thesis

Corollary

There is $\mathrm{Pr}_{\mathrm{CT}}(x) : \mathbb{F}$ *such that* $\mathrm{PA} \vdash \varphi$ *iff* $\mathrm{PA} \vdash \mathrm{Pr}_{\mathrm{CT}}(\overline{\varphi})$.

Lemma (Diagonal Lemma, Carnap (1934))

For all $\varphi(x) : \mathbb{F}$ *there is* $G : \mathbb{F}$ *s.t.* $\mathrm{PA} \vdash G \leftrightarrow \varphi(\overline{G})$.

- Gödel's first incompleteness theorem (1931), with Rosser's strengthening[1]
- Tarski's theorem (1935)
- Essential undecidability of PA

Problem

$\mathrm{CT}_{\mathrm{PA}}$ not strong enough for Löb's theorem (internal vs external provability).

[1] Needs variant of $\mathrm{CT}_{\mathrm{PA}}$ which also follows from EPF_{μ} (Kirst and Peters (2023)).

Defining a Provability Predicate (Continued)

- Proof '=' List of formulas
- List and syntax functions not native to PA $\rightarrow$ tedious to define (Boolos (1993))

Definition (Extended Signature of Peano Arithmetic, simplified)

EPA adds the following function symbols to PA:

$[\,]$ (nil)	$\lvert\ell\rvert$ (length)	$\ell + \ell'$ (append)
$x :: \ell$ (cons)	$\ell[i]$ (indexed access)	$x \rightsquigarrow y$ (implication)

Based on such a definition, we

1. defined a candidate for an internal provability predicate, and
2. mechanised necessitation as well as the distributivity law for it.

Contributions

Is there a proof of Löb's theorem à la Kirst and Peters? No!

- Mechanised proof of Löb's theorem
  - For first-order arithmetic in Rocq assuming HBL conditions and CT_{PA}
  - In Isabelle based on Paulson's development, axiom-free
- Mechanised diagonal lemma and key limitative theorems assuming CT_{PA}
- Analysed why CT_{PA} is too weak for Löb's theorem
- Mechanised extension of PA easing definition of internal provability predicates
- Gave candidate for internal provability predicate and parts of correctness proof

Rocq Code: `https://www.ps.uni-saarland.de/~bailitis/bachelor/Rocq_with_README.zip`
Isabelle Code: `https://www.ps.uni-saarland.de/~bailitis/bachelor/Isabelle_afp.zip`

Future Work

- Mechanise internal necessitation
- Decide whether to keep using extended PA
- Contribute Isabelle development to Archive of Formal Proofs[1]
- Contribute Rocq development to Rocq Library of First-Order Logic [Kir+22]
- Mechanise axiom-free proof of diagonal lemma and limitative theorems

Thank You!

[1] `https://www.isa-afp.org/`; Mechanisation has been submitted, decision is pending.

References i

[Bek10] Lev D Beklemishev. **'Gödel incompleteness theorems and the limits of their applicability. I'.** In: **Russian Mathematical Surveys** 65.5 (2010), p. 857. DOI: `10.1070/RM2010v065n05ABEH004703`.

[Boo93] George S. Boolos. **The Logic of Provability.** 5th. Cambridge University Press, 1993.

[Car34] Rudolf Carnap. **Logische Syntax der Sprache.** 1st. Schriften zur wissenschaftlichen Weltauffassung. Springer Berlin, Heidelberg, 1934.

[DT88] Dirk van Dalen and Anne S. Troelstra. **Constructivism in Mathematics. An Introduction.** Elsevier Science Publishers B.V., 1988. ISBN: 0-444-70266-0.

[For21] Yannick Forster. **'Church's Thesis and Related Axioms in Coq's Type Theory'.** In: **29th EACSL Annual Conference on Computer Science Logic, CSL 2021, January 25-28, 2021, Ljubljana, Slovenia (Virtual Conference)**. Ed. by Christel Baier and Jean Goubault-Larrecq. Vol. 183. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, 21:1–21:19. DOI: 10.4230/LIPICS.CSL.2021.21.

[GGF16] Jason Gross, Jack Gallagher and Benya Fallenstein. **Löb's theorem: A functional pearl of dependently typed quining.** Unpublished. 2016. URL: https://jasongross.github.io/papers/2016-lob-icfp-2016-draft.pdf.

[Gö31] Kurt Gödel. **'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I'.** In: **Monatshefte für Mathematik** 38.1 (1931), pp. 173–198.

[Har09] John Harrison. **Handbook of Practical Logic and Automated Reasoning.** Cambridge University Press, 2009. DOI: 10.1017/CBO9780511576430.

[HB39] David Hilbert and Paul Bernays. **Grundlagen der Mathematik.** 1st. Vol. 2. Berlin: Springer, 1939.

[Hen52] Leon Henkin. **'A problem concerning provability'.** In: **The Journal of Symbolic Logic** 17.2 (1952), p. 160. ISSN: 00224812. URL: http://www.jstor.org/stable/2266288.

[HK22] Marc Hermes and Dominik Kirst. **'An Analysis of Tennenbaum's Theorem in Constructive Type Theory'.** In: **7th International Conference on Formal Structures for Computation and Deduction (FSCD 2022)**. Ed. by Amy P. Felty. Vol. 228. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 9:1–9:19. ISBN: 978-3-95977-233-4. DOI: `10.4230/LIPIcs.FSCD.2022.9`.

[HKK21] Johannes Hostert, Mark Koch and Dominik Kirst. **'A Toolbox for Mechanised First-Order Logic'.** In: **The Coq Workshop** (2021).

[Kir+22] Dominik Kirst et al. **'A Coq Library for Mechanised First-Order Logic'.** In: **The Coq Workshop** (2022).

[Kle52] Stephen C. Kleene. **Introduction to Metamathematics.** North Holland, 1952.

References v

[Kle67] Stephen C. Kleene. **Mathematical Logic.** Dover Publications, 1967.

[KP23] Dominik Kirst and Benjamin Peters. **'Gödel's Theorem Without Tears - Essential Incompleteness in Synthetic Computability'.** In: **31st EACSL Annual Conference on Computer Science Logic (CSL 2023)**. Ed. by Bartek Klin and Elaine Pimentel. Vol. 252. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 30:1–30:18. ISBN: 978-3-95977-264-8. DOI: 10.4230/LIPIcs.CSL.2023.30.

[Kre65] Georg Kreisel. **'Mathematical logic'.** In: **Lectures on Modern Mathematics** 3 (1965), pp. 95–195.

[Lö55] Martin H. Löb. **'Solution of a Problem of Leon Henkin'.** In: **The Journal of Symbolic Logic** 20.2 (1955), pp. 115–118. DOI: 10.2307/2266895.

[O'C05] Russell O'Connor. **'Essential Incompleteness of Arithmetic Verified by Coq'.** In: **Theorem Proving in Higher Order Logics, 18th International Conference, TPHOLs 2005, Oxford, UK, August 22-25, 2005, Proceedings**. Ed. by Joe Hurd and Thomas F. Melham. Vol. 3603. Lecture Notes in Computer Science. Springer, 2005, pp. 245–260. DOI: 10.1007/11541868_16.

[P2́4] Pierre-Marie Pédrot. **' "Upon This Quote I Will Build My Church Thesis" '.** In: **Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science**. LICS '24. Tallinn, Estonia: Association for Computing Machinery, 2024. ISBN: 9798400706608. DOI: 10.1145/3661814.3662070.

[Pau15] Lawrence C. Paulson. **'A Mechanised Proof of Gödel's Incompleteness Theorems Using Nominal Isabelle'.** In: **Journal of Automated Reasoning** 55.1 (2015), pp. 1–37. DOI: 10.1007/S10817-015-9322-8.

[PT21] Andrei Popescu and Dmitriy Traytel. **'Distilling the Requirements of Gödel's Incompleteness Theorems with a Proof Assistant'.** In: **Journal of Automated Reasoning** 65.7 (2021), pp. 1027–1070. DOI: 10.1007/S10817-021-09599-8.

[Ric83] Fred Richman. **'Church's Thesis Without Tears'.** In: **The Journal of Symbolic Logic** 48.3 (1983), pp. 797–803. DOI: 10.2307/2273473.

[Ros36] J. Barkley Rosser. **'Extensions of Some Theorems of Gödel and Church'.** In: **The Journal of Symbolic Logic** 1.3 (1936), pp. 87–91. DOI: 10.2307/2269028.

[Sha86] Natarajan Shankar. **'Proof-checking metamathematics'.** PhD thesis. University of Texas, 1986.

[Sha94] Natarajan Shankar. **Metamathematics, Machines and Gödel's Proof.** Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1994. DOI: 10.1017/CBO9780511569883.

[SU19] Andrew Swan and Taichi Uemura. **'On Church's Thesis in Cubical Assemblies'.** In: **CoRR** abs/1905.03014 (2019). arXiv: 1905.03014. URL: http://arxiv.org/abs/1905.03014.

[Tar35] Alfred Tarski. **'Der Wahrheitsbegriff in den formalisierten Sprachen'.** In: **Studia Philosophica. Commentarii Societatis Philosophicae Polonorum** 1 (1935), pp. 261 –405. URL: https://www.sbc.org.pl/dlibra/publication/24411/edition/21615.

Mechanisation

Rocq

- 2600 lines of code (600 specification, 1900 proof, 100 comment)
- Most intricate proof: Distributivity law in EHA (about 400 lines of code)
- Koch's [HKK21] proof mode immensely helpful
- Lots of code dealing with substitutions

Isabelle

- 100 lines of code (60 for Löb proof, 40 for lemmas)
- Can still be shortened

Background: Used Hilbert System

$$\mathcal{H}(\varphi \to \psi \to \varphi) \qquad \mathcal{H}((\varphi \to \psi \to \tau) \to (\psi \to \tau) \to \varphi \to \tau)$$

$$\mathcal{H}(\varphi \to \psi \to \varphi \land \psi) \qquad \mathcal{H}(\varphi \land \psi \to \varphi)$$

$$\mathcal{H}(\varphi \to \varphi \lor \psi) \qquad \mathcal{H}(\varphi \land \psi \to \psi)$$

$$\mathcal{H}(\psi \to \varphi \lor \psi) \qquad \mathcal{H}(\varphi \lor \psi \to (\varphi \to \tau) \to (\psi \to \tau) \to \tau)$$

$$\mathcal{H}(\bot \to \varphi) \qquad \mathcal{H}(\varphi \to \forall x.\, \varphi) \qquad x \text{ fresh for } \varphi$$

$$\mathcal{H}((\forall x.\, \varphi) \to \varphi[x \mapsto t]) \qquad \mathcal{H}((\forall x.\, \varphi \to \psi) \to (\forall x.\, \varphi) \to \forall x.\, \psi)$$

$$\mathcal{H}(\varphi[x \mapsto t] \to \exists x.\, \varphi) \qquad \mathcal{H}((\exists x.\, \varphi) \to (\forall x.\, \varphi \to \psi) \to \psi) \qquad x \text{ fresh for } \psi$$

$$\frac{\mathrm{PA} \vdash_{\mathcal{H}} \varphi \to \psi \qquad \mathrm{PA} \vdash_{\mathcal{H}} \varphi}{\mathrm{PA} \vdash_{\mathcal{H}} \psi} \qquad \frac{\varphi \in \mathcal{H}}{\mathrm{PA} \vdash_{\mathcal{H}} \forall x_1. \ldots x_n.\, \varphi} \qquad \frac{\varphi \in \mathrm{PA}}{\mathrm{PA} \vdash_{\mathcal{H}} \varphi}$$

Elements from Rautenberg, Troelstra and Schwichtenberg, as well as both.

Extended PA

Definition (Extended Signature of Peano Arithmetic (EPA), simplified)

In addition to the symbols of PA, EPA contains the following function symbols:

$[\,]$ (nil)	$\lvert\ell\rvert$ (length)	$\ell +\!\!+ \ell'$ (append)
$x :: \ell$ (cons)	$\ell[i]$ (indexed access)	$x \rightsquigarrow y$ (implication)

Further, EPA adds the unary predicate symbol $\mathcal{A}$ to PA.

- $\mathrm{EPA} \vdash \overline{\varphi \to \psi} = \overline{\varphi} \rightsquigarrow \overline{\psi}$ (object level implication function)
- If $\varphi \in \mathcal{H}$, then $\mathrm{EPA} \vdash \mathcal{A}\,(\forall x_1. \ldots x_n. \varphi)$
- If $\varphi \in \mathrm{PA}$, then $\mathrm{EPA} \vdash \mathcal{A}\,\varphi$

Formal proofs: Spelling out (some of) the Details

Definition (Formal proofs)

A proof of φ is a nonempty list $\ell = [\psi_1, \ldots, \psi_n] : \mathcal{L}(\mathbb{F})$ with $\varphi = \psi_n$ s.t. for each i

- ψ_i is an axiom of PA, a generalisation of a Hilbert axiom, or
- there are $j, j' < i$ such that ψ_i follows from $\psi_j, \psi_{j'}$ by modus ponens.

Definition (Provability predicate)

$$\begin{aligned}\mathrm{Prf}(x, y) &:= (\exists z.\, |x| = S\, z \wedge x[z] = y) \wedge \forall i.\, i < |x| \rightarrow \mathrm{WellFormed}(x, i) \\ \mathrm{WellFormed}(x, i) &:= \mathcal{A}(x) \vee \exists j\, j'.\, j < i \wedge j' < i \wedge x[j] = x[j'] \rightsquigarrow x[i]\end{aligned}$$

Technical Background: Gödel Numberings

Problem

Let $\varphi(x), \psi : \mathbb{F}$.
We used $\varphi(\overline{\psi})$ for 'substituting some encoding of ψ for x in φ'.
ψ is not a **number**, but a **formula**.

Typical issue. Gödel faced it himself.

Remark (Gödelisation)

There are functions $\text{göd} : \mathbb{F} \to \mathbb{N}$, $\text{göd}^{-1} : \mathbb{N} \to \mathbb{F}$ *inverting each other.*

$\varphi(\overline{\psi}) \rightsquigarrow \varphi(\overline{\text{göd}(\psi)})$

Technical Background: $\mathsf{CT}_{\mathsf{PA}}$ is too Weak

Axiom ($\mathsf{CT}_{\mathsf{PA}}$)

For every $f : \mathbb{N} \to \mathbb{N}$, there is a formula $\varphi(x_1, x_2)$ such that for all $n : \mathbb{N}$

$$\mathsf{PA} \vdash \forall y.\, \varphi(\overline{n}, y) \leftrightarrow y = \overline{f\ n}.$$

Example

Suppose the successor function $\mathtt{S} : \mathbb{N} \to \mathbb{N}$ is represented by $\varphi_{\mathtt{S}}(x, y)$.

Question: Can we derive, for all $n \in \mathbb{N}$, that $\mathsf{PA} \vdash \varphi_{\mathtt{S}}(\overline{n}, \mathsf{S}\,\overline{n})$? **Yes!**

- Use property of $\varphi_{\mathtt{S}}$: $\mathsf{PA} \vdash \mathsf{S}\,\overline{n} = \overline{\mathtt{S}n}$
- By definition of numerals, $\mathsf{PA} \vdash \mathsf{S}\,\overline{n} = \mathsf{S}\,\overline{n}$, easy to finish

Question: Can we derive $\mathsf{PA} \vdash \forall x.\, \varphi_{\mathtt{S}}(x, \mathsf{S}\,x)$? **No!**

- Introduce x: $\mathsf{PA} \vdash \varphi_{\mathtt{S}}(x, \mathsf{S}\,x)$. No way to continue as x not a numeral