Proof theory of arithmetic

Lecture 1 – The rise of proof theory

Anupam Das

University of Birmingham

34TH European Summer School
in Logic, Language and Information

Ljubljana, Slovenia
7 August 2023

These slides are available at `http://www.anupamdas.com/esslli23`.
This lecture is based on slides from NASSLLI 2018, prepared with Thomas Powell.

Outline

Wherefore proof theory?

Proof theory is the study of mathematical proofs as formal objects.

Wherefore proof theory?

Proof theory is the study of mathematical proofs as formal objects.

Formally, a **proof system** defines what a proof may be. This allows not only the study of what is provable, but also of proofs themselves.

Wherefore proof theory?

Proof theory is the study of mathematical proofs as formal objects.

Formally, a **proof system** defines what a proof may be. This allows not only the study of what is provable, but also of proofs themselves.

But *why* do we study proofs?

Wherefore proof theory?

Proof theory is the study of mathematical proofs as formal objects.

Formally, a **proof system** defines what a proof may be. This allows not only the study of what is provable, but also of proofs themselves.

But *why* do we study proofs?

- *Foundational results*, which tell us something fundamental about reasoning itself. These lead to...
- *Applications*, which use insights and techniques from proof theory to accomplish something concrete in another discipline altogether.

Foundational results from the early days

Gödel's incompleteness theorem (1931), informally

We cannot prove every true sentence in a given proof system.

This result fundamentally reoriented the direction of logic research in the 20th century.

Foundational results from the early days

Gödel's incompleteness theorem (1931), informally

We cannot prove every true sentence in a given proof system.

This result fundamentally reoriented the direction of logic research in the 20th century.

Gentzen's *Hauptsatz* (1934), informally

Every purely logical theorem can be proved analytically, i.e. without making 'guesses'.

The translation of proofs to analytic form, cut-elimination, has become one of the most powerful tools in all of logic.

Proof theory and computation

Proof theory plays an essential role in understanding computation:

- Extracting *computational content* via normal forms. (cut-elimination, normalisation,…)
- …and by proof interpretations. (realisability, Dialectica, witnessing,…)
- Modelling non-deterministic complexity (proof complexity, proof search,…)
- Proof normalisation as computation. (Curry-Howard-Lambek,…)
- Proof search as computation. (focussing, logic programming,…)

References for this course

There are a number of excellent textbooks, including but not limited to:

1. [Takeuti, 1975] *Proof Theory*.
2. [Pohlers, 1989] *Proof Theory: an Introduction*.
3. [Buss, 1998] *Handbook of Proof Theory*.
4. [Troelstra and Schwichtenberg, 1996] *Basic Proof Theory*.
5. [Arai, 2020] *Ordinal Analysis with an Introduction to Proof Theory*.

For excellent high-level references, see also:

6. *Stanford Encyclopedia of Philosophy*. `http://plato.stanford.edu/`

NB: While 1 and 2 are a little dated, they remain seminal introductions to the field.

Sprechen Sie deutsch?

For this lecture, I recommend taking a look at original texts:

- [Frege, 1879] *Begriffsschrift*. (The first serious attempt to formalise mathematics in logic, unfortunately erroneously)
- [Whitehead and Russell, 1927] *Principia Mathematica*. (The second serious attempt, notably error-free)
- [Gödel, 1931] *Über Formal Unentscheidbare Sätze der Principia Mathematica Und Verwandter Systeme I*. (The incompleteness theorems)
- [Szabo, 1972] *The collected papers of Gerhard Gentzen*. English translations. (The foundations of modern proof theory)

Outline

"Concrete" mathematics

Up to the 19^{th} century, mathematics was primarily concerned with concrete objects which could be explicitly constructed. E.g.:

- Every number is either even or odd.
- There are infinitely many prime numbers.
- Every number can be written as the sum of four squares.
- Every continuous function can be integrated.
- Every non-constant polynomial over over the complex numbers has a root.

"Concrete" mathematics

Up to the 19th century, mathematics was primarily concerned with concrete objects which could be explicitly constructed. E.g.:

- Every number is either even or odd.
- There are infinitely many prime numbers.
- Every number can be written as the sum of four squares.
- Every continuous function can be integrated.
- Every non-constant polynomial over over the complex numbers has a root.

Proofs of these results yield concrete algorithms. E.g.:

- We can decide whether a number is even or odd.
- We can find the next prime number.
- For any number we can find four squares which sum to that number.
- We can compute the integral of f up to *any desired accuracy*.
- We can compute roots of polynomials up to *any desired accuracy*.

“Non-constructive” mathematics

With the advent of modern mathematics in the 19^{th} century, mathematicians started to reason about non-constructible objects.

- For every $f : \mathbb{N} \to \mathbb{N}$ there exists an n such that $f(n) \leqslant f(m)$ for all m.
- Every set of real numbers has a least upper bound.
- Every monotone, bounded sequence converges to a limit.
- Every ring has a maximal ideal.
- Every vector space has a basis.

"Non-constructive" mathematics

With the advent of modern mathematics in the 19$^{\text{th}}$ century, mathematicians started to reason about non-constructible objects.

- For every $f : \mathbb{N} \to \mathbb{N}$ there exists an n such that $f(n) \leqslant f(m)$ for all m.
- Every set of real numbers has a least upper bound.
- Every monotone, bounded sequence converges to a limit.
- Every ring has a maximal ideal.
- Every vector space has a basis.

In general, these existence results yield no effective algorithms.

They give the existence of **ideal objects**, based purely on formal reasoning.

"Non-constructive" mathematics

With the advent of modern mathematics in the 19$^{\text{th}}$ century, mathematicians started to reason about non-constructible objects.

- For every $f : \mathbb{N} \to \mathbb{N}$ there exists an n such that $f(n) \leqslant f(m)$ for all m.
- Every set of real numbers has a least upper bound.
- Every monotone, bounded sequence converges to a limit.
- Every ring has a maximal ideal.
- Every vector space has a basis.

In general, these existence results yield no effective algorithms.

They give the existence of **ideal objects**, based purely on formal reasoning.

Question

*Do these ideal objects really '*exist*'?*

Example: irrational powers

Proposition

There are irrational numbers a, b such that a^b is rational.

Example: irrational powers

Proposition

There are irrational numbers a, b such that a^b is rational.

Proof.

We know that $\sqrt{2}$ is irrational. What about $\sqrt{2}^{\sqrt{2}}$? We have two cases:

- If $\sqrt{2}^{\sqrt{2}}$ is rational, then set $a = b = \sqrt{2}$.
- Otherwise $\sqrt{2}^{\sqrt{2}}$ is irrational. Set $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$. We have,

$$a^b = (\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2$$

so a^b is rational as required. □

Example: irrational powers

Proposition

There are irrational numbers a, b such that a^b is rational.

Proof.

We know that $\sqrt{2}$ is irrational. What about $\sqrt{2}^{\sqrt{2}}$? We have two cases:

- If $\sqrt{2}^{\sqrt{2}}$ is rational, then set $a = b = \sqrt{2}$.
- Otherwise $\sqrt{2}^{\sqrt{2}}$ is irrational. Set $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$. We have,

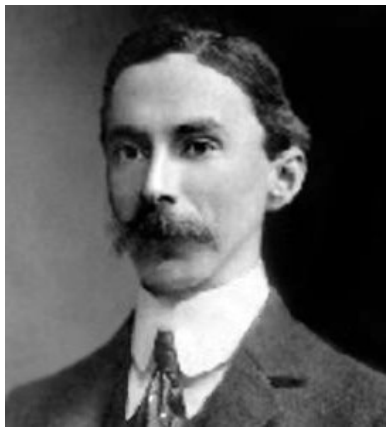
$$a^b = (\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2$$

so a^b is rational as required. □

Question: Is $\sqrt{2}^{\sqrt{2}}$ rational or irrational?!

Down the rabbit hole...

Even worse, this style of mathematical reasoning can lead us down fallacious paths.



Example (Russell's paradox (B. Russell, 1901))

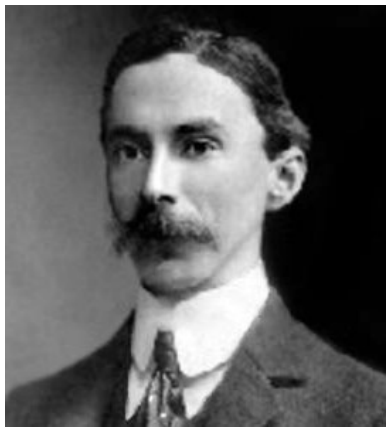
Let $R := \{x : x \notin x\}$. Is $R \in R$?

- If $R \in R$ then by definition we must have that $R \notin R$;
- But if $R \notin R$ then we must have $R \in R$.

We have a contradiction!

Down the rabbit hole…

Even worse, this style of mathematical reasoning can lead us down fallacious paths.



Example (Russell's paradox (B. Russell, 1901))

Let $R := \{x : x \notin x\}$. Is $R \in R$?

- If $R \in R$ then by definition we must have that $R \notin R$;
- But if $R \notin R$ then we must have $R \in R$.

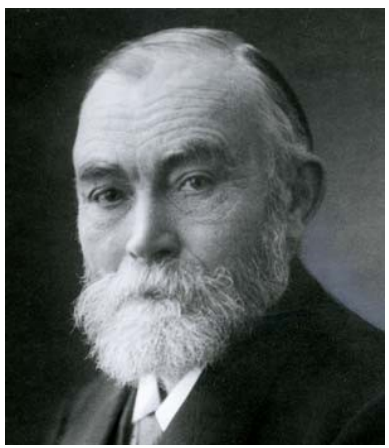
We have a contradiction!

Conclusion: Our naive formulation of mathematics, in particular set theory, is *inconsistent*: it cannot be trusted.

This one simple paradox destroyed a life's work:

Corollary

G. Frege's foundations of arithmetic are inconsistent.



The foundational crisis

The tragedy of Frege's foundations led to the so-called **foundational crisis**:

- Can we construct solid formal foundations for mathematics?
- Can we ensure that they are consistent?
- Can we give a *formal* treatment of set theory?

The foundational crisis

The tragedy of Frege's foundations led to the so-called **foundational crisis**:

- Can we construct solid formal foundations for mathematics?
- Can we ensure that they are consistent?
- Can we give a *formal* treatment of set theory?

These led to the resurgence of fundamental philosophical questions:

- What *is* a mathematical proof?
- Can mathematics/arithmetic be *reduced* to pure logic?
- Do mathematical objects exist, or are they just symbols?

The foundational crisis

The tragedy of Frege's foundations led to the so-called **foundational crisis**:

- Can we construct solid formal foundations for mathematics?
- Can we ensure that they are consistent?
- Can we give a *formal* treatment of set theory?

These led to the resurgence of fundamental philosophical questions:

- What *is* a mathematical proof?
- Can mathematics/arithmetic be *reduced* to pure logic?
- Do mathematical objects exist, or are they just symbols?

A HISTORICAL NOTE: Set-theoretic paradoxes were arguably known already by the 1880s, e.g. the *Burali-Forti* paradox. However, the sheer simplicity of Russell's paradox shook the mathematical world, questioning the most basic principles of mathematical reasoning.

Outline

Emerging philosophies

Two competing schools of thought emerged as a reaction to the foundational crisis:

Emerging philosophies

Two competing schools of thought emerged as a reaction to the foundational crisis:

INTUITIONISM (led by L. E. J. Brouwer, see [Iemhoff, 2016])

- Based on semantics.
- Mathematics is a mental construction: Something exists only if it can be exhibited.
- Infinite sets, maximals ideals, limits are all dubious unless they can be built explicitly.

Emerging philosophies

Two competing schools of thought emerged as a reaction to the foundational crisis:

Intuitionism (led by L. E. J. Brouwer, see [Iemhoff, 2016])

- Based on semantics.
- Mathematics is a mental construction: Something exists only if it can be exhibited.
- Infinite sets, maximals ideals, limits are all dubious unless they can be built explicitly.

Formalism (led by D. Hilbert, see [Weir, 2015])

- Based on syntax.
- Mathematics is a game of symbols: something 'exists' if it can be derived by mathematical axioms and logical rules.
- Infinite sets, maximal ideals, limits are all fine, as long as our underlying logical system can be trusted.

Emerging philosophies

Two competing schools of thought emerged as a reaction to the foundational crisis:

INTUITIONISM (led by L. E. J. Brouwer, see [Iemhoff, 2016])

- Based on semantics.
- Mathematics is a mental construction: Something exists only if it can be exhibited.
- Infinite sets, maximals ideals, limits are all dubious unless they can be built explicitly.

FORMALISM (led by D. Hilbert, see [Weir, 2015])

- Based on syntax.
- Mathematics is a game of symbols: something 'exists' if it can be derived by mathematical axioms and logical rules.
- Infinite sets, maximal ideals, limits are all fine, as long as our underlying logical system can be trusted.

We follow the formalist approach, but both have had deep impact on foundations.

Hilbert's Program

Hilbert's Program

In the early 20th century, Hilbert proposed benchmarks for satisfactory foundations:

Hilbert's Program, 1921

Find a collection of axioms and inference rules P for arithmetic which is:

1. **Complete:** all true statements in the language of arithmetic are provable in P.
2. **Consistency:** P does not prove a contradiction.

Hilbert's Program

In the early 20th century, Hilbert proposed benchmarks for satisfactory foundations:

Hilbert's Program, 1921

Find a collection of axioms and inference rules P for arithmetic which is:

1. **Complete:** all true statements in the language of arithmetic are provable in P.
2. **Consistency:** P does not prove a contradiction.

Aside: But this statement is circular! To show that P is consistent we need to work in some other system, which in turn needs to be shown to be consistent etc.

Hilbert's Program

In the early 20th century, Hilbert proposed benchmarks for satisfactory foundations:

Hilbert's Program, 1921

Find a collection of axioms and inference rules P for arithmetic which is:

1. **Complete:** all true statements in the language of arithmetic are provable in P.
2. **Consistency:** P does not prove a contradiction.

Aside: But this statement is circular! To show that P is consistent we need to work in some other system, which in turn needs to be shown to be consistent etc. Hilbert was well aware of this, so he asked for the following further refinement:

(continued)

3. **Finitary consistency:** the fact that P is consistent is demonstrable using only simple **finitary** methods, whose validity *cannot be questioned*.

Unwinding Hilbert's formalism

The notion of '*finitary*' in Hilbert's program is crucial (and controversial). It distinguishes object-level systems P from 'finitary' meta-level systems N:

- P can reason about crazy objects which cannot be constructed. These objects would be rejected by intuitionists.
- N is grounded in a world of numbers and simple arithmetic operations, which no reasonable person could doubt. It is unquestionably correct.

Unwinding Hilbert's formalism

The notion of '*finitary*' in Hilbert's program is crucial (and controversial). It distinguishes object-level systems P from 'finitary' meta-level systems N:

- P can reason about crazy objects which cannot be constructed. These objects would be rejected by intuitionists.
- N is grounded in a world of numbers and simple arithmetic operations, which no reasonable person could doubt. It is unquestionably correct.

In particular Hilbert's program requires:

$$\mathsf{N} \vdash \text{“}\mathsf{P} \text{ is consistent”} \tag{1}$$

Unwinding Hilbert's formalism

The notion of '*finitary*' in Hilbert's program is crucial (and controversial). It distinguishes object-level systems P from 'finitary' meta-level systems N:

- P can reason about crazy objects which cannot be constructed. These objects would be rejected by intuitionists.
- N is grounded in a world of numbers and simple arithmetic operations, which no reasonable person could doubt. It is unquestionably correct.

In particular Hilbert's program requires:

$$\mathsf{N} \vdash \text{“}\mathsf{P} \text{ is consistent”} \tag{1}$$

Intuition: to trust P, it is enough to trust N.

Unwinding Hilbert's formalism

The notion of '*finitary*' in Hilbert's program is crucial (and controversial). It distinguishes object-level systems P from 'finitary' meta-level systems N:

- P can reason about crazy objects which cannot be constructed. These objects would be rejected by intuitionists.
- N is grounded in a world of numbers and simple arithmetic operations, which no reasonable person could doubt. It is unquestionably correct.

In particular Hilbert's program requires:

$$\mathsf{N} \vdash \text{“P is consistent”} \tag{1}$$

Intuition: to trust P, it is enough to trust N.

Example: Hilbert's formalist philosophy identifies:

- Limits of monotone bounded sequences exist;
- There is a proof in P that every monotone bounded sequence has a limit;

for an appropriate system P, in particular satisfying (1), for an appropriate system N.

Chasing dreams

Hilbert's program is a great idea! It is the gold standard of formalist philosophy, guaranteeing mathematical practice that is free from contradictions and paradoxes.

There is only one tiny catch…

Chasing dreams

Hilbert's program is a great idea! It is the gold standard of formalist philosophy, guaranteeing mathematical practice that is free from contradictions and paradoxes.

There is only one tiny catch…

… it doesn't work.

Chasing dreams

Hilbert's program is a great idea! It is the gold standard of formalist philosophy, guaranteeing mathematical practice that is free from contradictions and paradoxes.

There is only one tiny catch…

… it doesn't work.

ENTER K. GÖDEL:

Outline

Gödel's First Incompletness Theorem, 1931

Any consistent formal system P *over the language of arithmetic is incomplete: there are true statements that* P *cannot prove.*

Gödel's First Incompletness Theorem, 1931

Any consistent formal system P *over the language of arithmetic is incomplete: there are true statements that* P *cannot prove.*

So Hilbert's program fails at the very first step! Namely, condition 1 fails.

How on earth did Gödel prove such a result?

Gödel's First Incompletness Theorem, 1931

Any consistent formal system P *over the language of arithmetic is incomplete: there are true statements that* P *cannot prove.*

So Hilbert's program fails at the very first step! Namely, condition 1 fails.

How on earth did Gödel prove such a result?

References for this part: [Raatikainen, 2018, Smullyan, 1992].

Fixed points in arithmetic

Remember Russell's paradox? It uses self-reference. But the language of arithmetic is quite expressive, and can formulate all sorts of **fixed points**.

Fixed points in arithmetic

Remember Russell's paradox? It uses self-reference. But the language of arithmetic is quite expressive, and can formulate all sorts of **fixed points**.

Fix a **coding** $\ulcorner \cdot \urcorner$ from arithmetical syntax to $\mathbb{N}$:

The Diagonal Lemma

For any formula $\varphi(x)$, there is a sentence ψ such that $\mathsf{P} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$.

Intuition: ψ is the sentence "I satisfy φ". The proof is recursion-theoretic.

Fixed points in arithmetic

Remember Russell's paradox? It uses self-reference. But the language of arithmetic is quite expressive, and can formulate all sorts of **fixed points**.

Fix a **coding** $\ulcorner \cdot \urcorner$ from arithmetical syntax to $\mathbb{N}$:

The Diagonal Lemma

For any formula $\varphi(x)$, there is a sentence ψ such that $\mathsf{P} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$.

Intuition: ψ is the sentence "I satisfy φ". The proof is recursion-theoretic.

In particular, by setting $\varphi(\cdot)$ to be "unprovability", we may construct the sentence:

This sentence is not provable in P.

Fixed points in arithmetic

Remember Russell's paradox? It uses self-reference. But the language of arithmetic is quite expressive, and can formulate all sorts of **fixed points**.

Fix a **coding** $\ulcorner \cdot \urcorner$ from arithmetical syntax to $\mathbb{N}$:

The Diagonal Lemma

For any formula $\varphi(x)$, there is a sentence ψ such that $\mathsf{P} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$.

Intuition: ψ is the sentence "I satisfy φ". The proof is recursion-theoretic.

In particular, by setting $\varphi(\cdot)$ to be "unprovability", we may construct the sentence:

This sentence is not provable in P.

- If it were false, then it would be provable, which is bad.
- So it must be true, but then by definition it is not provable!

Fixed points in arithmetic

Remember Russell's paradox? It uses self-reference. But the language of arithmetic is quite expressive, and can formulate all sorts of **fixed points**.

Fix a **coding** $\ulcorner \cdot \urcorner$ from arithmetical syntax to $\mathbb{N}$:

The Diagonal Lemma

For any formula $\varphi(x)$, there is a sentence ψ such that $\mathsf{P} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$.

Intuition: ψ is the sentence "I satisfy φ". The proof is recursion-theoretic.

In particular, by setting $\varphi(\cdot)$ to be "unprovability", we may construct the sentence:

This sentence is not provable in P.

- If it were false, then it would be provable, which is bad.
- So it must be true, but then by definition it is not provable!

Thus P *does not prove* every true sentence, yielding first incompleteness.

Things can only get worse

Things can only get worse

Gödel's results went *far deeper*: not only are theories like arithmetic incomplete, they are unable to verify their own consistency.

Things can only get worse

Gödel's results went *far deeper*: not only are theories like arithmetic incomplete, they are unable to verify their own consistency.

Gödel's Second Incompletess Theorem, 1931

If P is consistent, then its consistency cannot be proved within P itself.

Things can only get worse

Gödel's results went *far deeper*: not only are theories like arithmetic incomplete, they are unable to verify their own consistency.

Gödel's Second Incompletess Theorem, 1931

If P is consistent, then its consistency cannot be proved within P itself.

This is even *worse* for Hilbert's program. Not only does condition 1 fail, even if it satisfies condition 2, even a very weak version of condition 3 fails too.

Things can only get worse

Gödel's results went *far deeper*: not only are theories like arithmetic incomplete, they are unable to verify their own consistency.

Gödel's Second Incompletess Theorem, 1931

If P is consistent, then its consistency cannot be proved within P itself.

This is even *worse* for Hilbert's program. Not only does condition 1 fail, even if it satisfies condition 2, even a very weak version of condition 3 fails too.

Idea of argument: Formalise proof of the first incompleteness within P itself.

Understanding incompleteness: provability and modal logic

Understanding incompleteness: provability and modal logic

Implicit is that (even weak systems) P may formalise their own **provability predicate**:

Definition

Write $\Box\varphi$ for an arithmetical formula stating "there is a P-proof of φ".

Understanding incompleteness: provability and modal logic

Implicit is that (even weak systems) P may formalise their own **provability predicate**:

Definition

Write $\Box\varphi$ for an arithmetical formula stating "there is a P-proof of φ".

Lemma (Hilbert-Bernays-Löb conditions)

(nec) *If* $\mathsf{P} \vdash \varphi$ *then* $\mathsf{P} \vdash \Box\varphi$.

(k) $\mathsf{P} \vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.

(4) $\mathsf{P} \vdash \Box\varphi \to \Box\Box\varphi$.

Understanding incompleteness: provability and modal logic

Implicit is that (even weak systems) P may formalise their own **provability predicate**:

Definition

Write $\Box\varphi$ for an arithmetical formula stating "there is a P-proof of φ".

Lemma (Hilbert-Bernays-Löb conditions)

(nec) *If* $\mathsf{P} \vdash \varphi$ *then* $\mathsf{P} \vdash \Box\varphi$.

(k) $\mathsf{P} \vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.

(4) $\mathsf{P} \vdash \Box\varphi \to \Box\Box\varphi$.

Theorem (Löb, 1955)

If $\mathsf{P} \vdash \Box\varphi \to \varphi$ *then* $\mathsf{P} \vdash \varphi$.

Intuition: "I am provable" is provable!

Understanding incompleteness: provability and modal logic

Implicit is that (even weak systems) P may formalise their own **provability predicate**:

Definition

Write $\Box\varphi$ for an arithmetical formula stating "there is a P-proof of φ".

Lemma (Hilbert-Bernays-Löb conditions)

(nec) *If* $\mathsf{P} \vdash \varphi$ *then* $\mathsf{P} \vdash \Box\varphi$.

(k) $\mathsf{P} \vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.

(4) $\mathsf{P} \vdash \Box\varphi \to \Box\Box\varphi$.

Theorem (Löb, 1955)

If $\mathsf{P} \vdash \Box\varphi \to \varphi$ *then* $\mathsf{P} \vdash \varphi$.

Intuition: "I am provable" is provable!

This has a purely modal proof (using conditions above), after just one instance of the *Diagonal Lemma*, giving ψ such that:

$$\mathsf{P} \vdash \psi \leftrightarrow (\Box\psi \to \varphi) \tag{2}$$

Concluding second incompleteness

From *Löb's theorem* we immediately arrive at *Gödel's second incompleteness*:

Corollary (Gödel's Second Incompleteness Theorem, again)

If P *proves its own consistency, then* P *is, in fact, inconsistent*.

Concluding second incompleteness

From *Löb's theorem* we immediately arrive at *Gödel's second incompleteness*:

Corollary (Gödel's Second Incompleteness Theorem, again)

If P *proves its own consistency, then* P *is, in fact, inconsistent*.

Proof.

Writing $\mathrm{Con}(\mathsf{P})$ for $\neg\Box\bot$ ("there is no proof of a contradiction"), we have:

$$\begin{array}{lll} \mathsf{P} \vdash \mathrm{Con}(\mathsf{P}) & \Longrightarrow & \mathsf{P} \vdash \Box\bot \to \bot \quad \text{by pure logic} \\ & \Longrightarrow & \mathsf{P} \vdash \bot \quad \text{by Löb's theorem.} \end{array} \quad \square$$

Recap: what this means for Hilbert's program

Recap: what this means for Hilbert's program

For any consistent system P:

- P cannot be complete. If it were then it would prove elementary arithmetic, which would render it vulnerable to the first incompleteness theorem.

Recap: what this means for Hilbert's program

For any consistent system P:

- P cannot be complete. If it were then it would prove elementary arithmetic, which would render it vulnerable to the first incompleteness theorem.
- P cannot prove its own consistency, by the second incompleteness theorem. In particular, its consistency cannot be proved in any simpler subsystem N.

Recap: what this means for Hilbert's program

For any consistent system P:

- P cannot be complete. If it were then it would prove elementary arithmetic, which would render it vulnerable to the first incompleteness theorem.
- P cannot prove its own consistency, by the second incompleteness theorem. In particular, its consistency cannot be proved in any simpler subsystem N.

Hilbert's program, in its strictest sense, is unachievable.

Outline

Exercises

1. Use the *Diagonal Lemma* to derive Tarski's *Undefinability of Truth:* there is no formula $\mathrm{Tr}(x)$ of arithmetic such that $\mathsf{P} \vdash \mathrm{Tr}(\ulcorner\psi\urcorner) \leftrightarrow \psi$, for any sentence ψ.
2. Show that the following three formulations of 'consistency' are equivalent:
   a. P does not prove φ and $\neg\varphi$, for any φ.
   b. P does not prove $\bot$.
   c. There is something that P does not prove.
3. What does it mean to be a fixed point (i.e. the sentence ψ from the *Diagonal Lemma*) of the formula $\mathrm{Even}(x)$, stating "x is even".
4. **(Hard)** Prove *Löb's Theorem* using only the 'modal' HBL conditions from just the one instance of the *Diagonal Lemma* in (2).

   Hint: it is known that *Löb's Theorem* cannot be proved in this way without using the *contraction*: from $\varphi \to \varphi \to \psi$ infer $\varphi \to \psi$.
5. (Think about how to) construct an appropriate injection $\ulcorner\cdot\urcorner$: from logical syntax to $\mathbb{N}$. Reflect on what arithmetic you need to (de)code.

Answers to exercises I

1. Suppose otherwise, and use the Diagonal Lemma to construct a sentence ψ such theat $\mathsf{P} \vdash \psi \leftrightarrow \neg \mathrm{Tr}(\ulcorner \psi \urcorner)$. Now, by plugging the two equivalences together, we get $\mathsf{P} \vdash \mathrm{Tr}(\ulcorner \psi \urcorner) \leftrightarrow \neg \mathrm{Tr}(\ulcorner \psi \urcorner)$, which contradicts consistency of P.
2.
   - (b) $\Longrightarrow$ (a). By contraposition, if P proves both φ and $\neg\varphi$, which is equivalent to $\varphi \to \bot$, then $\mathsf{P} \vdash \bot$ simply by *modus ponens*.
   - (c) $\Longrightarrow$ (b). By contraposition, if $\mathsf{P} \vdash \bot$ then $\mathsf{P} \vdash \varphi$ for any φ by *ex falso quodlibet*.
   - (a) $\Longrightarrow$ (c). By contraposition, and choose any formula φ.
3. ψ is a fixed point of $\mathrm{Even}(x)$ just if:
   - $\mathsf{P} \vdash \psi$ and $\ulcorner \psi \urcorner$ is even; or,
   - $\mathsf{P} \vdash \neg\psi$ and $\ulcorner \psi \urcorner$ is odd.
4. Assume $\mathsf{P} \vdash \Box\varphi \to \varphi$, and by the Diagonal Lemma let ψ be a sentence with $\mathsf{P} \vdash \psi \leftrightarrow (\Box\psi \to \varphi)$. We derive φ in P using only HBL conditions as follows:

$\psi \to (\Box\psi \to \varphi)$	by (2)($\to$)	$\Box\psi \to \Box\varphi$	by pure logic
$\Box(\psi \to (\Box\psi \to \varphi))$	by (nec)	$\Box\psi \to \varphi$	by assumption
$\Box\psi \to \Box(\Box\psi \to \varphi)$	by (k)	ψ	by (2)($\leftarrow$)
$\Box\psi \to (\Box\Box\psi \to \Box\varphi)$	by (k)	$\Box\psi$	by (nec)
$\Box\psi \to (\Box\psi \to \Box\varphi)$	by (4)	φ	by *mp*.

Outline

Reformulating Hilbert's program

Gödel incompleteness refutes Hilbert's program, *but only in its strictest sense*.

From a broader perspective, Hilbert's program has been an extraordinary success, particularly in proof theory and theoretical computer science. Its legacy includes:

Reformulating Hilbert's program

Gödel incompleteness refutes Hilbert's program, *but only in its strictest sense*.

From a broader perspective, Hilbert's program has been an extraordinary success, particularly in proof theory and theoretical computer science. Its legacy includes:

- The establishment of formalism: proofs as mathematical objects.
- The connection between proofs and programs.

Reformulating Hilbert's program

Gödel incompleteness refutes Hilbert's program, *but only in its strictest sense*.

From a broader perspective, Hilbert's program has been an extraordinary success, particularly in proof theory and theoretical computer science. Its legacy includes:

- The establishment of formalism: proofs as mathematical objects.
- The connection between proofs and programs.

To achieve this, we must *tweak* Hilbert's program. In particular we must drop condition 1, but we can keep 2 under an alternative to 3:

3' **Computational consistency:** the fact that P is consistent is provable in a system N that is entirely **computational**.

Reformulating Hilbert's program

Gödel incompleteness refutes Hilbert's program, *but only in its strictest sense*.

From a broader perspective, Hilbert's program has been an extraordinary success, particularly in proof theory and theoretical computer science. Its legacy includes:

- The establishment of formalism: proofs as mathematical objects.
- The connection between proofs and programs.

To achieve this, we must *tweak* Hilbert's program. In particular we must drop condition 1, but we can keep 2 under an alternative to 3:

3' **Computational consistency:** the fact that P is consistent is provable in a system N that is entirely **computational**.

The trustworthyness of P is reduced to that of 'simple' computational principles.

Computational interpretations

Computational interpretations

The spirit of Hilbert's program has led to deep connections between:

- Induction on complex invariants (quantifiers over $\mathbb{N}$, $\mathbb{R}$,...);
- Recursion up to explicit ordinals. (e.g. Gentzen's cut-elimination);
- Recursion at higher types. (e.g. Gödel *Dialectica*, Kleene/Kreisel *realisability*).

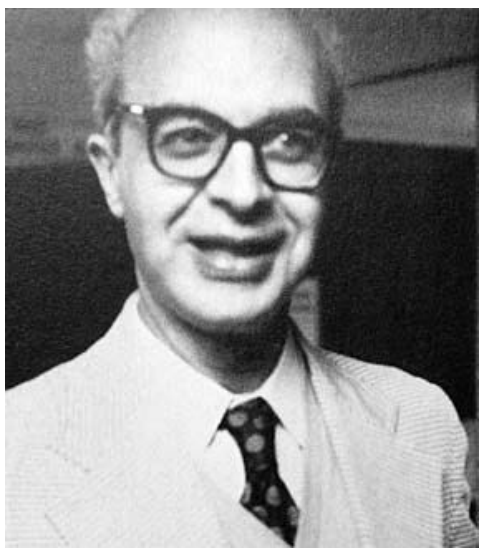
Computational interpretations

The spirit of Hilbert's program has led to deep connections between:

- Induction on complex invariants (quantifiers over $\mathbb{N}$, $\mathbb{R}$,...);
- Recursion up to explicit ordinals. (e.g. Gentzen's cut-elimination);
- Recursion at higher types. (e.g. Gödel *Dialectica*, Kleene/Kreisel *realisability*).

These have lead to one of the most successful applications of logic to mathematics:

Kreisel's unwinding program: **proof mining**



"What more do we know if we have proved a theorem by restricted mean than if we merely know that it is true?"

Outline

Goal of this course

This lecture course will focus on the connections between induction in arithmetic and recursion on higher ordinals, in the spirit of G. Gentzen.

We emphasise mathematical methods, but only sketch formalisations and explicit ordinal calculations.

Structure of this course

Background and motivation

Today: *The rise of proof theory*. Foundational crisis, Hilbert's program, Gödel incompleteness.

In-depth exploration

Tuesday: *Peano Arithmetic*. Syntax of first-order arithmetic, semantics, inductive reasoning, sequent calculus.

Wednesday: *Two consistency proofs*. Soundness (formalised), an infinitary system, cut-elimination, ordinal bounds.

Thursday: *Provably recursive functions*. Ordinal notations, transfinite recursion, Kleene realisability (time permitting).

Overview

Friday: *Further directions*. Fragments of arithmetic, extensions of arithmetic, second-order arithmetic.

References I

Arai, T. (2020).
Ordinal Analysis with an Introduction to Proof Theory.
Logic in Asia: Studia Logica Library. Springer Singapore.

Buss, S. R., editor (1998).
Handbook of Proof Theory, volume 137 of *Studies in Logic and the Foundations of Mathematics*.
Elsevier.

Frege, G. (1879).
Begriffsschrift: Eine Der Arithmetische Nachgebildete Formelsprache des Reinen Denkens.
L. Nebert.

Gödel, K. (1931).
über formal unentscheidbare sätze der principia mathematica und verwandter systeme i.
Monatshefte für Mathematik, 38(1):173–198.

Iemhoff, R. (2016).
Intuitionism in the philosophy of mathematics.
In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition.

References II

Pohlers, W. (1989).
Proof theory: An introduction, volume 1407.
Springer.

Raatikainen, P. (2018).
Gödel's incompleteness theorems.
In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2018 edition.

Smullyan, R. M. (1992).
Gödel's Incompleteness Theorems.
Oxford Logic Guides. Oxford University Press.

Szabo, M. E. (1972).
The Collected Papers of Gerhard Gentzen.
Philosophy of Science, 39(1):91–91.

Takeuti, G. (1975).
Proof Theory.
New York, N.Y., U.S.A.: Sole distributors for the U.S.A. and Canada, Elsevier Science Pub. Co.

References III

Troelstra, A. S. and Schwichtenberg, H. (1996).

Basic Proof Theory.

Cambridge University Press, New York, NY, USA.

Weir, A. (2015).

Formalism in the philosophy of mathematics.

In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2015 edition.

Whitehead, A. N. and Russell, B. (1925–1927).

Principia Mathematica.

Cambridge University Press.