```
DXXXXXXXXXXXXXX
XIXXXXXXXXXXXXX
XXAXXXXXXXXXXXX
XXXGXXXXXXXXXXX
XXXXOXXXXXXXXXX
XXXXXNXXXXXXXXX
XXXXXXAXXXXXXXX
XXXXXXXLXXXXXXX
XXXXXXXXIXXXXXX
XXXXXXXXXSXXXXX
XXXXXXXXXXAXXXX
XXXXXXXXXXXTXXX
XXXXXXXXXXXXIXX
XXXXXXXXXXXXXOX
XXXXXXXXXXXXXXN
```

Diagonalisation

The Road to Infinities, Truth and Gödel

Rasmus Bakken — Mahin Hossain

University of Oxford

Michaelmas 2025

Summary: We introduce provability inside a formal system, prove Gödel's diagonal lemma, and use it to show the undefinability of truth in a formal logical setting.

Quote

"All Cretans are liars."

— Epimenides, a Cretan

Lecture 3 Summary

- What is a proof: define Hilbert-style system and axioms; example derivations.
- The Diagonal Lemma: construction of self-referential sentences via substitution and the diagonal operator.
- Undefinability of truth (Tarski): apply diagonal lemma to rule out a truth-defining formula in $\mathcal{L}$.

Roadmap

Motivating the concept

- From last time: a proof of φ_k is finite sequence of $\mathcal{L}$-sentences $\varphi_1, \dots, \varphi_k$ with some condition.

Motivating the concept

- From last time: a proof of φ_k is finite sequence of $\mathcal{L}$-sentences $\varphi_1, \ldots, \varphi_k$ with some condition.
- The relevant condition: for all $1 \leq i \leq k$
  - φ_i is an axiom, or

Motivating the concept

- From last time: a proof of φ_k is finite sequence of $\mathcal{L}$-sentences $\varphi_1, \ldots, \varphi_k$ with some condition.
- The relevant condition: for all $1 \leq i \leq k$
  - φ_i is an axiom, or
  - φ_i follows from φ_j and φ_m by some rule of inference where $j \leq m < i$.
- We use a Hilbert-style calculus.

Motivating the concept

- From last time: a proof of φ_k is finite sequence of $\mathcal{L}$-sentences $\varphi_1, \ldots, \varphi_k$ with some condition.
- The relevant condition: for all $1 \leq i \leq k$
  - φ_i is an axiom, or
  - φ_i follows from φ_j and φ_m by some rule of inference where $j \leq m < i$.
- We use a Hilbert-style calculus.
- Logical vs non-logical axioms, and only two rules of inference.

Logical axioms

Definition (First-order Predicate Logic)

Let φ, ψ and ξ be $\mathcal{L}$ formula, then the following are axioms of first-order logic:

- C1 $\varphi \to (\psi \to \varphi)$.
- C2 $(\varphi \to (\psi \to \xi)) \to ((\varphi \to \psi) \to (\varphi \to \xi))$.
- C3 $(\varphi \to \psi) \to (\neg\psi \to \neg\varphi)$.

Logical axioms

Definition (First-order Predicate Logic)

Let φ, ψ and ξ be $\mathcal{L}$ formula, then the following are axioms of first-order logic:

- C1 $\varphi \to (\psi \to \varphi)$.
- C2 $(\varphi \to (\psi \to \xi)) \to ((\varphi \to \psi) \to (\varphi \to \xi))$.
- C3 $(\varphi \to \psi) \to (\neg\psi \to \neg\varphi)$.
- C4 $\forall v_i \varphi(v_i) \to \varphi(t)$, where t is a term of $\mathcal{L}$ such that it does not contain a variable v_j where v_i occurs free in the scope of $\forall v_j$ in φ.
- C5 $\forall v_i(\varphi \to \psi) \to (\varphi \to \forall v_i \psi)$, provided v_i does not occur free in ψ.
- C6 $\forall v_i(v_i = v_i)$.
- C7 If F and G are atomic formulae, where G results from replacing some but not necessarily all of v_i in F by v_j, then $\forall v_i \forall v_j(v_i = v_j \to (F \to G))$ is an axiom.

Rules of Inference

Definition

The rules of inference are the following:

MP From φ and $\varphi \to \psi$ you may infer ψ.

Gen From φ you may infer $\forall v_i \varphi$.

Example: proving $\varphi \to \varphi$

1 $(\varphi \to ((\varphi \to \varphi) \to \varphi)) \to ((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$ C2

Example: proving $\varphi \to \varphi$

1 $(\varphi \to ((\varphi \to \varphi) \to \varphi)) \to ((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$ C2

2 $\varphi \to ((\varphi \to \varphi) \to \varphi)$ C1

Example: proving $\varphi \to \varphi$

1 $(\varphi \to ((\varphi \to \varphi) \to \varphi)) \to ((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$ C2

2 $\varphi \to ((\varphi \to \varphi) \to \varphi)$ C1

3 $((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$ MP line 1 and line 2

Example: proving $\varphi \to \varphi$

1 $(\varphi \to ((\varphi \to \varphi) \to \varphi)) \to ((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$ C2

2 $\varphi \to ((\varphi \to \varphi) \to \varphi)$ C1

3 $((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$ MP line 1 and line 2

4 $\varphi \to (\varphi \to \varphi)$ C1

Example: proving $\varphi \to \varphi$

1 $(\varphi \to ((\varphi \to \varphi) \to \varphi)) \to ((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$ C2

2 $\varphi \to ((\varphi \to \varphi) \to \varphi)$ C1

3 $((\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi))$ MP line 1 and line 2

4 $\varphi \to (\varphi \to \varphi)$ C1

5 $\varphi \to \varphi$ MP line 3 and 4

Peano Arithmetic

Definition (Peano Arithmetic)

Peano Arithmetic, or *PA* for short, is the theory given by the following set of axioms:

PA1 $\forall v_i(S(v_i) \neq 0)$.

PA2 $\forall v_i \forall v_j(S(v_i) = S(v_j) \rightarrow v_i = v_j)$.

Peano Arithmetic

Definition (Peano Arithmetic)

Peano Arithmetic, or *PA* for short, is the theory given by the following set of axioms:

PA1 $\forall v_i(S(v_i) \neq 0)$.

PA2 $\forall v_i \forall v_j(S(v_i) = S(v_j) \rightarrow v_i = v_j)$.

PA3 $\forall v_i(v_i + 0 = v_i)$.

PA4 $\forall v_i \forall v_j(v_i + S(v_j) = S(v_i + v_j))$.

Peano Arithmetic

Definition (Peano Arithmetic)

Peano Arithmetic, or *PA* for short, is the theory given by the following set of axioms:

PA1 $\forall v_i(S(v_i) \neq 0)$.

PA2 $\forall v_i \forall v_j(S(v_i) = S(v_j) \to v_i = v_j)$.

PA3 $\forall v_i(v_i + 0 = v_i)$.

PA4 $\forall v_i \forall v_j(v_i + S(v_j) = S(v_i + v_j))$.

PA5 $\forall v_i(v_i * 0 = 0)$.

PA6 $\forall v_i \forall v_j(v_i * S(v_j) = v_i * v_j + v_i)$.

Peano Arithmetic

Definition (Peano Arithmetic)

Peano Arithmetic, or *PA* for short, is the theory given by the following set of axioms:

PA1 $\forall v_i(S(v_i) \neq 0)$.

PA2 $\forall v_i \forall v_j(S(v_i) = S(v_j) \rightarrow v_i = v_j)$.

PA3 $\forall v_i(v_i + 0 = v_i)$.

PA4 $\forall v_i \forall v_j(v_i + S(v_j) = S(v_i + v_j))$.

PA5 $\forall v_i(v_i * 0 = 0)$.

PA6 $\forall v_i \forall v_j(v_i * S(v_j) = v_i * v_j + v_i)$.

PA7 $\forall v_i(exp(v_i, 0) = \overline{1})$.

PA8 $\forall v_i \forall v_j(exp(v_i, S(v_j)) = exp(v_i, v_j) * v_i)$.

Peano Arithmetic

Definition (Peano Arithmetic)

Peano Arithmetic, or *PA* for short, is the theory given by the following set of axioms:

PA1 $\forall v_i(S(v_i) \neq 0)$.

PA2 $\forall v_i \forall v_j(S(v_i) = S(v_j) \rightarrow v_i = v_j)$.

PA3 $\forall v_i(v_i + 0 = v_i)$.

PA4 $\forall v_i \forall v_j(v_i + S(v_j) = S(v_i + v_j))$.

PA5 $\forall v_i(v_i * 0 = 0)$.

PA6 $\forall v_i \forall v_j(v_i * S(v_j) = v_i * v_j + v_i)$.

PA7 $\forall v_i(exp(v_i, 0) = \overline{1})$.

PA8 $\forall v_i \forall v_j(exp(v_i, S(v_j)) = exp(v_i, v_j) * v_i)$.

$Ind_{\mathcal{L}}$ For any $\mathcal{L}$ formula with exactly one free variable $\varphi(v_i)$,
$(\varphi(0) \wedge \forall v_i(\varphi(v_i) \rightarrow \varphi(S(v_i)))) \rightarrow \forall v_i \varphi(v_i)$ is an axiom.

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.1)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}.$

Proof.

1 $\forall v_i \forall v_j (v_i + S(v_j) = S(v_i + v_j))$ PA4

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.1)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}.$

Proof.

1 $\forall v_i \forall v_j (v_i + S(v_j) = S(v_i + v_j))$ PA4

2 $\forall v_i \forall v_j (v_i + S(v_j) = S(v_i + v_j)) \rightarrow \forall v_j (S(0) + S(v_j) = S(S(0) + v_j))$
C4

3 $\forall v_j (S(0) + S(v_j) = S(S(0) + v_j))$ MP with line 1 and line 2

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.1)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}.$

Proof.

1 $\forall v_i \forall v_j (v_i + S(v_j) = S(v_i + v_j))$ PA4

2 $\forall v_i \forall v_j (v_i + S(v_j) = S(v_i + v_j)) \rightarrow \forall v_j (S(0) + S(v_j) = S(S(0) + v_j))$ C4

3 $\forall v_j (S(0) + S(v_j) = S(S(0) + v_j))$ MP with line 1 and line 2

4 $\forall v_j (S(0) + S(v_j) = S(S(0) + v_j)) \rightarrow S(0) + S(0) = S(S(0) + 0)$ C4

5 $S(0) + S(0) = S(S(0) + 0)$ MP with line 3 and 4

$\square$

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.2)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}.$

Proof.

5 $S(0) + S(0) = S(S(0) + 0)$ MP with line 3 and 4

6 $\forall v_i(v_i + 0 = v_i)$ PA3

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.2)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}$.

Proof.

5 $S(0) + S(0) = S(S(0) + 0)$ MP with line 3 and 4

6 $\forall v_i(v_i + 0 = v_i)$ PA3

7 $\forall v_i(v_i + 0 = v_i) \rightarrow S(0) + 0 = S(0)$ C4

8 $S(0) + 0 = S(0)$ MP with line 6 and line 7

$\square$

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.3)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}$.

Proof.

5 $S(0) + S(0) = S(S(0) + 0)$ MP with line 3 and 4

8 $S(0) + 0 = S(0)$ MP with line 6 and line 7

9 $\forall v_i \forall v_j (v_i = v_j \rightarrow (S(0) + S(0) = S(v_i) \rightarrow S(0) + S(0) = S(v_j))$ C7

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.3)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}.$

Proof.

5 $S(0) + S(0) = S(S(0) + 0)$ MP with line 3 and 4

8 $S(0) + 0 = S(0)$ MP with line 6 and line 7

9 $\forall v_i \forall v_j (v_i = v_j \to (S(0) + S(0) = S(v_i) \to S(0) + S(0) = S(v_j))$ C7

10 $\forall v_i \forall v_j (v_i = v_j \to (S(0) + S(0) = S(v_i) \to S(0) + S(0) = S(v_j)) \to \forall v_j (S(0) + 0 = v_j \to (S(0) + S(0) = S(S(0) + 0) \to S(0) + S(0) = S(v_j))$ C4

11 $\forall v_j (S(0) + 0 = v_j \to (S(0) + S(0) = S(S(0) + 0) \to S(0) + S(0) = S(v_j))$
MP line 9 and line 10

$\square$

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.4)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}$.

Proof.

5 $S(0) + S(0) = S(S(0) + 0)$ MP with line 3 and 4

8 $S(0) + 0 = S(0)$ MP with line 6 and line 7

11 $\forall v_j (S(0) + 0 = v_j \to (S(0) + S(0) = S(S(0) + 0) \to S(0) + S(0) = S(v_j))$ MP line 9 and line 10

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.4)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}.$

Proof.

5 $S(0) + S(0) = S(S(0) + 0)$ MP with line 3 and 4

8 $S(0) + 0 = S(0)$ MP with line 6 and line 7

11 $\forall v_j(S(0) + 0 = v_j \to (S(0) + S(0) = S(S(0) + 0) \to S(0) + S(0) = S(v_j))$
MP line 9 and line 10

12 $\forall v_j(S(0) + 0 = v_j \to (S(0) + S(0) = S(S(0) + 0) \to S(0) + S(0) = S(v_j)) \to S(0) + 0 = S(0) \to (S(0) + S(0) = S(S(0)) \to S(0) + S(0) = S(S(0))$ C4

13 $S(0) + 0 = S(0) \to (S(0) + S(0) = S(S(0) + 0) \to S(0) + S(0) = S(S(0))$
MP line 11 and line 12

Example in PA: $\overline{1}+\overline{1}=\overline{2}$ (pt.4)

Theorem

$PA \vdash \overline{1}+\overline{1}=\overline{2}.$

Proof.

5 $S(0)+S(0)=S(S(0)+0)$ MP with line 3 and 4

8 $S(0)+0=S(0)$ MP with line 6 and line 7

11 $\forall v_j(S(0)+0=v_j \rightarrow (S(0)+S(0)=S(S(0)+0) \rightarrow S(0)+S(0)=S(v_j))$
MP line 9 and line 10

12 $\forall v_j(S(0)+0=v_j \rightarrow (S(0)+S(0)=S(S(0)+0) \rightarrow S(0)+S(0)=S(v_j)) \rightarrow S(0)+0=S(0) \rightarrow (S(0)+S(0)=S(S(0)) \rightarrow S(0)+S(0)=S(S(0))$ C4

13 $S(0)+0=S(0) \rightarrow (S(0)+S(0)=S(S(0)+0) \rightarrow S(0)+S(0)=S(S(0))$
MP line 11 and line 12

14 $(S(0)+S(0)=S(S(0)+0) \rightarrow S(0)+S(0)=S(S(0))$ MP line 8 and line 13

Example in PA: $\overline{1} + \overline{1} = \overline{2}$ (pt.5)

Theorem

$PA \vdash \overline{1} + \overline{1} = \overline{2}.$

Proof.

5 $S(0) + S(0) = S(S(0) + 0)$ MP with line 3 and 4

14 $(S(0) + S(0) = S(S(0) + 0) \rightarrow S(0) + S(0) = S(S(0))$ MP line 8 and line 13

15 $S(0) + S(0) = S(S(0))$ MP line 5 and line 14

$\square$

Roadmap

Gödel's Diagonal Lemma

Diagonal Lemma

If $\varphi(v_n)$ is an $\mathcal{L}$-formula with one free variable, then there is an $\mathcal{L}$-sentence γ such that

$$PA \vdash \gamma \leftrightarrow \varphi(\overline{\ulcorner \gamma \urcorner}).$$

- For any one-place formula, we can manufacture a sentence that says of itself that it has property φ.
- This turns Gödel codes $+$ substitution into genuine *self-reference*.

Substitution on Formulae

- Given a formula φ and a term t, $\varphi(t/v_i)$ means:
  - uniformly replace all *free* occurrences of v_i in φ by t.
- Example: φ is $S(v_5) = S(S(0))$.
  - $\varphi(S(0)/v_5)$ is $S(S(0)) = S(S(0))$.
  - $\varphi(0/v_5)$ is $S(0) = S(S(0))$.
  - $\varphi(v_3/v_5)$ is $S(v_3) = S(S(0))$.
- Beware **variable capture**:
  - quantifiers can accidentally bind new free variables,
  - we avoid this by renaming bound variables before substituting.

Internalising Substitution: The Function *sub*

Lemma (informal)

There is an $\mathcal{L}$-definable function $sub(v_1, v_2, v_3)$ such that

$$PA \vdash \overline{\ulcorner \varphi \urcorner} = sub(\overline{\ulcorner \psi \urcorner}, \overline{n}, \overline{\ulcorner t \urcorner})$$

iff $\varphi = \psi(t/v_n)$.

- So substitution on formulas can be mirrored by a calculation on Gödel numbers.
- We take the existence of such a *sub*-function for granted (construction is tedious but standard).

The Diagonal Operator

Definition (Diagonal Operator)

Let $\varphi(v_n)$ be a formula with v_n its only free variable. Define

$$\mathrm{dia}(\overline{\ulcorner\varphi(v_n)\urcorner}) := sub(\overline{\ulcorner\varphi\urcorner}, \overline{n}, \overline{\ulcorner\overline{\ulcorner\varphi\urcorner}\urcorner}).$$

- Intuition:

$$\mathrm{dia}(\overline{\ulcorner\varphi\urcorner}) = \overline{\ulcorner\varphi(\overline{\ulcorner\varphi\urcorner}/v_n)\urcorner}.$$

- It takes the code of φ and returns the code of the result of plugging that very code back into φ.
- This is the core mechanism behind self-reference.

Proof Idea of the Diagonal Lemma

- Fix a formula $\varphi(v_n)$.
- Consider the auxiliary formula

$$\theta(v_n) := \exists x \, (x = \mathrm{dia}(v_n) \wedge \varphi(x)).$$

- Now look at the sentence

$$\gamma := \exists x \, (x = \mathrm{dia}(\overline{\ulcorner \theta \urcorner}) \wedge \varphi(x)).$$

- Using the properties of *sub* and dia, one checks inside *PA* that

$$\mathrm{dia}(\overline{\ulcorner \theta \urcorner}) = \overline{\ulcorner \gamma \urcorner}.$$

- Hence *PA* proves

$$\gamma \leftrightarrow \varphi(\overline{\ulcorner \gamma \urcorner}),$$

  which is exactly the conclusion of the Diagonal Lemma.

Roadmap

What Should a Truth Predicate Do?

- Suppose $\varphi(v_i)$ is meant to say:

$$\varphi(\overline{\ulcorner\psi\urcorner}) \quad = \text{“}\psi \text{ is true”}.$$

- We would like φ to respect logical structure, e.g. disjunction:

$$PA \vdash \varphi(\overline{\ulcorner\psi_1 \vee \psi_2\urcorner}) \leftrightarrow \varphi(\overline{\ulcorner\psi_1\urcorner}) \vee \varphi(\overline{\ulcorner\psi_2\urcorner}).$$

- More strongly: there is a coding function $x \dot{\vee} y$ with

$$x = \ulcorner\psi_1\urcorner, y = \ulcorner\psi_2\urcorner \Rightarrow x \dot{\vee} y = \ulcorner\psi_1 \vee \psi_2\urcorner,$$

  and we might want

$$PA \vdash \forall x \forall y\, (\varphi(x \dot{\vee} y) \leftrightarrow \varphi(x) \vee \varphi(y)).$$

Tarski Biconditionals

Tarski Biconditionals (for φ)

For every $\mathcal{L}$-sentence ψ,

$$PA \vdash \varphi(\overline{\ulcorner \psi \urcorner}) \leftrightarrow \psi.$$

- This is a very natural *minimal* requirement for a truth definition.
- Read: "φ holds of the code of ψ iff ψ."
- Tarski's theorem: no such φ exists in the language of *PA*.

Tarski's Undefinability Theorem

Theorem (Tarski)

There is no $\mathcal{L}$-formula $\varphi(v_i)$ such that for all $\mathcal{L}$-sentences ψ,

$$PA \vdash \varphi(\overline{\ulcorner \psi \urcorner}) \leftrightarrow \psi.$$

- So: the intuitive notion of *truth in the standard model of PA* is not arithmetically definable.
- This contrasts with many other notions (e.g. "even", "provable in *PA*") which *are* definable in $\mathcal{L}$.

Proof Idea: The Liar Sentence

Proof sketch

Assume, for contradiction, that such a $\varphi(v_i)$ exists.

- Apply the Diagonal Lemma to $\neg\varphi(v_i)$.
- Obtain a sentence λ such that

$$PA \vdash \lambda \leftrightarrow \neg\varphi(\overline{\ulcorner\lambda\urcorner}).$$

- But by the Tarski biconditionals we also have

$$PA \vdash \lambda \leftrightarrow \varphi(\overline{\ulcorner\lambda\urcorner}).$$

- Combining:

$$PA \vdash \lambda \leftrightarrow \neg\lambda,$$

  contradicting the consistency of PA.

The Liar and Beyond

- The sentence λ above is a formal **liar sentence**:

$$\lambda \text{ says "I am not true (w.r.t. } \varphi\text{)".}$$

- Tarski's theorem shows:
  - No $\mathcal{L}$-formula can define a fully adequate truth predicate for arithmetic.
  - Truth is strictly more complex than many other arithmetical properties.
- One response: extend the language with a new primitive truth predicate T.
- Example theory:

$$TB := PA \cup \{ T\overline{\ulcorner\psi\urcorner} \leftrightarrow \psi \mid \psi \text{ an } \mathcal{L}\text{-sentence} \}.$$

- The liar reappears, but now at the *extended* level, and must be handled with care.