Epistemic Arithmetic

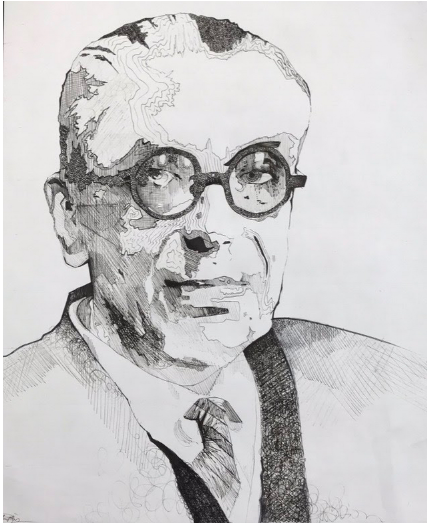
Eric Pacuit

July 28, 2025

Plan

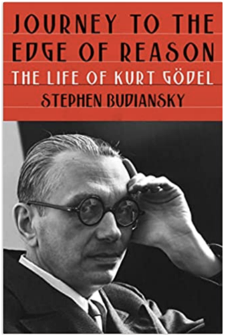
- Introduction: Smullyan's Machine
- Background
  - Formal Arithmetic
  - Gödel's Incompleteness Theorems
  - Names and Gödel numbering
  - Fixed Point Theorem
- Provability predicate and Löb's Theorem
- Provability logic
- Truth predicate and Tarski's Theorem
- A Primer on Epistemic and Doxastic Logic
- Anti-Expert Paradoxes
- Predicate approach to modality
- The Knower Paradox and variants
- Epistemic Arithmetic
- Gödel's Disjunction

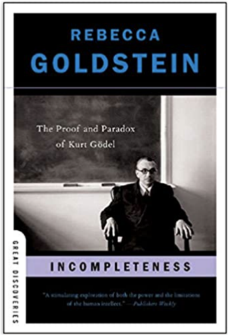
Introduction

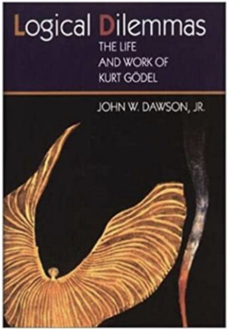


Kurt Gödel (1906 - 1978)

plato.stanford.edu/entries/goedel/


JOURNEY TO THE
EDGE OF REASON
THE LIFE OF KURT GÖDEL
STEPHEN BUDIANSKY


REBECCA
GOLDSTEIN
The Proof and Paradox
of Kurt Gödel
INCOMPLETENESS
GREAT DISCOVERIES


Logical Dilemmas
THE LIFE
AND WORK OF
KURT GÖDEL
JOHN W. DAWSON, JR.

1929 Completeness of First-Order Logic

1931 First and Second Incompleteness Theorems

1933 Translation of classical logic in intuitionistic logic

1936 Speed-up Theorems

1938 Consistency of the Continuum Hypothesis

1949 Work on General Relativity

1958 The “Dialectica interpretation”

1929 Completeness of First-Order Logic

1931 First and Second Incompleteness Theorems

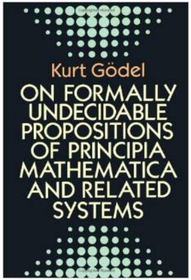
1933 Translation of classical logic in intuitionistic logic

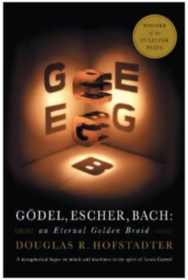
1936 Speed-up Theorems

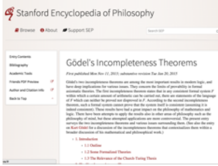
1938 Consistency of the Continuum Hypothesis

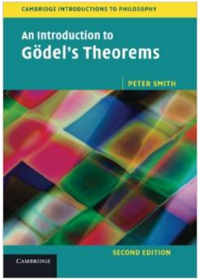
1949 Work on General Relativity

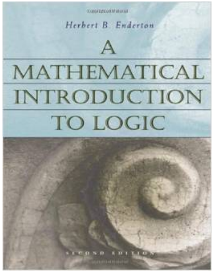
1958 The “Dialectica interpretation”

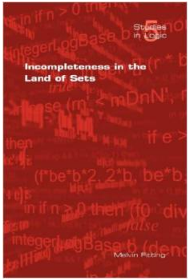

Kurt Gödel
ON FORMALLY UNDECIDABLE PROPOSITIONS OF PRINCIPIA MATHEMATICA AND RELATED SYSTEMS


WINNER of the PULITZER PRIZE
GÖDEL, ESCHER, BACH:
an Eternal Golden Braid
DOUGLAS R. HOFSTADTER
A metaphorical fugue on minds and machines in the spirit of Lewis Carroll

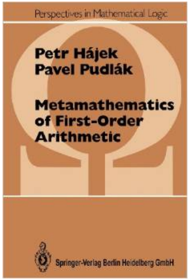

Stanford Encyclopedia of Philosophy
Gödel's Incompleteness Theorems


CAMBRIDGE INTRODUCTIONS TO PHILOSOPHY
An Introduction to
Gödel's Theorems
PETER SMITH
SECOND EDITION
CAMBRIDGE


Herbert B. Enderton
A MATHEMATICAL INTRODUCTION TO LOGIC
SECOND EDITION


Studies in Logic
5
Incompleteness in the Land of Sets
Melvin Fitting


GÖDEL'S INCOMPLETENESS THEOREMS
RAYMOND M. SMULLYAN

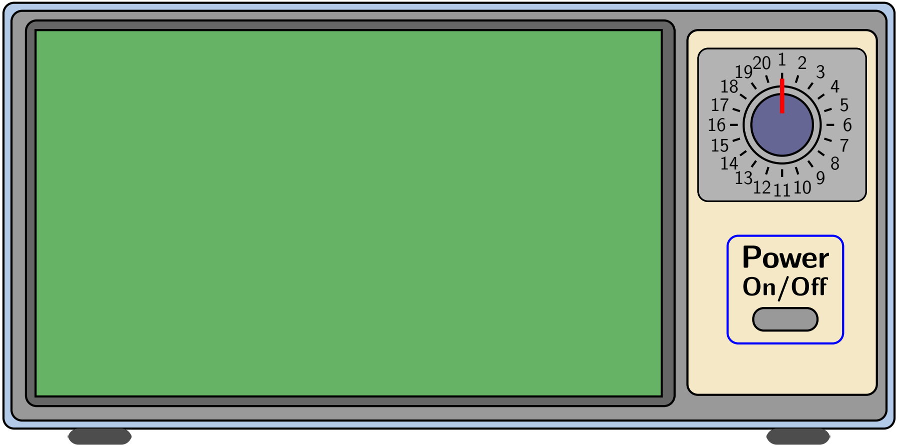

Perspectives in Mathematical Logic
Petr Hájek
Pavel Pudlák
Metamathematics of First-Order Arithmetic
Springer-Verlag Berlin Heidelberg GmbH

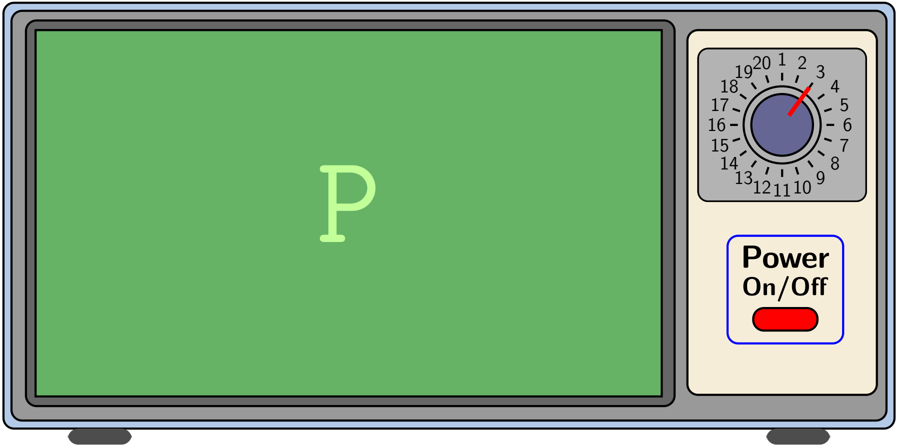
Smullyan's machine

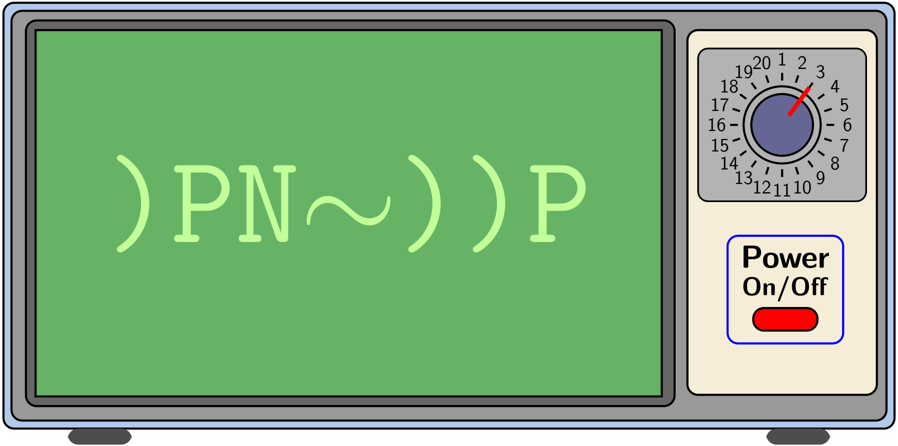
R. Smullyan. *Chapter 1: The General Idea Behind Gödel's Proof, In* Gödel's Incompleteness Theorems. Oxford University Press, 1992.

Consider a machine that displays strings of the following symbols:

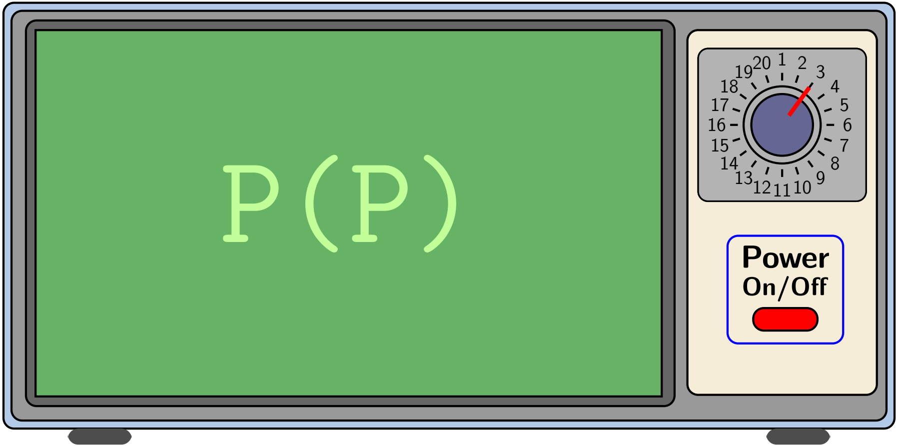
$$\texttt{)} \qquad \texttt{(} \qquad \texttt{P} \qquad \texttt{N} \qquad \sim$$


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Power
On/Off


P
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Power
On/Off



)PN~))P
20 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Power
On/Off


P(P)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Power
On/Off

An **expression** is any finite string of `)`,`(`, `P`,`N` or $\sim$.

An **expression** is any finite string of `)`, `(`, `P`, `N` or $\sim$.

Given an expression X, the **norm** of X is $X(X)$.

An **expression** is any finite string of `)`, `(`, `P`, `N` or $\sim$.

Given an expression X, the **norm** of X is $X(X)$.

Question: What are the norms of $\sim$`P`, `N)P`, `P(P)`, `PN` and $\sim$`PN`?

An **expression** is any finite string of `)`,`(`, `P`,`N` or $\sim$.

Given an expression X, the **norm** of X is $X(X)$.

Question: What are the norms of $\sim$`P`, `N)P`, `P(P)`, `PN` and $\sim$`PN`?

Answer:

1. The norm of $\sim$`P` is $\sim$`P(`$\sim$`P)`
2. The norm of `N)P` is `N)P(N)P)`
3. The norm of `P(P)` is `P(P)(P(P))`

An **expression** is any finite string of `)`, `(`, `P`, `N` or $\sim$.

Given an expression X, the **norm** of X is $X(X)$.

Question: What are the norms of $\sim$`P`, `N)P`, `P(P)`, `PN` and $\sim$`PN`?

Answer:

1. The norm of $\sim$`P` is $\sim$`P(`$\sim$`P)`
2. The norm of `N)P` is `N)P(N)P)`
3. The norm of `P(P)` is `P(P)(P(P))`
4. The norm of `PN` is `PN(PN)`
5. The norm of $\sim$`PN` is $\sim$`PN(`$\sim$`PN)`

A **statement** is any expression of the following form:

`P(X)`

`~P(X)`

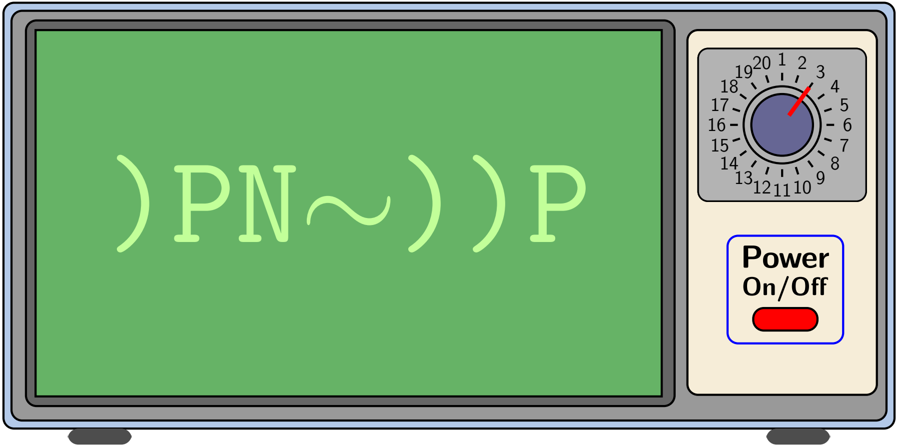
`PN(X)`

`~PN(X)`

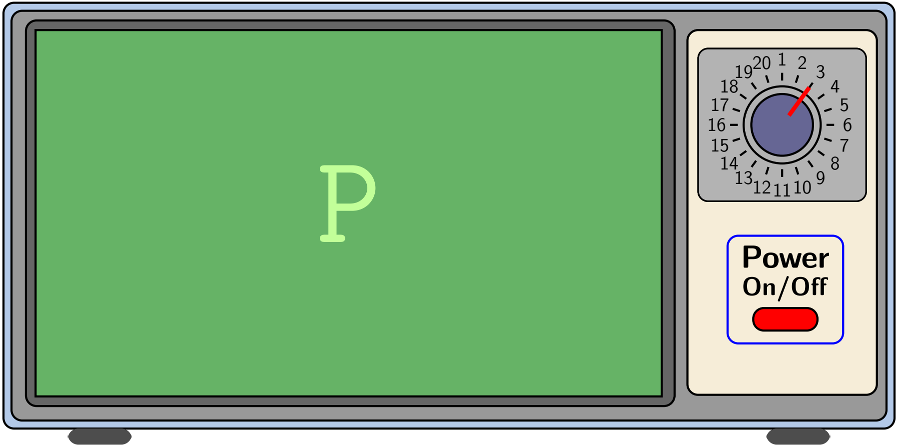
Statement	is true if...
`P(X)`	
~`P(X)`	
`PN(X)`	
~`PN(X)`	

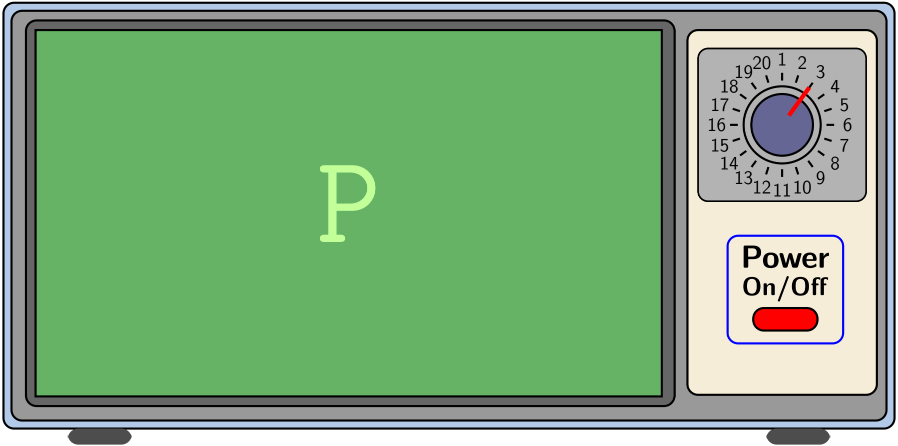
Statement	is true if...
`P(X)`	the expression `X` is printable.
$\sim$`P(X)`	the expression `X` is not printable.
`PN(X)`	
$\sim$`PN(X)`	

Statement	is true if...
`P(X)`	the expression `X` is printable.
$\sim$`P(X)`	the expression `X` is not printable.
`PN(X)`	the norm of `X` is printable.
$\sim$`PN(X)`	the norm of `X` is not printable.

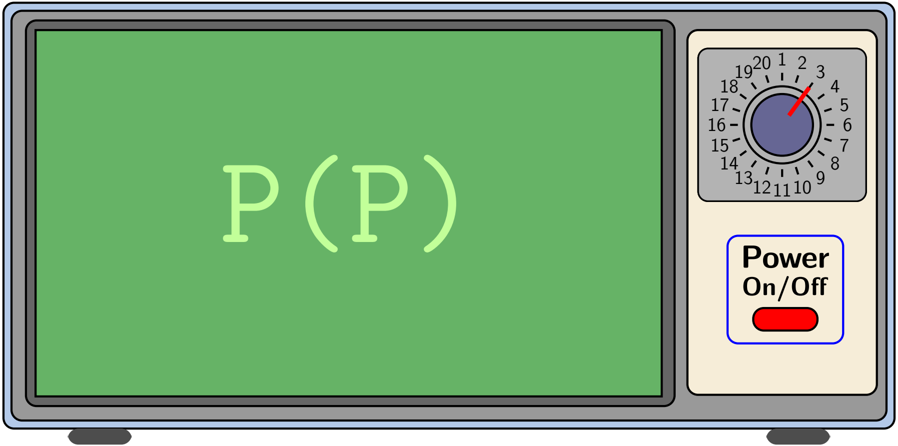


Not a statement, so neither true nor false.

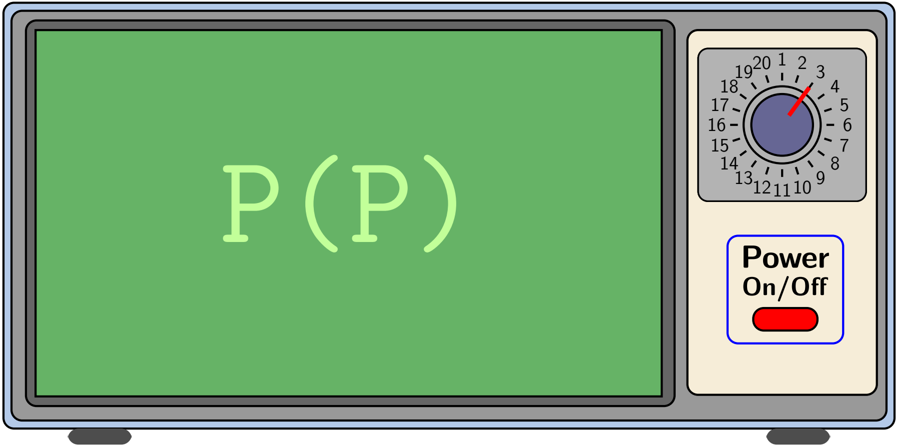

P
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Power
On/Off



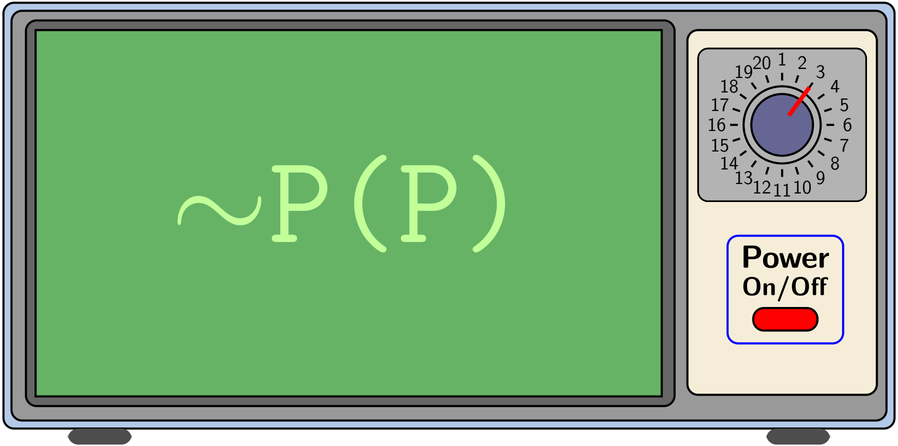
Not a statement, so neither true nor false.



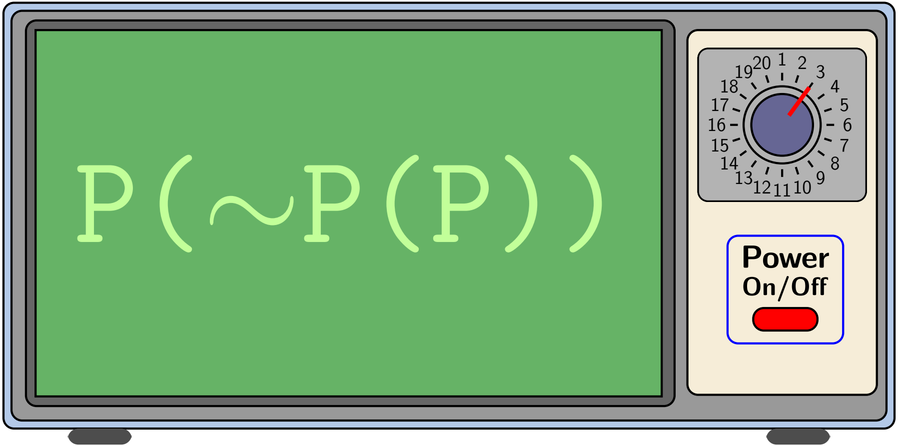
P(P)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Power
On/Off



This is true.



This is false.

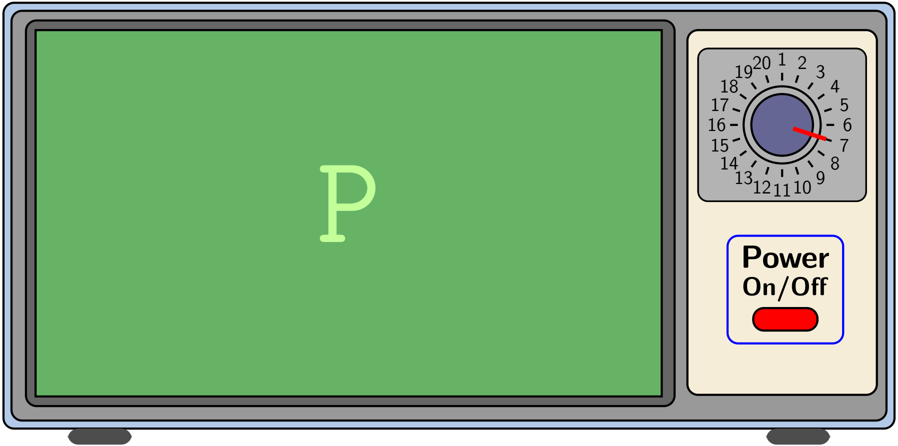


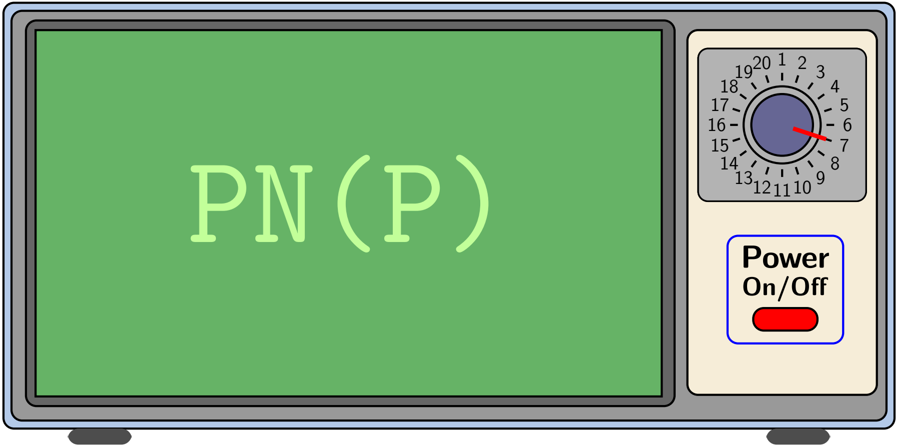
This is true.

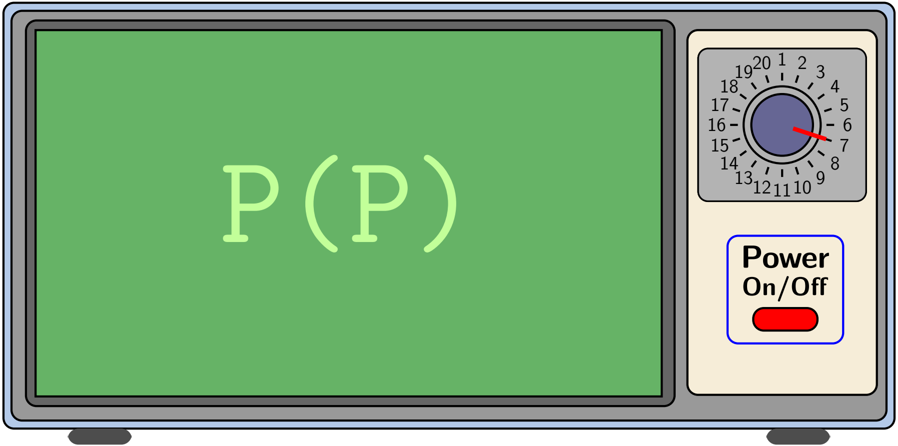
Assumption: The machine only prints true statements
(if the machine prints a statement, then it is true).

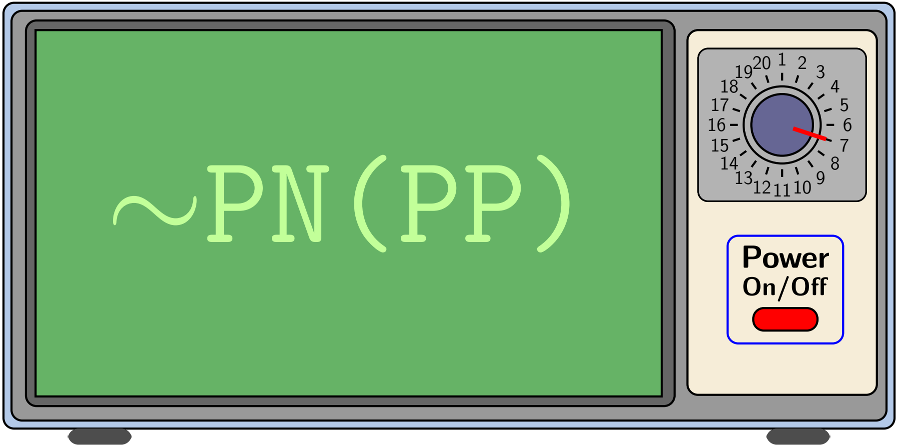
Assumption: The machine only prints true statements
(if the machine prints a statement, then it is true).

Is it possible to construct a machine that print *all* true statements?

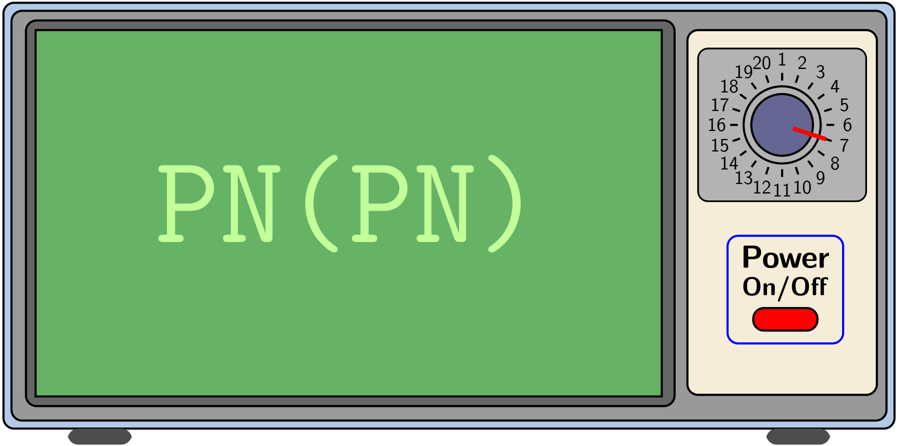

P
20 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Power
On/Off


PN(P)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Power
On/Off


P(P)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Power
On/Off


~PN(PP)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Power
On/Off

The machine is designed so that `PP(PP)` will not be printed


PN(PN)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Power
On/Off

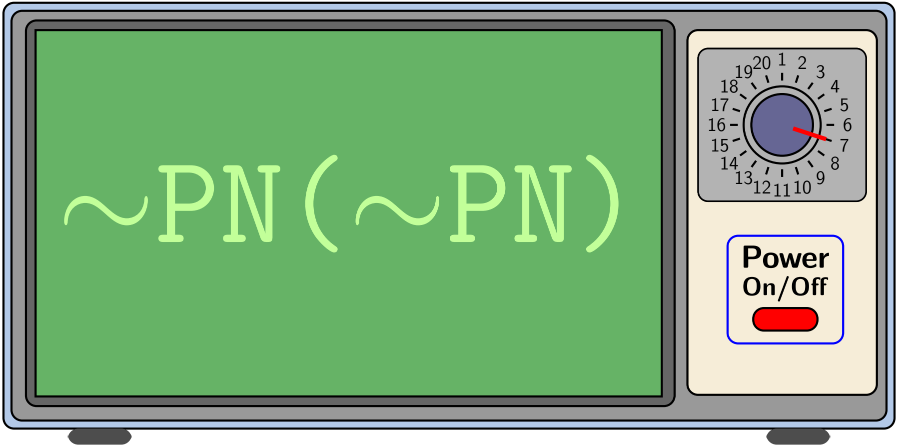
`PN(PN)` is true

if, and only if,

the norm of `PN` is printable

if, and only if,

`PN(PN)` is printable.


~PN(~PN)
20 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Power
On/Off

$\sim$`PN(`$\sim$`PN)` is true

if, and only if,

the norm of $\sim$`PN` is not printable

if, and only if,

$\sim$`PN(`$\sim$`PN)` is not printable.

$\sim$`PN(`$\sim$`PN)` is true if, and only if, $\sim$`PN(`$\sim$`PN)` is not printable.

$\sim$PN($\sim$PN) is true if, and only if, $\sim$PN($\sim$PN) is not printable.

Two possibilities:

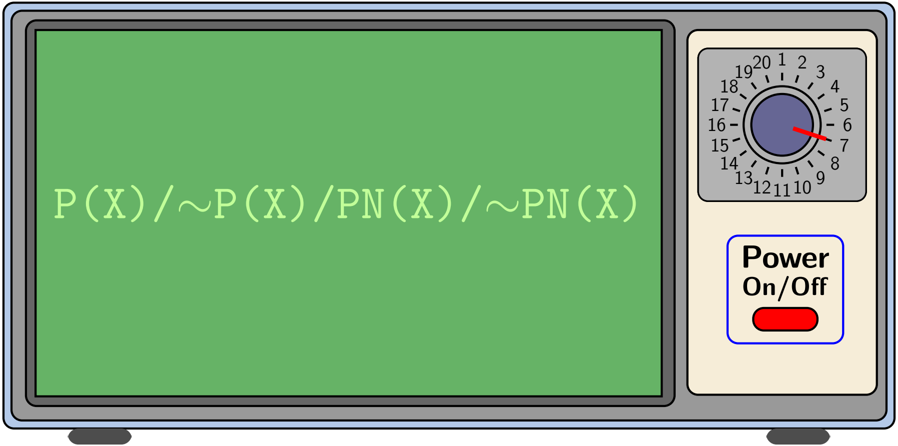
1. The machine is designed to print $\sim$PN($\sim$PN)

2. The machine is designed to not print $\sim$PN($\sim$PN)

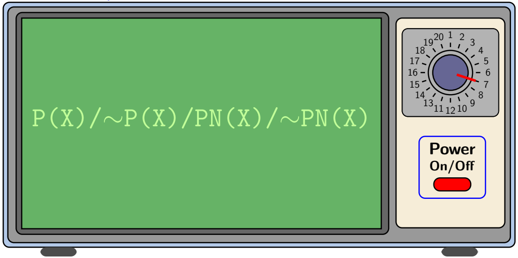
$\sim$`PN(`$\sim$`PN)` is true if, and only if, $\sim$`PN(`$\sim$`PN)` is not printable.

Two possibilities:

1. The machine is designed to print $\sim$`PN(`$\sim$`PN)`: There is a statement that is printable, but not true. (Contradicts the assumption.)
2. The machine is designed to not print $\sim$`PN(`$\sim$`PN)`: There is a statement that is true, but is not printable.

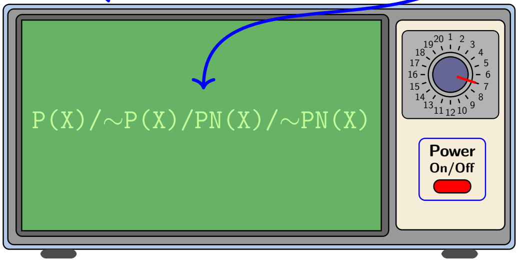

P(X)/∼P(X)/PN(X)/∼PN(X)
20 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Power
On/Off

Formal system producing statements about the natural numbers



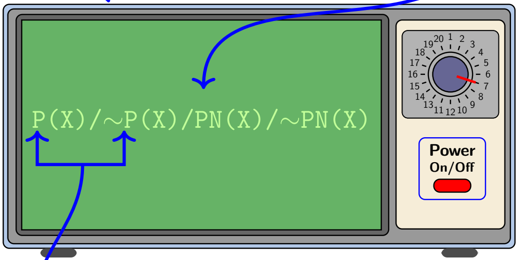
Formal system producing statements about the natural numbers

Statements about the natural numbers



Formal system producing statements about the natural numbers

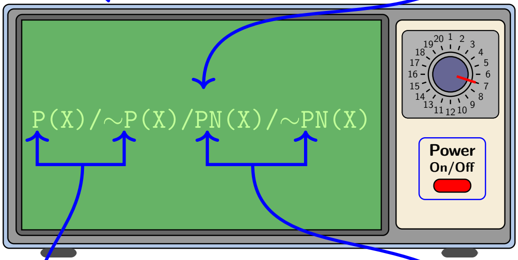
Statements about the natural numbers



It is/is not provable in the formal system that...

Formal system producing statements about the natural numbers

Statements about the natural numbers



It is/is not provable in the formal system that...

The *diagonal* of the expression is/is not provable...

Background

"...It would seem reasonable, therefore, to surmise that these **axioms and rules of inference are sufficient to decide all mathematical questions which can be formulated in the system concerned**.

"...It would seem reasonable, therefore, to surmise that these **axioms and rules of inference are sufficient to decide all mathematical questions which can be formulated in the system concerned**. In what follows it will be shown that this is not the case, but rather that, in both cited systems, there exists relatively simple problems of the theory of ordinary whole numbers which cannot be decided on the basis of the axioms." (Gödel)

K. Gödel. *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I*. Monatshefte für Mathematik und Physik, v. 38 n. 1, pp. 173 - 198, 1931.

Hilbert's Program

Hilbert's Program had two goals:

1. A complete axiomatization of mathematics, one which will settle every question in mathematics.
2. A proof using strictly finitary means to analyze the formal aspects of the above theory that the axiomatization is *reliable* (i.e., consistent).

R. Zach. *Hilbert's Program*. Stanford Encyclopedia of Philosophy, 2019, https://plato.stanford.edu/entries/hilbert-program/.

- $0 \neq 1$
- $2 \times (1 + 4) = 5 + 3 \times 1 + 2 \times 1 + 0$
- $3 \times 2 = 2 \times 3$

- for all n, if $n \neq 0$, then there is a m such that $m + 1 = n$

- for all n, m, $n \times m = m \times n$

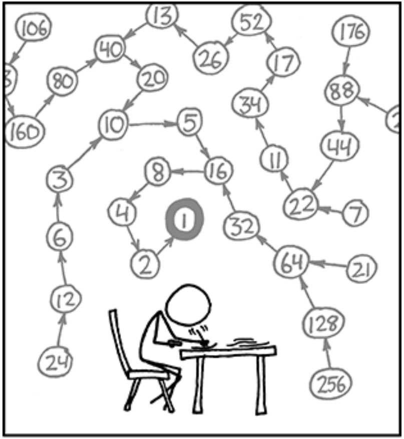
- There is no number smaller than 0

- There is no biggest prime number
- there are no a, b, c such that $a^n + b^n = c^n$ for $n > 2$

- every even number is the sum of two prime numbers

- there are infinitely many primes that differ by 2

- for every number n there is a sequence of numbers $k_0, k_1, \ldots, k_m$ such that $k_0 = n$, for each $0 < i \leq m$, $k_m = k_{m-1}/2$ if k_{m-1} is even and $k_m = 3k_{m-1} + 1$ if k_{m-1} is odd, and $k_m = 1$



THE COLLATZ CONJECTURE STATES THAT IF YOU PICK A NUMBER, AND IF IT'S EVEN DIVIDE IT BY TWO AND IF IT'S ODD MULTIPLY IT BY THREE AND ADD ONE, AND YOU REPEAT THIS PROCEDURE LONG ENOUGH, EVENTUALLY YOUR FRIENDS WILL STOP CALLING TO SEE IF YOU WANT TO HANG OUT.

Language of Arithmetic $\mathcal{L}_A$

Each of these statements can be expressed in the **language of arithmetic**.

Language of Arithmetic $\mathcal{L}_A$

Each of these statements can be expressed in the **language of arithmetic**.

Terms	$0 \mid x \mid \mathsf{S}(x) \mid (x + y) \mid x \times y$
Formulas of $\mathcal{L}_A$	$(t = s) \mid (t < s) \mid \neg\varphi \mid (\varphi \wedge \psi) \mid (\forall x)\varphi$

Language of Arithmetic $\mathcal{L}_A$

Each of these statements can be expressed in the **language of arithmetic**.

Terms	$0 \mid x \mid \mathsf{S}(x) \mid (x + y) \mid x \times y$
Formulas of $\mathcal{L}_A$	$(t = s) \mid (t < s) \mid \neg\varphi \mid (\varphi \wedge \psi) \mid (\forall x)\varphi$

If we could specify some axioms and inference rules that pin down the number sequence and characterize S, $+$ and $\times$, then we should be able to *decide* any statement about the natural numbers.

The Standard Model

$$\mathcal{N} = (\mathbb{N}, 0, \mathsf{S}, +, *, <)$$

- $0^{\mathcal{N}} = 0$
- $\mathsf{S}^{\mathcal{N}} : \mathbb{N} \to \mathbb{N}$ is the successor function: for all $n \in \mathbb{N}$, $\mathsf{S}^{\mathcal{N}}(n) = n + 1$
- $+^{\mathcal{N}} : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is addition: for all $n, m \in \mathbb{N}$, $+^{\mathcal{N}}(n, m) = n + m$
- $\times^{\mathcal{N}} : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is multiplication: for all $n, m \in \mathbb{N}$, $\times^{\mathcal{N}}(n, m) = n * m$
- $<^{\mathcal{N}} \subseteq \mathbb{N} \times \mathbb{N}$ is less-than: for all $n, m \in \mathbb{N}$, $(n, m) \in <^{\mathcal{N}}(n, m)$ provided that $n < m$.

Numerals

For each $n \in \mathbb{N}$ we write $\overline{n}$ for the term representing n:

$$\overline{n} \text{ is } \underbrace{\mathsf{S}(\cdots(\mathsf{S}}_{n \text{ times}}(0))\cdots)$$

For instance, $\overline{3}$ is S(S(S(0)))

To simplify the notation, we often drop the parentheses in the terms $\overline{n}$. For instance, we write SSS(0) instead of S(S(S(0))).

Robinson's **Q**

$S1.$ $\forall x(0 \neq \mathsf{S}(x))$

$S2.$ $\forall x \forall y(\mathsf{S}(x) = \mathsf{S}(y) \to x = y)$

$S3.$ $\forall x(x \neq 0 \to \exists y(x = \mathsf{S}(y)))$

Robinson's **Q**

$S1.$ $\forall x(0 \neq \mathsf{S}(x))$

$S2.$ $\forall x \forall y(\mathsf{S}(x) = \mathsf{S}(y) \rightarrow x = y)$

$S3.$ $\forall x(x \neq 0 \rightarrow \exists y(x = \mathsf{S}(y)))$

$A1.$ $\forall x(x + 0 = x)$

$A2.$ $\forall x \forall y(x + \mathsf{S}(y) = \mathsf{S}(x + y))$

Robinson's $\mathbf{Q}$

$$
\begin{array}{ll}
S1. & \forall x(0 \neq \mathsf{S}(x)) \\
S2. & \forall x \forall y(\mathsf{S}(x) = \mathsf{S}(y) \to x = y) \\
S3. & \forall x(x \neq 0 \to \exists y(x = \mathsf{S}(y))) \\
A1. & \forall x(x + 0 = x) \\
A2. & \forall x \forall y(x + \mathsf{S}(y) = \mathsf{S}(x + y)) \\
M1. & \forall x(x \times 0 = 0) \\
M2. & \forall x \forall y(x \times \mathsf{S}(y) = x \times y + x)
\end{array}
$$

We write $\mathbf{Q} \vdash A$ when there is a derivation of A in which the only open assumptions are the axioms of $\mathbf{Q}$.

Defining $<$

$$x < y \leftrightarrow \exists z(x + \mathsf{S}(z) = y)$$

Exercises

- $\mathbf{Q} \vdash \overline{1} \neq \overline{2}$
- $\mathbf{Q} \vdash \overline{1} + \overline{1} = \overline{2}$
- $\mathbf{Q} \vdash 0 + \overline{3} = \overline{3} + 0$

Exercises

- $\mathbf{Q} \vdash \overline{1} \neq \overline{2}$
- $\mathbf{Q} \vdash \overline{1} + \overline{1} = \overline{2}$
- $\mathbf{Q} \vdash 0 + \overline{3} = \overline{3} + 0$

- For all closed terms s, t,
  - if $\mathcal{N} \models s = t$, then $\mathbf{Q} \vdash s = t$
  - if $\mathcal{N} \models s \neq t$, then $\mathbf{Q} \vdash s \neq t$

Exercises

- $\mathbf{Q} \vdash \overline{1} \neq \overline{2}$
- $\mathbf{Q} \vdash \overline{1} + \overline{1} = \overline{2}$
- $\mathbf{Q} \vdash 0 + \overline{3} = \overline{3} + 0$

- For all closed terms s, t,
  - if $\mathcal{N} \models s = t$, then $\mathbf{Q} \vdash s = t$
  - if $\mathcal{N} \models s \neq t$, then $\mathbf{Q} \vdash s \neq t$

- $\mathbf{Q} \vdash \forall x(x + 0 = x)$

Exercises

- $\mathbf{Q} \vdash \overline{1} \neq \overline{2}$
- $\mathbf{Q} \vdash \overline{1} + \overline{1} = \overline{2}$
- $\mathbf{Q} \vdash 0 + \overline{3} = \overline{3} + 0$

- For all closed terms s, t,
  - if $\mathcal{N} \models s = t$, then $\mathbf{Q} \vdash s = t$
  - if $\mathcal{N} \models s \neq t$, then $\mathbf{Q} \vdash s \neq t$

- $\mathbf{Q} \vdash \forall x(x + 0 = x)$

- For all $n \in \mathbb{N}$, $\mathbf{Q} \vdash 0 + \overline{n} = \overline{n}$

Exercises

- $\mathbf{Q} \vdash \overline{1} \neq \overline{2}$
- $\mathbf{Q} \vdash \overline{1} + \overline{1} = \overline{2}$
- $\mathbf{Q} \vdash 0 + \overline{3} = \overline{3} + 0$

- For all closed terms s, t,
  - if $\mathcal{N} \models s = t$, then $\mathbf{Q} \vdash s = t$
  - if $\mathcal{N} \models s \neq t$, then $\mathbf{Q} \vdash s \neq t$

- $\mathbf{Q} \vdash \forall x(x + 0 = x)$

- For all $n \in \mathbb{N}$, $\mathbf{Q} \vdash 0 + \overline{n} = \overline{n}$

- $\mathbf{Q} \nvdash \forall x(0 + x = x)$

Peano Arithmetic (**PA**)

The axioms of **PA** (Peano Arithmetic) are all the axioms of **Q** with every instance of the following axiom schema:

Induction Scheme: For all formulas φ of $\mathcal{L}_A$,

$$(\varphi(0) \land \forall x(\varphi(x) \to \varphi(S(x)))) \to \forall x\, \varphi(x)$$

We write **PA** $\vdash \varphi$ when there is a derivation of φ in which the only open assumptions are the axioms of **PA**.

Exercises, continued

- $\mathbf{PA} \vdash \forall x(0 + x = x)$

The Theory of True Arithmetic

True arithmetic: $Th(\mathcal{N}) = \{\varphi \mid \mathcal{N} \models \varphi\}$

1. Is there a computational procedure we can use to test if a sentence is in $Th(\mathcal{N})$?

The Theory of True Arithmetic

True arithmetic: $Th(\mathcal{N}) = \{\varphi \mid \mathcal{N} \models \varphi\}$

1. Is there a computational procedure we can use to test if a sentence is in $Th(\mathcal{N})$?

A **theory** is a set of sentences that is closed under entailment, i.e., $\mathbf{T}$ is a theory if $\mathbf{T} = \{\varphi \mid \mathbf{T} \models \varphi\}$

A theory is **axiomatizable** if there is a *decidable* set of sentences $\mathbf{T}_0$ such that $\mathbf{T} = \{\varphi \mid \mathbf{T}_0 \models \varphi\}$

The Theory of True Arithmetic

True arithmetic: $Th(\mathcal{N}) = \{\varphi \mid \mathcal{N} \models \varphi\}$

1. Is there a computational procedure we can use to test if a sentence is in $Th(\mathcal{N})$?

A **theory** is a set of sentences that is closed under entailment, i.e., $\mathbf{T}$ is a theory if $\mathbf{T} = \{\varphi \mid \mathbf{T} \models \varphi\}$

A theory is **axiomatizable** if there is a *decidable* set of sentences $\mathbf{T}_0$ such that $\mathbf{T} = \{\varphi \mid \mathbf{T}_0 \models \varphi\}$

2. Is there an axiomatizable theory $\mathbf{T}$ such that $\mathbf{T} = Th(\mathcal{N})$? This is equivalent to asking whether $\mathbf{T}$ is **complete**: For every sentence φ, either $\mathbf{T} \models \varphi$ or $\mathbf{T} \models \neg\varphi$.

The answer to both questions is no.

The answer to both questions is no.

Gödel's first incompleteness theorem (informal statement):
Any consistent formal theory within which a certain amount of elementary arithmetic can be carried out is **incomplete**.

Arithmetic Hierarchy

- A quantifier is **bounded** if it is the form '$\forall x \leq t$' or '$\exists x \leq t$', where t is a term not involving x.

- A formula is a **bounded formula** (denoted Δ_0^0) if all of its quantifiers are bounded.

Arithmetic Hierarchy

- A quantifier is **bounded** if it is the form '$\forall x \leq t$' or '$\exists x \leq t$', where t is a term not involving x.
- A formula is a **bounded formula** (denoted Δ_0^0) if all of its quantifiers are bounded.
- For $n \geq 0$, the classes of formulas Σ_n^0 and Π_n^0 are defined as follows:
  - $\Sigma_0^0 = \Pi_0^0 = \Delta_0^0$.
  - Σ_{n+1}^0 is the set of formulas of the form $\exists \vec{x} \varphi$ where φ is a Π_n^0 formula and $\vec{x}$ is a (possibly empty) list of variables.
  - Π_{n+1}^0 is the set of formulas of the form $\forall \vec{x} \varphi$ where φ is a Σ_n^0 formula and $\vec{x}$ is a (possibly empty) list of variables.

Definition

Σ^0_1-sound A theory **T** is Σ^0_1-**sound** iff for every Σ^0_1-formula φ, if $\mathbf{T} \vdash \varphi$, then φ is true (in the standard model).

Definition

Σ^0_1-complete A theory **T** is Σ^0_1-**complete** iff for every Σ^0_1-formula φ, if φ is true (in the standard model), then $\mathbf{T} \vdash \varphi$.

Proposition

PA (in fact, even **Q**) is Σ^0_1-complete.

Theorem (Gödel's First Incompleteness Theorem)

Assume that **PA** is Σ^0_1-sound. Then there is a Π^0_1-sentence φ such that **PA** neither proves φ nor $\neg\varphi$.

Theorem (Gödel's Second Incompleteness Theorem)

Assume that **PA** is consistent. Then **PA** cannot prove $\mathrm{Con}_{\mathbf{PA}}$.

$\mathrm{Con}_{\mathbf{PA}}$ *is a* Π^0_1*-statement that informally asserts "for all* x*,* x *does not code a proof of a contradiction from the axioms of* **PA***"*

- Gödel numbering and naming systems
- Gödel-Carnap Fixed Point Theorem
- Representing functions/relations
- Provability predicate
- Löb's Theorem

Gödel Numbering

Gödel-numbering assigns numbers to the syntactic objects of a logic (i.e., the terms, the formulas, and the derivations).

Suppose that χ is a syntactic object (i.e., a term, formula or a derivation). We use the following notation:

> $gn(\chi)$: The Gödel number of χ (an integer)
>
> $\ulcorner \chi \urcorner$: The numeral of the Gödel number of χ (a numeral). That is:

$$\ulcorner \chi \urcorner \equiv \overline{gn(\chi)}$$

Fixed-Point Theorem

Theorem (Gödel-Carnap Fixed-Point Theorem)

Let $A(x)$ be any formula of $\mathcal{L}_A$ with one free variable x. Then there is a sentence B such that

$$\mathbf{Q} \vdash B \leftrightarrow A(\ulcorner B \urcorner).$$

Substitution

Suppose that A is a formula where x is a free variable. We write $A(x)$ to when the formula A has at most one free variable x.

If t is a term, then $A(x)[x/t]$ is A with every instance of x replaced with t. We sometimes abuse notation and write $A(t)$ instead of $A(x)[x/t]$.

Substitution

Suppose that A is a formula where x is a free variable. We write $A(x)$ to when the formula A has at most one free variable x.

If t is a term, then $A(x)[x/t]$ is A with every instance of x replaced with t. We sometimes abuse notation and write $A(t)$ instead of $A(x)[x/t]$.

Let $Sub : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be a function, where for each $n, m \in \mathbb{N}$, $Sub(n, m)$ is the code of $\alpha(x)[x/\overline{m}]$ where n is the code of $\alpha(x)$. So, for any formula $\alpha(x)$ and $m \in \mathbb{N}$:

$$Sub(gn(\alpha(x)), m) = gn(\alpha(\overline{m}))$$

We sketch a proof under the assumption that sub is a function symbol in the language $\mathcal{L}_A$ and the theory $\mathbf{Q}$ "represents" *Sub* in the following sense:

For any formula $A(x)$ and $n \in \mathbb{N}$,

$$\mathbf{Q} \vdash \mathsf{sub}(\ulcorner A(x) \urcorner, \overline{n}) = \ulcorner A(\overline{n}) \urcorner$$

- Let $A^*(x)$ be $A(\mathrm{sub}(x, x))$
  Let B be $A^*(\ulcorner A^*(x) \urcorner)$

- $A^*(\ulcorner A^*(x) \urcorner)$ is the formula $A(\mathrm{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner))$

- Let $A^*(x)$ be $A(\mathrm{sub}(x,x))$
  Let B be $A^*(\ulcorner A^*(x) \urcorner)$

- $A^*(\ulcorner A^*(x) \urcorner)$ is the formula $A(\mathrm{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner))$

$$\mathbf{Q} \vdash \qquad \mathrm{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner) \qquad = \qquad \ulcorner A^*(\ulcorner A^*(x) \urcorner) \urcorner$$

- Let $A^*(x)$ be $A(\mathrm{sub}(x, x))$
  Let B be $A^*(\ulcorner A^*(x) \urcorner)$

- $A^*(\ulcorner A^*(x) \urcorner)$ is the formula $A(\mathrm{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner))$

$$\mathbf{Q} \vdash \quad \mathrm{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner) \quad = \quad \ulcorner A^*(\ulcorner A^*(x) \urcorner) \urcorner$$

$$\mathbf{Q} \vdash \quad A(\mathrm{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner)) \quad \leftrightarrow \quad A(\ulcorner A^*(\ulcorner A^*(x) \urcorner) \urcorner)$$

- Let $A^*(x)$ be $A(\text{sub}(x, x))$
  Let B be $A^*(\ulcorner A^*(x) \urcorner)$

- $A^*(\ulcorner A^*(x) \urcorner)$ is the formula $A(\text{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner))$

$$\begin{aligned}
\mathbf{Q} &\vdash \quad \text{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner) &&= \quad \ulcorner A^*(\ulcorner A^*(x) \urcorner) \urcorner \\
\mathbf{Q} &\vdash \quad \text{A}(\text{sub}(\ulcorner A^*(x) \urcorner, \ulcorner A^*(x) \urcorner)) &&\leftrightarrow \quad A(\ulcorner A^*(\ulcorner A^*(x) \urcorner) \urcorner) \\
\mathbf{Q} &\vdash \quad B &&\leftrightarrow \quad A(\ulcorner B \urcorner)
\end{aligned}$$

From Cantor to Gödel...

H. Gaifman (2006). *Naming and Diagonalization, From Cantor to Gödel to Kleene*. Logic Journal of the IGPL, pp. 709 - 728.

What's in a name?

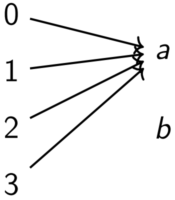
Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3

What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

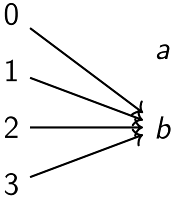
	0	1	2	3
	a	a	a	a



What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

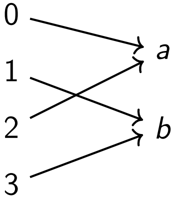
	0	1	2	3
	a	a	a	a
	b	b	b	b



What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

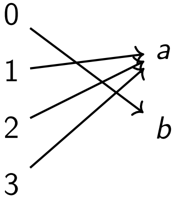
	0	1	2	3
	a	a	a	a
	b	b	b	b
	a	b	a	b



What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

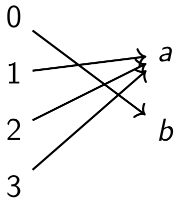
	0	1	2	3
	a	a	a	a
	b	b	b	b
	a	b	a	b
	b	a	a	a



What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
α	a	a	a	a
β	b	b	b	b
γ	a	b	a	b
δ	b	a	a	a



What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
α	a	a	a	a
β	b	b	b	b
γ	a	b	a	b
δ	b	a	a	a

$$g(n) = \begin{cases} b & \text{if } \gamma(n) = a \\ a & \text{if } \gamma(n) = b \end{cases}$$

0

1

2

3

a

b

What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
α	a	a	a	a
β	b	b	b	b
γ	a	b	a	b
δ	b	a	a	a

$$g(n) = \begin{cases} b & \text{if } \gamma(n) = a \\ a & \text{if } \gamma(n) = b \end{cases}$$

0

1

2

3

a

b

What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
α	a	a	a	a
β	b	b	b	b
γ	a	b	a	b
δ	b	a	a	a

$$g(n) = \begin{cases} b & \text{if } \gamma(n) = a \\ a & \text{if } \gamma(n) = b \end{cases}$$

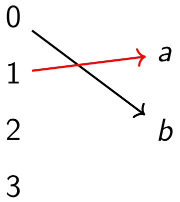
$$\begin{tikzcd} 0 \arrow[ddr] & \\ 1 & a \\ 2 & b \\ 3 & \end{tikzcd}$$

What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
α	a	a	a	a
β	b	b	b	b
γ	a	b	a	b
δ	b	a	a	a

$$g(n) = \begin{cases} b & \text{if } \gamma(n) = a \\ a & \text{if } \gamma(n) = b \end{cases}$$

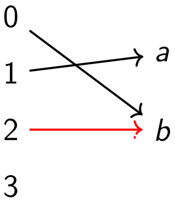


What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
α	a	a	a	a
β	b	b	b	b
γ	a	b	a	b
δ	b	a	a	a

$$g(n) = \begin{cases} b & \text{if } \gamma(n) = a \\ a & \text{if } \gamma(n) = b \end{cases}$$

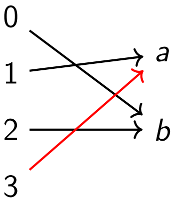


What's in a name?

Functions from $\{0,1,2,3\}$ to $\{a,b\}$

	0	1	2	3
α	a	a	a	a
β	b	b	b	b
γ	a	b	a	b
δ	b	a	a	a

$$g(n) = \begin{cases} b & \text{if } \gamma(n) = a \\ a & \text{if } \gamma(n) = b \end{cases}$$



What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
[0]	a	a	a	a
[1]	b	b	b	b
[2]	a	b	a	b
[3]	b	a	a	a

0	
1	a
2	b
3	

What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
[0]	a	a	a	a
[1]	b	b	b	b
[2]	a	b	a	b
[3]	b	a	a	a

$$diag(n) = \begin{cases} b & \text{if } [n](n) = a \\ a & \text{if } [n](n) = b \end{cases}$$

0

1 $\qquad a$

2 $\qquad b$

3

What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
[0]	a	a	a	a
[1]	b	b	b	b
[2]	a	b	a	b
[3]	b	a	a	a

$$diag(n) = \begin{cases} b & \text{if } [n](n) = a \\ a & \text{if } [n](n) = b \end{cases}$$

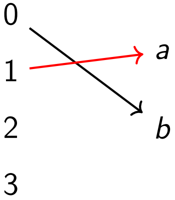
$$\begin{tikzcd} 0 \arrow[ddr] & \\ 1 & a \\ 2 & b \\ 3 & \end{tikzcd}$$

What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
[0]	a	a	a	a
[1]	b	b	b	b
[2]	a	b	a	b
[3]	b	a	a	a

$$diag(n) = \begin{cases} b & \text{if } [n](n) = a \\ a & \text{if } [n](n) = b \end{cases}$$

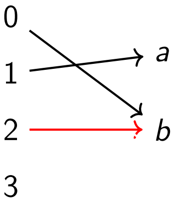


What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
[0]	a	a	a	a
[1]	b	b	b	b
[2]	a	b	a	b
[3]	b	a	a	a

$$diag(n) = \begin{cases} b & \text{if } [n](n) = a \\ a & \text{if } [n](n) = b \end{cases}$$

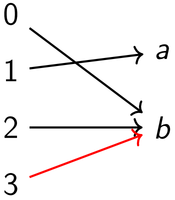


What's in a name?

Functions from $\{0, 1, 2, 3\}$ to $\{a, b\}$

	0	1	2	3
[0]	a	a	a	a
[1]	b	b	b	b
[2]	a	b	a	b
[3]	b	a	a	a

$$diag(n) = \begin{cases} b & \text{if } [n](n) = a \\ a & \text{if } [n](n) = b \end{cases}$$



Cantor's Diagonalization Proof

Functions from $\mathbb{N}$ to $\{0, 1\}$

	0	1	2	3	$\cdots$	n	$\cdots$
	0	0	0	0	$\cdots$	0	$\cdots$
	0	1	0	1	$\cdots$	1	$\cdots$
	0	1	1	0	$\cdots$	0	$\cdots$
	1	0	1	0	$\cdots$	1	$\cdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	0	0	1	0	$\cdots$	1	$\cdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Cantor's Diagonalization Proof

Functions from $\mathbb{N}$ to $\{0, 1\}$

	0	1	2	3	$\cdots$	n	$\cdots$
[0]	$\boxed{0}$	0	0	0	$\cdots$	0	$\cdots$
[1]	0	$\boxed{1}$	0	1	$\cdots$	1	$\cdots$
[2]	0	1	$\boxed{1}$	0	$\cdots$	0	$\cdots$
[3]	1	0	1	$\boxed{0}$	$\cdots$	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[n]$	0	0	1	0	$\cdots$	$\boxed{1}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$d : \mathbb{N} \to \{0, 1\}$ $\qquad$ $d(n) = 1 - [n](n)$

Cantor's Diagonalization Proof

Functions from $\mathbb{N}$ to $\{0,1\}$

$$
\begin{array}{c|cccccccc}
 & 0 & 1 & 2 & 3 & \cdots & n & \cdots \\
\hline
[0] & \boxed{0} & 0 & 0 & 0 & \cdots & 0 & \cdots \\
[1] & 0 & \boxed{1} & 0 & 1 & \cdots & 1 & \cdots \\
[2] & 0 & 1 & \boxed{1} & 0 & \cdots & 0 & \cdots \\
[3] & 1 & 0 & 1 & \boxed{0} & \cdots & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
[n] & 0 & 0 & 1 & 0 & \cdots & \boxed{1} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}
$$

$d : \mathbb{N} \to \{0,1\} \qquad d(n) = 1 - [n](n)$

Then, $d \neq [n]$ for any $n \in \mathbb{N}$.

Cantor's original statement is phrased as a non-existence claim: there is no function mapping all the members of a set S onto the set of all $0, 1$-valued functions over S. But the proof establishes a positive result: given any way of correlating functions with members of S, one can construct a function not correlated with any member of S.

(Gaiffman, p. 711)

Richard's Paradox (1905)

Consider all the definitions (in English) of real numbers.

Richard’s Paradox (1905)

Consider all the definitions (in English) of real numbers. Since any such definition is a finite sequence of letters, the definitions can be listed in order.

Richard's Paradox (1905)

Consider all the definitions (in English) of real numbers. Since any such definition is a finite sequence of letters, the definitions can be listed in order.

Let u_i be the real number defined by the ith definition and $f_i(n)$ be the nth member of the decimal expansion of u_i.

Richard's Paradox (1905)

Consider all the definitions (in English) of real numbers. Since any such definition is a finite sequence of letters, the definitions can be listed in order.

Let u_i be the real number defined by the ith definition and $f_i(n)$ be the nth member of the decimal expansion of u_i.

Let u^* be the number who's decimal expansion is $0.g(1)g(2)\cdots g(n)\cdots$ where g is defined by $g(n) = f_n(n) + 1$ if $f_n(n) < 8$, $g(n) = 1$ otherwise.

Richard's Paradox (1905)

Consider all the definitions (in English) of real numbers. Since any such definition is a finite sequence of letters, the definitions can be listed in order.

Let u_i be the real number defined by the ith definition and $f_i(n)$ be the nth member of the decimal expansion of u_i.

Let u^* be the number who's decimal expansion is $0.g(1)g(2)\cdots g(n)\cdots$ where g is defined by $g(n) = f_n(n) + 1$ if $f_n(n) < 8$, $g(n) = 1$ otherwise.

But the previous description defines a number, so $u^* = u_i$ for some i. But, this is impossible.

Richard's Paradox (1905)

1. Let A be the set of all positive integers that can be defined in under 100 words. Since there are only finitely many of these, there must be a smallest positive integer n that does not belong to A.

Richard's Paradox (1905)

1. Let A be the set of all positive integers that can be defined in under 100 words. Since there are only finitely many of these, there must be a smallest positive integer n that does not belong to A.

   But haven't I just defined n in under 100 words?

Richard's Paradox (1905)

1. Let A be the set of all positive integers that can be defined in under 100 words. Since there are only finitely many of these, there must be a smallest positive integer n that does not belong to A.

   But haven't I just defined n in under 100 words?

2. Let B be the set of all reasonably interesting positive integers. Let n be the smallest integer not belonging to B.

Richard's Paradox (1905)

1. Let A be the set of all positive integers that can be defined in under 100 words. Since there are only finitely many of these, there must be a smallest positive integer n that does not belong to A.

   But haven't I just defined n in under 100 words?

2. Let B be the set of all reasonably interesting positive integers. Let n be the smallest integer not belonging to B.

   But surely this defining property of n makes it reasonably interesting.

Let f be a function that associates each number $x \in \mathbb{N}$ with a subset of $\mathbb{N}$, i.e., for all $x \in \mathbb{N}$, $f(x) \subseteq \mathbb{N}$.

Let f be a function that associates each number $x \in \mathbb{N}$ with a subset of $\mathbb{N}$, i.e., for all $x \in \mathbb{N}$, $f(x) \subseteq \mathbb{N}$.

Define S^* by:

$$x \in S^* \Leftrightarrow x \notin f(x)$$

Let f be a function that associates each number $x \in \mathbb{N}$ with a subset of $\mathbb{N}$, i.e., for all $x \in \mathbb{N}$, $f(x) \subseteq \mathbb{N}$.

Define S^* by:

$$x \in S^* \Leftrightarrow x \notin f(x)$$

The assumption that there is some z such that $f(z) = S^*$ leads to a contradiction.

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$f(0)$	0	0	0	0	$\cdots$	0	$\cdots$	
$f(1)$	0	1	0	1	$\cdots$	1	$\cdots$	
$f(2)$	0	1	1	0	$\cdots$	0	$\cdots$	
$f(3)$	1	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$f(n)$	0	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$f(0)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$f(1)$	0	1	0	1	$\cdots$	1	$\cdots$	
$f(2)$	0	1	1	0	$\cdots$	0	$\cdots$	
$f(3)$	1	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$f(n)$	0	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$f(0)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$f(1)$	0	1	0	1	$\cdots$	1	$\cdots$	$\{1, 3, \ldots, n, \ldots\}$
$f(2)$	0	1	1	0	$\cdots$	0	$\cdots$	
$f(3)$	1	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$f(n)$	0	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$f(0)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$f(1)$	0	1	0	1	$\cdots$	1	$\cdots$	$\{1, 3, \ldots, n, \ldots\}$
$f(2)$	0	1	1	0	$\cdots$	0	$\cdots$	$\{1, 2\}$
$f(3)$	1	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$f(n)$	0	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$f(0)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$f(1)$	0	1	0	1	$\cdots$	1	$\cdots$	$\{1, 3, \ldots, n, \ldots\}$
$f(2)$	0	1	1	0	$\cdots$	0	$\cdots$	$\{1, 2\}$
$f(3)$	1	0	1	0	$\cdots$	1	$\cdots$	$\{0, 2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$f(n)$	0	0	1	0	$\cdots$	1	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$f(0)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$f(1)$	0	1	0	1	$\cdots$	1	$\cdots$	$\{1, 3, \ldots, n, \ldots\}$
$f(2)$	0	1	1	0	$\cdots$	0	$\cdots$	$\{1, 2\}$
$f(3)$	1	0	1	0	$\cdots$	1	$\cdots$	$\{0, 2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$f(n)$	0	0	1	0	$\cdots$	1	$\cdots$	$\{2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$n \in S^* \text{ iff } n \notin f(n)$$

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$\varphi_0(x)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$\varphi_1(x)$	0	1	0	1	$\cdots$	1	$\cdots$	$\{1, 3, \ldots, n, \ldots\}$
$\varphi_2(x)$	0	1	1	0	$\cdots$	0	$\cdots$	$\{1, 2\}$
$\varphi_3(x)$	1	0	1	0	$\cdots$	1	$\cdots$	$\{0, 2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\varphi_n(x)$	0	0	1	0	$\cdots$	1	$\cdots$	$\{2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$n \in S^* \text{ iff } n \notin \text{ set defined by } \varphi_n(x)$$

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$\varphi_0(x)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$\varphi_1(x)$	0	1	0	1	$\cdots$	1	$\cdots$	$\{1, 3, \ldots, n, \ldots\}$
$\varphi_2(x)$	0	1	1	0	$\cdots$	0	$\cdots$	$\{1, 2\}$
$\varphi_3(x)$	1	0	1	0	$\cdots$	1	$\cdots$	$\{0, 2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\varphi_n(x)$	0	0	1	0	$\cdots$	1	$\cdots$	$\{2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$n \in S^*$ iff $n \notin$ set defined by $\varphi_n(x)$

Suppose that S^* is definable in our language (say by $\varphi_m(x)$)

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$\varphi_0(x)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$\varphi_1(x)$	0	1	0	1	$\cdots$	1	$\cdots$	$\{1, 3, \ldots, n, \ldots\}$
$\varphi_2(x)$	0	1	1	0	$\cdots$	0	$\cdots$	$\{1, 2\}$
$\varphi_3(x)$	1	0	1	0	$\cdots$	1	$\cdots$	$\{0, 2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\varphi_n(x)$	0	0	1	0	$\cdots$	1	$\cdots$	$\{2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$n \in S^*$ iff $n \notin$ set defined by $\varphi_n(x)$

Write $\varphi_m(\overline{\mathrm{n}})$ for "$\varphi_m(x)$ is true of n"

	0	1	2	3	$\cdots$	n	$\cdots$	$S \subseteq \mathbb{N}$
$\varphi_0(x)$	0	0	0	0	$\cdots$	0	$\cdots$	$\emptyset$
$\varphi_1(x)$	0	1	0	1	$\cdots$	1	$\cdots$	$\{1, 3, \ldots, n, \ldots\}$
$\varphi_2(x)$	0	1	1	0	$\cdots$	0	$\cdots$	$\{1, 2\}$
$\varphi_3(x)$	1	0	1	0	$\cdots$	1	$\cdots$	$\{0, 2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\varphi_n(x)$	0	0	1	0	$\cdots$	1	$\cdots$	$\{2, \ldots, n, \ldots\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$n \in S^*$ iff $n \notin$ set defined by $\varphi_n(x)$

$$\varphi_m(\overline{n}) \leftrightarrow \neg \mathrm{True}(\ulcorner \varphi_n(\overline{n}) \urcorner)$$

where $\ulcorner \varphi_n(\overline{n}) \urcorner$ is the term in the language representing the code of $\varphi_n(\overline{n})$

D-Liar

$$\varphi_m(\overline{\mathrm{m}}) \leftrightarrow \neg\mathrm{True}(\ulcorner \varphi_m(\overline{\mathrm{m}}) \urcorner)$$

"m is true of $\varphi_m(x)$ iff it is not true that m is true of $\varphi_m(x)$"

Gödel's Idea

$$\varphi_m(\overline{\mathsf{m}}) \leftrightarrow \neg\mathsf{True}(\ulcorner \varphi_m(\overline{\mathsf{m}}) \urcorner)$$

Gödel's Idea

$$\varphi_m(\overline{\mathsf{m}}) \leftrightarrow \neg\mathsf{True}(\ulcorner \varphi_m(\overline{\mathsf{m}}) \urcorner)$$

$$\varphi_m(\overline{\mathsf{m}}) \leftrightarrow \neg\mathsf{Prov}(\ulcorner \varphi_m(\overline{\mathsf{m}}) \urcorner)$$

"$\varphi_m(\overline{\mathsf{m}})$ is true iff $\varphi_m(\overline{\mathsf{m}})$ is not provable."

Gödel's Idea

$$\varphi_m(\overline{\mathsf{m}}) \leftrightarrow \neg\mathsf{True}(\ulcorner \varphi_m(\overline{\mathsf{m}}) \urcorner)$$

$$\varphi_m(\overline{\mathsf{m}}) \leftrightarrow \neg\mathsf{Prov}(\ulcorner \varphi_m(\overline{\mathsf{m}}) \urcorner)$$

"$\varphi_m(\overline{\mathsf{m}})$ is true iff $\varphi_m(\overline{\mathsf{m}})$ is not provable."

$$\varphi_m(\overline{\mathsf{m}}) \leftrightarrow \neg\mathsf{Prov}(\ulcorner \varphi_m(\overline{\mathsf{m}}) \urcorner)$$

$\varphi_m(\overline{\mathsf{m}}) \leftrightarrow \neg\mathsf{Prov}(\ulcorner\varphi_m(\overline{\mathsf{m}})\urcorner)$

$\varphi_m(\overline{\mathsf{m}})$ is not provable: Suppose $\varphi_m(\overline{\mathsf{m}})$ is provable. Then, since we can only prove true statements, $\varphi_m(\overline{\mathsf{m}})$ is true. This means that $\neg\mathsf{Prov}(\ulcorner\varphi_m(\overline{\mathsf{m}})\urcorner)$ is true. So, $\varphi_m(\overline{\mathsf{m}})$ is not provable. Contradiction.

$\varphi_m(\overline{\mathsf{m}}) \leftrightarrow \neg\mathsf{Prov}(\ulcorner\varphi_m(\overline{\mathsf{m}})\urcorner)$

$\varphi_m(\overline{\mathsf{m}})$ is not provable: Suppose $\varphi_m(\overline{\mathsf{m}})$ is provable. Then, since we can only prove true statements, $\varphi_m(\overline{\mathsf{m}})$ is true. This means that $\neg\mathsf{Prov}(\ulcorner\varphi_m(\overline{\mathsf{m}})\urcorner)$ is true. So, $\varphi_m(\overline{\mathsf{m}})$ is not provable. Contradiction.

$\neg\varphi_m(\overline{\mathsf{m}})$ is not provable: Suppose that $\neg\varphi_m(\overline{\mathsf{m}})$ is provable. Since our system only proves true statements, $\neg\varphi_m(\overline{\mathsf{m}})$ is true. Then $\neg\neg\mathsf{Prov}(\ulcorner\varphi_m(\overline{\mathsf{m}})\urcorner)$ is true. So, $\varphi_m(\overline{\mathsf{m}})$ is provable. This contradicts the assumption that the system is consistent.

$\varphi_m(\overline{\mathrm{m}}) \leftrightarrow \neg\mathrm{Prov}(\ulcorner\varphi_m(\overline{\mathrm{m}})\urcorner)$

$\varphi_m(\overline{\mathrm{m}})$ is not provable: Suppose $\varphi_m(\overline{\mathrm{m}})$ is provable. Then, since we can only prove true statements, $\varphi_m(\overline{\mathrm{m}})$ is true. This means that $\neg\mathrm{Prov}(\ulcorner\varphi_m(\overline{\mathrm{m}})\urcorner)$ is true. So, $\varphi_m(\overline{\mathrm{m}})$ is not provable. Contradiction.

$\neg\varphi_m(\overline{\mathrm{m}})$ is not provable: Suppose that $\neg\varphi_m(\overline{\mathrm{m}})$ is provable. Since our system only proves true statements, $\neg\varphi_m(\overline{\mathrm{m}})$ is true. Then $\neg\neg\mathrm{Prov}(\ulcorner\varphi_m(\overline{\mathrm{m}})\urcorner)$ is true. So, $\varphi_m(\overline{\mathrm{m}})$ is provable. This contradicts the assumption that the system is consistent.

Conclusion: Neither $\varphi_m(\overline{\mathrm{m}})$ nor $\neg\varphi_m(\overline{\mathrm{m}})$ is provable.

$$\varphi_m(\overline{\mathrm{m}}) \leftrightarrow \neg \mathrm{Prov}(\ulcorner \varphi_m(\overline{\mathrm{m}}) \urcorner)$$

1. Apply Richard's move to Cantor's construction to get the D-Liar
2. Replace 'true' with 'provable' on the right-hand side of the sentence
3. Proceed with the difficult task of *arithmetizing syntax* to construct the right-side of the sentence ($\mathrm{Prov}(v)$).
4. Show that the above sentence is provable within the formal system eliminating any appeal to the concept of "truth". The assumption that provable implies truth is replaced with (ω-)consistency.

H. Gaifman (2006). *Naming and Diagonalization, From Cantor to Gödel to Kleene*. Logic Journal of the IGPL, pp. 709 - 728.

Naming systems

Naming systems are intended as a basic framework for studying situations in which functions can be applied to their names....In a naming system we do not specify how the names are attached to functions, we assume only that there is such a correlation and that it satisfies certain minimal requirements.

H. Gaifman (2006). *Naming and Diagonalization, From Cantor to Gödel to Kleene*. Logic Journal of the IGPL, pp. 709 - 728.

Naming systems I

$$\mathcal{D} = (D, type, \{\ \})$$

such that:

- D is a non-empty set.
- $type$ assigns to each $a \in D$ its type: $type(a)$ tells us if a is a name (of a function) and, if it is, the function's arity.

  A name of arity n, or n-ary name, is one that names an n-ary function.

  Types can be construed as tuples: (0)—if a is not a name, $(1, n)$—if it is an n-ary name.
- $\{\ \}$ is a mapping that assigns to every n-ary name, a, a function:

$$\{a\} : D^n \to D$$

Naming systems II

- There is at least one named function of arity greater than 0

Naming systems II

- There is at least one named function of arity greater than 0
- Substitution of names (SN): If f is an n-ary named function, where $n > 0$, then, for every name a:

$$\lambda x_2, \ldots x_n f(a, x_2, \ldots, x_n) \text{ is named}$$

Naming systems II

- There is at least one named function of arity greater than 0
- Substitution of names (SN): If f is an n-ary named function, where $n > 0$, then, for every name a:

$$\lambda x_2, \ldots x_n f(a, x_2, \ldots, x_n) \text{ is named}$$

- Variable permutation (VP): If f is an n-ary named function, where $n > 0$, and π is a permutation of $\{1, \ldots, n\}$, then

$$\lambda x_1, \ldots x_n f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) \text{ is named}$$

n-Diagonal Function

For $n > 0$, an n-diagonal function, denoted dl_n, is a function that maps each n-ary name a to a name of the function:

$$\lambda x_2, \ldots, x_n\{a\}(a, x_2, \ldots, x_n)$$

Thus, $dl_n(a)$ is the name of the above function.

n-Diagonal Function

For $n > 0$, an n-diagonal function, denoted dl_n, is a function that maps each n-ary name a to a name of the function:

$$\lambda x_2, \ldots, x_n \{a\}(a, x_2, \ldots, x_n)$$

Thus, $dl_n(a)$ is the name of the above function.

For all n-ary names a,

$$\{dl_n(a)\}(x_2, \ldots, x_n) = \{a\}(a, x_2, \ldots, x_n)$$

General Fixed-Point Theorem

GFP Theorem. If F is an $(n+1)$-ary named function, $n \geq 0$, and the composition $F(dI_{n+1}(x_0), x_1, \ldots, x_n)$ is named, then there is an n-ary name, e, such that:

$$\{e\}(x_1, \ldots, x_n) = F(e, x_1, \ldots, x_n)$$

General Fixed-Point Theorem

GFP Theorem. If F is an $(n+1)$-ary named function, $n \geq 0$, and the composition $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ is named, then there is an n-ary name, e, such that:

$$\{e\}(x_1, \ldots, x_n) = F(e, x_1, \ldots, x_n)$$

Proof. Let c be the name of $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ and let $e = dl_{n+1}(c)$. Then for and $\vec{x} = (x_1, \ldots, x_n)$,

General Fixed-Point Theorem

GFP Theorem. If F is an $(n+1)$-ary named function, $n \geq 0$, and the composition $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ is named, then there is an n-ary name, e, such that:

$$\{e\}(x_1, \ldots, x_n) = F(e, x_1, \ldots, x_n)$$

Proof. Let c be the name of $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ and let $e = dl_{n+1}(c)$. Then for and $\vec{x} = (x_1, \ldots, x_n)$,

$$\{e\}(\vec{x}) \quad = \quad \{dl_{n+1}(c)\}(\vec{x}) \quad \text{(definition of } e\text{)}$$

General Fixed-Point Theorem

GFP Theorem. If F is an $(n+1)$-ary named function, $n \geq 0$, and the composition $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ is named, then there is an n-ary name, e, such that:

$$\{e\}(x_1, \ldots, x_n) = F(e, x_1, \ldots, x_n)$$

Proof. Let c be the name of $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ and let $e = dl_{n+1}(c)$. Then for and $\vec{x} = (x_1, \ldots, x_n)$,

$$\begin{array}{rcll} \{e\}(\vec{x}) & = & \{dl_{n+1}(c)\}(\vec{x}) & \text{(definition of } e\text{)} \\ & = & \{c\}(c, \vec{x}) & \text{(definition of } dl_{n+1}(c)\text{)} \end{array}$$

General Fixed-Point Theorem

GFP Theorem. If F is an $(n+1)$-ary named function, $n \geq 0$, and the composition $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ is named, then there is an n-ary name, e, such that:

$$\{e\}(x_1, \ldots, x_n) = F(e, x_1, \ldots, x_n)$$

Proof. Let c be the name of $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ and let $e = dl_{n+1}(c)$. Then for and $\vec{x} = (x_1, \ldots, x_n)$,

$$\begin{array}{rcll} \{e\}(\vec{x}) & = & \{dl_{n+1}(c)\}(\vec{x}) & \text{(definition of } e\text{)} \\ & = & \{c\}(c, \vec{x}) & \text{(definition of } dl_{n+1}(c)\text{)} \\ & = & F(dl_{n+1}(c), \vec{x}) & \text{(definition of } c\text{)} \end{array}$$

General Fixed-Point Theorem

GFP Theorem. If F is an $(n+1)$-ary named function, $n \geq 0$, and the composition $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ is named, then there is an n-ary name, e, such that:

$$\{e\}(x_1, \ldots, x_n) = F(e, x_1, \ldots, x_n)$$

Proof. Let c be the name of $F(dl_{n+1}(x_0), x_1, \ldots, x_n)$ and let $e = dl_{n+1}(c)$. Then for and $\vec{x} = (x_1, \ldots, x_n)$,

$$\begin{aligned}
\{e\}(\vec{x}) &= \{dl_{n+1}(c)\}(\vec{x}) && \text{(definition of } e\text{)} \\
&= \{c\}(c, \vec{x}) && \text{(definition of } dl_{n+1}(c)\text{)} \\
&= F(dl_{n+1}(c), \vec{x}) && \text{(definition of } c\text{)} \\
&= F(e, \vec{x}) && \text{(definition of } e\text{)}
\end{aligned}$$

- ✓ Gödel numbering and naming systems
- ✓ Gödel-Carnap Fixed Point Theorem
- ▶ Representing functions/relations
- ▶ Provability predicate
- ▶ Löb's Theorem

Representability

Definition

Suppose that $f : \mathbb{N}^k \to \mathbb{N}$. We say that f is **representable** in **Q** when there is a formula $A_f(x_0, \ldots, x_{k-1}, y)$ such that for all $n_0, \ldots, n_{k-1} \in \mathbb{N}$: if $f(n_0, \ldots, n_{k-1}) = m$ then

1. $\mathbf{Q} \vdash A_f(\overline{n_0}, \ldots, \overline{n_{k-1}}, \overline{m})$
2. $\mathbf{Q} \vdash \forall y(A_f(\overline{n_0}, \ldots, \overline{n_{k-1}}, y) \to y = \overline{m})$

Equivalent definitions of representability

- f is representable in $\mathbf{Q}$ iff there is a formula $A_f(x_0, \ldots, x_{k-1}, y)$ such that for all $n_0, \ldots, n_{k-1} \in \mathbb{N}$, if $f(n_0, \ldots, n_{k-1}) = m$ then:

$$\mathbf{Q} \vdash \forall y(A_f(\overline{n_0}, \ldots, \overline{n_{k-1}}, y) \leftrightarrow y = \overline{m})$$

Equivalent definitions of representability

- f is representable in $\mathbf{Q}$ iff there is a formula $A_f(x_0, \dots, x_{k-1}, y)$ such that for all $n_0, \dots, n_{k-1} \in \mathbb{N}$, if $f(n_0, \dots, n_{k-1}) = m$ then:

$$\mathbf{Q} \vdash \forall y(A_f(\overline{n_0}, \dots, \overline{n_{k-1}}, y) \leftrightarrow y = \overline{m})$$

- f is representable in $\mathbf{Q}$ iff there is a formula $A_f(x_0, \dots, x_{k-1}, y)$ such that for all $n_0, \dots, n_{k-1} \in \mathbb{N}$:
  1. If $f(n_0, \dots, n_{k-1}) = m$, then $\mathbf{Q} \vdash A_f(\overline{n_0}, \dots, \overline{n_{k-1}}, \overline{m})$
  2. If $f(n_0, \dots, n_{k-1}) \neq m$, then $\mathbf{Q} \vdash \neg A_f(\overline{n_0}, \dots, \overline{n_{k-1}}, \overline{m})$

Equivalent definitions of representability

- f is representable in $\mathbf{Q}$ iff there is a formula $A_f(x_0, \ldots, x_{k-1}, y)$ such that for all $n_0, \ldots, n_{k-1} \in \mathbb{N}$, if $f(n_0, \ldots, n_{k-1}) = m$ then:

$$\mathbf{Q} \vdash \forall y(A_f(\overline{n_0}, \ldots, \overline{n_{k-1}}, y) \leftrightarrow y = \overline{m})$$

- f is representable in $\mathbf{Q}$ iff there is a formula $A_f(x_0, \ldots, x_{k-1}, y)$ such that for all $n_0, \ldots, n_{k-1} \in \mathbb{N}$:
  1. If $f(n_0, \ldots, n_{k-1}) = m$, then $\mathbf{Q} \vdash A_f(\overline{n_0}, \ldots, \overline{n_{k-1}}, \overline{m})$
  2. If $f(n_0, \ldots, n_{k-1}) \neq m$, then $\mathbf{Q} \vdash \neg A_f(\overline{n_0}, \ldots, \overline{n_{k-1}}, \overline{m})$

- f is representable in $\mathbf{Q}$ iff there is a formula $A_f(x_0, \ldots, x_{k-1}, y)$ such that for all $n_0, \ldots, n_{k-1} \in \mathbb{N}$:
  1. if $f(n_0, \ldots, n_{k-1}) = m$ then $\mathbf{Q} \vdash A_f(\overline{n_0}, \ldots, \overline{n_{k-1}}, \overline{m})$
  2. $\mathbf{Q} \vdash \exists! y A_f(\overline{n_0}, \ldots, \overline{n_{k-1}}, y)$

Exercise

Prove that all of the definitions of representability are equivalent.

Representing Relations

A relation $R \subseteq \mathbb{N}^k$ is **representable** in **Q** provided that the characteristic function χ_R is representable in **Q**. It is not hard to see that this is equivalent to saying that $R \subseteq \mathbb{N}^k$ is representable in **Q** provided that there is a formula A_R such that for all $n_0, \ldots, n_{k-1} \in \mathbb{N}$:

1. if $(n_0, \ldots, n_{k-1}) \in R$, then $\mathbf{Q} \vdash A_R(\overline{n_0}, \ldots, \overline{n_{k-1}})$
2. if $(n_0, \ldots, n_{k-1}) \notin R$, then $\mathbf{Q} \vdash \neg A_R(\overline{n_0}, \ldots, \overline{n_{k-1}})$

All of the following relations are representable in **Q**:

- $Sent(x)$: x is the Gödel number of a sentence of $\mathcal{L}_A$
- $Form(x)$: x is the Gödel number of a formula of $\mathcal{L}_A$
- $Term(x)$: x is the Gödel number of a term of $\mathcal{L}_A$
- $Axiom(x)$: x is the Gödel number of an axiom of **Q**
- $Prf_{\mathbf{PA}}(x, y)$: x is the Gödel number of a derivation in **PA** of a formula with Gödel number y.
- $\ldots$

Proof Predicate

The proof relation $Prf_{\mathbf{PA}}(x, y)$ is represented by a formula $\mathsf{Prf}_{\mathbf{PA}}$.

Proof Predicate

The proof relation $Prf_{\mathbf{PA}}(x, y)$ is represented by a formula $\mathrm{Prf}_{\mathbf{PA}}$.

The *proof predicate*, denoted $\mathrm{Prov}_{\mathbf{PA}}(y)$, is defined as follows:

$$\exists x \mathrm{Prf}_{\mathbf{PA}}(x, y)$$

Derivability Conditions

It can be shown that the provability predicate $\mathrm{Prov}_{\mathbf{PA}}$ satisfies the following:

*D*1. If $\mathbf{PA} \vdash A$, then $\mathbf{PA} \vdash \mathrm{Prov}_{\mathbf{PA}}(\ulcorner A \urcorner)$

*D*2. $\mathbf{PA} \vdash \mathrm{Prov}_{\mathbf{PA}}(\ulcorner A \to B \urcorner) \to (\mathrm{Prov}_{\mathbf{PA}}(\ulcorner A \urcorner) \to \mathrm{Prov}_{\mathbf{PA}}(\ulcorner B \urcorner))$

*D*3. $\mathbf{PA} \vdash \mathrm{Prov}_{\mathbf{PA}}(\ulcorner A \urcorner) \to \mathrm{Prov}_{\mathbf{PA}}(\ulcorner \mathrm{Prov}_{\mathbf{PA}}(\ulcorner A \urcorner) \urcorner)$

Derivability Conditions

A provability predicate for $\mathbf{T}$, denoted $\mathrm{Prov}_{\mathbf{T}}$, satisfies the following:

*D*1. If $\mathbf{T} \vdash A$, then $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner)$

*D*2. $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \to B \urcorner) \to (\mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to \mathrm{Prov}_{\mathbf{T}}(\ulcorner B \urcorner))$

*D*3. $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to \mathrm{Prov}_{\mathbf{T}}(\ulcorner \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \urcorner)$

Reflection Principle

The reflection principle for **T** is the schema

$$\mathsf{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to A$$

Monotonicity Inference for the Provability Predicate

Lemma

For any theory $\mathbf{T}$, if $\mathrm{Prov}_{\mathbf{T}}$ satisfies $D1$ and $D2$, then:

$$\text{From } \mathbf{T} \vdash A \to B, \text{ infer } \mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to \mathrm{Prov}(\ulcorner B \urcorner).$$

Löb's Theorem

Theorem (Löb's Theorem)

Let $\mathbf{T}$ be an axiomatizable theory extending $\mathbf{Q}$, and suppose $\mathrm{Prov}_{\mathbf{T}}(y)$ is a formula satisfying conditions $D1$-$D3$.

$$\text{If } \mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \rightarrow A, \text{ then } \mathbf{T} \vdash A.$$

Suppose A is a sentence such that $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to A$. Let $B(y)$ be the formula

$$\mathrm{Prov}_{\mathbf{T}}(y) \to A$$

Suppose A is a sentence such that $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \rightarrow A$. Let $B(y)$ be the formula

$$\mathrm{Prov}_{\mathbf{T}}(y) \rightarrow A$$

By the Fixed-Point Theorem, there is a sentence D such that

$$\mathbf{T} \vdash D \leftrightarrow B(\ulcorner D \urcorner)$$

Suppose that $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \rightarrow A$.

Suppose A is a sentence such that $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to A$. Let $B(y)$ be the formula

$$\mathrm{Prov}_{\mathbf{T}}(y) \to A$$

By the Fixed-Point Theorem, there is a sentence D such that

$$\mathbf{T} \vdash D \leftrightarrow B(\ulcorner D \urcorner)$$

Suppose that $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to A$.

To simplify the notation, we write $\mathrm{Prov}(\cdot)$ instead of $\mathrm{Prov}_{\mathbf{T}}$

1.	$D \leftrightarrow (\mathrm{Prov}(\ulcorner D \urcorner) \to A)$	FPT
2.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \to A \urcorner)$	Lemma: 1
3.	$\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \to A \urcorner) \to (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner))$	D2
4.	$\mathrm{Prov}(\ulcorner D \urcorner) \to (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner))$	PC: 2, 3

1.	$D \leftrightarrow (\mathrm{Prov}(\ulcorner D \urcorner) \to A)$	FPT
$\vdots$	$\vdots$	$\vdots$
4.	$\mathrm{Prov}(\ulcorner D \urcorner) \to (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner))$	PC: 2, 3
5.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner)$	D3

1.	$D \leftrightarrow (\mathrm{Prov}(\ulcorner D \urcorner) \to A)$	FPT
$\vdots$	$\vdots$	$\vdots$
4.	$\mathrm{Prov}(\ulcorner D \urcorner) \to (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner))$	PC: 2, 3
5.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner)$	D3
6.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner)$	PC: 4, 5

1.	$D \leftrightarrow (\mathrm{Prov}(\ulcorner D \urcorner) \to A)$	FPT
$\vdots$	$\vdots$	$\vdots$
4.	$\mathrm{Prov}(\ulcorner D \urcorner) \to (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner))$	PC: 2, 3
5.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner)$	D3
6.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner)$	PC: 4, 5
7.	$\mathrm{Prov}(\ulcorner A \urcorner) \to A$	Assumption

1.	$D \leftrightarrow (\mathrm{Prov}(\ulcorner D \urcorner) \to A)$	FPT
$\vdots$	$\vdots$	$\vdots$
4.	$\mathrm{Prov}(\ulcorner D \urcorner) \to (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner))$	PC: 2, 3
5.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner)$	D3
6.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner)$	PC: 4, 5
7.	$\mathrm{Prov}(\ulcorner A \urcorner) \to A$	Assumption
8.	$\mathrm{Prov}(\ulcorner D \urcorner) \to A$	PC: 6, 7

1.	$D \leftrightarrow (\mathrm{Prov}(\ulcorner D \urcorner) \rightarrow A)$	FPT
$\vdots$	$\vdots$	$\vdots$
4.	$\mathrm{Prov}(\ulcorner D \urcorner) \rightarrow (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \rightarrow \mathrm{Prov}(\ulcorner A \urcorner))$	PC: 2, 3
5.	$\mathrm{Prov}(\ulcorner D \urcorner) \rightarrow \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner)$	D3
6.	$\mathrm{Prov}(\ulcorner D \urcorner) \rightarrow \mathrm{Prov}(\ulcorner A \urcorner)$	PC: 4, 5
7.	$\mathrm{Prov}(\ulcorner A \urcorner) \rightarrow A$	Assumption
8.	$\mathrm{Prov}(\ulcorner D \urcorner) \rightarrow A$	PC: 6, 7
9.	D	PC: 1, 8

1.	$D \leftrightarrow (\mathrm{Prov}(\ulcorner D \urcorner) \to A)$	FPT
$\vdots$	$\vdots$	$\vdots$
4.	$\mathrm{Prov}(\ulcorner D \urcorner) \to (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner))$	PC: 2, 3
5.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner)$	D3
6.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner)$	PC: 4, 5
7.	$\mathrm{Prov}(\ulcorner A \urcorner) \to A$	Assumption
8.	$\mathrm{Prov}(\ulcorner D \urcorner) \to A$	PC: 6, 7
9.	D	PC: 1, 8
10.	$\mathrm{Prov}(\ulcorner D \urcorner)$	D1 from 9

1.	$D \leftrightarrow (\mathrm{Prov}(\ulcorner D \urcorner) \to A)$	FPT
$\vdots$	$\vdots$	$\vdots$
4.	$\mathrm{Prov}(\ulcorner D \urcorner) \to (\mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner))$	PC: 2, 3
5.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner D \urcorner) \urcorner)$	D3
6.	$\mathrm{Prov}(\ulcorner D \urcorner) \to \mathrm{Prov}(\ulcorner A \urcorner)$	PC: 4, 5
7.	$\mathrm{Prov}(\ulcorner A \urcorner) \to A$	Assumption
8.	$\mathrm{Prov}(\ulcorner D \urcorner) \to A$	PC: 6, 7
9.	D	PC: 1, 8
10.	$\mathrm{Prov}(\ulcorner D \urcorner)$	D1 from 9
11.	A	PC: 8, 10

1.	$D \leftrightarrow (\mathsf{Prov}(\ulcorner D \urcorner) \to A)$	FPT
$\vdots$	$\vdots$	$\vdots$
4.	$\mathsf{Prov}(\ulcorner D \urcorner) \to (\mathsf{Prov}(\ulcorner \mathsf{Prov}(\ulcorner D \urcorner) \urcorner) \to \mathsf{Prov}(\ulcorner A \urcorner))$	PC: 2, 3
5.	$\mathsf{Prov}(\ulcorner D \urcorner) \to \mathsf{Prov}(\ulcorner \mathsf{Prov}(\ulcorner D \urcorner) \urcorner)$	D3
6.	$\mathsf{Prov}(\ulcorner D \urcorner) \to \mathsf{Prov}(\ulcorner A \urcorner)$	PC: 4, 5
7.	$\mathsf{Prov}(\ulcorner A \urcorner) \to A$	Assumption
8.	$\mathsf{Prov}(\ulcorner D \urcorner) \to A$	PC: 6, 7
9.	D	PC: 1, 8
10.	$\mathsf{Prov}(\ulcorner D \urcorner)$	D1 from 9
11.	A	PC: 8, 10

‘**PA** couldn't be more modest about its own veracity’

By Löb's Theorem, it is not true that for all sentences φ,

$$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$$

‘**PA** couldn't be more modest about its own veracity’

By Löb's Theorem, it is not true that for all sentences φ,

$$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner \varphi \urcorner) \rightarrow \varphi \urcorner)$$

Statement	It is not true that...
$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \varphi \urcorner)$ implies $\mathbf{PA} \vdash \varphi$	$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \varphi \urcorner) \rightarrow \varphi$

‘**PA** couldn’t be more modest about its own veracity’

By Löb’s Theorem, it is not true that for all sentences φ,

$$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$$

Statement	It is not true that...
$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \varphi \urcorner)$ implies $\mathbf{PA} \vdash \varphi$	$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \varphi \urcorner) \to \varphi$
$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \neg\varphi \urcorner)$ implies $\mathbf{PA} \nvdash \mathrm{Prov}(\ulcorner \varphi \urcorner)$	$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \neg\varphi \urcorner) \to \neg\mathrm{Prov}(\ulcorner \varphi \urcorner)$

‘**PA** couldn’t be more modest about its own veracity’

By Löb’s Theorem, it is not true that for all sentences φ,

$$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$$

Statement	It is not true that...
$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \varphi \urcorner)$ implies $\mathbf{PA} \vdash \varphi$	$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \varphi \urcorner) \to \varphi$
$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \neg\varphi \urcorner)$ implies $\mathbf{PA} \nvdash \mathrm{Prov}(\ulcorner \varphi \urcorner)$	$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \neg\varphi \urcorner) \to \neg\mathrm{Prov}(\ulcorner \varphi \urcorner)$
$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \neg\mathrm{Prov}(\ulcorner \varphi \urcorner) \urcorner)$ implies $\mathbf{PA} \vdash \neg\mathrm{Prov}(\varphi)$	$\mathbf{PA} \vdash \mathrm{Prov}(\ulcorner \neg\mathrm{Prov}(\ulcorner \varphi \urcorner) \urcorner) \to \neg\mathrm{Prov}(\ulcorner \varphi \urcorner)$

Plan

- ✓ Introduction: Smullyan's Machine
- ✓ Background
  - ✓ Formal Arithmetic
  - ✓ Gödel's Incompleteness Theorems
  - ✓ Names and Gödel numbering
  - ✓ Fixed Point Theorem
- ✓ Provability predicate and Löb's Theorem
- Provability logic
- Truth predicate and Tarski's Theorem
- A Primer on Epistemic and Doxastic Logic
- Anti-Expert Paradoxes
- Predicate approach to modality
- The Knower Paradox and variants
- Epistemic Arithmetic
- Gödel's Disjunction

Propositional Modal Logic

$$p \mid \neg\varphi \wedge \varphi \wedge \psi \mid \Box\varphi$$

where $p \in \mathsf{AT}$ (at set of atomic propositions).

The intended interpretation of $\Box\varphi$ is "there is a proof (in **PA**) of φ".

Propositional Modal Logic

$$p \mid \neg\varphi \wedge \varphi \wedge \psi \mid \Box\varphi$$

where $p \in \mathrm{AT}$ (at set of atomic propositions).

The intended interpretation of $\Box\varphi$ is "there is a proof (in **PA**) of φ".

A **frame** is a tuple (W, R) such that $W \neq \emptyset$ and $R \subseteq W \times W$.

Propositional Modal Logic

$$p \mid \neg\varphi \wedge \varphi \wedge \psi \mid \Box\varphi$$

where $p \in \mathrm{AT}$ (at set of atomic propositions).

The intended interpretation of $\Box\varphi$ is "there is a proof (in **PA**) of φ".

A **frame** is a tuple (W, R) such that $W \neq \emptyset$ and $R \subseteq W \times W$.

A **model** is a tuple (W, R, V) where (W, R) is a frame and $V : \mathrm{AT} \to \wp(W)$.

Truth/Validity

For a model $\mathcal{M} = (W, R, V)$ and $w \in W$, we write $\mathcal{M} \models \varphi$ when φ is true at w in $\mathcal{M}$.

- $\mathcal{M}, w \models p$ iff $w \in V(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models \Box\varphi$ iff for all $v \in W$, if $w\, R\, v$, then $\mathcal{M}, v \models \varphi$

For a frame $\mathcal{F} = (W, R)$, φ is **valid on** $\mathcal{F}$, denoted $\mathcal{F} \models \varphi$, when $\mathcal{M}, w \models \varphi$ for all models $\mathcal{M}$ based on $\mathcal{F}$ and $w \in W$.

Provability Logic: **GL**

K	$\Box(\varphi \to \psi) \to (\Box\varphi \to \psi)$
L	$\Box(\Box\varphi \to \varphi) \to \Box\varphi$
MP	$\varphi, \varphi \to \psi \ \therefore\ \psi$
NEC	$\varphi \ \therefore\ \Box\varphi$

Some Results

- $\mathbf{GL} \vdash \Box\varphi \to \Box\Box\varphi$.

Some Results

- $\mathbf{GL} \vdash \Box\varphi \to \Box\Box\varphi$.
- $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ is valid on a frame (W, R) if, and only if, R is transitive and converse well-founded (there are no infinite ascending sequences, that is sequences of the form $w_1 \; R \; w_2 \; R \; w_3 \cdots$).

Some Results

- $\mathbf{GL} \vdash \Box\varphi \to \Box\Box\varphi$.
- $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ is valid on a frame (W, R) if, and only if, R is transitive and converse well-founded (there are no infinite ascending sequences, that is sequences of the form $w_1 \; R \; w_2 \; R \; w_3 \cdots$).
- The logic **GL** is not compact:

$$\Gamma = \{\Diamond p_0, \Box(p_0 \to \Diamond p_1), \Box(p_1 \to \Diamond p_2), \ldots, \Box(p_n \to \Diamond p_{n+1}), \ldots\}.$$

is finitely satisfiable, but not satisfiable.

Some Results

- $\mathbf{GL} \vdash \Box\varphi \to \Box\Box\varphi$.
- $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ is valid on a frame (W, R) if, and only if, R is transitive and converse well-founded (there are no infinite ascending sequences, that is sequences of the form $w_1 \; R \; w_2 \; R \; w_3 \cdots$).
- The logic **GL** is not compact:

  $$\Gamma = \{\Diamond p_0, \Box(p_0 \to \Diamond p_1), \Box(p_1 \to \Diamond p_2), \ldots, \Box(p_n \to \Diamond p_{n+1}), \ldots\}.$$

  is finitely satisfiable, but not satisfiable.
- The logic **GL** is sound and weakly complete with respect to the class of frames that are transitive and converse well-founded.

Arithmetic Completeness

An **arithmetic translation** is a function t such that

1. For all $p \in \mathrm{At}$, $t(p)$ is a sentence of $\mathcal{L}_A$
2. t commutes with the boolean connectives: $t(\neg\varphi) = \neg t(\varphi)$, $t(\varphi \wedge \psi) = t(\varphi) \wedge t(\psi)$, etc.
3. $t(\Box\varphi) = \mathrm{Prov}_{\mathbf{PA}}(\ulcorner t(\varphi) \urcorner)$

Theorem (Solovay 1976).

$$\mathbf{GL} \vdash \varphi \text{ iff for every arithmetic translation } t,\ \mathbf{PA} \vdash t(\varphi).$$