Logic in Computer Science IV

Lesson 7, the incompleteness theorem,
1. formalization of arithmetic

Richard Dedekind 1831–1916
Giuseppe Peano 1858–1932

Lesson 7, the incompleteness theorem, 1. formalization of arithmetic

Richard Dedekind 1831–1916
Giuseppe Peano 1858–1932

$(N; 0, S)$

1. for every x, $S(x) \neq 0$,
2. if $x \neq y$, then $S(x) \neq S(y)$,
3. for every set $X \subseteq N$, if $0 \in N$ and $x \in N$ implies $S(x) \in N$, then $X = N$.

Lesson 7, the incompleteness theorem, 1. formalization of arithmetic

Richard Dedekind 1831–1916
Giuseppe Peano 1858–1932

$(N; 0, S)$

1. for every x, $S(x) \neq 0$,
2. if $x \neq y$, then $S(x) \neq S(y)$,
3. for every set $X \subseteq N$, if $0 \in N$ and $x \in N$ implies $S(x) \in N$, then $X = N$.

This formalization of arithmetic uses a second order concept of a set of numbers.

Robinson's Arithmetic **Q**

(Raphael M. Robinson)

Robinson's Arithmetic **Q**

(Raphael M. Robinson)

Language 0, S, $+$, $\times$ and $\leq$.

Axioms – universal closure of the following formulas:

1. successor function
   - $S(x) \neq 0$
   - $x \neq y \to S(x) \neq S(y)$
   - $x \neq 0 \to \exists y(y = S(x))$
2. addition
   - $x + 0 = x$
   - $x + S(y) = S(x + y)$
3. multiplication
   - $x \times 0 = 0$
   - $x \times S(y) = x \times y + x$
4. inequality
   - $x \leq y \equiv \exists z(z + x = y)$

Robinson's Arithmetic **Q**

(Raphael M. Robinson)

Language 0, S, $+$, $\times$ and $\leq$.
Axioms – universal closure of the following formulas:

1. successor function
   - $S(x) \neq 0$
   - $x \neq y \rightarrow S(x) \neq S(y)$
   - $x \neq 0 \rightarrow \exists y(y = S(x))$
2. addition
   - $x + 0 = x$
   - $x + S(y) = S(x + y)$
3. multiplication
   - $x \times 0 = 0$
   - $x \times S(y) = x \times y + x$
4. inequality
   - $x \leq y \equiv \exists z(z + x = y)$

weak, but essentially undecidable–every consistent extension is undecidable

Σ-completeness

Q cannot even prove the commutativity of addition, but

Σ-completeness

Q cannot even prove the commutativity of addition, but

Theorem (and Definition)

Q *is Σ-complete, which means that for every* Σ_1 *sentence* ϕ

$$\mathbb{N} \models \phi \ \Rightarrow \ \mathsf{Q} \vdash \phi.$$

I.e. **Q** *proves all true* Σ_1 *sentences.*

A $\mathbf{\Sigma_1}$ sentence has a prefix of existential quantifiers followed by bounded quantifiers $\exists x \leq t$ and $\forall y \leq s$.

example

$\mathbf{Q}$ proves $1 \leq 2$.

example

Q proves $1 \leq 2$.

We will use axioms

1. $x + 0 = x$
2. $x + S(y) = S(x + y)$
3. $x \leq y \equiv \exists z(z + x = y)$

- $S0 + S0 = S(S0 + 0)$
- $S0 + S0 = S(S0)$
- $\exists z(S0 + z = SS0)$
- $S0 \leq SS0$

example

Q proves $1 \leq 2$.

We will use axioms

1. $x + 0 = x$
2. $x + S(y) = S(x + y)$
3. $x \leq y \equiv \exists z(z + x = y)$

- $S0 + S0 = S(S0 + 0)$
- $S0 + S0 = S(S0)$
- $\exists z(S0 + z = SS0)$
- $S0 \leq SS0$

Exercise

Prove $2 \times 2 = 4$.

Proof idea (of Σ-completeness of $\mathbf{Q}$)

Proof idea (of Σ-completeness of **Q**)

- Eliminate existential quantifiers by substituting numerals
- Elminate bounded universal quantifiers using
  $\mathsf{Q} \quad \vdash \quad x \leq S^n 0 \quad \equiv \quad (x = 0 \vee x = S0 \vee \cdots \vee x = S^n 0)$
  —(**Exercise**)
- Show that every closed term equals to a numeral provably in **Q**.
- Show that every true quantifier-free sentence is provable.

Peano Arithmetic **PA**

= Robinson's Arithmetic + induction formulas for all formulas in the language 0, S, $+$, $\times$ and $\leq$.

$$\phi(0) \wedge \forall x(\phi(x) \to \phi(S(x))) \to \forall y.\phi(y)$$

Peano Arithmetic **PA**

= Robinson's Arithmetic + induction formulas for all formulas in the language 0, S, $+$, $\times$ and $\leq$.

$$\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(S(x))) \rightarrow \forall y.\phi(y)$$

Remarks

1. $\phi(x)$ may have other free variables,
2. $x \neq 0 \rightarrow \exists y(y = S(x))$ is redundant if we have induction,
3. $x \leq y \equiv \exists z(z + x = y)$ is a definition of $\leq$, so it can be omitted.
4. Peano Arithmetic is incomplete, because it only has induction for arithmetical formulas.

Finite Set Theory

Zermelo-Fraenkel Set Theory

- without the axiom of infinity
- plus the axiom *"every set is finite"*

Theorem

Peano Arithmetic and Finite Set Theory are mutually interpretable, i.e.,

1. *there are arithmetical formulas $V(x)$ for the universe of sets and $E(x, y)$ for the relation of being an element such that translations of all axioms of Finite Set Theory are provable in Peano Arithmetic,*
2. *there are set theoretical formulas ... such that all axioms of Peano Arithmetic are provable in Finite Set Theory.*

Finite Set Theory

Zermelo-Fraenkel Set Theory

- without the axiom of infinity
- plus the axiom *"every set is finite"*

Theorem

Peano Arithmetic and Finite Set Theory are mutually interpretable, i.e.,

1. *there are arithmetical formulas $V(x)$ for the universe of sets and $E(x, y)$ for the relation of being an element such that translations of all axioms of Finite Set Theory are provable in Peano Arithmetic,*
2. *there are set theoretical formulas ... such that all axioms of Peano Arithmetic are provable in Finite Set Theory.*
3. **Moreover,** *the interpretations are faithful, i.e., Finite Set Theory proves exactly the same sentences about numbers as Peano Arithmetic and vice versa Peano Arithmetic ...*

Proof.

The essence is coding pairs and sequences in PA.

Proof.

The essence is coding pairs and sequences in PA.

- coding pairs, Cantor's pairing function

$$\langle x, y \rangle := \frac{(x+y)^2 + x + 1}{2}$$

Proof.

The essence is coding pairs and sequences in PA.

- coding pairs, Cantor's pairing function

$$\langle x, y\rangle := \frac{(x+y)^2 + x + 1}{2}$$

- Gödel's function

  $$\beta(a, i) := \min\{x < a\ ;\ \exists y < a \exists z < a(a = \langle y, z\rangle \wedge 1 + (\langle x, i\rangle + 1) | y\}.$$

  Given $a_1, \ldots, a_n$, there exists a such that $\beta(a, i) = a_i$, for $i = 1, \ldots, n$.

Proof.

The essence is coding pairs and sequences in PA.

- coding pairs, Cantor's pairing function

$$\langle x, y\rangle := \frac{(x+y)^2 + x + 1}{2}$$

- Gödel's function

  $$\beta(a, i) := \min\{x < a\ ;\ \exists y < a \exists z < a(a = \langle y, z\rangle \wedge 1 + (\langle x, i\rangle + 1)|y\}.$$

  Given $a_1, \ldots, a_n$, there exists a such that $\beta(a, i) = a_i$, for $i = 1, \ldots, n$. This fact cannot be expressed in PA. We have to prove some properties of β. In particular:

  - the empty sequence has a code,
  - given a and b, one can extend the sequence a by adding b at the end.

□

- Alternatively, one can define bits of the numbers. One can define "x is a power of 2" by

$$\forall y(y|x \rightarrow (y = 1 \wedge 2|y)$$

etc.

Exercise

Define $x|y$ (x divides y).

Corollary

It is possible to formalize all standard syntactical concepts in PA.

Corollary

It is possible to formalize all standard syntactical concepts in PA.

So why do we use Peano Arithmetic?

Corollary

It is possible to formalize all standard syntactical concepts in PA.

So why do we use Peano Arithmetic?

- tradition
- linearly ordered models
- numerals for denoting elements
- hierarchies of arithmetical formulas
- fragments of PA defined by restricting the induction schema

yet, it is not so simple!

yet, it is not so simple!

We have to distinguish

1. concepts (in the metatheory)
2. formalized concepts (in the theory)
3. names representing formalized concepts (in the theory)

yet, it is not so simple!

We have to distinguish

1. concepts (in the metatheory)
2. formalized concepts (in the theory)
3. names representing formalized concepts (in the theory)

Example

Consider PA and numbers.

1. *in metatheory,* 0 *is* $\emptyset$, 1 *is* $\{0\}$, 2 *is* $\{0, 1\}$ *etc.*
2. *in PA, every element is a number*
3. *terms* 0, $S(0)$, $SS(0)$,... *are names for 0,1,2,...*

yet, it is not so simple!

We have to distinguish

1. concepts (in the metatheory)
2. formalized concepts (in the theory)
3. names representing formalized concepts (in the theory)

Example

Consider PA and numbers.

1. *in metatheory,* 0 *is* $\emptyset$, 1 *is* $\{0\}$, 2 *is* $\{0, 1\}$ *etc.*
2. *in PA, every element is a number*
3. *terms* 0, $S(0)$, $SS(0)$,... *are names for 0,1,2,...*

We call terms 0, $S(0)$, $SS(0)$,... numerals.

Formulas in PA.

1. in metatheory ϕ is a string of symbols
2. we assign numbers to symbols

Formulas in PA.

1. in metatheory ϕ is a string of symbols
2. we assign numbers to symbols
3. in PA, ϕ is a number that codes the string of numbers of the symbols, this is the Gödel number of ϕ

Formulas in PA.

1. in metatheory ϕ is a string of symbols
2. we assign numbers to symbols
3. in PA, ϕ is a number that codes the string of numbers of the symbols, this is the Gödel number of ϕ
4. if we need to talk about a concrete ϕ in PA, we use the numeral representing the Gödel number of ϕ

Formulas in PA.

1. in metatheory ϕ is a string of symbols
2. we assign numbers to symbols
3. in PA, ϕ is a number that codes the string of numbers of the symbols, this is the Gödel number of ϕ
4. if we need to talk about a concrete ϕ in PA, we use the numeral representing the Gödel number of ϕ

Notation The the numeral representing the Gödel number of ϕ will be denoted by

$$\ulcorner \phi \urcorner$$

.

Example

1. *let* ϕ *be* $x + 0 = x$
2. $x \mapsto 1, + \mapsto 2, = \mapsto 3$
3. *the Gödel number of* ϕ *is the number that encodes the sequence* $(1, 2, 3, 1)$, *say* 2500
4. $\ulcorner \phi \urcorner$ *is* $SS \ldots S0$ *with 2500 symbols* S.

Example

1. *let* ϕ *be* $x + 0 = x$
2. $x \mapsto 1, + \mapsto 2, = \mapsto 3$
3. *the Gödel number of* ϕ *is the number that encodes the sequence* $(1, 2, 3, 1)$, *say* 2500
4. $\ulcorner \phi \urcorner$ *is* $SS \ldots S0$ *with 2500 symbols* S.

Let ϕ and $\psi(x)$ be formulas. Then

- $\psi(\phi)$ is not a well-formed formula, but
- $\psi(\ulcorner \phi \urcorner)$ is, because $\ulcorner \phi \urcorner$ is a term.

self-reference

Lemma (diagonal, or fixed-point)

Let $\psi(x)$ be an arithmetical formula with one free variable. Then there exists a sentence ϕ such that

$$\mathsf{Q} \ \vdash \ \phi \equiv \psi(\ulcorner \phi \urcorner)$$

self-reference

Lemma (diagonal, or fixed-point)

Let $\psi(x)$ be an arithmetical formula with one free variable. Then there exists a sentence ϕ such that

$$\mathsf{Q} \;\vdash\; \phi \equiv \psi(\ulcorner \phi \urcorner)$$

ϕ says: "I have property ψ"

proof-idea

First attempt:

- The following formula has property ψ: *The following formula has property* ψ.

proof-idea

First attempt:

- The following formula has property ψ: *The following formula has property* ψ.

Not good, it only refers to its part.

proof-idea

First attempt:

- The following formula has property ψ: *The following formula has property ψ.*

Not good, it only refers to its part.

Second attempt:

- The following formula written twice has property ψ: *The following formula written twice has property ψ.*

proof-idea

First attempt:

- The following formula has property ψ: *The following formula has property* ψ.

Not good, it only refers to its part.

Second attempt:

- The following formula written twice has property ψ: *The following formula written twice has property* ψ.

Good! (except for : and .)

Proof

Consider the numerical function:

G. number of $\alpha(x) \mapsto$ G. number of $\alpha(\ulcorner\alpha(x)\urcorner)$

which is

$\ulcorner\alpha(x)\urcorner \mapsto \ulcorner\alpha(\ulcorner\alpha(x)\urcorner)\urcorner$

Proof

Consider the numerical function:

G. number of $\alpha(x) \mapsto$ G. number of $\alpha(\ulcorner\alpha(x)\urcorner)$

which is

$\ulcorner\alpha(x)\urcorner \mapsto \ulcorner\alpha(\ulcorner\alpha(x)\urcorner)\urcorner$

Suppose we have a term $t(y)$ that represents this function.

Proof

Consider the numerical function:

G. number of $\alpha(x) \mapsto$ G. number of $\alpha(\ulcorner \alpha(x) \urcorner)$

which is

$\ulcorner \alpha(x) \urcorner \mapsto \ulcorner \alpha(\ulcorner \alpha(x) \urcorner) \urcorner$

Suppose we have a term $t(y)$ that represents this function.

Define the fixed-point of $\psi(x)$ by

$$\phi \;:=\; \psi(t(\ulcorner \psi(t(x)) \urcorner))$$

Proof

Consider the numerical function:

G. number of $\alpha(x) \mapsto$ G. number of $\alpha(\ulcorner\alpha(x)\urcorner)$

which is

$\ulcorner\alpha(x)\urcorner \mapsto \ulcorner\alpha(\ulcorner\alpha(x)\urcorner)\urcorner$

Suppose we have a term $t(y)$ that represents this function.

Define the fixed-point of $\psi(x)$ by

$$\phi \; := \; \psi(t(\ulcorner\psi(t(x))\urcorner))$$

Note that

$$t(\ulcorner\psi(t(x))\urcorner) = \ulcorner\psi(t(\ulcorner\psi(t(x))\urcorner))\urcorner = \ulcorner\phi\urcorner$$

Thus

$$\phi \equiv \psi(\ulcorner\phi\urcorner)$$

$\square$

$$\psi(t(\ulcorner\psi(t(x))\urcorner))$$

"The following formula..."

- ψ — "has property ψ..."
- t — "if written twice:"
- $\ulcorner\psi(t(x))\urcorner$ — "$\psi(t(x))$."

1st incompleteness theorem

Theorem

Let T be a theory such that

1. *the set of axioms is r.e. (computably enumerable),*
2. *T extends $\mathbf{Q}$,*
3. *T is consistent.*

Then there exists a true sentence γ_T which is not provable in T.

1st incompleteness theorem

Theorem

Let T be a theory such that

1. *the set of axioms is r.e. (computably enumerable),*
2. *T extends $\mathbf{Q}$,*
3. *T is consistent.*

Then there exists a true sentence γ_T which is not provable in T.

Corollary

If moreover

4. *$N \models T$, i.e., T only proves true arithmetical sentences,*

then T is incomplete.

Proof

1. As T is r.e., there is a Σ_1 formula $Pr_T(x)$ that formalizes "x is provable in T".

Proof

1. As T is r.e., there is a Σ_1 formula $Pr_T(x)$ that formalizes "x is provable in T".

2. Since $\mathsf{Q} \subseteq T$, we can apply the diagonal lemma and get a formula γ_T such that

$$T \ \vdash \ \gamma_T \ \equiv \ \neg Pr_T(\ulcorner \gamma_T \urcorner).$$

Proof

1. As T is r.e., there is a Σ_1 formula $Pr_T(x)$ that formalizes "x is provable in T".

2. Since $\mathsf{Q} \subseteq T$, we can apply the diagonal lemma and get a formula γ_T such that

$$T \;\vdash\; \gamma_T \;\equiv\; \neg Pr_T(\ulcorner \gamma_T \urcorner).$$

3. Suppose that $T \vdash \gamma_T$. This means

$$\mathbb{N} \models Pr_T(\ulcorner \gamma_T \urcorner).$$

Since this is a Σ_1 formula and T is Σ_1 complete, we have

$$T \vdash Pr_T(\ulcorner \gamma_T \urcorner).$$

Proof

1. As T is r.e., there is a Σ_1 formula $Pr_T(x)$ that formalizes "x is provable in T".

2. Since $\mathsf{Q} \subseteq T$, we can apply the diagonal lemma and get a formula γ_T such that

$$T \vdash \gamma_T \equiv \neg Pr_T(\ulcorner \gamma_T \urcorner).$$

3. Suppose that $T \vdash \gamma_T$. This means

$$\mathbb{N} \models Pr_T(\ulcorner \gamma_T \urcorner).$$

Since this is a Σ_1 formula and T is Σ_1 complete, we have

$$T \vdash Pr_T(\ulcorner \gamma_T \urcorner).$$

But $T \vdash \gamma_T$ also means

$$T \vdash \neg Pr_T(\ulcorner \gamma_T \urcorner).$$

So T would be inconsistent. Hence $T \nvdash \gamma_T$.

4. We prove that $\mathbb{N} \models \gamma_T$ (i.e., γ_T is true).

4. We prove that $\mathbb{N} \models \gamma_T$ (i.e., γ_T is true).

We know that $T \not\vdash \gamma_T$, which means

$$\mathbb{N} \models \neg Pr_T(\ulcorner \gamma_T \urcorner).$$

But this is, by the definition of γ_T,

$$\mathbb{N} \models \gamma_T.$$

$\square$

Theorem

Let T be a theory such that

1. *the set of axioms is r.e. (computably enumerable),*
2. *T extends $\mathbf{Q}$,*
3. *T is consistent, and moreover*
4. *$Pr_T(x)$ is "properly formalized".*

Then

$$T \nvdash \neg Pr_T(\ulcorner 0 = 1 \urcorner),$$

i.e., T does not prove its own consistency.

proof–idea: formalize the 1st incompleteness theorem in T

proof–idea: formalize the 1st incompleteness theorem in T

Let's denote by $Con_T := \neg Pr_T(\ulcorner 0 = 1 \urcorner)$.

proof–idea: formalize the 1st incompleteness theorem in T

Let's denote by $Con_T := \neg Pr_T(\ulcorner 0 = 1 \urcorner)$.

We proved

- If T is consistent,
- then $T \nvdash \gamma_T$, which is $\mathbb{N} \models \neg Pr_T(\ulcorner \gamma_T \urcorner)$, which is also $\mathbb{N} \models \gamma_T$.

proof–idea: formalize the 1st incompleteness theorem in T

Let's denote by $Con_T := \neg Pr_T(\ulcorner 0 = 1 \urcorner)$.

We proved

- If T is consistent,
- then $T \nvdash \gamma_T$, which is $\mathbb{N} \models \neg Pr_T(\ulcorner \gamma_T \urcorner)$, which is also $\mathbb{N} \models \gamma_T$.

In other words

$$T \text{ consistent } \Rightarrow \gamma_T \text{ is true.}$$

proof–idea: formalize the 1st incompleteness theorem in T

Let's denote by $Con_T := \neg Pr_T(\ulcorner 0 = 1 \urcorner)$.

We proved

- If T is consistent,
- then $T \nvdash \gamma_T$, which is $\mathbb{N} \models \neg Pr_T(\ulcorner \gamma_T \urcorner)$, which is also $\mathbb{N} \models \gamma_T$.

In other words

$$T \text{ consistent } \Rightarrow \gamma_T \text{ is true.}$$

We will formalize this proof in T and get

$$T \vdash Con_T \rightarrow \gamma_T.$$

proof–idea: formalize the 1st incompleteness theorem in T

Let's denote by $Con_T := \neg Pr_T(\ulcorner 0 = 1 \urcorner)$.

We proved

- If T is consistent,
- then $T \nvdash \gamma_T$, which is $\mathbb{N} \models \neg Pr_T(\ulcorner \gamma_T \urcorner)$, which is also $\mathbb{N} \models \gamma_T$.

In other words

$$T \text{ consistent } \Rightarrow \gamma_T \text{ is true.}$$

We will formalize this proof in T and get

$$T \vdash Con_T \rightarrow \gamma_T.$$

Since $T \nvdash \gamma_T$, we also have $T \nvdash Con_T$.

$\square$

proof–idea: formalize the 1st incompleteness theorem in T

Let's denote by $Con_T := \neg Pr_T(\ulcorner 0 = 1 \urcorner)$.

We proved

- If T is consistent,
- then $T \nvdash \gamma_T$, which is $\mathbb{N} \models \neg Pr_T(\ulcorner \gamma_T \urcorner)$, which is also $\mathbb{N} \models \gamma_T$.

In other words

$$T \text{ consistent } \Rightarrow \gamma_T \text{ is true.}$$

We will formalize this proof in T and get

$$T \vdash Con_T \to \gamma_T.$$

Since $T \nvdash \gamma_T$, we also have $T \nvdash Con_T$.

$\square$

In fact $T \vdash \gamma_T \equiv Con_T$.

proper formalizations of provability in T

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner \phi \urcorner)$
   — $Pr_T(x)$ defines correctly provability in $\mathbb{N}$

proper formalizations of provability in T

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner \phi \urcorner)$
   — $Pr_T(x)$ defines correctly provability in $\mathbb{N}$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner \phi \urcorner)$
   — satisfied if $Pr_T(x)$ is a Σ_1 formula and T is Σ complete, the latter is satified if T contains Robinson's Q

proper formalizations of provability in T

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner\phi\urcorner)$
   — $Pr_T(x)$ defines correctly provability in $\mathbb{N}$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner\phi\urcorner)$
   — satisfied if $Pr_T(x)$ is a Σ_1 formula and T is Σ complete, the latter is satified if T contains Robinson's Q
3. $T \vdash Pr_T(\ulcorner\phi\urcorner) \rightarrow Pr_T(\ulcorner Pr_T(\ulcorner\phi\urcorner)\urcorner)$
   — means that T is able to prove that it is Σ-complete

proper formalizations of provability in T

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner\phi\urcorner)$
   — $Pr_T(x)$ defines correctly provability in $\mathbb{N}$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner\phi\urcorner)$
   — satisfied if $Pr_T(x)$ is a Σ_1 formula and T is Σ complete, the latter is satified if T contains Robinson's Q
3. $T \vdash Pr_T(\ulcorner\phi\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\phi\urcorner)\urcorner)$
   — means that T is able to prove that it is Σ-complete
4. $T \vdash Pr_T(\ulcorner\phi\urcorner) \wedge Pr_T(\ulcorner\phi \to \psi\urcorner) \to Pr_T(\ulcorner\psi\urcorner)$
   — means that provable formulas are closed under modus ponens

proper formalizations of provability in T

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\lceil\phi\rceil)$
   — $Pr_T(x)$ defines correctly provability in $\mathbb{N}$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\lceil\phi\rceil)$
   — satisfied if $Pr_T(x)$ is a Σ_1 formula and T is Σ complete, the latter is satified if T contains Robinson's Q
3. $T \vdash Pr_T(\lceil\phi\rceil) \to Pr_T(\lceil Pr_T(\lceil\phi\rceil)\rceil)$
   — means that T is able to prove that it is Σ-complete
4. $T \vdash Pr_T(\lceil\phi\rceil) \wedge Pr_T(\lceil\phi \to \psi\rceil) \to Pr_T(\lceil\psi\rceil)$
   — means that provable formulas are closed under modus ponens

Natural formalizations satisfy 1.-4. Exception: cut-free proofs. If T does not prove the cut-elimination theorem, then formalizations based on cut-free proofs do not satisfy 4.

proper formalizations of provability in T

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner \phi \urcorner)$
   — $Pr_T(x)$ defines correctly provability in $\mathbb{N}$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner \phi \urcorner)$
   — satisfied if $Pr_T(x)$ is a Σ_1 formula and T is Σ complete, the latter is satified if T contains Robinson's Q
3. $T \vdash Pr_T(\ulcorner \phi \urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner \phi \urcorner) \urcorner)$
   — means that T is able to prove that it is Σ-complete
4. $T \vdash Pr_T(\ulcorner \phi \urcorner) \wedge Pr_T(\ulcorner \phi \to \psi \urcorner) \to Pr_T(\ulcorner \psi \urcorner)$
   — means that provable formulas are closed under modus ponens

Natural formalizations satisfy 1.-4. Exception: cut-free proofs. If T does not prove the cut-elimination theorem, then formalizations based on cut-free proofs do not satisfy 4.

For 1st inco. thm. we only needed 1. and 2.

a wrong formalization

Let $Pr_T(x)$ be a natural formalization and define

$$Pr'(x) \equiv Pr_T(x) \wedge Con_T.$$

a wrong formalization

Let $Pr_T(x)$ be a natural formalization and define

$$Pr'(x) \equiv Pr_T(x) \wedge Con_T.$$

Then

$$T \vdash \neg Pr'(\ulcorner 0 = 1 \urcorner).$$

a wrong formalization

Let $Pr_T(x)$ be a natural formalization and define

$$Pr'(x) \equiv Pr_T(x) \wedge Con_T.$$

Then

$$T \vdash \neg Pr'(\ulcorner 0 = 1 \urcorner).$$

Indeed,

$$\neg Pr'(\ulcorner 0 = 1 \urcorner) \equiv (\neg Pr_T(\ulcorner 0 = 1 \urcorner) \vee \neg Con_T) \equiv \neg Pr_T(\ulcorner 0 = 1 \urcorner) \vee Pr_T(\ulcorner 0 = 1 \urcorner).$$

a wrong formalization

Let $Pr_T(x)$ be a natural formalization and define

$$Pr'(x) \equiv Pr_T(x) \wedge Con_T.$$

Then

$$T \vdash \neg Pr'(\lceil 0 = 1 \rceil).$$

Indeed,

$$\neg Pr'(\lceil 0 = 1 \rceil) \equiv (\neg Pr_T(\lceil 0 = 1 \rceil) \vee \neg Con_T) \equiv \neg Pr_T(\lceil 0 = 1 \rceil) \vee Pr_T(\lceil 0 = 1 \rceil).$$

Note that if T is consistent, then we do have 1.:

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr'(\lceil \phi \rceil)$ i.e., $Pr'(x)$ defines correctly provability in $\mathbb{N}$

because $\mathbb{N} \models Pr'(x) \equiv Pr_T(x)$.

a wrong formalization

Let $Pr_T(x)$ be a natural formalization and define

$$Pr'(x) \;\equiv\; Pr_T(x) \wedge Con_T.$$

Then

$$T \vdash \neg Pr'(\lceil 0 = 1 \rceil).$$

Indeed,

$$\neg Pr'(\lceil 0 = 1 \rceil) \;\equiv\; (\neg Pr_T(\lceil 0 = 1 \rceil) \vee \neg Con_T) \;\equiv\; \neg Pr_T(\lceil 0 = 1 \rceil) \vee Pr_T(\lceil 0 = 1 \rceil).$$

Note that if T is consistent, then we do have 1.:

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr'(\lceil \phi \rceil)$ i.e., $Pr'(x)$ defines correctly provability in $\mathbb{N}$

because $\mathbb{N} \models Pr'(x) \;\equiv\; Pr_T(x)$.

What is wrong?

a wrong formalization

Let $Pr_T(x)$ be a natural formalization and define

$$Pr'(x) \equiv Pr_T(x) \wedge Con_T.$$

Then

$$T \vdash \neg Pr'(\ulcorner 0 = 1 \urcorner).$$

Indeed,

$$\neg Pr'(\ulcorner 0 = 1 \urcorner) \equiv (\neg Pr_T(\ulcorner 0 = 1 \urcorner) \vee \neg Con_T) \equiv \neg Pr_T(\ulcorner 0 = 1 \urcorner) \vee Pr_T(\ulcorner 0 = 1 \urcorner).$$

Note that if T is consistent, then we do have 1.:

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr'(\ulcorner \phi \urcorner)$ i.e., $Pr'(x)$ defines correctly provability in $\mathbb{N}$

because $\mathbb{N} \models Pr'(x) \equiv Pr_T(x)$.

What is wrong? $Pr'(x)$ is not Σ_1 and does not satisfy 2; in fact T does not prove $Pr'_T(\ulcorner \phi \urcorner)$ for any formula ϕ.

Proof of 2nd incompleteness theorem

If we assume that

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner \phi \urcorner)$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner \phi \urcorner)$
3. $T \vdash Pr_T(\ulcorner \phi \urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner \phi \urcorner) \urcorner)$
4. $T \vdash Pr_T(\ulcorner \phi \urcorner) \wedge Pr_T(\ulcorner \phi \to \psi \urcorner) \to Pr_T(\ulcorner \psi \urcorner)$

then T proves:

Proof of 2nd incompleteness theorem

If we assume that

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner\phi\urcorner)$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner\phi\urcorner)$
3. $T \vdash Pr_T(\ulcorner\phi\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\phi\urcorner)\urcorner)$
4. $T \vdash Pr_T(\ulcorner\phi\urcorner) \wedge Pr_T(\ulcorner\phi \to \psi\urcorner) \to Pr_T(\ulcorner\psi\urcorner)$

then T proves:

i. $\neg\gamma \to Pr_T(\ulcorner\gamma\urcorner)$ – by definition of γ

Proof of 2nd incompleteness theorem

If we assume that

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner\phi\urcorner)$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner\phi\urcorner)$
3. $T \vdash Pr_T(\ulcorner\phi\urcorner) \rightarrow Pr_T(\ulcorner Pr_T(\ulcorner\phi\urcorner)\urcorner)$
4. $T \vdash Pr_T(\ulcorner\phi\urcorner) \wedge Pr_T(\ulcorner\phi \rightarrow \psi\urcorner) \rightarrow Pr_T(\ulcorner\psi\urcorner)$

then T proves:

i. $\neg\gamma \rightarrow Pr_T(\ulcorner\gamma\urcorner)$ – by definition of γ

ii. $Pr_T(\ulcorner\gamma\urcorner) \rightarrow Pr_T(\ulcorner Pr_T(\ulcorner\gamma\urcorner)\urcorner)$ – by 3.

Proof of 2nd incompleteness theorem

If we assume that

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner\phi\urcorner)$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner\phi\urcorner)$
3. $T \vdash Pr_T(\ulcorner\phi\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\phi\urcorner)\urcorner)$
4. $T \vdash Pr_T(\ulcorner\phi\urcorner) \wedge Pr_T(\ulcorner\phi \to \psi\urcorner) \to Pr_T(\ulcorner\psi\urcorner)$

then T proves:

i. $\neg\gamma \to Pr_T(\ulcorner\gamma\urcorner)$ – by definition of γ

ii. $Pr_T(\ulcorner\gamma\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\gamma\urcorner)\urcorner)$ – by 3.

iii. $Pr_T(\ulcorner Pr_T(\ulcorner\gamma\urcorner) \to \neg\gamma\urcorner)$ – by definition and 2.

Proof of 2nd incompleteness theorem

If we assume that

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner\phi\urcorner)$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner\phi\urcorner)$
3. $T \vdash Pr_T(\ulcorner\phi\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\phi\urcorner)\urcorner)$
4. $T \vdash Pr_T(\ulcorner\phi\urcorner) \wedge Pr_T(\ulcorner\phi \to \psi\urcorner) \to Pr_T(\ulcorner\psi\urcorner)$

then T proves:

i. $\neg\gamma \to Pr_T(\ulcorner\gamma\urcorner)$ – by definition of γ

ii. $Pr_T(\ulcorner\gamma\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\gamma\urcorner)\urcorner)$ – by 3.

iii. $Pr_T(\ulcorner Pr_T(\ulcorner\gamma\urcorner) \to \neg\gamma\urcorner)$ – by definition and 2.

iv. $Pr_T(\ulcorner\gamma\urcorner) \to Pr_T(\ulcorner\neg\gamma\urcorner)$ – by 4.

Proof of 2nd incompleteness theorem

If we assume that

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner\phi\urcorner)$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner\phi\urcorner)$
3. $T \vdash Pr_T(\ulcorner\phi\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\phi\urcorner)\urcorner)$
4. $T \vdash Pr_T(\ulcorner\phi\urcorner) \wedge Pr_T(\ulcorner\phi \to \psi\urcorner) \to Pr_T(\ulcorner\psi\urcorner)$

then T proves:

i. $\neg\gamma \to Pr_T(\ulcorner\gamma\urcorner)$ – by definition of γ

ii. $Pr_T(\ulcorner\gamma\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\gamma\urcorner)\urcorner)$ – by 3.

iii. $Pr_T(\ulcorner Pr_T(\ulcorner\gamma\urcorner) \to \neg\gamma\urcorner)$ – by definition and 2.

iv. $Pr_T(\ulcorner\gamma\urcorner) \to Pr_T(\ulcorner\neg\gamma\urcorner)$ – by 4.

v. $\neg\gamma \to Pr_T(\ulcorner\gamma\urcorner) \wedge Pr_T(\ulcorner\neg\gamma\urcorner)$ – from i. and iv.

Proof of 2nd incompleteness theorem

If we assume that

1. $T \vdash \phi \Leftrightarrow \mathbb{N} \models Pr_T(\ulcorner \phi \urcorner)$
2. $T \vdash \phi \Rightarrow T \vdash Pr_T(\ulcorner \phi \urcorner)$
3. $T \vdash Pr_T(\ulcorner \phi \urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner \phi \urcorner) \urcorner)$
4. $T \vdash Pr_T(\ulcorner \phi \urcorner) \wedge Pr_T(\ulcorner \phi \to \psi \urcorner) \to Pr_T(\ulcorner \psi \urcorner)$

then T proves:

i. $\neg\gamma \to Pr_T(\ulcorner \gamma \urcorner)$ – by definition of γ
ii. $Pr_T(\ulcorner \gamma \urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner \gamma \urcorner) \urcorner)$ – by 3.
iii. $Pr_T(\ulcorner Pr_T(\ulcorner \gamma \urcorner) \to \neg\gamma \urcorner)$ – by definition and 2.
iv. $Pr_T(\ulcorner \gamma \urcorner) \to Pr_T(\ulcorner \neg\gamma \urcorner)$ – by 4.
v. $\neg\gamma \to Pr_T(\ulcorner \gamma \urcorner) \wedge Pr_T(\ulcorner \neg\gamma \urcorner)$ – from i. and iv.
vi. $Con_T \to \gamma$ – from v.

$\square$

Exercise

1. *Check that this is a formalization of the proof of the 1st inco. thm.*
2. *Explain why vi. follows from v.*
3. *Prove that* $\gamma \to Con_T$.

Rossers's theorem

Let $\mathsf{Q} \subseteq T$ and T be consistent, e.g., $T := PA$.
Define $S := T + \neg Con_T$. Then

- $S \vdash \neg Con_S$.

Rossers's theorem

Let $\mathsf{Q} \subseteq T$ and T be consistent, e.g., $T := PA$.
Define $S := T + \neg Con_T$. Then

- $S \vdash \neg Con_S$.

By the 1st inco. thm.

- $S \nvdash Con_S$,

but we cannot conclude that S is incomplete!

Rossers's theorem

Let $\mathsf{Q} \subseteq T$ and T be consistent, e.g., $T := PA$.
Define $S := T + \neg Con_T$. Then

- $S \vdash \neg Con_S$.

By the 1st inco. thm.

- $S \nvdash Con_S$,

but we cannot conclude that S is incomplete!

Can we weaken the condition of soundness to consistency?

yes, we can

Theorem (Rosser)

Suppose

1. $\mathsf{Q} \subseteq T$,
2. T *is consistent,*
3. T *computably axiomatized (the axioms of* T *are a computably enumerable set)*

Then T *is incomplete.*

yes, we can

Theorem (Rosser)

Suppose

1. $\mathsf{Q} \subseteq T$,
2. T *is consistent,*
3. T *computably axiomatized (the axioms of* T *are a computably enumerable set)*

Then T *is incomplete.*

Example

PA is incomplete, because $PA \nvdash Con_{PA}$*, but* $PA + \neg Con_{PA}$ *is still incomplete.*

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation

Let $Prf(x, y)$ formalize *"x is a T-proof of y".*
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation
Let $Prf(x, y)$ formalize *"x is a T-proof of y"*.
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.
Define

$$\rho \ \equiv_{df} \forall x(Prf(x, \ulcorner\rho\urcorner) \rightarrow \exists y(y < x \wedge (Prf(y, \ulcorner\neg\rho\urcorner))$$

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation

Let $Prf(x, y)$ formalize *"x is a T-proof of y"*.
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.
Define

$$\rho \equiv_{df} \forall x(Prf(x, \ulcorner\rho\urcorner) \to \exists y(y < x \wedge (Prf(y, \ulcorner\neg\rho\urcorner))$$

1. Suppose $T \vdash \rho$. Then

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation

Let $Prf(x, y)$ formalize *"x is a T-proof of y"*.
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.
Define

$$\rho \ \equiv_{df} \forall x(Prf(x, \ulcorner \rho \urcorner) \rightarrow \exists y(y < x \wedge (Prf(y, \ulcorner \neg\rho \urcorner))$$

1. Suppose $T \vdash \rho$. Then

    1. $T \vdash Prf(\bar{n}, \ulcorner \rho \urcorner)$, where n is the G. number of the proof.

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation

Let $Prf(x, y)$ formalize *"x is a T-proof of y"*.
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.
Define

$$\rho \equiv_{df} \forall x(Prf(x, \ulcorner\rho\urcorner) \to \exists y(y < x \wedge (Prf(y, \ulcorner\neg\rho\urcorner))$$

1. Suppose $T \vdash \rho$. Then

   1. $T \vdash Prf(\bar{n}, \ulcorner\rho\urcorner)$, where n is the G. number of the proof.
   2. $T \vdash \exists y(y < \bar{n} \wedge Prf(\bar{y}, \ulcorner\neg\rho\urcorner)$

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation

Let $Prf(x, y)$ formalize *"x is a T-proof of y"*.
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.
Define

$$\rho \ \equiv_{df} \forall x(Prf(x, \ulcorner \rho \urcorner) \rightarrow \exists y(y < x \wedge (Prf(y, \ulcorner \neg\rho \urcorner))$$

1. Suppose $T \vdash \rho$. Then

   1. $T \vdash Prf(\bar{n}, \ulcorner \rho \urcorner)$, where n is the G. number of the proof.
   2. $T \vdash \exists y(y < \bar{n} \wedge Prf(\bar{y}, \ulcorner \neg\rho \urcorner)$
   3. but $T \vdash \neg Prf(\bar{m}, \ulcorner \neg\rho \urcorner)$ for all $m < n$, because T is consistent.

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation

Let $Prf(x, y)$ formalize *"x is a T-proof of y"*.
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.
Define

$$\rho \ \equiv_{df} \forall x(Prf(x, \ulcorner\rho\urcorner) \to \exists y(y < x \wedge (Prf(y, \ulcorner\neg\rho\urcorner))$$

1. Suppose $T \vdash \rho$. Then

   1. $T \vdash Prf(\bar{n}, \ulcorner\rho\urcorner)$, where n is the G. number of the proof.
   2. $T \vdash \exists y(y < \bar{n} \wedge Prf(\bar{y}, \ulcorner\neg\rho\urcorner)$
   3. but $T \vdash \neg Prf(\bar{m}, \ulcorner\neg\rho\urcorner)$ for all $m < n$, because T is consistent.
   4. using $T \vdash y < \bar{n} \equiv y = \bar{0} \vee \cdots \vee \overline{n-1}$, we get

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation

Let $Prf(x, y)$ formalize *"x is a T-proof of y"*.
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.
Define

$$\rho \equiv_{df} \forall x(Prf(x, \ulcorner \rho \urcorner) \rightarrow \exists y(y < x \wedge (Prf(y, \ulcorner \neg\rho \urcorner))$$

1. Suppose $T \vdash \rho$. Then

   1. $T \vdash Prf(\bar{n}, \ulcorner \rho \urcorner)$, where n is the G. number of the proof.
   2. $T \vdash \exists y(y < \bar{n} \wedge Prf(\bar{y}, \ulcorner \neg\rho \urcorner)$
   3. but $T \vdash \neg Prf(\bar{m}, \ulcorner \neg\rho \urcorner)$ for all $m < n$, because T is consistent.
   4. using $T \vdash y < \bar{n} \equiv y = \bar{0} \vee \cdots \vee \overline{n-1}$, we get
   5. $T \vdash \neg\exists y(y < \bar{n} \wedge Prf(y, \ulcorner \neg\rho \urcorner)$

Proof

Define a sentence ρ that says:

"For every proof of myself, there exists a shorter proof of my negation."

Notation

Let $Prf(x, y)$ formalize *"x is a T-proof of y"*.
Let $\bar{n}$ denote $S^n(0)$, the n-th numeral.
Define

$$\rho \equiv_{df} \forall x(Prf(x, \ulcorner\rho\urcorner) \to \exists y(y < x \wedge (Prf(y, \ulcorner\neg\rho\urcorner))$$

1. Suppose $T \vdash \rho$. Then

   1. $T \vdash Prf(\bar{n}, \ulcorner\rho\urcorner)$, where n is the G. number of the proof.
   2. $T \vdash \exists y(y < \bar{n} \wedge Prf(\bar{y}, \ulcorner\neg\rho\urcorner)$
   3. but $T \vdash \neg Prf(\bar{m}, \ulcorner\neg\rho\urcorner)$ for all $m < n$, because T is consistent.
   4. using $T \vdash y < \bar{n} \equiv y = \bar{0} \vee \cdots \vee \overline{n-1}$, we get
   5. $T \vdash \neg\exists y(y < \bar{n} \wedge Prf(y, \ulcorner\neg\rho\urcorner)$
   6. T is consistent is in contradiction with 2. and 5.

Thus $T \nvdash \rho$.

2. Suppose $T \vdash \neg\rho \quad (\equiv \exists x(Prf(x, \ulcorner\rho\urcorner) \wedge \forall y(y < x \wedge (\neg Prf(y, \ulcorner\neg\rho\urcorner))$
Then

2. Suppose $T \vdash \neg\rho \quad (\equiv \exists x(Prf(x, \ulcorner\rho\urcorner) \wedge \forall y(y < x \wedge (\neg Prf(y, \ulcorner\neg\rho\urcorner))$
Then

1. $T \vdash Prf(\bar{n}, \ulcorner\neg\rho\urcorner)$, where n is the G. number of the proof.

2. Suppose $T \vdash \neg\rho \quad (\equiv \exists x(Prf(x, \ulcorner\rho\urcorner) \wedge \forall y(y < x \wedge (\neg Prf(y, \ulcorner\neg\rho\urcorner))$
Then

1. $T \vdash Prf(\bar{n}, \ulcorner\neg\rho\urcorner)$, where n is the G. number of the proof.
2. $T \vdash \exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from $\neg\rho$ and 1.

2. Suppose $T \vdash \neg\rho \quad (\equiv \exists x(Prf(x, \ulcorner\rho\urcorner) \wedge \forall y(y < x \wedge (\neg Prf(y, \ulcorner\neg\rho\urcorner))$
Then

1. $T \vdash Prf(\bar{n}, \ulcorner\neg\rho\urcorner)$, where n is the G. number of the proof.
2. $T \vdash \exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from $\neg\rho$ and 1.
3. $\mathbb{N} \models \neg\exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from consistency of T and $T \vdash \neg\rho$.

2. Suppose $T \vdash \neg\rho \quad (\equiv \exists x(Prf(x, \ulcorner\rho\urcorner) \wedge \forall y(y < x \wedge (\neg Prf(y, \ulcorner\neg\rho\urcorner))$
Then

1. $T \vdash Prf(\bar{n}, \ulcorner\neg\rho\urcorner)$, where n is the G. number of the proof.
2. $T \vdash \exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from $\neg\rho$ and 1.
3. $\mathbb{N} \models \neg\exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from consistency of T and $T \vdash \neg\rho$.
4. $T \vdash \neg\exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from 3. by Σ-completeness.

2. Suppose $T \vdash \neg\rho \quad (\equiv \exists x(Prf(x, \ulcorner\rho\urcorner) \wedge \forall y(y < x \wedge (\neg Prf(y, \ulcorner\neg\rho\urcorner))$
Then

1. $T \vdash Prf(\bar{n}, \ulcorner\neg\rho\urcorner)$, where n is the G. number of the proof.
2. $T \vdash \exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from $\neg\rho$ and 1.
3. $\mathbb{N} \models \neg\exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from consistency of T and $T \vdash \neg\rho$.
4. $T \vdash \neg\exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from 3. by Σ-completeness.
5. T is consistent is in contradiction with 2. and 4.

Thus $T \nvdash \neg\rho$.

$\square$

2. Suppose $T \vdash \neg\rho \quad (\equiv \exists x(Prf(x, \ulcorner\rho\urcorner) \wedge \forall y(y < x \wedge (\neg Prf(y, \ulcorner\neg\rho\urcorner)))$
Then

1. $T \vdash Prf(\bar{n}, \ulcorner\neg\rho\urcorner)$, where n is the G. number of the proof.
2. $T \vdash \exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from $\neg\rho$ and 1.
3. $\mathbb{N} \models \neg\exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from consistency of T and $T \vdash \neg\rho$.
4. $T \vdash \neg\exists x \leq \bar{n} \;\; Prf(x, \ulcorner\rho\urcorner)$, from 3. by Σ-completeness.
5. T is consistent is in contradiction with 2. and 4.

Thus $T \nvdash \neg\rho$.

$\square$

We know that $T \vdash \gamma_T \equiv Con_T$. What about ρ_T?

Exercise

1. *Prove* $T \vdash Con_T \rightarrow \rho_T$.
2. *Prove* $T \nvdash \rho_T \rightarrow Con_T$.

unpredictable algorithms

Theorem

Let $T \supseteq \mathsf{Q}$ be consistent and computably axiomatizable. Then one can write a program P_T such that for every $n \in \mathbb{N}$,

$$T + P_T \text{ outputs } \bar{n}$$

is consistent.

T is not able to predict the output of P_T.

Proof.

Define P_T using the fixpoint lemma so that P_T systematically searches all T-proofs until it finds a T-proof of $\neg(P_T := \bar{n})$; then it prints n.

unpredictable algorithms

Theorem

Let $T \supseteq \mathrm{Q}$ be consistent and computably axiomatizable. Then one can write a program P_T such that for every $n \in \mathbb{N}$,

$$T + P_T \textit{ outputs } \bar{n}$$

is consistent.

T is not able to predict the output of P_T.

Proof.

Define P_T using the fixpoint lemma so that P_T systematically searches all T-proofs until it finds a T-proof of $\neg(P_T := \bar{n})$; then it prints n.

$T + P_T := \bar{n}$ is consistent, because if $T \vdash \neg(P_T := \bar{n})$, then $P_T := n$ and, by Σ-completeness, $T \vdash P_T := \bar{n}$; so T would be inconsistent. $\square$

unpredictable algorithms

Theorem

Let $T \supseteq \mathsf{Q}$ be consistent and computably axiomatizable. Then one can write a program P_T such that for every $n \in \mathbb{N}$,

$$T + P_T \text{ outputs } \bar{n}$$

is consistent.

T is not able to predict the output of P_T.

Proof.

Define P_T using the fixpoint lemma so that P_T systematically searches all T-proofs until it finds a T-proof of $\neg(P_T := \bar{n})$; then it prints n.

$T + P_T := \bar{n}$ is consistent, because if $T \vdash \neg(P_T := \bar{n})$, then $P_T := n$ and, by Σ-completeness, $T \vdash P_T := \bar{n}$; so T would be inconsistent. $\square$

Exercise

What does the program output?

flexible formula

Theorem (A. Mostowski)

Let $T \supseteq \mathsf{Q}$ be consistent and computably axiomatizable. Then there exists a formula $\phi(x)$ such that for every set $S \subseteq \mathbb{N}$

$$T \cup \{\phi(\bar{n}) \mid n \in S\} \cup \{\neg\phi(\bar{n}) \mid n \in \mathbb{N} \setminus S\}$$

is consistent.

flexible formula

Theorem (A. Mostowski)

Let $T \supseteq \mathrm{Q}$ be consistent and computably axiomatizable. Then there exists a formula $\phi(x)$ such that for every set $S \subseteq \mathbb{N}$

$$T \cup \{\phi(\bar{n}) \mid n \in S\} \cup \{\neg\phi(\bar{n}) \mid n \in \mathbb{N} \setminus S\}$$

is consistent.

Proof.

By compactness, it suffices to prove for every $m \in \mathbb{N}$ and every $S \subseteq [0, m]$,

$$T \cup \{\phi(\bar{n}) \mid n \in S\} \cup \{\neg\phi(\bar{n}) \mid n \in [0, m] \setminus S\}$$

is consistent.

flexible formula

Theorem (A. Mostowski)

Let $T \supseteq \mathsf{Q}$ be consistent and computably axiomatizable. Then there exists a formula $\phi(x)$ such that for every set $S \subseteq \mathbb{N}$

$$T \cup \{\phi(\bar{n}) \mid n \in S\} \cup \{\neg\phi(\bar{n}) \mid n \in \mathbb{N} \setminus S\}$$

is consistent.

Proof.

By compactness, it suffices to prove for every $m \in \mathbb{N}$ and every $S \subseteq [0, m]$,

$$T \cup \{\phi(\bar{n}) \mid n \in S\} \cup \{\neg\phi(\bar{n}) \mid n \in [0, m] \setminus S\}$$

is consistent.

Let P_T be an unpredictable algorithm in T. Define $\phi(x)$:

- $\exists y(P_T := y \wedge (y)_x = 1)$ (think of y as a code of $(y_0, \ldots, y_m)$)

flexible formula

Theorem (A. Mostowski)

Let $T \supseteq \mathsf{Q}$ be consistent and computably axiomatizable. Then there exists a formula $\phi(x)$ such that for every set $S \subseteq \mathbb{N}$

$$T \cup \{\phi(\bar{n}) \mid n \in S\} \cup \{\neg\phi(\bar{n}) \mid n \in \mathbb{N} \setminus S\}$$

is consistent.

Proof.

By compactness, it suffices to prove for every $m \in \mathbb{N}$ and every $S \subseteq [0, m]$,

$$T \cup \{\phi(\bar{n}) \mid n \in S\} \cup \{\neg\phi(\bar{n}) \mid n \in [0, m] \setminus S\}$$

is consistent.

Let P_T be an unpredictable algorithm in T. Define $\phi(x)$:

- $\exists y(P_T := y \wedge (y)_x = 1)$ (think of y as a code of $(y_0, \ldots, y_m)$)

By the previous theorem, it is consistent that P_T prints an arbitrary string of 0s and 1s. □

Exercise

1. *Generalize the fixpoint lemma to formulas with one free variable:*
$$\mathsf{Q} \ \vdash \ \phi(x) \equiv \psi(\ulcorner \phi(\bar{x}) \urcorner, x),$$
where $\ulcorner \phi(\bar{x}) \urcorner$ *denotes a formalization of the function that given a number* n*, constructs a godel number of* $\phi(\bar{n})$*.*
2. *Construct a formula* $\psi(x)$ *such that for every* n*,*
$$T + \psi(\bar{n}) \wedge \forall x < \bar{n} \ \neg\psi(x)$$
is consistent.
3. *construct a flexible formula from* ψ*.*

Kolmogorov complexity and incompleteness

Definition

U is a universal Turing machine if for every Turing machine M there exists a string p (program) such that for all x, $M(x) = U(px)$.

All strings are binary.

px is the concatenation of p and x.

$M(x) = U(px)$ means that M stops iff U stops and if they stop, they print the same string.

Kolmogorov complexity and incompleteness

Definition

U is a universal Turing machine if for every Turing machine M there exists a string p (program) such that for all x, $M(x) = U(px)$.

All strings are binary.

px is the concatenation of p and x.

$M(x) = U(px)$ means that M stops iff U stops and if they stop, they print the same string.

Definition

The Kolmogorov complexity of a string x (w.r.t. to U), $K_U(x)$, is the length of the shortest string p such that $U(p) = x$.

basic facts

- $\exists c \forall x \; K_U(x) \leq |x| + c$
- For U, U' universal Turing machines, there exists c such that for all x

$$K_U(x) \leq K_{U'}(x) + c \text{ and } K_{U'}(x) \leq K_U(x) + c$$

- $K_U(x)$ is not computable.

basic facts

- $\exists c \forall x\ K_U(x) \leq |x| + c$
- For U, U' universal Turing machines, there exists c such that for all x

$$K_U(x) \leq K_{U'}(x) + c \text{ and } K_{U'}(x) \leq K_U(x) + c$$

- $K_U(x)$ is not computable.

Proposition (and Definition)

For every n there exists a string x, $|x| = n$ such that $K_U(x) \geq n$. Such a string is called Kolmogorov random or incompressible.

Proof.

The number of Kolmogorov non-random strings of length n is

$$\leq 1 + 2 + 4 + \cdots + 2^{n-1} < 2^n.$$

$\square$

Theorem (Chaitin)

For every theory T, $T \supseteq \mathsf{Q}$, sound,[1] and computably axiomatizable, there exists a number k_T such that for no string a, T proves $K_U(\bar{a}) > \bar{k}_T$.[2]

[1] Proves only true arithmetical sentences.

[2] Bars are used to represent strings and numbers by numerals in theory T.

Theorem (Chaitin)

For every theory T, $T \supseteq \mathrm{Q}$, sound,[1] and computably axiomatizable, there exists a number k_T such that for no string a, T proves $K_U(\bar{a}) > \bar{k}_T$.[2]

Proof.

Let M be a Turing machine that on input k, a number in binary, systematically checks all strings and

- if it finds a T-proof of $K_U(\bar{a}) > \bar{k}$ for some a, it prints a,
- otherwise it does not stop.

[1] Proves only true arithmetical sentences.

[2] Bars are used to represent strings and numbers by numerals in theory T.

Theorem (Chaitin)

For every theory T, $T \supseteq \mathrm{Q}$, sound,[1] *and computably axiomatizable, there exists a number k_T such that for no string a, T proves $K_U(\bar{a}) > \bar{k}_T$.*[2]

Proof.

Let M be a Turing machine that on input k, a number in binary, systematically checks all strings and

- if it finds a T-proof of $K_U(\bar{a}) > \bar{k}$ for some a, it prints a,
- otherwise it does not stop.

Let Π be the first T-proof of a sentence of the form $K_U(\bar{a}) > \bar{k}$. Hence

$$K_U(a) \leq C + \log_2 k$$

for some constant C. Since T is sound, such a proof does not exist if

$$C + \log_2 k \leq k.$$

Take (any/least) k_T that satisfies this inequality. $\square$

[1] Proves only true arithmetical sentences.

[2] Bars are used to represent strings and numbers by numerals in theory T.

Corollary

The first incompleteness theorem.

with a proof that does not use self-reference.

Corollary

The first incompleteness theorem.

with a proof that does not use self-reference.

But are such independent sentences *interesting?*

Corollary

The first incompleteness theorem.

with a proof that does not use self-reference.

But are such independent sentences *interesting?*

If we take k_T the least number that satisfies Chaitin's Theorem, we get a measure of the strength of theory T.

Corollary

The first incompleteness theorem.

with a proof that does not use self-reference.

But are such independent sentences *interesting?*

If we take k_T the least number that satisfies Chaitin's Theorem, we get a measure of the strength of theory T.

Clearly, if $Thm(S) \subseteq Thm(T)$ then $k_S \leq k_T$, regardless the complexity of the axiomatizations of S and T.

Corollary

The first incompleteness theorem.

with a proof that does not use self-reference.

But are such independent sentences *interesting?*

If we take k_T the least number that satisfies Chaitin's Theorem, we get a measure of the strength of theory T.

Clearly, if $Thm(S) \subseteq Thm(T)$ then $k_S \leq k_T$, regardless the complexity of the axiomatizations of S and T.

But it is almost impossible to determine k_T. Moreover, it depends on U.

Corollary

The first incompleteness theorem.

with a proof that does not use self-reference.

But are such independent sentences *interesting?*

If we take k_T the least number that satisfies Chaitin's Theorem, we get a measure of the strength of theory T.

Clearly, if $Thm(S) \subseteq Thm(T)$ then $k_S \leq k_T$, regardless the complexity of the axiomatizations of S and T.

But it is almost impossible to determine k_T. Moreover, it depends on U.

Paradox A sufficiently strong T can prove that there exist Kolmogorov random strings for every n, but for large enough n, it is unable to prove it for any concrete string.

2nd incompleteness theorem using Kolmogorov complexity

Proposition

Let $T \supseteq \mathrm{Q}$ be sound and computably axiomatizable. Then for every $n > k_T$, if

$$T \vdash \exists \geq \bar{M} \text{ Kolmogorov random strings of length } \bar{n},$$

then, in fact, there are $> M$ Kolmogorov random strings of length n.

Proof.

Let M be given and let N be the actual number of Kolmogorov random strings of length n.

2nd incompleteness theorem using Kolmogorov complexity

Proposition

Let $T \supseteq \mathrm{Q}$ be sound and computably axiomatizable. Then for every $n > k_T$, if

$$T \vdash \exists \geq \bar{M} \text{ Kolmogorov random strings of length } \bar{n},$$

then, in fact, there are $> M$ Kolmogorov random strings of length n.

Proof.

Let M be given and let N be the actual number of Kolmogorov random strings of length n.

1. $M > N$ is impossible: if $K(w) < n$, then $T \vdash K(\bar{w}) \leq \bar{n}$ (by Σ-completeness), whence T proves that $M \leq N$.

2nd incompleteness theorem using Kolmogorov complexity

Proposition

Let $T \supseteq \mathsf{Q}$ be sound and computably axiomatizable. Then for every $n > k_T$, if

$$T \vdash \exists \geq \bar{M} \text{ Kolmogorov random strings of length } \bar{n},$$

then, in fact, there are $> M$ Kolmogorov random strings of length n.

Proof.

Let M be given and let N be the actual number of Kolmogorov random strings of length n.

1. $M > N$ is impossible: if $K(w) < n$, then $T \vdash K(\bar{w}) \leq \bar{n}$ (by Σ-completeness), whence T proves that $M \leq N$.
2. $M = N$ is impossible: For all w such that $K(w) < n$, T can prove that they are not K. random, so T knows that any of the remaining M must be K. random—contradiction with $k_T < n$.

2nd incompleteness theorem using Kolmogorov complexity

Proposition

Let $T \supseteq \mathsf{Q}$ be sound and computably axiomatizable. Then for every $n > k_T$, if

$$T \vdash \exists \geq \bar{M} \text{ Kolmogorov random strings of length } \bar{n},$$

then, in fact, there are $> M$ Kolmogorov random strings of length n.

Proof.

Let M be given and let N be the actual number of Kolmogorov random strings of length n.

1. $M > N$ is impossible: if $K(w) < n$, then $T \vdash K(\bar{w}) \leq \bar{n}$ (by Σ-completeness), whence T proves that $M \leq N$.
2. $M = N$ is impossible: For all w such that $K(w) < n$, T can prove that they are not K. random, so T knows that any of the remaining M must be K. random—contradiction with $k_T < n$.
3. $M < N$ is the only remaining possibility.

Proof of 2nd inco. thm.

Suppose that T is sufficiently strong and proves its own consistency. Then

- T proves that for every n there exists at least one K. random string and not all strings are K. random.

Proof of 2nd inco. thm.

Suppose that T is sufficiently strong and proves its own consistency. Then

- T proves that for every n there exists at least one K. random string and not all strings are K. random.
- T can formalize the argument of the Proposition and we get: If

  $$T \vdash \exists \geq \bar{M} \text{ Kolmogorov random strings of length } \bar{n},$$

  then

  $$T \vdash \exists \geq \overline{M+1} \text{ Kolmogorov random strings of length } \bar{n}.$$

Proof of 2nd inco. thm.

Suppose that T is sufficiently strong and proves its own consistency. Then

- T proves that for every n there exists at least one K. random string and not all strings are K. random.
- T can formalize the argument of the Proposition and we get: If

$$T \vdash \exists \geq \bar{M} \text{ Kolmogorov random strings of length } \bar{n},$$

then

$$T \vdash \exists \geq \overline{M+1} \text{ Kolmogorov random strings of length } \bar{n}.$$

- Thus T would prove that all strings of length n are K. random $\Rightarrow$ T is not consistent.

$\square$

Exercise

Where did we use the assumption that T proves its own consistency?

Lesson 9, Peano Arithmetic and Bounded Arithmetic

see Chapter 2, by Buss