# Gödel's Incompleteness Theorems

Reinhard Kahle

CMA & Departamento de Matemática
FCT, Universidade Nova de Lisboa

Hilbert Bernays Summer School 2015
Göttingen

Partially funded by FCT project PTDC/MHC-FIL/5363/2012 and FCT project UID/MAT/00297/2013.



## References

Jon Barwise.
An introduction to first-order logic.
In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 5–46. North-Holland, 1977.

C. Smorynski.
The incompleteness theorems.
In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 821–865. North-Holland, 1977.

---

R. Kahle and W. Keller.
Syntax versus Semantics.
In M. Antonia Huertas et al., editors, *4th International Conference on Tools for Teaching Logic*, pages 75–84. University of Rennes 1, 2015.

# First-order languages

## Definition

A *first-order language* $\mathcal{L}$ is a set of symbols which can be divided in the following six (disjunctive) subsets:

- logical symbols: $\{\neg, \wedge, \vee, \rightarrow, \forall, \exists, =\}$;
- constant symbols: $\mathcal{C} \subseteq \{c_i | i \in \mathbb{N}\}$,

  examples: $c_0 = 0$, $c_1 = 1$, $c_2 = \pi$.
- function symbols: $\mathcal{F} \subseteq \{f_i^j | i \in \mathbb{N}, j \in \mathbb{N}, j > 0\}$, where $f_i^j$ is the $i$-th function symbol of arity $j$;

  examples: $f_0^2 = +$, $f_1^2 = \cdot$, $f_2^2 = -$, $f_0^1 = -$ (change of sign).
- relation symbols $\mathcal{R} \subseteq \{R_i^j | i \in \mathbb{N}, j \in \mathbb{N}\}$, where $R_i^j$ is the $i$-th relation symbol of arity $j$;

  examples: $R_0^2 = <$, $R_1^2 = >$, $R_0^3 = \cdot \equiv \cdot \bmod \cdot$, $R_0^1 = \mathrm{Prim}(\cdot)$.
- variables: $\{x, y, z, w, \ldots, x_0, x_1, x_2, \ldots\}$;
- auxiliary signs: $\{$"(", ")", ",", "."$\}$.



# First-order languages

According to the definition, for a concrete first-order language we have only to specify only the sets $\mathcal{C}$, $\mathcal{F}$, and $\mathcal{R}$.

## Examples

1. For the language $\mathcal{L}_{\mathrm{PA}}$ of the *Peano arithmetic* we have: $\mathcal{C} = \{0\}$, $\mathcal{F} = \{s, +, \cdot\}$, and $\mathcal{R} = \emptyset$, where $s$ is a unary function symbol for the successor function.
2. The language of *set theory* (without urelements) can be given by $\mathcal{C} = \mathcal{F} = \emptyset$ and $\mathcal{R} = \{\in\}$.

# Terms

## Definition

The *terms* of $\mathcal{L}$ are defined *inductively* as following:

1. Each variable is a term.
2. Each constant symbol is a term.
3. If $t_1, t_2, \ldots, t_n$ are terms and $f^n$ is a $n$-ary function symbol ($n > 0$), then the expression $f^n(t_1, t_2, \ldots, t_n)$ is also a term.



# Formulae

## Definition

The *formulae* of $\mathcal{L}$ are defined inductively as follows:

1. If $t_1$ and $t_2$ are terms, then the expression $t_1 = t_2$ is a formula.
2. If $t_1, t_2, \ldots, t_n$ are terms and $R^n$ is a $n$-ary relation symbol ($n \geq 0$), then the expression $R^n(t_1, t_2, \ldots, t_n)$ is a formula.
3. If $\varphi$ and $\psi$ are formulae, then the following expressions are also formulae:
$$(\neg\varphi),\ (\varphi \wedge \psi),\ (\varphi \vee \psi),\ (\varphi \rightarrow \psi).$$
4. If $\varphi$ is a formula and $x$ a variable, then the expressions $(\forall x.\varphi)$ and $(\exists x.\varphi)$ are also formulae.

The formulas, constructed according 1 and 2 are also called *atomic formulae*.

# Free variables; sentences

## Definition

The set $FV(\varphi)$ of the *free variables* of a formula $\varphi$ is *recursively* defined as follows:

1. If $\varphi$ is an atomic formula, then $FV(\varphi)$ is the set of variables which occur in $\varphi$;
2. $FV(\neg\varphi) = FV(\varphi)$;
3. $FV(\varphi \wedge \psi) = FV(\varphi \vee \psi) = FV(\varphi \rightarrow \psi) = FV(\varphi) \cup FV(\psi)$;
4. $FV(\exists x.\varphi) = FV(\forall x.\varphi) = FV(\varphi) \setminus \{x\}$.

A *(first-order) sentence* of the language $\mathcal{L}$ is a formula $\varphi$ without free variables, i.e., $FV(\varphi) = \emptyset$.



# Semantics

- So far, we only considered finite sequences of symbols which we call *terms* or *formulae*; among the *formulae* we distinguished, in particular, the *sentences*.
- Up to this point, these sequences of symbols have to be considered as "meaningless".
- In the following, we will describe how one can relates a meaning *in the usual mathematical sense* to these sequences of symbols.

## Structure

### Definition

An $\mathcal{L}$-*structure* is a pair $\mathfrak{M} = \langle M, F\rangle$, with $M$ a non-empty set and $F$ a function whose domain consists of the constants symbols, function symbols, and relation symbols of $\mathcal{L}$ such that:

1. If $c \in \mathcal{C}$, then $F(c) \in M$.
2. If $f^j \in \mathcal{F}$, with $j > 0$, then $F(f^j) : M^j \longrightarrow M$, i.e., a $j$-ary function from $M^j$ to $M$.
3. If $R^0 \in \mathcal{R}$, then $F(R^0)$ is one of the two truth values $\mathsf{t}$ (true) or $\mathsf{f}$ (false).
4. If $R^j \in \mathcal{R}$, with $j > 0$, then $F(R^j) \subseteq M^j$.

In the following, we will write, in general, $I^{\mathfrak{M}}$ instead of $F(I)$, $I \in \mathcal{R} \cup \mathcal{F} \cup \mathcal{C}$. We also give a structure for languages, for which we use only finitely many constant symbols, function symbols, and relation symbols, by the tuple $\langle M, {c_1}^{\mathfrak{M}}, \ldots, {c_n}^{\mathfrak{M}}, {f_1}^{\mathfrak{M}}, \ldots, {f_k}^{\mathfrak{M}}, {R_1}^{\mathfrak{M}}, \ldots, {R_l}^{\mathfrak{M}}\rangle$.



## The structure of the natural numbers

### Example

For the language of the Peano arithmetik $\mathcal{L}_{\mathrm{PA}}$, we can define the *structure of the natural numbers* by $\mathcal{N} = \langle \mathbb{N}, 0, \_ + 1, +, \cdot\rangle$.
Notice that the *functions* are usual mathematical (set-theoretical) objects. For example, $+$ is the (infinite) set

$$\begin{array}{l} \{(0,0,0),(0,1,1),(0,2,2),(0,3,3),\ldots \\ \ (1,0,1),(1,1,2),(1,2,3),(1,3,4),\ldots \\ \ (2,0,2),(2,1,3),(2,2,4),(2,3,5),\ldots \\ \qquad\qquad \vdots \qquad\qquad\qquad\qquad\qquad\qquad \} \end{array}$$

In other words, $+$ is the subset of $\mathbb{N}^3$ consisting of the triples $(x, y, z)$ with $x + y = z$.

# Assignment

## Definition

An *assignment* in $\mathfrak{M}$ is a function $s$, which has as domain the variables of $\mathcal{L}$ and as range a subset of $M$.

## Definition

Let $\mathcal{L}$ and $\mathfrak{M}$ be given and let $s$ be an assignment in $\mathfrak{M}$. We define $(t)^{\mathfrak{M}}(s)$ recursively for every term $t$ of $\mathcal{L}$:

1. If $t$ is a variable $x$ then $(x)^{\mathfrak{M}}(s) = s(x)$.
2. If $t$ is a constant symbol $c$, then $(c)^{\mathfrak{M}}(s) = (c)^{\mathfrak{M}}$.
3. If $t$ is a term of the form $f^j(t_1, \ldots, t_j)$, then

$$(t)^{\mathfrak{M}}(s) = (f^j)^{\mathfrak{M}}((t_1)^{\mathfrak{M}}(s), \ldots, (t_j)^{\mathfrak{M}}(s)).$$



# Modified assignment

For the following definition we need the possibility to *modify* assignments (i.e., a function from variables to elements of $M$).

Given an assignment $s$ and an element $a \in M$, we designate by $s(\substack{a \\ x})$ the assignment which coincides with $s$ for all variables except $x$; independently of the value of $s(x)$, we fix $s(\substack{a \\ x})(x) = a$. More exactly:

$$s(\substack{a \\ x})(y) = \begin{cases} s(y), & \text{if } y \text{ is a variable different from } x, \\ a, & \text{if } y \text{ is the variable } x. \end{cases}$$

## Definition

Let $\mathfrak{M}$ be a $\mathcal{L}$ structure. We define, for every assignment $s$ and every formula $\varphi$ the relation $\mathfrak{M} \models \varphi[s]$:

1. $\mathfrak{M} \models (t_1 = t_2)[s]$ if and only if $t_1^{\mathfrak{M}}(s) = t_2^{\mathfrak{M}}(s)$,
2. $\mathfrak{M} \models R_i^0[s]$ if and only if $(R_i^0)^{\mathfrak{M}} = \mathrm{t}$,
3. $\mathfrak{M} \models R_i^j(t_1, \ldots, t_j)[s]$, $j > 0$, if and only if $$(t_1^{\mathfrak{M}}(s), \ldots, t_j^{\mathfrak{M}}(s)) \in (R_i^j)^{\mathfrak{M}},$$
4. $\mathfrak{M} \models (\neg\varphi)[s]$ if and only if **it is not the case that** $\mathfrak{M} \models \varphi[s]$,
5. $\mathfrak{M} \models (\varphi \wedge \psi)[s]$ if and only if $\mathfrak{M} \models \varphi[s]$ **and** $\mathfrak{M} \models \psi[s]$,
6. $\mathfrak{M} \models (\varphi \vee \psi)[s]$ if and only if $\mathfrak{M} \models \varphi[s]$ **or** $\mathfrak{M} \models \psi[s]$,
7. $\mathfrak{M} \models (\varphi \rightarrow \psi)[s]$ if and only if, it is not the case that $\mathfrak{M} \models \varphi[s]$ or it is the case that $\mathfrak{M} \models \psi[s]$,
8. $\mathfrak{M} \models (\exists x.\varphi)[s]$ if and only if **there exists an element** $a \in M$, such that $\mathfrak{M} \models \varphi[s(\substack{a\\x})]$,
9. $\mathfrak{M} \models (\forall x.\varphi)[s]$ if and only if **for all elements** $a \in M$ it holds that $\mathfrak{M} \models \varphi[s(\substack{a\\x})]$.



# Semantic consequence

The assignment $s$ is necessary to assign elements of $M$ to the *free* variables of a formula. For *sentences* $\varphi$ (i.e., formulas without free variables) $s$ does not matter and can be surpressed in the relation $\mathfrak{M} \models \varphi[s]$:

## Definition

Let $\Phi$ be a set of $\mathcal{L}$-sentences and $\mathfrak{M}$ be a $\mathcal{L}$ structure. $\mathfrak{M}$ is a *model* of $\Phi$, written as $\mathfrak{M} \models \Phi$, if for every sentence $\varphi \in \Phi$ we have $\mathfrak{M} \models \varphi$.
*Semantic consequence* is now defined as follows:
For a sentence $\psi$ we say that it *follows (semantically) from* $\Phi$, written as $\Phi \models \psi$, if for every model $\mathfrak{M}$ of $\Phi$ it holds that $\mathfrak{M} \models \psi$.

If $\mathfrak{M} \models \varphi$ holds, we also say that $\varphi$ *is true in* $\mathfrak{M}$.
If $\mathfrak{M} \models \varphi$ holds for every structure $\mathfrak{M}$, we also write $\models \varphi$.

## Compactness Theorem

**Theorem (Compactness Theorem)**

Let $\Phi$ be a set of first-order sentences.

If every finite subset $\Phi_0$ of $\Phi$ has a model, then there exists also a model of $\Phi$.

Alternative formulation:

**Theorem (Compactness Theorem)**

Let $\Phi \cup \{\varphi\}$ be a set of first-order sentences.

If $\Phi \models \varphi$, then there exists already a finite subset $\Phi_0$ of $\Phi$, such that $\Phi_0 \models \varphi$.



## Hilbert-style calculus I

**Definition**

We define the *Hilbert-style calculus* **H** as a derivation system with the following (logical) axioms and rules:

1. The following formulae are axioms:
   - $\vdash \varphi \to (\psi \to \varphi)$
   - $\vdash (\varphi \to (\chi \to \psi)) \to (\varphi \to \chi) \to (\varphi \to \psi)$
   - $\vdash (\neg\varphi \to \neg\psi) \to \psi \to \varphi$
   - $\vdash \varphi \to (\varphi \vee \psi)$
   - $\vdash \psi \to (\varphi \vee \psi)$
   - $\vdash (\varphi \to \chi) \to ((\psi \to \chi) \to (\varphi \vee \psi \to \chi))$
   - $\vdash (\varphi \wedge \psi) \to \varphi$
   - $\vdash (\varphi \wedge \psi) \to \psi$
   - $\vdash \varphi \to (\psi \to (\varphi \wedge \psi))$

# Hilbert-style calculus II

## Definition

2. Equality axioms.
   - $(u = u)$,
   - $(u = w) \to (w = u)$,
   - $(u_1 = u_2 \wedge u_2 = u_3) \to (u_1 = u_3)$,
   - $(u_1 = w_1 \wedge \cdots \wedge u_n = w_n) \to (R(u_1, \ldots, u_n) \to R(w_1, \ldots, w_n))$,
   - $(u_1 = w_1 \wedge \cdots \wedge u_m = w_m) \to (t[u_1, \ldots, u_m] = t[w_1, \ldots, w_m])$,

   where $u, w, u_1, \ldots$ are variables and constant symbols, $R$ a $n$-ary relation symbol, and $t$ a term, in which $u_1, \ldots, u_m$ or $w_1, \ldots, w_m$ may occur.
3. Quantifier axioms:
   - $\vdash (\forall x.\varphi(x)) \to \varphi(t)$
   - $\vdash \varphi(t) \to (\exists x.\varphi(x))$



# Hilbert-style calculus III

## Definition

As rules we have:

4. Modus Ponens.

$$\frac{\vdash \varphi \to \psi \qquad \vdash \varphi}{\vdash \psi}$$

5. Generalisation; let $x$ be a variable not free in $\varphi$.

$$\frac{\vdash \varphi \to \psi(x)}{\vdash \varphi \to \forall y.\psi(y)}$$

$$\frac{\vdash \psi(x) \to \varphi}{\vdash (\exists y.\psi(y)) \to \varphi}$$

## Proof in **H**

### Definition

A *proof of* $\varphi$ *starting from a set of formulae* $\Phi$ (in the Hilbert-style calculus **H**), is a *finite* sequence of formulae $\psi_1, \psi_2, \ldots, \psi_n$ with $\psi_n = \varphi$, and each of these formulae $\psi_i$ is either

- an axiom of **H**,
- an element of $\Phi$, or
- is obtained from the previous formulae $\psi_j$, $j < i$, by an application of a rule.

We say that $\varphi$ *is provable from* $\Phi$ (in the Hilbert-style calculus **H**), and write $\Phi \vdash \varphi$, if there exists a proof of $\varphi$ starting from $\Phi$.



## Example

$\varphi \to \varphi$ is not an axiom in our calculus.

### Example

| | |
|---|---|
| $\vdash (\varphi \to ((\varphi \to \varphi) \to \varphi)) \to (\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi)$ | Second axiom |
| $\vdash \varphi \to ((\varphi \to \varphi) \to \varphi)$ | First axiom |
| $\vdash (\varphi \to (\varphi \to \varphi)) \to (\varphi \to \varphi)$ | Modus Ponens |
| $\vdash \varphi \to (\varphi \to \varphi)$ | First axiom |
| $\vdash \varphi \to \varphi$ | Modus Ponens |

# Deduction theorem

**Proposition (Deduction theorem)**

Let $\Phi$ be a set of sentences, $\varphi$ a sentence, and $\psi$ a formula.

$$\Phi \vdash \varphi \to \psi \quad \text{if, and only if} \quad \Phi \cup \{\varphi\} \vdash \psi.$$

**Example**

| | |
|---|---|
| $\{\varphi\} \vdash \varphi$ | Definition of proof |
| $\vdash \varphi \to \varphi$ | Deduction theorem |



# Correctness

**Lemma (Correctness lemma)**

Let $\Phi$ be a set of sentences and $\mathfrak{M}$ a model of $\Phi$.

If $\varphi(x_1, x_2, \ldots, x_n)$ is provable from $\Phi$, then

$$\mathfrak{M} \models \forall x_1.\forall x_2.\ldots.\forall x_n.\varphi(x_1, x_2, \ldots, x_n).$$

# Completeness of predicate logic

**Theorem (Gödel's completeness theorem (for H))**

Let $\Phi$ be a set of sentences of a first-order language $\mathcal{L}$.

A sentence $\varphi$ is provable from $\Phi$ if and only if $\varphi$ is true in all structures which are models of $\Phi$. Formally:

$$\Phi \vdash \varphi \quad \text{if and only if} \quad \Phi \models \varphi.$$

- This theorem speaks about *semantic completeness*.
- It ensures that the logical symbols ($\neg, \wedge, \vee, \rightarrow, \forall, \exists, =$) are treated by our calculus exactly in the way we have attributed a meaning to them (in the definition of the notion of structure).
- Please note the implicit universal quantification on the right hand side: $\Phi \models \varphi$ stands for:

  For **all** models $\mathfrak{M}$ of $\Phi$ it holds that $\mathfrak{M} \models \varphi$.



# Completeness: duality

- The equivalence proven in the completeness theorem:

$$\Phi \vdash \varphi \quad \Leftrightarrow \quad \Phi \models \varphi$$

  results in an interesting duality:
- On the left side we have a statement of the form:

  *It exists a proof ...*

  while on the right hand side we a statement of the form:

  *For all models ...*
- Thus, the completeness theorem allows to replace the universal quantification over models (which, in general, is not easy to handle) by an existential quantification over proofs.
- To show the semantic consequence $\Phi \models \varphi$ we do not need to "search" for $\varphi$ in all models of $\Phi$, but we can simply give *one* proof.

# Completeness: syntax vs. semantics

- In this perspective, (syntactic) proofs seem to be superior to semantic arguments.
- But we may ask how can we show that a formula is *not* provable or, equivalently, that it does not hold semantically, i.e.,

$$\Phi \nvdash \varphi \quad \text{or} \quad \Phi \not\models \varphi.$$

- In this case, we obtain a *negated* existential quantification on the syntactic side, which is equivalent to a universal quantification:

  *For all proofs it is the case, that $\varphi$ is not the last formula.*

- Now, the semantic side has the "advantage"; its negated universal quantifier turns into a existential quantifier:

  *It exists a model in which $\varphi$ is false.*

- Such a model can be called *counter model* for $\varphi$.



- There is a known historical example for this case: for more than 2000 years mathematicians where looking for a *proof* of the parallel axiom from the other euclidean axioms.
- We know today, that it is not provable from these axioms.
- This was shown "semantically": by construction of a counter model.
- The syntactic side may compensate its disadvantage to show "negative" propositions, if it is possible to prove $\Phi \vdash \neg\varphi$.
- Assuming the consistency of $\Phi$, this implies immediately $\Phi \nvdash \varphi$.
- However, in general, $\Phi \nvdash \varphi$ does *not* imply $\Phi \vdash \neg\varphi$.
- This follows, for instance, from the geometry example: Let the *absolute Geometry* $\Phi_{\mathrm{Geo}}$ be the euclidean axioms without the parallel axiom $\varphi_{\mathrm{Par}}$.
- Of course, $\Phi_{\mathrm{Geo}}$ does not imply the negation of the $\varphi_{\mathrm{Par}}$.
- In this sense, this axiom system $\Phi_{\mathrm{Geo}}$ is *syntactically incomplete*: It exists a formula, namely $\varphi_{\mathrm{Par}}$, such that:

$$\Phi_{\mathrm{Geo}} \nvdash \varphi_{\mathrm{Par}} \quad \text{and} \quad \Phi_{\mathrm{Geo}} \nvdash \neg\varphi_{\mathrm{Par}}.$$

# Peano arithmetic

We use the language of Peano arithmetic $\mathcal{L}_{\mathsf{PA}} = \{0, s, +, \cdot\}$.

## Definition (Peano arithmetic)

Peano arithmetic PA comprises the following six non-logical axioms and the following axiom scheme:

(PA1) $\forall x.\neg(s(x) = 0)$,

(PA2) $\forall x, y.s(x) = s(y) \to x = y$,

(PA3) $\forall x.x + 0 = x$,

(PA4) $\forall x, y.x + s(y) = s(x + y)$,

(PA5) $\forall x.x \cdot 0 = 0$,

(PA6) $\forall x, y.x \cdot s(y) = (x \cdot y) + x$.

The axiom scheme of complete induction:

$$\varphi(0) \wedge (\forall y.\varphi(y) \to \varphi(s(y))) \to \forall x.\varphi(x).$$



# Syntactic completeness

- The *standard model* of Peano arithmetic is given by the structure of the natural numbers:
$$\mathcal{N} = \langle \mathbb{N}, 0, +1, +, \cdot \rangle.$$
- "By construction", $\mathcal{N}$ is a model of PA, i.e. for every sentence $\varphi$ it holds
$$\mathsf{PA} \vdash \varphi \quad \Rightarrow \quad \mathcal{N} \models \varphi.$$
- The obvious question is whether the other direction also holds:
$$\mathcal{N} \models \varphi \quad \overset{?}{\Rightarrow} \quad \mathsf{PA} \vdash \varphi.$$
- Gödel's First Incompleteness theorem shows that this implication does not hold.
- It is easy to observe, that this implication is equivalent to the *syntactical completeness* of PA, i.e., the question whether for every formula $\varphi$ it holds that:
$$\mathsf{PA} \vdash \varphi \quad \text{or} \quad \mathsf{PA} \vdash \neg\varphi\ ?$$

# Primitive-recursive functions

## Definition (Primitive-recursive function)

A function $f$, which maps (a tupel of) natural numbers on natural numbers, is called *primitive-recursive*, if can be is given by a finite numbers of steps of the following rules:

1. $Z(x) = 0$, the *zero function*, is primitive-recursive (PrimR);
2. $S(x) = x + 1$, the *successor function* is PrimR;
3. $P_i^n(x_1, x_2, \dots, x_n) = x_i$, $1 \leq i < n \in \mathbb{N}$, the *projection functions* are PrimR;
4. If $g, h_1, \dots, h_n$ are PrimR, then the *composition* $f(\vec{x}) = g(h_1(\vec{x}), \dots, h_n(\vec{x}))$ is also PrimR;
5. If $g$ and $h$ are PrimR, then the following function $f$, defined by *primitive recursion* is also PrimR:

$$f(0, \vec{x}) = g(\vec{x}),$$
$$f(x + 1, \vec{x}) = h(f(x, \vec{x}), x, \vec{x})$$



# Primitive-recursive relations

## Definition

A relation $R \subseteq N^n$ is called *primitive-recursive* if its *characteristic function*

$$\chi_R(\vec{x}) = \begin{cases} 0, & \text{if } R(\vec{x}), \\ 1, & \text{if } \neg R(\vec{x}) \end{cases}$$

is primitive-recursive.

## Lemma

If $R(x, \vec{x})$ is a primitive-recursive relation, then the relations $S_1$ and $S_2$ are also primitive-recursive with:

$$S_1(x, \vec{x}) \leftrightarrow \exists y \leq x . R(y, \vec{x}); \quad S_2(x, \vec{x}) \leftrightarrow \forall y \leq x . R(y, \vec{x}).$$

## Sequence numbers

- To define a primitive-recursive relation, expressing that $x$ codes a proof of a formula encoded by $y$, we will encode formulae, as sequences of symbols, by natural numbers. Then, a proof can be coded by a sequence of natural numbers (which encode the formulae of the proofs) respecting the formal definition of proof.
- Gödel had the tricky idea to encode sequences by a product of prime number powers:

$[a_1, a_2, \ldots, a_n]$ will be encoded by $2^{a_0+1} 3^{a_1+1} \cdot \ldots \cdot p(n)^{a_n+1}$.

- The prime number powers are particularly convenient for the *decoding*: using the fundamental theorem of arithmetic (uniqueness of prime factorization), one can find the $i^{\text{th}}$ element in the sequence by simply counting the occurencies of the $i^{\text{th}}$ prime number in the prime factorization of a sequence number (and subtracting 1).



## Representability

### Definition

Let T be an arbitrary theory.

- A relation $R \subseteq \mathbb{N}^n$ is *numeralwise representable* in T by a formula $\varphi$ if one has, for all natural numbers $m_1, \ldots, m_n$:
  $R(m_1, \ldots, m_n)$ is true if and only if $\mathrm{T} \vdash \varphi(\bar{m}_1, \ldots, \bar{m}_n)$,
  where $\bar{n}$ is a term of the language of T representing the natural number $n$.
  We also say $\varphi$ *numerates* the relation $R$ in $T$.
- $\varphi$ *binumerates* $R$ in $T$ if it numerates it and one has also:
  $R(m_1, \ldots, m_n)$ is false if and only if $\mathrm{T} \vdash \neg\varphi(\bar{m}_1, \ldots, \bar{m}_n)$.

### Theorem (Representation Theorem)

*PA binumerates all primitive-recursive relations.*

# Gödel's First Incompleteness Theorem

- The first incompleteness theorem shows that the Peano Arithmetic is *syntactically* incomplete. That means, there is a formula $\varphi$ such that

$$\mathsf{PA} \nvdash \varphi \quad \text{and} \quad \mathsf{PA} \nvdash \neg\varphi.$$

- The idea of the proof is quite simple. Consider the classical paradox of the *liar*:

  *This sentence is false.*

  Obviously, the sentence can neither be *true* nor *false* without provocating a contradiction.
- In analogy, consider now the following *Gödel sentence*:

  *This sentence is not provable.*

  If this sentence can be represented *faithfully* in the language of Peano-Arithmetic, it can neither be provable nor refutable (i.e., its negation would be provable).



# Two challenges

To formalize the Gödel sentence "This sentence is not provable." in PA we have to solve two problems:

1. Formalizing *provability*.
2. Expressing the self-reference ("*This* sencence ...").

# The provability predicate

*"The details of an encoding are fascinating to work out and boring to read."* (Smoryński)

- Let's first work within the realm of the primitive-recursive functions.
- To simplify matters we assume in the following, that our first-order language of the Peano Arithmetic comprises only the two propositional connectives $\neg$ and $\to$, and the universal quantifier $\forall$. All other connectives can be introduced as abbreviations.
- We may introduce (numerical) codes for the symbols of the language of Peano-Arithmetic:

$$
\begin{array}{lll}
0 \mapsto \langle 0, 0\rangle & x_i \mapsto \langle 1, i\rangle & s \mapsto \langle 2, \langle 1, 0\rangle\rangle \\
+ \mapsto \langle 2, \langle 2, 0\rangle\rangle & \cdot \mapsto \langle 2, \langle 2, 1\rangle\rangle & = \mapsto \langle 3, \langle 2, 0\rangle\rangle \\
\neg \mapsto \langle 4, 4\rangle & \to \mapsto \langle 5, 5\rangle & \forall \mapsto \langle 6, 6\rangle
\end{array}
$$

- For a given symbol $l$ we denote its code by $\ulcorner l \urcorner$.



# Codes for terms and formulae

## Definition

The codes for complex terms and formulae are recursively defined as follows:

- $\ulcorner f_i^n(t_1, \ldots, t_n)\urcorner = [\ulcorner f_i^n \urcorner, \ulcorner t_1 \urcorner, \ldots, \ulcorner t_n \urcorner]$, with $\ulcorner f_i^n \urcorner$ the code attributed to the respective function symbols $s$, $+$, and $\cdot$.
- $\ulcorner t_1 = t_2 \urcorner = [\ulcorner = \urcorner, \ulcorner t_1 \urcorner, \ulcorner t_n \urcorner]$;
- $\ulcorner \neg\varphi \urcorner = [\ulcorner \neg \urcorner, \ulcorner \varphi \urcorner]$;
- $\ulcorner \varphi \to \psi \urcorner = [\ulcorner \to \urcorner, \ulcorner \varphi \urcorner, \ulcorner \psi \urcorner]$;
- $\ulcorner \forall x_i.\varphi \urcorner = [\ulcorner \forall \urcorner, \ulcorner x_i \urcorner, \ulcorner \varphi \urcorner]$.

## Lemma

There are primitive-recursive relations $\mathrm{Term}(x)$ and $\mathrm{Formula}(x)$ which are true if and only if $x$ is the code of term or a formula, respectively.

Proof.

$$\mathrm{Term}(x) = \begin{cases} 0 & \text{if } x = \langle 0,0\rangle \vee \exists i \le x.x = \langle 1,i\rangle, \\ 0 & \text{if } \mathrm{Seq}(x) \wedge \exists n \le x.\exists i \le x.(x)_0 = \langle 2, \langle n,i\rangle\rangle \wedge \\ & \qquad \mathrm{ln}(x) = n \wedge \forall y < n.\mathrm{Term}((x)_{y+1}) = 0, \\ 1 & \text{otherwise} \end{cases}$$

$$\mathrm{Formula}(x) = \begin{cases} 0 & \text{if } \mathrm{Seq}(x) \wedge \exists n \le x.\exists i \le x.(x)_0 = \langle 3, \langle n,i\rangle\rangle \wedge \\ & \qquad \mathrm{ln}(x) = n \wedge \forall y < n.\mathrm{Term}((x)_{y+1}) = 0, \\ 0 & \text{if } \mathrm{Seq}(x) \wedge (x)_0 = \ulcorner\neg\urcorner \wedge \mathrm{Formula}((x)_1) = 0 \wedge \\ & \qquad \mathrm{ln}(x) = 1, \\ 0 & \text{if } \mathrm{Seq}(x) \wedge (x)_0 = \ulcorner\to\urcorner \wedge \mathrm{Formula}((x)_1) = 0 \wedge \\ & \qquad \mathrm{Formula}((x)_2) = 0 \wedge \mathrm{ln}(x) = 2, \\ 0 & \text{if } \mathrm{Seq}(x) \wedge (x)_0 = \ulcorner\forall\urcorner \wedge \exists i \le x.((x)_1 = \langle 1,i\rangle) \wedge \\ & \qquad \mathrm{Formula}((x)_2) = 0 \wedge \mathrm{ln}(x) = 2, \\ 1 & \text{otherwise} \end{cases}$$



## Substitution function

- We need a primitive-recursive function with the property:
$$\mathrm{sub}(\ulcorner\varphi(x_i)\urcorner, i, \ulcorner t\urcorner) = \ulcorner\varphi(t)\urcorner.$$
- It can be defined by course-of-value recursion satisfying the clauses:
  - $\mathrm{sub}(\ulcorner x_i\urcorner, i, y) = y$;
  - $\mathrm{sub}(\ulcorner x_j\urcorner, i, y) = \ulcorner x_j\urcorner$ if $j \neq i$;
  - $\mathrm{sub}(\ulcorner f_j^n(t_1,\ldots,t_n)\urcorner, i, y) = [\ulcorner f_j^n\urcorner, \mathrm{sub}(\ulcorner t_1\urcorner, i, y), \ldots, \mathrm{sub}(\ulcorner t_n\urcorner, i, y)]$;
  - $\mathrm{sub}(\ulcorner t_1 = t_2\urcorner, i, y) = [\ulcorner=\urcorner, \mathrm{sub}(\ulcorner t_1\urcorner, i, y), \mathrm{sub}(\ulcorner t_2\urcorner, i, y)]$;
  - $\mathrm{sub}(\ulcorner\neg\varphi\urcorner, i, y) = [\ulcorner\neg\urcorner, \mathrm{sub}(\ulcorner\varphi\urcorner, i, y)]$;
  - $\mathrm{sub}(\ulcorner\varphi\to\psi\urcorner, i, y) = [\ulcorner\to\urcorner, \mathrm{sub}(\ulcorner\varphi\urcorner, i, y), \mathrm{sub}(\ulcorner\psi\urcorner, i, y)]$;
  - $\mathrm{sub}(\ulcorner\forall x_i.\varphi\urcorner, i, y) = \ulcorner\forall x_i.\varphi\urcorner$;
  - $\mathrm{sub}(\ulcorner\forall x_j.\varphi\urcorner, i, y) = [\ulcorner\forall\urcorner, \ulcorner x_i\urcorner, \mathrm{sub}(\ulcorner\varphi\urcorner, i, y)]$;
  - $\mathrm{sub}(x, i, y) = 0$ if $x$ is not a code of any of the terms or formulae in the preceding clauses.

## Substitution function; numerals

- For a formula $\varphi(x)$ with exactly one free variable $x$ we need a primitive recursive function with the property:

$$\mathrm{Sub}(\ulcorner\varphi(x)\urcorner, \ulcorner t\urcorner) = \ulcorner\varphi(t)\urcorner.$$

- It can be defined by:

$$\mathrm{Sub}(x,y) = \begin{cases} \mathrm{sub}(x,i,y) \text{ with } i = \mu j < x[x_j \text{ is free in } x] & \text{if } i \neq x \\ x & \text{if } i = x \end{cases}$$

- We need a primitive recursive function with the property:

$$\mathrm{Num}(n) = \ulcorner\bar{n}\urcorner.$$

- It can be defined by:

$$\mathrm{Num}(0) = \ulcorner 0\urcorner,$$
$$\mathrm{Num}(x+1) = [\ulcorner s\urcorner, \mathrm{Num}(x)].$$



## The proof predicate

- To simplify matters we assume in the following, that our Hilbert-style calculus has as only rule *Modus Ponens*; it is possible to replace the generalization rules by (an infinite list of) axioms.
- We first define a primitive recursive relation $\mathrm{Bew}_{\mathsf{PA}}$ such that $\mathrm{Bew}_{\mathsf{PA}}(x,y)$ is true, if and only if $x$ is the Gödel number of a proof in PA of the formula with the Gödel number $y$:

$$\begin{aligned} \mathrm{Bew}_{\mathsf{PA}}(x,y) \Leftrightarrow \mathrm{Seq}(x) \wedge {} & \\ & y = (x)_{\mathrm{ln}(x)} \wedge \\ & \forall i \leq \mathrm{ln}(x).((x)_i \text{ is a logical axiom}) \vee \\ & \qquad ((x)_i \text{ is an axiom of PA}) \vee \\ & \qquad (\exists j < i.\exists k < i.(x)_k = [\ulcorner\rightarrow\urcorner, (x)_j, (x)_i]) \end{aligned}$$

## The proof predicate

The representation theorem says that PA binumerates all primitive-recursive relations. Thus, it applies to $\mathrm{Bew}_{\mathsf{PA}}$ and we have that there is a formula $\mathrm{Bew}_{\mathsf{PA}}$ in the language of PA with:

$$\begin{array}{rl} \mathrm{Bew}_{\mathsf{PA}}(m_1, m_2) \text{ is true} & \text{if and only if } \mathsf{PA} \vdash \mathrm{Bew}_{\mathsf{PA}}(\bar{m}_1, \bar{m}_2) \\ \mathrm{Bew}_{\mathsf{PA}}(m_1, m_2) \text{ is false} & \text{if and only if } \mathsf{PA} \vdash \neg\mathrm{Bew}_{\mathsf{PA}}(\bar{m}_1, \bar{m}_2). \end{array}$$



## A provability predicate

- $\mathrm{Bew}_{\mathsf{PA}}(x, y)$ is a *proof predicate*.
- If we try to define a **provability** *predicate*, first in the recursion theoretic realm, we would like to define:

$$\mathrm{B}_{\mathsf{PA}}(y) \Leftrightarrow \exists x.\mathrm{Bew}_{\mathsf{PA}}(x, y).$$

- But this definition involves an **unbounded** existential quantification, which is not expressible by primitive-recursive functions.
- It is, however, expressible by a *partial recursive* function. Doing so, one could show that there is corresponding formula in PA which *numerates*—but not *binumerates*—this provability predicate.
- Here, we don't need to consider partial recursive predicates explicitly, but we define in PA simply:

$$\mathrm{B}_{\mathsf{PA}}(y) \Leftrightarrow \exists x.\mathrm{Bew}_{\mathsf{PA}}(x, y).$$

## A provability predicate

- Bei definition of the relation $\mathrm{Bew}_{\mathsf{PA}}$ we have for its representation $\mathrm{Bew}_{\mathsf{PA}}$ in PA:

$$\begin{aligned} \mathsf{PA} \vdash \varphi &\iff \mathsf{PA} \vdash \mathrm{Bew}_{\mathsf{PA}}(t, \ulcorner\varphi\urcorner) \qquad \text{for a closed term } t \\ &\implies \mathsf{PA} \vdash \exists x.\mathrm{Bew}_{\mathsf{PA}}(x, \ulcorner\varphi\urcorner) \\ &\iff \mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \end{aligned}$$

  $t$ is a sequence number of the form $[\ulcorner\varphi_0\urcorner, \ulcorner\varphi_1\urcorner, \ldots, \ulcorner\varphi_{n-1}\urcorner, \ulcorner\varphi\urcorner]$.
- In short:

$$\mathsf{PA} \vdash \varphi \implies \mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \tag{1}$$

- Note that we don't have immediately the "missing" direction:
$$\mathsf{PA} \vdash \exists x.\mathrm{Bew}_{\mathsf{PA}}(x, \ulcorner\varphi\urcorner) \implies \mathsf{PA} \vdash \mathrm{Bew}_{\mathsf{PA}}(t, \ulcorner\varphi\urcorner)$$
- In general, one cannot conclude from an existential statement like $\exists x.\mathrm{Bew}_{\mathsf{PA}}(x, \ulcorner\varphi\urcorner)$ that there is also a *closed term* which exemplifies such an $x$.



## Diagonalization lemma

### Theorem (Diagonalization lemma)

Let $\varphi(x)$ be a formula with exactly one free variable $x$. Then there is a sentence $\psi$ such that:
$$\mathsf{PA} \vdash \psi \leftrightarrow \varphi(\ulcorner\psi\urcorner).$$

### Proof.

Define $\vartheta(x)$ as $\varphi(\mathrm{Sub}(x, \mathrm{Num}(x)))$. Let $\bar{m}$ be $\ulcorner\vartheta(x)\urcorner$ and let $\psi$ be $\vartheta(\bar{m})$.

$$\begin{aligned} \psi &\leftrightarrow \vartheta(\bar{m}) \\ &\leftrightarrow \varphi(\mathrm{Sub}(\bar{m}, \mathrm{Num}(\bar{m}))) \\ &\leftrightarrow \varphi(\mathrm{Sub}(\ulcorner\vartheta(x)\urcorner, \ulcorner\bar{m}\urcorner)) \\ &\leftrightarrow \varphi(\ulcorner\vartheta(\bar{m})\urcorner) \\ &\leftrightarrow \varphi(\ulcorner\psi\urcorner) \end{aligned}$$

$\psi$ expresses "I have the property $\varphi$".

## Gödel's First Incompleteness Theorem

**Theorem (First Incompleteness Theorem; Gödel 1931)**

Assume, PA is consistent. Then, there is a sentence $\varphi$ such that:

1. $\mathsf{PA} \nvdash \varphi$;
2. If $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \Rightarrow \mathsf{PA} \vdash \varphi$, then $\mathsf{PA} \nvdash \neg\varphi$.

**Proof.**

According to the diagonalization lemma, there is a sentence $\varphi$ such that

$$\mathsf{PA} \vdash \varphi \leftrightarrow \neg\mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner). \tag{$*$}$$

1. Assume $\mathsf{PA} \vdash \varphi$. With (1) we have $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$. With $(*)$ it follows $\mathsf{PA} \vdash \neg\mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ in contradiction to the consistency of PA.
2. Assume $\mathsf{PA} \vdash \neg\varphi$. With $(*)$ we have $\mathsf{PA} \vdash \neg\neg\mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ and also $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$. Because of the additional premise this gives $\mathsf{PA} \vdash \varphi$, again in contradiction to the consistency of PA.



## $\omega$-consistency

- The premise $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \Rightarrow \mathsf{PA} \vdash \varphi$ in the second case corresponds to the *$\omega$-consistency* which was assumed by Gödel in his original paper.
- In 1936, B. J. Rosser found a trick to avoid this condition, using a modified proof predicate $\mathrm{Bew}^R$ "on top" of Gödel's proof.

**Theorem (First Incompleteness Theorem; Gödel 1931, Rosser 1936)**

Assume, PA is consistent. Then, there is a sentence $\varphi$ such that:

1. $\mathsf{PA} \nvdash \varphi$;
2. $\mathsf{PA} \nvdash \neg\varphi$.

## Completing PA?

- So far, we showed the first incompleteness theorem "only" for PA.
- As for the parallel axiom with respect to "absolute Geometry" one could wonder whether there was only an axiom missing which should be added to make PA complete.
- The particular formula $\varphi$ of our proof, which is indepedent of PA, expresses "I'm not provable"; more exactly:

"I'm not provable in PA."

- As we just proved PA $\nvdash \varphi$ it is obviously *true* according to its reading.
- Thus, it would be fully justified to add it as a new axiom to define the theory $\mathsf{PA}' = \mathsf{PA} + \{\varphi\}$.
- This does not lead to a contradition, as $\varphi$ does not say "I'm not provable in $\mathsf{PA}'$"!
- But we can repeat Gödel's proof, now for $\mathsf{PA}'$ obtaining a new independent formula $\varphi'$.



## First Incompleteness Theorem: generic form

**Theorem (First Incompleteness Theorem)**

Assume, that $T$ is a consistent, *recursive* extension of PA. Then, there is a sentence $\varphi$ such that:

1. $T \nvdash \varphi$;
2. $T \nvdash \neg\varphi$.

- For primitive-recursive extensions, the proof for the general case works *literally* along the lines of the proof given above, except that we have to modify the clause for the non-logical axioms in the definition of the proof predicate $\mathrm{Bew}_{\mathsf{T}}$:

$$\ldots \vee ((x)_i \text{ is an axiom of } \mathsf{T}) \vee \ldots$$

- For $\mathrm{Bew}_{\mathsf{T}}$ being still primitive-recursive, we only need that the set of (codes of) axioms of $T$ is primitive recursive.
- The theorem also holds for the recursive extensions.

# A complete "axiomatization" and Tarski's theorem

- Let PA* the theory consisting of the following set of axioms:

$$\{\varphi \mid \mathcal{N} \models \varphi\}$$

- PA* is trivially complete.
- However, the set of axioms is not any longer *recursive*.

**Theorem (Tarski's theorem)**

*Arithmetical truth is not recursive.*



# Gödel's second incompleteness theorem

- Gödel's second incompleteness theorem says that a theory, which has at least the expressive power of Peano Arithmetic, cannot prove its own consistency.
- Using the techniques developed so far, consistency of a theory $T$ can be easily expressed as:

$$\mathrm{Con}_T \iff \neg \mathrm{B}_T(\ulcorner \Lambda \urcorner)$$

where $\Lambda$ is an arbitrary contradictory (false) formula, for instance, $0 = s(0)$.
- We say that a theory does not prove it own consistency if we have:

$$T \nvdash \mathrm{Con}_T.$$

## The idea of the proof of Gödel II

- First we consider, again, only PA.
- In a sloppy formulation, the idea for the proof of the second incompleteness theorem is to formalize the proof of the first incompleteness theorem in PA.

1. If $\mathsf{PA} \nvdash \varphi$, PA is obviously consistent (as an inconsistent theory proves every formula). Thus:

$$\mathsf{PA} \nvdash \varphi \Longrightarrow \mathsf{PA} \text{ is consistent.}$$

2. The first incompleteness theorem states, for the chosen $\varphi$:

$$\mathsf{PA} \text{ is consistent} \Longrightarrow \mathsf{PA} \nvdash \varphi.$$

- The formalization of both arguments *within* PA will show that this $\varphi$ is equivalent to the consistency statement of PA:

$$\begin{aligned} \mathsf{PA} &\vdash \neg \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \leftrightarrow \mathrm{Con}_{\mathsf{PA}} \\ \mathsf{PA} &\vdash \varphi \leftrightarrow \mathrm{Con}_{\mathsf{PA}}. \end{aligned}$$



## Provability conditions

- For the proof of the first incompleteness theorem we used the following property of $\mathrm{B}$:

$$\mathsf{PA} \vdash \varphi \Longrightarrow \mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \tag{1}$$

- For the proof of the second incompleteness theorem, we need the two additional properties of $\mathrm{B}_{\mathsf{PA}}$:

$$\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \mathrm{B}_{\mathsf{PA}}(\ulcorner \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \urcorner) \tag{2}$$

$$\mathsf{PA} \vdash [\mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \wedge \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \to \psi \urcorner)] \to \mathrm{B}_{\mathsf{PA}}(\ulcorner \psi \urcorner) \tag{3}$$

- (2) and (3) do not follow any longer directly from the representability theorem. But they can be proven for $\mathrm{B}_{\mathsf{PA}}$ (with some hard work).
- The three conditions are called *Hilbert-Bernays-Löb derivablity conditions*. They can be studied independently, and in an abstract from they are the base of *provability logic*.

# Gödel's second incompleteness theorem

**Theorem (Second incompleteness theorem)**

Assume PA is consistent. Then we have:

$$\mathsf{PA} \nvdash \mathrm{Con}_{\mathsf{PA}}.$$



# Proof of Gödel's second incompleteness theorem

- Let $\varphi$ be such that: $\mathsf{PA} \vdash \varphi \leftrightarrow \neg \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$ $(\star)$
- | | |
  |---|---|
  | $\mathsf{PA} \vdash \Lambda \to \varphi$ | Ex-falso-quodlibet |
  | $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \Lambda \urcorner) \to \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$ | (1) and (3) |
  | $\mathsf{PA} \vdash \neg \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \neg \mathrm{B}_{\mathsf{PA}}(\ulcorner \Lambda \urcorner)$ | Contrapositive |
  | $\mathsf{PA} \vdash \varphi \to \neg \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$ | $(\star)$ |
  | $\mathsf{PA} \vdash \varphi \to \neg \mathrm{B}_{\mathsf{PA}}(\ulcorner \Lambda \urcorner)$ | Logical reasoning |
  | $\mathsf{PA} \vdash \varphi \to \mathrm{Con}_{\mathsf{PA}}$ | Definition of $\mathrm{Con}_{\mathsf{PA}}$ |
- | | |
  |---|---|
  | $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \neg \varphi$ | Contrapositive of $(\star)$ |
  | $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \urcorner) \to \mathrm{B}_{\mathsf{PA}}(\ulcorner \neg \varphi \urcorner)$ | (1) and (3) |
  | $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \mathrm{B}_{\mathsf{PA}}(\ulcorner \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \urcorner)$ | (2) |
  | $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \mathrm{B}_{\mathsf{PA}}(\ulcorner \neg \varphi \urcorner)$ | Logical reasoning |
  | $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \wedge \neg \varphi \urcorner)$ | (1), (3) and logical reasoning |
  | $\mathsf{PA} \vdash \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \mathrm{B}_{\mathsf{PA}}(\ulcorner \Lambda \urcorner)$ | Definition of $\Lambda$ |
  | $\mathsf{PA} \vdash \neg \mathrm{B}_{\mathsf{PA}}(\ulcorner \Lambda \urcorner) \to \neg \mathrm{B}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$ | Contrapositive |
  | $\mathsf{PA} \vdash \mathrm{Con}_{\mathsf{PA}} \to \varphi$ | Definition of $\mathrm{Con}_{\mathsf{PA}}$ and $(\star)$ |
- As $\mathsf{PA} \nvdash \varphi$ we have also $\mathsf{PA} \nvdash \mathrm{Con}_{\mathsf{PA}}$.

## Gödel's second incompleteness theorem; generic version

**Theorem (Second incompleteness theorem; Gödel 1931)**

Assume, that $T$ is a consistent, *recursive* extension of PA. Then

$$T \nvdash \mathrm{Con}_T.$$



## Why reasoning in PA about PA?

- Assume, the second incompleteness theorem would not hold, and it would be the case that $\mathsf{PA} \vdash \mathrm{Con}_{\mathsf{PA}}$.
- Obviously, such a proof would not give any evidence for the consistency of PA: if PA would be incosistent, every formula would be provable, in particular also $\mathrm{Con}_{\mathsf{PA}}$.
- The significance of the second incompleteness theorem (as given here) is based on an immediate corollary: if PA cannot prove its consistency, no *weaker* theory—in particular, any subsystem of PA—could do so.
- But this was the idea in **Hilbert's programme**: using *finitistic mathematics*—which is is supposed to be a subsystem of PA—to prove the consistency of PA (and other theories).

**Note**

It is possible to uncouple the two uses of PA in $\mathsf{PA} \nvdash \mathrm{Con}_{\mathsf{PA}}$ and to use a separate theory $T$ for the "meta-reasoning"; for $T$ we need only the representation theorem ($T$ binumerates all primitive-recursive functions). Then one can prove $T \nvdash \mathrm{Con}_{\mathsf{PA}}$. (See the Handbook article of Smorynski.)

## Hilbert's programme

- How to prove that a theory, given by a set of axioms $\Phi$, is *consistent*?
- First (traditional) method (Frege): Giving a model (blue object).
- Problem: where does the blue object "live"? How to define it rigidly? Mathematics had some "bad experiences" (paradoxes in naive set theory!)
- New method — proposed by Hilbert:
  Proving on a *metalevel* that there is **no** formula $\varphi$ such that
  $$\Phi \vdash \varphi \text{ and } \Phi \vdash \neg\varphi.$$
- To do so: *Formalizing mathematically the relation* $\vdash$ and studying it with "ordinary mathematical tools".
- New problem: these "ordinary mathematical tools" are safe? Wouldn't I run into a vicious circle?
- Solution: restricting yourself to "safe mathematics".
- Hilbert's proposal: *finitistic mathematics*.



## Consistency proofs revisited

- Gödel's second incompleteness theorem shows that Hilbert's programme cannot work (in its original formulation).
- The second incompleteness theorem does not exclude that the consistency of PA can be proven by use of *stronger* or *"other"* means.
- These means do not need to be extensions of PA and it is possible that one takes the consistency of such other means for granted—depending on the respective mathematico-philosophical viewpoint.
- For PA, we may consider the following three alternative approaches (all of them already discussed by Gödel as early as 1938):
  - Intuitionistic Arithmetic: double negation interpretation. (Kolmogorov 1925; Gödel 1933; Gentzen 1936)
  - Primitive-recursive arithmetic with *transfinite* induction up to the ordinal $\varepsilon_0$ (Gentzen 1936)
  - Functionals of higher type: *Gödel's $\mathcal{T}$; Dialectica interpretation* (Gödel 1958)
- Ordinal analysis for stronger and stronger systems...

## The fate of Hilbert's programme

- By aiming for a consistency proof of "higher" mathematics in "finitistic" mathematics, Hilbert's programme can be coupled with *conservativity*.
- Conservativity means that methods of a stronger theory will not prove *new results* in the weaker theory—although proofs might become easier (shorter).
- Gödel I refutes conservativity of stronger theories over *PA*.
- *This* is a surprising and amazing result, opening a completely new perspective on mathematical strength.

  It is reported that, for instance, Gauß expressed a strong conviction concerning conservativity.
- The failure of Hilbert's programme with respect to conservativity cannot be "revised".
  - There is no way to reduce all higher Mathematics to finitistic Mathematics; higher Mathematics may prove finitistic statements which are not provable with pure finitistic methods.



## The fate of Hilbert's programme

- Let us draw on a comparison here:
  - nobody will deny that Columbus failed to find the sea route to India;
  - but he didn't sink in the Ocean,
  - he discovered America.

  - In the same way, Hilbert's Programme, aiming for consistency and conservativity, didn't succeed;
  - but it didn't sink in inconsistency,
  - it discovered Non-Conservativity.
- Exploring this new phenomena in Mathematics is the driving force of modern proof theory.

R. Kahle. Gentzen's theorem in context.
In: R. Kahle and M. Rathjen (eds.). Gentzen's Centenary: The quest for consistency. Springer, 2015. To appear.