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 φ and ψ are formulae, then the following expressions are also formulae:
$$(\neg\varphi),\ (\varphi \wedge \psi),\ (\varphi \vee \psi),\ (\varphi \rightarrow \psi).$$
4. If φ 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 φ is *recursively* defined as follows:

1. If φ is an atomic formula, then $FV(\varphi)$ is the set of variables which occur in φ;
2. $FV(\neg\varphi) = FV(\varphi)$;
3. $FV(\varphi \wedge \psi) = FV(\varphi \vee \psi) = FV(\varphi \to \psi) = FV(\varphi) \cup FV(\psi)$;
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 φ 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 t (true) or 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 $\langle M, {c_1}^{\mathfrak{M}}, \dots, {c_n}^{\mathfrak{M}}, {f_1}^{\mathfrak{M}}, \dots, {f_k}^{\mathfrak{M}}, {R_1}^{\mathfrak{M}}, \dots, {R_l}^{\mathfrak{M}} \rangle$.

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

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

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 $s(\frac{a}{x})$ the assignment which coincides with s for all variables except x; independently of the value of $s(x)$, we fix $s(\frac{a}{x})(x) = a$. More exactly:

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

Definition

Let $\mathfrak{M}$ be a $\mathcal{L}$ structure. We define, for every assignment s and every formula φ 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 $\mathfrak{M} \models \varphi[s(\tfrac{a}{x})]$,
9. $\mathfrak{M} \models (\forall x.\varphi)[s]$ if and only if **for all elements** $a \in M$ it holds that $\mathfrak{M} \models \varphi[s(\tfrac{a}{x})]$.

Semantic consequence

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

Definition

Let Φ be a set of $\mathcal{L}$-sentences and $\mathfrak{M}$ be a $\mathcal{L}$ structure. $\mathfrak{M}$ is a *model* of Φ, 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 ψ we say that it *follows (semantically) from* Φ, written as $\Phi \models \psi$, if for every model $\mathfrak{M}$ of Φ it holds that $\mathfrak{M} \models \psi$.

If $\mathfrak{M} \models \varphi$ holds, we also say that φ *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 Φ be a set of first-order sentences.

If every finite subset Φ_0 of Φ has a model, then there exists also a model of Φ.

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 Φ_0 of Φ, 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 φ.

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

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

Proof in **H**

Definition

A *proof of* φ *starting from a set of formulae* Φ (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 ψ_i is either

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

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

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 Φ be a set of sentences, φ a sentence, and ψ 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 Φ be a set of sentences and $\mathfrak{M}$ a model of Φ.

If $\varphi(x_1, x_2, \ldots, x_n)$ is provable from Φ, 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 Φ be a set of sentences of a first-order language $\mathcal{L}$.

A sentence φ is provable from Φ if and only if φ is true in all structures which are models of Φ. 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 Φ 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 φ in all models of Φ, 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 φ 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 φ is false.

- Such a model can be called *counter model* for φ.

- 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 Φ, 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* Φ_{Geo} be the euclidean axioms without the parallel axiom φ_{Par}.
- Of course, Φ_{Geo} does not imply the negation of the φ_{Par}.
- In this sense, this axiom system Φ_{Geo} is *syntactically incomplete*: It exists a formula, namely φ_{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 $\mathcal{L}_{\mathrm{PA}} = \{0, s, +, \cdot\}$.

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) $\forall x, y.s(x) = s(y) \rightarrow x = y$,

(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:

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

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 φ it holds
$$\mathrm{PA} \vdash \varphi \quad \Rightarrow \quad \mathcal{N} \models \varphi.$$
- The obvious question is whether the other direction also holds:
$$\mathcal{N} \models \varphi \quad \overset{?}{\Rightarrow} \quad \mathrm{PA} \vdash \varphi.$$
- 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 φ it holds that:
$$\mathrm{PA} \vdash \varphi \quad \text{or} \quad \mathrm{PA} \vdash \neg\varphi\ ?$$