Gödel's First Incompleteness Theorem

Unit 1: An introduction

Lavinia Picollo

1 Foundations for mathematics

On what grounds are we allowed to accept mathematical – and, in particular, arithmetical – claims?

Hilbert and Frege agreed that mathematical statements are acceptable just in case they follow logically from an adequate (e.g. consistent, correct, basic, simple, etc.) collection of axioms. For these foundational projects to succeed it should be possible to find a theory of arithmetic which we can know to be

(a) consistent (so it doesn't entail a sentence and its negation, or every sentence); and

(b) complete (so it entails all truths about natural numbers).

2 Formal theories and the Entscheidungsproblem

Definition (formal theory). *Given a deductive proof-system $\mathcal{D}$ for a formal language L, a formal theory T in $\mathcal{D}$ is a set of sentences of L closed under derivability in $\mathcal{D}$. A set $\Sigma \subseteq L$ counts as a set of axioms for T in $\mathcal{D}$ just in case (i) $\Sigma \subseteq T$ and (ii) every member of T is derivable from Σ in $\mathcal{D}$.*

Definition (effectively decidable property). *A property P (set Σ) (defined over some domain of objects D) is effectively decidable iff there's an algorithm (a finite set of instructions for a deterministic computation, a step-by-step mechanical routine) for settling in a finite number of steps, for any $o \in D$, whether o has property P (whether $o \in \Sigma$).*

Definition (effectively formalized language). *An effectively formalized language L has an effectively decidable set of basic symbols and the syntactic rules of L are such that the properties of being a term, a wff, a wff with one free variable, a sentence (a formula with no free variables), etc. are effectively decidable.*

Definition (effectively axiomatized theory). *An effectively axiomatized formal theory T has an effectively formalized language L, an effectively decidable set of axioms, and a proof-system $\mathcal{D}$ such that it is effectively decidable whether a given sequence of wffs of L is indeed a proof from the axioms according to the rules of $\mathcal{D}$.*

Warning: It is one thing to be able to effectively check a proof once proposed, it is another thing to be able to effectively decide in advance whether there exists a proof to be discovered.

Our effectively formalized languages will be first-order languages with identity. For these languages there are many effectively decidable proof-systems. We assume one of them, a Hilbert-style one (see the appendix for the details). Let Γ be a set of sentences and φ be a sentence. '$\Gamma \vdash \varphi$' – φ is a proof-theoretic consequence of Γ – means that there is a derivation of φ from the members of Γ. If Γ is a theory (i.e. if it's closed under derivability), we say that φ is a theorem of Γ. '$\Gamma \vDash \varphi$' – φ is a semantic consequence of Γ – says that all models/interpretations of the language in which all members of Γ are true are such that φ is true in them. In the chosen proof-system the following holds for every Γ and every φ:

Theorem (Gödel's Completeness)**.** $\Gamma \vdash \varphi$ *iff* $\Gamma \vDash \varphi$.

Notation. L, L', etc. are metavariables ranging over formal languages, T, T', etc. over formal theories, Γ, Σ, etc. over arbitrary sets of wff of a given formal language, φ, ψ, etc. over wffs of a given formal language.

Question 1 (Entscheidungsproblem (decision problem))**.** *Let T be an arbitrary effectively axiomatized theory. Is there an algorithm that decides, for each sentence φ of the language of the theory, whether φ is a theorem of T?*

Exercise 1. *How would an answer to the Entscheidungsproblem impact Hilbert and Frege's foundational projects? In particular, what impact would it have on the search for consistency proofs for effectively axiomatized theories?*

In 1936, Alonzo Church and Alan Turing published independent papers answering the question posed by the Entscheidungsproblem negatively, assuming that the intuitive notion of effective decidability is captured by the functions computable by a Turing machine or equivalently, by those expressible in the lambda calculus. This assumption is now known as the Church–Turing thesis.

3 Decidability and number theory

3.1 Basic notions

Definition (formal decidability)**.** *Let φ be some sentence of the language of T. Then, T formally decides φ iff either $T \vdash \varphi$ or $T \vdash \neg\varphi$; and φ is formally undecidable by T iff $T \nvdash \varphi$ and $T \nvdash \neg\varphi$.*

Definition (negation completeness)**.** *T is negation complete iff it formally decides every sentence of its language – i.e. for every sentence φ, $T \vdash \varphi$ or $T \vdash \neg\varphi$.*

Trivially, then, there are 'formally undecidable propositions' in T iff T isn't negation complete.

Exercise 2. *Give an example of a negation complete theory and one of a negation incomplete one. (We specify a theory by laying down the set of its axioms.)*

Definition (consistency)**.** *T is consistent iff there is no φ such that $T \vdash \varphi$ and $T \vdash \neg\varphi$. (Alternatively, iff there is a sentence φ of the language of T such that $T \nvdash \varphi$.)*

Definition (soundness)**.** *T is sound iff its axioms are true (on the interpretation built into T's language). Since the logic is truth preserving, all of T's theorems are true too.*

Exercise 3. *Prove that soundness implies consistency. Show that the converse is not true.*

3.2 Undecidability of arithmetic

Definition (language of arithmetic). *L_A is the first-order language with identity that has as non-logical symbols an individual constant, $\mathsf{0}$, a 1-place function symbol (for the successor function), S, and two 2-place function symbols, $+$ and $\times$ (L_A's only primitive predicate is identity).*

Closed terms of L_A are, e.g. $\mathsf{0}, \mathsf{S0}, \mathsf{S0} \times (\mathsf{S0} + \mathsf{0})$. For each natural number $0, 1, 2, 3, \ldots$ L_A has a term denoting it. We use the terms $\mathsf{0}, \mathsf{S0}, \mathsf{SS0}, \mathsf{SSS0}, \ldots$ as the canonical ones, which we call '(standard) numerals'. We'll use $\overline{\mathsf{n}}$ to abbreviate the standard numeral for n.

Sentences of L_A are, e.g. $(\mathsf{0}+\mathsf{0} = \mathsf{S0} \wedge \mathsf{S0}+\mathsf{S0} = \mathsf{SS0})$, $\overline{\mathsf{n}}+\overline{\mathsf{1}} = \overline{\mathsf{n+1}}$, $\forall \mathsf{x} \exists \mathsf{y}\, (\mathsf{x} \times \mathsf{y} = \mathsf{0})$. Presumably, all sentences of L_A are either true or false.

Notation. Particular expressions from formal languages – and abbreviations of them – will be in sans serif type. Expressions in informal mathematics will be in serif font.

Question 2. *Is there an effectively axiomatized theory formulated in L_A that is sound and negation complete (i.e. that proves all the truths expressible in the language)?*

Exercise 4. *Show that if an effectively axiomatized theory is negation complete then it is formally decidable (tip: consider the cases in which the theory is consistent and inconsistent separately). Is the converse true?*

In 1931, Gödel published a paper titled 'On formally undecidable propositions of Principia Mathematica and related systems I', where he proves his First Incompleteness Theorem.

Theorem 1 (Gödel's First Incompleteness (semantic)). *If T is a sound effectively axiomatized theory whose language contains contains L_A, then there will be a true sentence G_T of L_A such that $T \nvdash \mathsf{G}_T$ and $T \nvdash \neg\mathsf{G}_T$, so T is negation incomplete.*

Theorem 2 (Gödel's First Incompleteness (syntactic)). *If T is a consistent effectively axiomatized theory whose language contains L_A, it can prove certain modest amount of arithmetic, and if it has a certain additional desirable syntactically definable property that any sensible formalized arithmetic will share, then there will be a sentence G_T of L_A such that $T \nvdash \mathsf{G}_T$ and $T \nvdash \neg\mathsf{G}_T$, so T is negation incomplete.*

Exercise 5. *Let T be a theory formulated in L_A. Is $T + \mathsf{G}_T$ a negation complete system?*[1]

3.3 Proof strategy for the semantic version

Definition (expressibility/definabilty). *Let φ be a formula of the (standardly) interpreted language L containing L_A.*

- *$\varphi(\mathsf{x})$ expresses the numerical property P iff, for every natural number n, $\varphi(\overline{\mathsf{n}})$ is true on interpretation just when n has property P.*
- *$\varphi(\mathsf{x_1}, \ldots, \mathsf{x_m})$ expresses the numerical relation R iff, for every n-tuple of natural numbers $n_1, \ldots, n_m$, $\varphi(\overline{\mathsf{n_1}}, \ldots, \overline{\mathsf{n_m}})$ is true on interpretation just when $n_1, \ldots, n_m$ are related via R.*
- *$\varphi(\mathsf{x_1}, \ldots, \mathsf{x_m}, \mathsf{y})$ expresses the numerical function f iff, for every $n+1$-tuple of natural numbers $n_1, \ldots, n_m, k$, $\varphi(\overline{\mathsf{n_1}}, \ldots, \overline{\mathsf{n_m}}, \overline{\mathsf{k}})$ is true on interpretation just in case $f(n_1, \ldots, n_m) = k$.*

[1] Recall a theory is a set of sentences closed under logical consequence. If φ is a sentence of the language of T, we write $T + \varphi$ for the set of sentences containing all members of T, φ, and closed under logical consequence.

Then Gödel's proof of Theorem 2 in outline form goes as follows:

1. *Set up a Gödel numbering*: we set up an effective way of coding wffs and sequences of wffs by natural numbers (i.e. we assign a unique natural number to every wff and every sequence of wffs of L_A). Then, given an effectively axiomatized theory T, we can define the following effectively decidable numerical properties/relation: Wff_T, Sent_T, and Prf_T where

$$\mathit{Wff}_T(n) \text{ iff } n \text{ is the code of a wff of the language of } T$$
$$\mathit{Sent}_T(n) \text{ iff } n \text{ is the code of a sentence of the language of } T$$
$$\mathit{Prf}_T(n,m) \text{ iff } m \text{ is the code of a } T\text{-proof of the sentence with code } n$$

2. *Express such properties/relations in L_A*: we show how to build wffs $\mathsf{Wff}_T(\mathsf{x})$, $\mathsf{Sent}_T(\mathsf{x})$, and $\mathsf{Prf}_T(\mathsf{x},\mathsf{y})$ of L_A expressing the corresponding properties by, more generally, showing that every effectively decidable set is expressible in L_A.

3. *Construct a Gödel sentence*: let $\mathsf{Prov}_T(\mathsf{x})$ abbreviate $\exists \mathsf{y}\, \mathsf{Prf}_T(\mathsf{y},\mathsf{x})$. This says that some number is the code of a proof of the wff with Gödel-number x – i.e. the wff with number x is a theorem. Next we show how to use $\mathsf{Prov}_T(\mathsf{x})$ to build a 'Gödel' sentence G_T whose Gödel number is g in such a way that G_T is equivalent to $\neg\mathsf{Prov}_T(\overline{\mathsf{g}})$. So G_T is true on interpretation iff $\neg\mathsf{Prov}_T(\overline{\mathsf{g}})$ is true, i.e. iff G_T is not a theorem.

4. *Reason as follows*: Assume T is sound. Suppose $T \vdash \mathsf{G}_T$. Then G_T would be a theorem, and hence G_T would be false, so T would have a false theorem and hence not be sound, contrary to the hypothesis that T is sound. So $T \nvdash \mathsf{G}_T$. So G_T is true, but unprovable. Thus, $\neg\mathsf{G}_T$ is false and T, being sound, can't prove it either. Hence, we also have $T \nvdash \neg\mathsf{G}_T$, so G_T is undecidable in T; T is negation incomplete.

Appendix

First-order languages with identity

Wffs of a first-order language are strings of symbols borrowing from the first-order alphabet.

Definition (first-order alphabet)**.** *The universal alphabet is a set consisting of the following basic symbols:*

- *Logical terms:* $\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \forall, \exists, =$
- *Individual variables:* $\mathsf{x}_0, \mathsf{x}_1, \ldots, \mathsf{x}_n, \ldots$
- *Individual constants:* $\mathsf{f}^0_0, \mathsf{f}^0_1, \ldots, \mathsf{f}^0_n, \ldots$
- *Function symbols:* $\mathsf{f}^1_0, \mathsf{f}^1_1, \ldots, \mathsf{f}^1_n, \ldots, \mathsf{f}^2_0, \mathsf{f}^2_1, \ldots, \mathsf{f}^2_n, \ldots, \mathsf{f}^m_0, \mathsf{f}^m_1, \ldots, \mathsf{f}^m_n, \ldots$
- *Predicate/relation symbols:* $\mathsf{P}^1_0, \mathsf{P}^1_1, \ldots, \mathsf{P}^1_n, \ldots, \mathsf{P}^2_0, \mathsf{P}^2_1, \ldots, \mathsf{P}^2_n, \ldots, \mathsf{P}^m_0, \mathsf{P}^m_1, \ldots, \mathsf{P}^m_n, \ldots$
- *Auxiliary symbols:* (,)

A first-order language with identity L is given by an effectively decidable subset of non-logical terms of the first-order alphabet, $V(L)$.

Definition (term of L)**.**

- *(Individual) variables are terms of L.*
- *Individual constants in $V(L)$ are terms of L.*
- *If ξ is an n-place function symbol in $V(L)$ and $\tau_1, \ldots, \tau_n$ are terms of L, $\xi(\tau_1, \ldots, \tau_n)$ is also a term of L.*
- *Nothing else is a term of L.*

Definition (formula of L)**.**

- *If τ_1 and τ_2 are terms of L, then $\tau_1 = \tau_2$ is a wff of L.*
- *If ρ is an n-place predicate symbol in $V(L)$ and $\tau_1, \ldots, \tau_n$ are terms of L, $\rho\tau_1, \ldots, \tau_n$ is a wff of L.*
- *If φ is a wff of L, $\neg\varphi$ is also a wff of L.*
- *If φ and ψ are wffs of L, then $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \to \psi)$, $(\varphi \leftrightarrow \psi)$ are also wffs of L.*
- *If φ is a wff of L and υ is a variable, $\forall\upsilon\,\varphi$ and $\exists\upsilon\,\varphi$ are also wffs of L.*
- *Nothing else is a wff of L.*

We identify a first-order language with the set of its wffs. The notions of scope of a quantifier, free and bound variable, free for, and sentence are (effectively) defined as usual. The usual bracketing conventions apply.

A calculus for first-order logic with identity

Let φ, ψ, and χ be wffs, υ, υ_1 and υ_2 variables, and τ a term of a first-order language with identity, L.

Definition (calculus for L)**.** *A calculus for L consists of the following axiom-schemas and rules:*

Axiom 1 $\varphi \to (\psi \to \varphi)$

Axiom 2 $(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi))$

Axiom 3 $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$

Axiom 4 $\neg(((\varphi \wedge \psi) \to \neg(\varphi \to \neg\psi)) \to \neg(\neg(\varphi \to \neg\psi) \to (\varphi \wedge \psi)))$

Axiom 5 $(\varphi \leftrightarrow \psi \to (\varphi \to \psi) \wedge (\psi \to \varphi)) \wedge ((\varphi \to \psi) \wedge (\psi \to \varphi) \to (\varphi \leftrightarrow \psi))$

Axiom 6 $\varphi \vee \psi \leftrightarrow \neg(\neg\varphi \wedge \neg\psi)$

Axiom 7 $\forall\upsilon\,(\varphi \to \psi) \to (\forall\upsilon\,\varphi \to \forall\upsilon\,\psi)$

Axiom 8 $\varphi \to \forall\upsilon\,\varphi$*, if υ is not free in φ*

Axiom 9 $\forall\upsilon\,\varphi \to \varphi[\tau/\upsilon]$

Axiom 10 $\exists\upsilon\,\varphi \leftrightarrow \neg\forall\upsilon\,\neg\varphi$

Axiom 11 $\forall\upsilon\,\upsilon = \upsilon$

Axiom 12 $\forall\upsilon_1\forall\upsilon_2\,(\upsilon_1 = \upsilon_2 \to (\varphi \to \varphi[\upsilon_2/\upsilon_1])$

Rule 1 *From φ and $\varphi \to \psi$, infer ψ*

Rule 2 *From φ infer $\forall \upsilon\, \varphi$*

Definition (derivation in the calculus for L)**.** *A derivation of a $\varphi \in L$ from a $\Gamma \subseteq L$ in the calculus for L is a sequence of wffs of L s.t. each of them is either an instance of an axiom of the calculus, a member of Γ, or follows by one of the rules from previous entries in the sequence, and the last wff of the sequence is φ.*

Definition (proof-theoretic consequence in the calculus for L)**.** *A $\varphi \in L$ is a proof-theoretic consequence of a $\Gamma \subseteq L$ in the calculus for L – $\Gamma \vdash \varphi$ – just in case there is a derivation of φ from Γ in the calculus.*