

GÖDEL'S SECOND INCOMPLETENESS THEOREM.

Sequent Calculus \mathfrak{S}

$$\begin{array}{ll}
(\text{Assm}) \frac{}{\Gamma \quad \varphi} \text{ if } \varphi \in \Gamma & (\text{Ant}) \frac{\Gamma \quad \varphi}{\Gamma' \quad \varphi} \text{ if } \Gamma \subset \Gamma' \\
(\text{PC}) \frac{\Gamma \quad \psi \quad \varphi}{\Gamma \quad \neg\psi \quad \varphi} & (\text{Ctr}) \frac{\Gamma \quad \neg\varphi \quad \psi}{\Gamma \quad \neg\varphi \quad \neg\psi} \\
(\vee A) \frac{\Gamma \quad \varphi \quad \chi}{\Gamma \quad (\varphi \vee \psi) \quad \chi} & (\vee S) \frac{\Gamma \quad \varphi}{\Gamma \quad (\varphi \vee \psi)}, \frac{\Gamma \quad \varphi}{\Gamma \quad (\psi \vee \varphi)} \\
(\exists A) \frac{\Gamma \quad \varphi \frac{y}{x} \quad \psi}{\Gamma \quad \exists x \varphi \quad \psi} \text{ if } y \text{ is not free in } \Gamma \exists x \varphi \psi & \\
(\exists S) \frac{\Gamma \quad \varphi \frac{t}{x}}{\Gamma \quad \exists x \varphi} & \\
(\equiv) \frac{}{t \equiv t} & (\text{Sub}) \frac{\Gamma \quad \varphi \frac{t}{x}}{\Gamma \quad t \equiv t' \quad \varphi \frac{t'}{x}}
\end{array}$$

The Peano Arithmetics Φ_{PA} .

$$\begin{array}{ll}
\forall x \neg x + 1 \equiv 0 & \forall x \forall y (x + 1 \equiv y + 1 \rightarrow x \equiv y) \\
\forall x x + 0 \equiv x & \forall x \forall y x + (y + 1) \equiv (x + y) + 1 \\
\forall x x \cdot 0 \equiv 0 & \forall x \forall y x \cdot (y + 1) \equiv x \cdot y + x
\end{array}$$

for all x_1, \dots, x_n, y and all $\varphi \in L^{S_{\text{ar}}}$ such that $\text{free}(\varphi) \subset \{x_1, \dots, x_n, y\}$ the sentence $\forall x_1 \dots \forall x_n \left((\varphi \frac{0}{y} \wedge \forall y (\varphi \rightarrow \varphi \frac{y+1}{y})) \rightarrow \forall y \varphi \right)$.

Encoding Finite Sequences over \mathbb{N}

Given a sequence (a_0, \dots, a_r) , let p be a prime larger than $a_0, \dots, a_r, r + 1$, and let

$$\begin{aligned}
(*) \quad t &:= 1 + a_0 p + 2p^2 + a_1 p^3 + 3p^4 + \dots + (r+1)p^{2r} + a_r p^{2r+1} \\
&= \sum_{i=0}^r [(i+1)p^{2i} + a_i p^{2i+1}].
\end{aligned}$$

By choice of p the right-hand side is the p -adic representation of t .

First, we show that for all i , $0 \leq i \leq r$,

$$\begin{aligned}
 a = a_i \quad \text{iff} \quad & \text{there are } b_0, b_1, b_2 \text{ such that} \\
 (**) \quad & \begin{aligned}
 & \text{(i) } t = b_0 + b_1((i+1) + ap + b_2p^2), \\
 & \text{(ii) } a < p, \\
 & \text{(iii) } b_0 < b_1, \\
 & \text{(iv) } b_1 = p^{2m} \text{ for a suitable } m.
 \end{aligned}
 \end{aligned}$$

The implication from left to right follows immediately from (*) with

$$\begin{aligned}
 b_0 &:= 1 \cdot p^0 + \dots + a_{i-1}p^{2i-1}, \\
 b_1 &:= p^{2i}, \\
 b_2 &:= (i+2) + a_{i+1}p + \dots + a_r p^{2(r-i)-1}.
 \end{aligned}$$

Conversely, suppose (i) - (iv) hold for b_0, b_1, b_2 and let $b_1 = p^{2m}$. From (i) we obtain

$$t = b_0 + (i+1)p^{2m} + ap^{2m+1} + b_2p^{2m+2}.$$

Since $b_0 < p^{2m}$, $a < p$, and $i+1 < p$, and since the p -adic representation of t is unique, a comparison with (*) yields $m = i$ and $a = a_i$.

Obviously, (iv) from (**) is equivalent to

$$(iv)' \quad b_1 \text{ is a square and for all } d \neq 1 \text{ with } d|b_1 \text{ we have } p|d.$$

We define $\beta(t, p, i)$ to be the uniquely determined (and hence the smallest) a for which the right-hand side of (**) (with (iv)' instead of (iv)) holds. We extend this definition to arbitrary triples of natural numbers by specifying:

Let $\beta(u, q, j)$ be the smallest a such that there are b_0, b_1, b_2 with

$$\begin{aligned}
 (i) \quad & u = b_0 + b_1((j+1) + aq + b_2q^2), \\
 (ii) \quad & a < q, \\
 (iii) \quad & b_0 < b_1, \\
 (iv)' \quad & b_1 \text{ is a square, and for all } d \neq 1 \text{ with } d|b_1 \text{ we have } q|d.
 \end{aligned}$$

If no such a exists let $\beta(u, q, j) = 0$.

Gödel's Second Incompleteness Theorem

In the following let $\Phi \subset L_0^{S_{\text{ar}}}$ be decidable and allow representations.

We choose an effective enumeration of all derivations in the sequent calculus associated with S_{ar} and define a relation H by

$$\begin{aligned}
 Hnm \quad \text{iff} \quad & \text{the } m\text{th derivation ends with a sequent of the form} \\
 & \psi_0 \dots \psi_{k-1} \varphi, \text{ where } \psi_0, \dots, \psi_{k-1} \in \Phi \text{ and } n = n^\varphi.
 \end{aligned}$$

Since Φ is decidable, so is H , and clearly

$\Phi \vdash \varphi$ iff there is $m \in \mathbb{N}$ such that $Hn^\varphi m$.

Since Φ allows representations, H can be represented in Φ by a suitable formula $\varphi_H(v_0, v_1) \in L_2^{S_{\text{ar}}}$. Again we write x, y for v_0, v_1 and set

$$\text{Der}_\Phi(x) := \exists y \varphi_H(x, y).$$

For $\psi = \neg \text{Der}_\Phi(x)$ we choose with 7.5 a fixed point $\varphi \in L_0^{S_{\text{ar}}}$, i.e. an S_{ar} -sentence φ with

$$(*) \quad \Phi \vdash \varphi \leftrightarrow \neg \text{Der}_\Phi(\mathbf{n}^\varphi).$$

φ says intuitively “I am not provable from Φ ”.

7.9 Lemma. *If $\text{Con } \Phi$ then not $\Phi \vdash \varphi$.*

Proof. Suppose $\Phi \vdash \varphi$ holds. Choose m such that $Hn^\varphi m$. Then $\Phi \vdash \varphi_H(\mathbf{n}^\varphi, \mathbf{m})$, and so $\Phi \vdash \text{Der}_\Phi(\mathbf{n}^\varphi)$. From $(*)$ we have $\Phi \vdash \neg \varphi$, and hence $\text{Inc } \Phi$. \square

Since $\Phi \vdash 0 \equiv 0$, we have

$$\text{Con } \Phi \quad \text{iff} \quad \text{not } \Phi \vdash \neg 0 \equiv 0.$$

The S_{ar} -sentence

$$\text{Consis}_\Phi := \neg \text{Der}_\Phi(\mathbf{n}^{-0 \equiv 0})$$

thus expresses the consistency of Φ . Lemma 7.9 may then be formalized as

$$(**) \quad \text{Consis}_\Phi \rightarrow \neg \text{Der}_\Phi(\mathbf{n}^\varphi).$$

An argument which is in principle simple, though technically rather tedious, could now be used to show that for $(**)$ the proof of 7.9 can be carried out on the basis of Φ , i.e., one can show that

$$(***) \quad \Phi \vdash \text{Consis}_\Phi \rightarrow \neg \text{Der}_\Phi(\mathbf{n}^\varphi),$$

in case $\Phi \supset \Phi_{\text{PA}}$ (and if a sufficiently simple formula $\varphi_H(x, y)$ to be used in Der_Φ has been chosen; cf. Exercise 7.12). Thus we obtain:

7.10 Gödel’s Second Incompleteness Theorem. *Let Φ be consistent and R -decidable with $\Phi \supset \Phi_{\text{PA}}$. Then*

$$\text{not } \Phi \vdash \text{Consis}_\Phi.$$

Proof. If $\Phi \vdash \text{Consis}_\Phi$ then by $(***)$ $\Phi \vdash \neg \text{Der}_\Phi(\mathbf{n}^\varphi)$. Since $\Phi \vdash \varphi \leftrightarrow \neg \text{Der}_\Phi(\mathbf{n}^\varphi)$ (cf. $(*)$) it would follow that $\Phi \vdash \varphi$, in contradiction to 7.9. \square

For $\Phi = \Phi_{\text{PA}}$, Gödel’s Second Incompleteness Theorem says intuitively that the consistency of Φ_{PA} cannot be proved on the basis of Φ_{PA} . This result shows that Hilbert’s program cannot be carried out in its original form. In particular, this

program aimed at a consistency proof for Φ_{PA} with elementary, so-called *finitistic* means. The concept “finitistic”, though not defined precisely, was taken in a very narrow sense; in particular it was meant that finitistic proof methods be carried out on the basis of Φ_{PA} .

The above argument can be transferred to other systems where there is a substitute for the natural numbers and where R-decidable relations and R-computable functions are representable. In particular, it applies to systems of axioms for set theory such as ZFC. One uses the natural numbers as defined in ZFC (cf. VII.7). Then one can define an $\{\in\}$ -sentence $\text{Consis}_{\text{ZFC}}$, which expresses the consistency of ZFC, to obtain:

7.11 Theorem. *If Con ZFC then not $\text{ZFC} \vdash \text{Consis}_{\text{ZFC}}$.* □

Since contemporary mathematics can be based on the ZFC axioms, and since “not $\text{ZFC} \vdash \text{Consis}_{\text{ZFC}}$ ” says that the consistency of ZFC cannot be proved using only means available within ZFC, we can formulate 7.11 as follows: If mathematics is consistent, we cannot prove its consistency by mathematical means.

In a similar way also Tarski’s Theorem and Gödel’s First Incompleteness Theorem can be transferred to axiom systems for set theory. For example, 7.8 would then assert that for every decidable and consistent system Φ of axioms for set theory which contains ZFC, there is an $\{\in\}$ -sentence φ such that neither $\Phi \vdash \varphi$ nor $\Phi \vdash \neg\varphi$. Intuitively this means that there is no decidable consistent system of axioms for mathematics which, for every mathematical statement, allows us to either prove or disprove it. In this fact an inherent limitation of the axiomatic method is manifested.

With the results of Matijasevič mentioned at the end of Section 6 we can formulate 7.11 in the following form which is easy to remember: One can write down a polynomial p in finitely many indeterminates with integer coefficients for which the following holds: Mathematics is consistent if and only if p has no (integer) root. By 7.11 we have therefore: If p has no root, then mathematics cannot prove it.

7.12 Exercise. For the (effectively given) symbol set S , fix a Gödel numbering of the S -formulas; let n^φ be the Gödel number of φ . Furthermore, for $n \in \mathbb{N}$, let \underline{n} be a variable free S -term.

For $\Phi \subset L_0^S$ let the S -formula $\text{der}(v_0)$ (“ v_0 is derivable from Φ ”) satisfy the so-called *Löb axioms*, i.e. for arbitrary $\varphi, \psi \in L^S$,

- (L1) If $\Phi \vdash \varphi$ then $\Phi \vdash \text{der}(\underline{n}^\varphi)$;
- (L2) $\Phi \vdash (\text{der}(\underline{n}^\varphi) \wedge \text{der}(\underline{n}^{(\varphi \rightarrow \psi)})) \rightarrow \text{der}(\underline{n}^\psi)$;

(L3) $\Phi \vdash (\text{der}(\underline{n}^\varphi) \rightarrow \text{der}(\underline{n^{\text{der}(\underline{n}^\varphi)}})).$

Show: If Φ is consistent and if there is an S -sentence φ_0 such that $\Phi \vdash (\varphi_0 \leftrightarrow \neg \text{der}(\underline{n^{\varphi_0}}))$, then not $\Phi \vdash \neg \text{der}(\underline{n^{\neg 0 \equiv 0}})$. (Hint: Show that, if (L1), (L2), (L3) hold for all $\varphi, \psi \in L^S$, then also $\Phi \vdash (\text{der}(\underline{n}^\varphi) \wedge \text{der}(\underline{n}^\psi) \rightarrow \text{der}(\underline{n^{(\varphi \wedge \psi)}}))$ and $\Phi \vdash (\text{der}(\underline{n^{\varphi_0}}) \rightarrow \text{der}(\underline{n^{\neg \varphi_0}})).$)