Chapter 10
The First Incompleteness Theorem

Gödel's First Incompleteness Theorem essentially states that PA is incomplete, *i.e.* there is a $\mathscr{L}_{\mathsf{PA}}$-sentence σ such that $\mathsf{PA} \nvdash \sigma$ and $\mathsf{PA} \nvdash \neg\sigma$. We prove the First Incompleteness Theorem not only for PA but also for weaker and stronger theories.

The provability predicate

In this section we state some properties of the provability predicate that we have introduced in Chapter **??**.

LEMMA 10.1. *The following statements hold:*

(a) $\mathsf{PA} \vdash \mathrm{prv}(x) \wedge \mathrm{prv}(\mathrm{imp}(x,y)) \to \mathrm{prv}(y)$

(b) $\mathsf{PA} \vdash \mathrm{prv}(x) \wedge \mathrm{prv}(y) \to \mathrm{prv}(\mathrm{and}(x,y))$.

Proof. For (a) note that $\mathrm{prv}(x)$ and $\mathrm{prv}(\mathrm{imp}(x,y))$ imply $\mathrm{mp}(x,\mathrm{imp}(x,y),y)$. Now if c, c' satisfy $\mathrm{c\_prv}(c,x)$ and $\mathrm{c\_prv}(c',\mathrm{imp}(x,y))$ then the concatenation of the codes yields $\mathrm{c\_prv}(c^\frown c'^\frown\langle y\rangle, y)$ and hence $\mathrm{prv}(y)$ as desired.

For (b) assume $\mathrm{prv}(x)$ and $\mathrm{prv}(y)$. In particular, this implies $\mathrm{fml}(x)$ and $\mathrm{fml}(y)$. note that using the formalised version of the axiom L_5 we get

$$\mathsf{PA} \vdash \mathrm{prv}(\mathrm{imp}(y,\mathrm{imp}(x,\mathrm{and}(x,y)))).$$

Using $\mathrm{prv}(y)$ and (a) we get $\mathrm{prv}(\mathrm{imp}(x,\mathrm{and}(x,y)))$ and a further application of (a) yields $\mathrm{prv}(\mathrm{and}(x,y))$. ⊣

An immediate consequence of Lemma 10.1 is the following:

COROLLARY 10.2. *Let φ and ψ be $\mathscr{L}_{\mathsf{PA}}$-formulae. Then we have*

(a) $\mathsf{PA} \vdash \mathrm{prv}(\ulcorner\varphi \to \psi\urcorner) \to (\mathrm{prv}(\ulcorner\varphi\urcorner) \to \mathrm{prv}(\ulcorner\psi\urcorner))$

(b) $\mathsf{PA} \vdash \mathrm{prv}(\ulcorner\varphi\urcorner) \wedge \mathrm{prv}(\ulcorner\psi\urcorner) \to \mathrm{prv}(\ulcorner\varphi \wedge \psi\urcorner)$.

Note that (a) corresponds to a formalised version of the inference rule (MP).

COROLLARY 10.3. *Let φ and ψ be $\mathscr{L}_{\mathsf{PA}}$-formulae. Then the following statements hold:*

(a) *If $\varphi \Leftrightarrow_{\mathsf{PA}} \psi$, then $\mathrm{prv}(\ulcorner\varphi\urcorner) \Leftrightarrow_{\mathsf{PA}} \mathrm{prv}(\ulcorner\psi\urcorner)$.*

(b) $\mathrm{prv}(\ulcorner\varphi\urcorner) \wedge \mathrm{prv}(\ulcorner\psi\urcorner) \Leftrightarrow_{\mathsf{PA}} \mathrm{prv}(\ulcorner\varphi \wedge \psi\urcorner)$.

Proof. For (a) assume that $\varphi \Leftrightarrow_{\mathsf{PA}} \psi$. By symmetry, it suffices to verify that $\mathsf{PA} \vdash \mathrm{prv}(\ulcorner\varphi\urcorner) \to \mathrm{prv}(\ulcorner\psi\urcorner)$. Since $\mathsf{PA} \vdash \varphi \to \psi$, Corollary 9.13 yields $\mathsf{PA} \vdash \mathrm{prv}(\ulcorner\varphi \to \psi\urcorner)$. The assertion then follows from Corollary 10.2 using MODUS PONENS. For (b) note that by part (b) Corollary 10.2 it suffices to prove $\mathsf{PA} \vdash \mathrm{prv}(\ulcorner\varphi \wedge \psi\urcorner) \to \mathrm{prv}(\ulcorner\varphi\urcorner) \wedge \mathrm{prv}(\ulcorner\psi\urcorner)$. But this is a direct consequence of Corollary 10.2 (a) using L_3 and L_4. ⊣

The Diagonalisation Lemma

Standard natural numbers are either 0 or the successor $\mathsf{s}n$ of a standard natural number n. Hence we can introduce a binary relation which states that x codes the natural number n in the following way:

$$\begin{aligned}\mathrm{c\_nat}(c,n,x) :\Longleftrightarrow\ & \mathrm{seq}(c) \wedge \mathrm{lh}(c) = \mathsf{s}n \wedge c_{\mathsf{0}} = \ulcorner\mathsf{0}\urcorner \wedge \forall i < n \\ & (c_{\mathsf{s}i} = \mathrm{succ}(c_i) \wedge c_n = x) \\ \mathrm{nat}(n,x) :\Longleftrightarrow\ & \exists c(\mathrm{c\_nat}(c,n,x)).\end{aligned}$$

Clearly, it follows from the definition that

$$\mathsf{PA} \vdash \mathrm{c\_nat}(c,n,x) \to \mathrm{c\_nat}(c^\frown\langle\mathrm{succ}(x)\rangle, \mathsf{s}n, \mathrm{succ}(x)).$$

LEMMA 10.4. *For any natural number $n \in \mathbb{N}$ we have $\mathsf{PA} \vdash \mathrm{nat}(\underline{n}, \ulcorner\underline{n}\urcorner)$. In particular, if φ is an $\mathscr{L}_{\mathsf{PA}}$-formula, then $\mathsf{PA} \vdash \mathrm{nat}(\ulcorner\varphi\urcorner, \ulcorner\ulcorner\varphi\urcorner\urcorner)$.*

Proof. We proceed by metainduction on n. For $n = 0$ the term $\underline{0}$ is the same as $\mathsf{0}$ and clearly the singleton sequence $c = \langle\ulcorner\mathsf{0}\urcorner\rangle$ witnesses $\mathrm{c\_nat}(c, \mathsf{0}, \ulcorner\mathsf{0}\urcorner)$. Now suppose that the claim holds for some $n \in \mathbb{N}$. Then there is c such that $\mathrm{c\_nat}(c, \underline{n}, \ulcorner\underline{n}\urcorner)$. We put $c' = c^\frown\langle\ulcorner\underline{\mathsf{s}n}\urcorner\rangle$. Notice that then $\mathrm{lh}(c') = \mathsf{s}\mathsf{s}\underline{n}$ and $(c')_{\mathsf{s}\underline{n}} = \ulcorner\underline{\mathsf{s}n}\urcorner = \mathrm{succ}(\ulcorner\underline{n}\urcorner)$. Using the induction hypothesis and the observation above we obtain $\mathrm{c\_nat}(c', \underline{\mathsf{s}n}, \ulcorner\underline{\mathsf{s}n}\urcorner)$. ⊣

We define

$$\mathrm{gn}(n) = x :\Longleftrightarrow \mathrm{nat}(n,x) \vee \neg\exists y(\mathrm{nat}(n,y) \wedge x = \mathsf{0}).$$

This indeed defines a function, since one can easily prove that $\mathsf{PA} \vdash \mathrm{nat}(n,x) \wedge \mathrm{nat}(n,y) \rightarrow x = y$ using the definition of the predicate seq. In particular, by Lemma 10.4 we have

$$\mathsf{PA} \vdash \mathrm{gn}(\ulcorner\varphi\urcorner) = \ulcorner\ulcorner\varphi\urcorner\urcorner. \tag{$*$}$$

Now we have assembled all the ingredients to prove the DIAGONALISATION LEMMA which is an important tool for the proof of Gödel's Incompleteness Theorems.

In order to simplify the notation, for $\mathscr{L}_{\mathsf{PA}}$-formulae φ and ψ we shall write $\varphi \Leftrightarrow_{\mathsf{PA}} \varphi$ to denote that $\mathsf{PA} \vdash \varphi \leftrightarrow \psi$.

THEOREM 10.5 (DIAGONALISATION LEMMA). *Let $\varphi(\nu)$ be an $\mathscr{L}_{\mathsf{PA}}$-formula with one free variable ν which does not occur bound in φ. Then there exists an $\mathscr{L}_{\mathsf{PA}}$-sentence σ_φ such that*

$$\mathsf{PA} \vdash \sigma \leftrightarrow \varphi(\nu/\ulcorner\sigma_\varphi\urcorner)\,,$$

i.e., $\sigma \Leftrightarrow_{\mathsf{PA}} \varphi(\ulcorner\sigma_\varphi\urcorner)$.

Proof. Recall that $\ulcorner v_0\urcorner = \mathtt{s0}$ and define

$$\psi(v_0) :\equiv \forall v_1 \big(\mathrm{sb\_fml}(\mathtt{s0}, \mathrm{gn}(v_0), v_0, v_1) \rightarrow \varphi(\nu/v_1)\big)$$

and

$$\sigma_\varphi :\equiv \psi(v_0/\ulcorner\psi\urcorner).$$

In other words, $\sigma_\varphi \equiv \psi(\ulcorner\psi\urcorner)$ and $\ulcorner\sigma_\varphi\urcorner = \ulcorner\psi(\ulcorner\psi\urcorner)\urcorner$. Then we have

$$\begin{aligned}
\sigma_\varphi &\equiv \forall v_1 \big(\mathrm{sb\_fml}(\mathtt{s0}, \mathrm{gn}(\ulcorner\psi(v_0)\urcorner), \ulcorner\psi(v_0)\urcorner, v_1) \rightarrow \varphi(\nu/v_1)\big)\\
&\Leftrightarrow_{\mathsf{PA}} \forall v_1 \big(\mathrm{sb\_fml}(\mathtt{s0}, \ulcorner\ulcorner\psi(v_0)\urcorner\urcorner, \ulcorner\psi(v_0)\urcorner, v_1) \rightarrow \varphi(v_1)\big)\\
&\Leftrightarrow_{\mathsf{PA}} \forall v_1 \big(v_1 = \ulcorner\psi(v_0/\ulcorner\psi(v_0)\urcorner)\urcorner \rightarrow \varphi(v_1)\big)\\
&\Leftrightarrow_{\mathsf{PA}} \varphi(\ulcorner\psi(v_0/\ulcorner\psi(v_0)\urcorner)\urcorner)\\
&\equiv \varphi(\ulcorner\psi(\ulcorner\psi\urcorner)\urcorner)\\
&\equiv \varphi(\ulcorner\sigma_\varphi\urcorner),
\end{aligned}$$

where the first equivalence follows from ($*$) and the second equivalence follows from Lemma 9.11. ⊣

The DIAGONALISATION LEMMA is often called FIXPOINT LEMMA, since the sentence σ can be conceived as a fixed point of σ. It is a powerful tool, since it allows us to make self-referential statements, *i.e.* for a formula φ with one free variable it provides a sentence σ which states "I have the property φ".

The First Incompleteness Theorem

Now we are ready to prove a first version of Gödel's FIRST INCOMPLETENESS THEOREM:

THEOREM 10.6 (FIRST INCOMPLETENESS THEOREM FOR PA). *PA is incomplete.*

Proof. By the DIAGONALISATION LEMMA there is an $\mathscr{L}_{\mathsf{PA}}$-sentence σ such that

$$\sigma \Leftrightarrow_{\mathsf{PA}} \neg\,\mathrm{prv}(\ulcorner\sigma\urcorner).$$

To see this, let $\varphi(v_0) :\equiv \neg\,\mathrm{prv}(v_0)$. Then $\sigma_\varphi \Leftrightarrow_{\mathsf{PA}} \varphi(\ulcorner\sigma_\varphi\urcorner)$ and $\varphi(\ulcorner\sigma_\varphi\urcorner) \equiv \varphi(v_0/\ulcorner\sigma_\varphi\urcorner) \equiv \neg\,\mathrm{prv}(v_0/\ulcorner\sigma_\varphi\urcorner) \equiv \neg\,\mathrm{prv}(\ulcorner\sigma_\varphi\urcorner)$.

Suppose for a contradiction that PA is complete. Then there are two cases:

Case 1. $\mathsf{PA} \vdash \sigma$. Then by Corollary 9.13 we have $\mathsf{PA} \vdash \mathrm{prv}(\ulcorner\sigma\urcorner)$. On the other hand, since $\sigma \Leftrightarrow_{\mathsf{PA}} \neg\,\mathrm{prv}(\ulcorner\sigma\urcorner)$, we have $\mathsf{PA} \vdash \neg\,\mathrm{prv}(\ulcorner\sigma\urcorner)$ and so $\mathsf{PA} \vdash$ ⧉. But since $\mathbb{N} \vDash \mathsf{PA}$, this contradicts the SOUNDNESS THEOREM.

Case 2. $\mathsf{PA} \vdash \neg\sigma$. From

$$\neg\sigma \Leftrightarrow_{\mathsf{PA}} \neg\neg\,\mathrm{prv}(\ulcorner\sigma\urcorner) \Leftrightarrow_{\mathsf{PA}} \mathrm{prv}(\ulcorner\sigma\urcorner)$$

we obtain $\mathsf{PA} \vdash \mathrm{prv}(\ulcorner\sigma\urcorner)$. In particular, $\mathbb{N} \vDash \mathrm{prv}(\#\sigma)$ and so there exists $n \in \mathbb{N}$ with $\mathbb{N} \vDash \mathrm{c\_prv}(n, \#\sigma)$. But then by Lemma 9.12, n codes a formal proof of σ and so $\mathsf{PA} \vdash \sigma$, a contradiction.

Since both cases lead to a contradiction, PA is incomplete. $\dashv$

In the proof of Theorem 10.6 above we proved that a sentence σ with the property $\sigma \Leftrightarrow_{\mathsf{PA}} \neg\,\mathrm{prv}(\ulcorner\sigma\urcorner)$ witnesses the incompleteness of PA. In $\mathbb{N}$ however, σ is true: Note that if $\mathbb{N} \vDash \neg\sigma$, then $\mathbb{N} \vDash \mathrm{prv}(\#\sigma)$. But then Lemma 9.12 would imply $\mathsf{PA} \vdash \sigma$ and hence also $\mathbb{N} \vDash \sigma$, a contradiction. Observe that in $\mathbb{N}$ the sentence σ expresses "I am not provable" – where provable is meant with respect to prv – which is, of course, true.

Completeness and Incompleteness of Theories of Arithmetic

A first attempt to deal with the incompleteness phenomenon might be to replace PA with $\mathsf{T} \equiv \mathsf{PA} + \sigma$, since $\mathbb{N} \vDash \mathsf{T}$. Moreover, the gödelisation process could be done in the same way, where one would just need to code an additional axiom, namely σ. However this would lead to a modified provability predicate $\mathrm{prv}_{\mathsf{T}}$ which additionally allows formal proofs to be initialised with σ. One could then prove a version of the DIAGONALISATION LEMMA which would allow us to define a version σ_{T} of σ with the property

$$\mathsf{T} \vdash \sigma_{\mathsf{T}} \leftrightarrow \neg\,\mathrm{prv}_{\mathsf{T}}(\ulcorner\sigma_{\mathsf{T}}\urcorner).$$

But then we obtain a version of the FIRST INCOMPLETENESS THEOREM, since $\mathsf{T} \nvdash \sigma_{\mathsf{T}}$ and $\mathsf{T} \nvdash \neg\sigma_{\mathsf{T}}$. This suggests that Theorem 10.6 can be generalised. This is exactly the goal of this section, whereby we consider both weaker and stronger theories than PA.

We investigate how much of PA is really needed for the incompleteness proof. As we have seen that exponentiation can be expressed using addition and multiplication, one idea might be to leave out multiplication and thus delete PA_4 and PA_5. The resulting theory, called **Presburger Arithmetic**, will however turn out to be complete (see Chapter **??**). The most critical axiom is certainly the induction schema PA_6, so we might consider the theory with PA_6 deleted. This is still not strong enough, but as we will see, one instance of PA_6 actually suffices. **Robinson Arithmetic** RA is the axiom system consisting of PA_0-PA_5 and the additional axiom

$$\forall x(x = \mathsf{0} \vee \exists y(x = \mathsf{s}y)).$$

The language of RA is also $\mathscr{L}_{\mathsf{PA}}$, so we can express the same statements as in PA but prove less theorems. Thus it is clear that RA must be incomplete. In fact, RA is so weak that it fails to prove $\forall(\mathsf{0} + x = x)$:

Example 10.1. We show that $\mathsf{RA} \nvdash \forall x(\mathsf{0}+x = x)$ and hence, in particular, RA fails to prove that addition is commutative. To achieve this, we provide a model $\mathbf{M}$ of RA in which $\forall x(\mathsf{0}+x = x)$ is false. The domain of the model is $M = \mathbb{N}\cup\{a, b\}$, where a and b are any two distinct mathematical objects which are not in $\mathbb{N}$. Furthermore, let $\bar{a} \equiv b$ and $\bar{b} \equiv a$. Then we can interpret $\mathsf{0}^{\mathbf{M}}$ by 0 and the functions by

$$\mathsf{s}^{\mathbf{M}}(x) \equiv \begin{cases} \mathsf{s}^{\mathbb{N}}(x) & x \in \mathbb{N} \\ x & x \in \{a, b\} \end{cases}$$

$$x +^{\mathbf{M}} y \equiv \begin{cases} x +^{\mathbb{N}} y & x, y \in \mathbb{N} \\ x & y \in \mathbb{N} \text{ and } x \notin \mathbb{N} \\ \bar{y} & y \notin \mathbb{N}. \end{cases}$$

$$x \cdot^{\mathbf{M}} y \equiv \begin{cases} x \cdot^{\mathbb{N}} y & x, y \in \mathbb{N} \\ y & y \in \{0, a, b\} \\ \bar{x} & y \not\equiv 0 \text{ and } x \in \{a, b\}. \end{cases}$$

It is easy to check that $\mathbf{M}$ is a model of RA, and $0 +^{\mathbf{M}} b \equiv a \not\equiv b \equiv b +^{\mathbf{M}} 0$.

Note that N_0–N_5 in Proposition 9.1 are also provable in RA, since the proof uses metainduction rather than induction in PA and the only non-trivial argument uses Lemma 8.6 which can easily be seen to hold in RA.

In the following, we prove that all relations and functions that are introduced in Chapters **??** and **??** are $\mathbb{N}$-conform. To achieve this, we prove that each such relation and function can be defined both by an $\exists$-formula and a $\forall$-formula. The representations with an $\exists$-formula are already given, and functions defined by an $\exists$-formula always have an equivalent definition by a $\forall$-formula by part(b) of Corollary 9.3.

The only relations whose representation by a $\forall$-formula is non-trivial, are term, fml as well as all relations used to formalise substitution and formal proofs. Note that if we are able to show that the existential quantifiers in term and fml can be replaced by a bounded existential quantifier, then the same can be achieved for all subsequent relations.

LEMMA 10.7. *If ψ is a formula of the form $\psi \equiv \exists c(\mathrm{seq}(c) \wedge \varphi(c))$ for some Δ-formula φ, and there is a term τ whose variables are among $\mathrm{free}(\psi)$ such that*

$$\mathsf{PA} \vdash \mathrm{seq}(c) \wedge \varphi(c) \to (\mathrm{lh}(c) < \tau \wedge \forall i < \mathrm{lh}(c)(c_i < \tau))$$

then ψ is also a Δ-formula.

Proof. We go once more through the proof of Theorem 9.8 and show that the quantifier $\exists c$ can be replaced by a bounded quantifier.

Suppose that $F(i)$ is a function defined by a Δ-formula. Let $F'(i) = \mathrm{op}(\tau, i) + 1$ and $m = \max_{i<\tau} F'(i)$. Moreover, note that by Exercise 9.1 we can define factorials in PA. Let $y = m!$. Furthermore, put $G(j) = 1 + (j+1)y$. By Lemma 8.14, we have that for all $i, j < m$, $G(i)$ and $G(j)$ are coprime. Now Lemma 9.7 allows us to pick x with $\chi(x)$, where

$$\chi(x) \equiv \forall j < m(G(j) \mid x \leftrightarrow \exists i < \tau(j = \mathrm{op}(\tau, i)).$$

We check that if $F(i) < \tau$ for every $i < \tau$ then we can find an upper bound τ' whose variables coincide with the variables of τ such that there is $c < \tau'$ with $\beta(c, i) = F(i)$ for all $i < \tau$. If this can be accomplished, then we have

$$\psi \Leftrightarrow_{\mathsf{PA}} \exists c < \tau'(\mathrm{seq}(c) \wedge \varphi(c)) :$$

To see this, suppose that $\mathrm{seq}(c) \wedge \varphi(c)$ with $c \geq \tau'$. Now take $F(i) := \beta(c, i) < \tau$. By our assumption, there is $c' < \tau' \leq c$ with $\beta(c', i) = F(i) = \beta(c, i)$ for all $i < \tau$. Moreover, note that $\mathrm{lh}(c') = \beta(c', 0) = \beta(c, 0) = \mathrm{lh}(c)$ and $\mathrm{lh}(c') = F(0) < \tau$ and hence $c'_i = c_i$ for all $i < \mathrm{lh}(c)$, contradicting $\mathrm{seq}(c)$.

It remains to find τ'. Note that we clearly have $m \leq \tau_1$ with $\tau_1 \equiv \mathrm{op}(\tau, \tau) + 1$ and hence $y \leq \tau_1!$. Furthermore, we have $G(j) < 1 + (\tau_1 + 1)!$ for each $j < m$. Therefore, since $G(i)$ and $G(j)$ are coprime for all $i, j < m$, we can find x which satisfies $\chi(x)$ such that $x < \tau_2$ with $\tau_2 \equiv (1 + (\tau_1 + 1)!)^{\tau_1}$. In particular, there is $c = \mathrm{op}(x, y)$ with $\mathrm{seq}(c) \wedge \varphi(c)$ and $c < \mathrm{op}(\tau_1, \tau_2)$.

$\dashv$

LEMMA 10.8. *The relations term and fml are $\mathbb{N}$-conform.*

Proof. We want to apply Lemma 10.7 to the defining formulae of term and $\mathrm{formula}$. Since both cases are similar, we only consider term. We prove that $\exists c\,\mathrm{c\_term}(c, t)$ is equivalent to the formula

$$\varphi(t) \equiv \exists c(\mathrm{c\_term}(c,t) \wedge \forall i < \mathrm{lh}(c) \forall j < i(c_j < c_i)).$$

Then Lemma 10.7 for $\tau \equiv t + 1$ concludes the proof. We proceed by strong induction on $\mathrm{lh}(c)$. If $\mathrm{lh}(c) = \mathtt{1}$, then there is nothing to prove. Suppose now that for all $t' < t$, $\mathrm{term}(t) \to \varphi(t)$ holds and assume $\mathrm{c\_term}(c,t)$. If $t = \mathtt{0}$ or $\mathrm{var}(t)$, then $\mathrm{c\_term}(\langle t \rangle, t)$ and hence $\varphi(t)$ holds. Hence we either have $t = \mathrm{succ}(c_i), t = \mathrm{add}(c_i, c_j)$ or $t = \mathrm{mult}(c_i, c_j)$ for $i, j < \mathrm{lh}(c)$. We only focus on the first case, since the others can be handled in the same way. Note that by Exercise 9.6 we can restrict c to $\langle c_j \mid j \leq i \rangle$ which we denote by $c \restriction \mathtt{s}c_i$. Clearly, $c_i < c$ and $\mathrm{c\_term}(c \restriction \mathtt{s}i, c_i)$. Hence by our induction hypothesis, there is d with $\mathrm{c\_term}(d, c_i)$ and $d_k < d_j$ for all $j < \mathrm{lh}(d)$ and $k < j$. But then $d^\frown\langle t \rangle$ witnesses $\varphi(t)$. ⊣

Lemma 10.8 implies that if $n \in \mathbb{N}$ is a natural number which is not the Gödel number of a term or formula, then

$$\mathsf{RA} \vdash \neg\, \mathrm{term}(\underline{n})$$
$$\mathsf{RA} \vdash \neg\, \mathrm{fml}(\underline{n}).$$

Moreover, the relation c_prv is also a Δ-formula and hence

$$\mathsf{RA} \vdash \neg\, \mathrm{c\_prv}(\underline{n}, \ulcorner \varphi \urcorner)$$

whenever n does not encode a formal proof of φ. However, the existential quantifier in the definition of the provability relation prv cannot be bounded: Otherwise $\mathsf{RA} \nvdash \varphi$ would imply $\mathsf{RA} \vdash \neg\, \mathrm{prv}(\ulcorner \varphi \urcorner)$, contradicting the incompleteness of RA.

There are two ways to generalise the FIRST INCOMPLETENESS THEOREM: Firstly, one can modify the underlying language and, secondly, one can use a different axiom system. If the language satisfies $\mathscr{L} \supseteq \mathscr{L}_{\mathsf{PA}}$ and we have $\mathbb{N}$-conformity of all relevant relations, then, as we shall see, the proof can easily be transferred to the new setting. However, there are two issues which are affected: The gödelisation of the language, and the gödelisation of the axioms. The coding of terms, formulae and proofs can then be realised in the same way as in Chapter **??**.

A language $\mathscr{L} \supseteq \mathscr{L}_{\mathsf{PA}}$ is said to be **gödelisable**, if it is countable. Note that if $\mathscr{L}$ is gödelisable, then its constant, relation and function symbols admit Gödel coding as described in Chapter **??**. A theory T in some gödelisable language $\mathscr{L} \supseteq \mathscr{L}_{\mathsf{PA}}$ is **gödelisable**, if there is a Δ-formula ax_{T} in the language $\mathscr{L}_{\mathsf{PA}}$ with the property that $\mathbb{N} \vDash \mathrm{ax}_{\mathsf{T}}(\#\varphi)$ if and only if $\varphi \in \mathsf{T}$, where $\#\varphi$ is the Gödel code of φ. As in the case of PA, we introduce Gödel codes on the formal level by $\ulcorner \varphi \urcorner :\equiv \underline{\#\varphi}$.

Refer somehow to recursion theory

Note that if T is gödelisable and satisfies $\mathsf{N}_0 - \mathsf{N}_5$, then by Corollary 9.3 every Δ-formula φ in the language $\mathscr{L}_{\mathsf{PA}}$ is $\mathbb{N}$-conform. In particular, by Lemma 10.8 is possible to define Δ-formulae $\mathrm{term}_{\mathsf{T}}$ and $\mathrm{fml}_{\mathsf{T}}$ such that

$$\mathbb{N} \vDash \mathrm{term}_{\mathsf{T}}(n) \quad \Longleftrightarrow \quad n \equiv \#\tau \text{ for some } \mathscr{L}\text{-term } \tau$$

$$\mathbb{N} \vDash \mathrm{fml}_{\mathsf{T}}(n) \quad \Longleftrightarrow \quad n \equiv \#\varphi \text{ for some } \mathscr{L}\text{-formula } \varphi.$$

Moreover, by gödelisability of T, the axioms can be coded by some Δ-formula ax_{T}. One can then proceed to define a Δ-formula $\mathrm{c\_prv}_{\mathsf{T}}$ and an $\exists$-formula $\mathrm{prv}_{\mathsf{T}}$ such that

$$\begin{aligned} \mathbb{N} \vDash \mathrm{c\_prv}_{\mathsf{T}}(n, \#\varphi) \quad &\Longleftrightarrow \quad n \text{ codes a formal proof of } \varphi \\ \mathbb{N} \vDash \mathrm{prv}_{\mathsf{T}}(\#\varphi) \quad &\Longleftrightarrow \quad \mathsf{T} \vdash \varphi \end{aligned}$$

for every $n \in \mathbb{N}$ and $\mathscr{L}$-formula φ. Notice that it is crucial that $\mathrm{c\_prv}_{\mathsf{T}}$ and $\mathrm{prv}_{\mathsf{T}}$ are $\mathscr{L}_{\mathsf{PA}}$-formulae, since otherwise we would have to specify how to interpret them in the standard model $\mathbb{N}$. Moreover, using Corollary 9.3, we obtain

$$\begin{aligned} \mathsf{P}_0: \qquad \mathbb{N} \vDash \mathrm{c\_prv}_{\mathsf{T}}(n, \#\varphi) \quad &\Longrightarrow \quad \mathsf{T} \vdash \mathrm{c\_prv}_{\mathsf{T}}(\underline{n}, \ulcorner\varphi\urcorner) \\ \mathsf{P}_1: \qquad \mathbb{N} \vDash \neg\, \mathrm{c\_prv}_{\mathsf{T}}(n, \#\varphi) \quad &\Longrightarrow \quad \mathsf{T} \vdash \neg\, \mathrm{c\_prv}_{\mathsf{T}}(\underline{n}, \ulcorner\varphi\urcorner). \end{aligned}$$

In the following, we present two proofs of the FIRST INCOMPLETENESS THEOREM for gödelisable theories $\mathsf{T} \supseteq \mathsf{RA}$. The restriction to extensions of RA ensures that N_0 –N_5 and hence also Corollary 9.3 hold.

Gödel's original proof uses the assumption of a slightly stronger property than consistency: An $\mathscr{L}_{\mathsf{PA}}$-theory T is said to be ***ω*-consistent**, if whenever $\mathsf{T} \vdash \exists x \varphi(x)$ for some $\mathscr{L}_{\mathsf{PA}}$-formula $\varphi(x)$, then there exists $n \in \mathbb{N}$ such that $\mathsf{T} \nvdash \neg\varphi(\underline{n})$.

FACT 10.9. *If* T *is an* $\mathscr{L}_{\mathsf{PA}}$*-theory with* $\mathbb{N} \vDash \mathsf{T}$, *then* T *is* ω*-consistent. In particular,* PA *and* RA *are* ω*-consistent.*

Proof. If $\mathsf{T} \vdash \exists x \varphi(x)$, then $\mathbb{N} \vDash \exists x \varphi(x)$. Hence there is $n \in \mathbb{N}$ with $\mathbb{N} \vDash \varphi(n)$. But then $\mathsf{T} + \varphi(\underline{n})$ is consistent and so $\mathsf{T} \nvdash \neg\varphi(\underline{n})$. ⊣

THEOREM 10.10 (FIRST INCOMPLETENESS THEOREM, GÖDEL'S VERSION). *Let* $\mathsf{T} \supseteq \mathsf{RA}$ *be a gödelisable* $\mathscr{L}_{\mathsf{PA}}$*-theory. If* T *is* ω*-consistent, then* T *is incomplete.*

Proof. Observe that the proof of DIAGONALISATION LEMMA still works if we replace PA by T. Take a sentence σ such that

$$\sigma \Leftrightarrow_{\mathsf{PA}} \neg\, \mathrm{prv}_{\mathsf{T}}(\ulcorner\sigma\urcorner).$$

Suppose for a contradiction that T is complete. Then we have that either $\mathsf{T} \vdash \sigma$ or $\mathsf{T} \vdash \neg\sigma$.

Case 1. $\mathsf{T} \vdash \sigma$. In this case the argument is the same as in Theorem 10.6.

Case 2. $\mathsf{T} \vdash \neg\sigma$. Then $\mathsf{T} \vdash \mathrm{prv}_{\mathsf{T}}(\ulcorner\neg\sigma\urcorner)$. On the other hand, by assumption $\neg\sigma \Leftrightarrow_{\mathsf{T}} \neg\neg\, \mathrm{prv}_{\mathsf{T}}(\ulcorner\sigma\urcorner) \Leftrightarrow_{\mathsf{T}} \mathrm{prv}_{\mathsf{T}}(\ulcorner\sigma\urcorner)$ and so $\mathsf{T} \vdash \mathrm{prv}_{\mathsf{T}}(\ulcorner\sigma\urcorner)$. By Corollary 10.2 we have $\mathsf{T} \vdash \mathrm{prv}_{\mathsf{T}}(\ulcorner\sigma \wedge \neg\sigma\urcorner)$ and so by ω-consistency there is $n \in \mathbb{N}$ such that $\mathsf{T} \nvdash \neg\, \mathrm{c\_prv}_{\mathsf{T}}(\underline{n}, \ulcorner\sigma \wedge \neg\sigma\urcorner)$. By ω-consistency there is $n \in \mathbb{N}$ such that $\mathsf{T} \nvdash \neg\, \mathrm{c\_prv}_{\mathsf{T}}(\underline{n}, \ulcorner\sigma \wedge \neg\sigma\urcorner)$. However, since T is consistent, $\mathsf{T} \nvdash \sigma \wedge \neg\sigma$ and so $\mathbb{N} \models \neg\, \mathrm{c\_prv}_{\mathsf{T}}(n, \#(\sigma \wedge \neg\sigma))$. But then P_1 implies $\mathsf{T} \vdash \neg\, \mathrm{c\_prv}_{\mathsf{T}}(\underline{n}, \ulcorner \sigma \wedge \neg\sigma \urcorner)$, a contradiction. ⊣

Rosser showed in [**?**] how to get rid of this dependency on ω-consistency by modifying slightly the provability predicate:

$$\begin{aligned}
\mathrm{c\_prv}^{\mathsf{R}}_{\mathsf{T}}(c, x) &:\Longleftrightarrow \mathrm{c\_prv}_{\mathsf{T}}(c, x) \wedge \neg\exists c' < c(\mathrm{c\_prv}_{\mathsf{T}}(c', \mathrm{not}(x))) \\
\mathrm{prv}^{\mathsf{R}}_{\mathsf{T}}(x) &:\Longleftrightarrow \exists c(\mathrm{c\_prv}^{\mathsf{R}}_{\mathsf{T}}(c, x)).
\end{aligned}$$

THEOREM 10.11 (FIRST INCOMPLETENESS THEOREM, USING ROSSER'S TRICK). *Let $\mathscr{L} \supseteq \mathscr{L}_{\mathsf{PA}}$ be a gödelisable language and let T be a gödelisable $\mathscr{L}$-theory. If T is consistent, then it is incomplete.*

Proof. As before, we want to apply the DIAGONALISATION LEMMA; this time to the formula $\neg\, \mathrm{prv}^{\mathsf{R}}(x)$. Thus we obtain an $\mathscr{L}$-sentence σ with

$$\sigma \Leftrightarrow_{\mathsf{PA}} \neg\, \mathrm{prv}^{\mathsf{R}}(\ulcorner \sigma \urcorner).$$

Again, we want to prove that neither σ nor its negation can be inferred from T. Observe first that our assumption on σ implies

$$\sigma \Leftrightarrow_{\mathsf{PA}} \forall c(\mathrm{c\_prv}(c, \ulcorner \sigma \urcorner) \to \exists c' < c(\mathrm{c\_prv}(c', \ulcorner \neg\sigma \urcorner)))$$

since $\mathrm{not}(\ulcorner \sigma \urcorner) = \ulcorner \neg\sigma \urcorner$. Assume, towards a contradiction, that T is complete. As before, we have two cases:

Case 1. $\mathsf{T} \vdash \sigma$. Then by P_0 there is $n \in \mathbb{N}$ such that $\mathsf{T} \vdash \mathrm{c\_prv}_{\mathsf{T}}(\underline{n}, \ulcorner \sigma \urcorner)$. On the other hand, by our computation above we have $\mathsf{T} \vdash \exists c < \underline{n}(\mathrm{c\_prv}_{\mathsf{T}}(c, \ulcorner \neg\sigma \urcorner))$. Since T satisfies N_5, this means that there exists $k < n$ in $\mathbb{N}$ such that $\mathsf{T} \vdash \mathrm{c\_prv}_{\mathsf{T}}(\underline{k}, \ulcorner \neg\sigma \urcorner)$. But then there is $m \in \mathbb{N}$ with $\mathsf{T} \vdash \mathrm{c\_prv}_{\mathsf{T}}(\underline{m}, \ulcorner \sigma \wedge \neg\sigma \urcorner)$. But then by $\mathbb{N}$-conformity of $\mathrm{c\_prv}_{\mathsf{T}}$, $\mathbb{N} \models \mathrm{c\_prv}_{\mathsf{T}}(m, \#(\sigma \wedge \neg\sigma))$ and so $\mathsf{T} \vdash \sigma \wedge \neg\sigma$, contradicting our assumption that T is consistent.

Case 1. $\mathsf{T} \vdash \neg\sigma$. Then there is $n \in \mathbb{N}$ such that $\mathsf{T} \vdash \mathrm{c\_prv}_{\mathsf{T}}(\underline{n}, \ulcorner \neg\sigma \urcorner)$. On the other hand, we have $\mathsf{T} \vdash \mathrm{prv}^{\mathsf{R}}_{\mathsf{T}}(\ulcorner \sigma \urcorner)$ and hence there is c with $\mathrm{c\_prv}^{\mathsf{R}}_{\mathsf{T}}(c, \ulcorner \sigma \urcorner)$. By definition of $\mathrm{c\_prv}^{\mathsf{R}}_{\mathsf{T}}$, we get $c < \underline{n}$. Now we can use N_5 to reach the same contradiction as in Case 1. ⊣

Tarski's Theorem

The DIAGONALISATION LEMMA allows us to make self-referential statements such as the Gödel sentence which formalizes the sentence "This sentence is not provable". Recall that we call an $\mathscr{L}_{\mathsf{PA}}$-sentence φ **true** in $\mathbb{N}$, if $\mathbb{N} \models \varphi$. Is it possible to express truth in the standard model $\mathbb{N}$ by a formula, *i.e.* is there a formula $\mathrm{truth}(x)$ with one free variable x such that for every $\mathscr{L}_{\mathsf{PA}}$-sentence φ,

$$\mathbb{N} \vDash \mathrm{truth}(\#\varphi) \quad \Longleftrightarrow \quad \mathbb{N} \vDash \varphi$$

which is equivalent to

$$\mathbb{N} \vDash \mathrm{truth}(\#\varphi) \leftrightarrow \varphi \,?$$

Using the DIAGONALISATION LEMMA we provide a negative answer.

THEOREM 10.12 (TARSKI'S THEOREM). *There is no $\mathscr{L}_{\mathsf{PA}}$-formula $\mathrm{truth}(x)$ with one free variable x such that $\mathbb{N} \vDash \mathrm{truth}(\#\varphi) \leftrightarrow \varphi$.*

Proof. Suppose for a contradiction that such a formula truth exists. By the DIAGONALISATION LEMMA there exists an $\mathscr{L}_{\mathsf{PA}}$-sentence σ such that

$$\mathsf{PA} \vdash \sigma \leftrightarrow \neg\, \mathrm{truth}(\ulcorner \sigma \urcorner).$$

But then

$$\begin{aligned} \mathbb{N} \vDash \mathrm{truth}(\#\sigma) \quad &\Longleftrightarrow \quad \mathbb{N} \vDash \sigma \\ &\Longleftrightarrow \quad \mathbb{N} \vDash \neg\, \mathrm{truth}(\#\sigma) \end{aligned}$$

which is impossible. ⊣

Note that we have solved the so-called **Liar paradox** concerned with the sentence

"This sentence is false"

which is obviously true iff it is false. Clearly, the above sentence corresponds to the sentence σ in the proof of TARSKI'S THEOREM. In order to express it (in PA) one would need to be able to define truth which is impossible by TARSKI'S THEOREM.

NOTES

The FIRST INCOMPLETENESS THEOREM war first proven by Kurt Gödel [**?**] in 1931. Rather than using Peano Arithmetic in first-order logic as we did, he based his proof on type theory in the system of Principia Mathematica [**?**] introduced by Russell and Whitehead. Gödel's original proof makes use of the stronger assumption of ω-consistency, which Barkely Rosser [**?**] showed to be negligible. The observation that all proof steps of the FIRST INCOMPLETENESS THEOREM can in fact be carried out in Robinson Arithmetic was made by R.M. Robinson [**?**] in 1950. Although TARSKI'S THEOREM is usually attributed to Alfred Tarksi and was first published by him [**?**], Gödel already mentioned this result in a 1931 letter to Paul Bernays; previously he had been trying to come up with a definition of a truth predicate (see [**?**]).

EXERCISES

10.0 Prove $\mathsf{PA} \vdash \forall x(\mathrm{term}(\mathrm{gn}(x)))$.

10.1 Let φ and ψ be $\mathscr{L}_{\mathsf{PA}}$-formulae. Show that

$$\mathsf{PA} \vdash \operatorname{prv}(\ulcorner\varphi\urcorner) \vee \operatorname{prv}(\ulcorner\psi\urcorner) \to \operatorname{prv}(\ulcorner\varphi \vee \psi\urcorner).$$

Does the converse also hold?

10.2 A theory T with signature $\mathscr{L}_{\mathsf{PA}}$ is said to be ω**-incomplete**, if it holds that $\mathsf{T} \vdash \varphi(\underline{n})$ for every $n \in \mathbb{N}$ but $\mathsf{T} \nvdash \forall x\varphi$.

(a) Show that PA is ω-incomplete.

(b) Show that every ω-complete $\mathscr{L}_{\mathsf{PA}}$-theory has an extension which is consistent but ω-inconsistent.

10.3 Let $\varphi(x, y)$ and $\psi(x, y)$ be $\mathscr{L}_{\mathsf{PA}}$-formulae with at most two free variables. Show that there are $\mathscr{L}_{\mathsf{PA}}$-sentences such that

$$\sigma \Leftrightarrow_{\mathsf{PA}} \varphi(\ulcorner\sigma\urcorner, \ulcorner\tau\urcorner) \quad \text{and} \quad \tau \Leftrightarrow_{\mathsf{PA}} \psi(\ulcorner\sigma\urcorner, \ulcorner\tau\urcorner).$$

Note that this is a generalisation of the DIAGONALISATION LEMMA.

10.4 A famous paradox, denoted as Curry's paradox, states informally: "If this sentence is true, then $0 = 1$ holds." Explain why this is contradictory and formalise the paradox in PA. Use this to give an alternative proof of Gödel's version of the FIRST INCOMPLETENESS THEOREM.