CHAPTER 3

Gödel's Theorems

We now bring proof and recursion together. A principal object of study in this chapter are the elementary functions, which are adequate for the arithmetization of syntax leading to Gödel's two incompleteness theorems.

3.1. The notion of truth in formal theories

We consider the question whether there is a truth formula $B(z)$ such that in appropriate theories T we have $T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner})$ for all sentences A. Here $\ulcorner A \urcorner$ is the "Gödel number" of A, and $\underline{a}$ is the "numeral" denoting $a \in \mathbb{N}$; both notions are defined in Section 3.1.1 below. The result will be that this is impossible, under rather weak assumptions on the theory T. Technically, the issue will be to have a syntactic substitute of the notion of definability by "representability" within a formal theory. This notion is defined in Section 3.1.2.

3.1.1. Gödel numbers. We will assign numbers – so-called Gödel numbers, GN for short – to the syntactical constructs developed in Chapter 1: terms, formulas and derivations. Using the elementary sequence-coding and decoding machinery developed earlier we will be able to construct the code number of a composed object from its parts, and conversely to disassemble the code number of a composed object into the code numbers of its parts.

Let $\mathcal{L}$ be a countable first-order language. Assume that we have injectively assigned to every n-ary relation symbol R a *symbol number* $\mathrm{sn}(R)$ of the form $\langle 1, n, i \rangle$ and to every n-ary function symbol f a symbol number $\mathrm{sn}(f)$ of the form $\langle 2, n, j \rangle$. Call $\mathcal{L}$ *elementarily presented* if the set $\mathrm{Symb}_{\mathcal{L}}$ of all these symbol numbers is elementary. In what follows we shall always assume that the languages $\mathcal{L}$ considered are elementarily presented. In particular this applies to every language with finitely many relation and function symbols.

Let $\operatorname{sn}(\mathrm{Var}) := \langle 0 \rangle$. For every $\mathcal{L}$-term t we define recursively its Gödel number $\ulcorner t \urcorner$ by

$$\begin{aligned}
\ulcorner x_i \urcorner &:= \langle \operatorname{sn}(\mathrm{Var}), i \rangle, \\
\ulcorner f t_1 \dots t_n \urcorner &:= \langle \operatorname{sn}(f), \ulcorner t_1 \urcorner, \dots, \ulcorner t_n \urcorner \rangle.
\end{aligned}$$

Assign numbers to the logical symbols by $\operatorname{sn}(\to) := \langle 3, 0 \rangle$ and $\operatorname{sn}(\forall) := \langle 3, 1 \rangle$. For simplicity we leave out the logical connective $\wedge$ here; it could be treated similarly. We define for every $\mathcal{L}$-formula A its Gödel number $\ulcorner A \urcorner$ by

$$\begin{aligned}
\ulcorner R t_1 \dots t_n \urcorner &:= \langle \operatorname{sn}(R), \ulcorner t_1 \urcorner, \dots, \ulcorner t_n \urcorner \rangle, \\
\ulcorner A \to B \urcorner &:= \langle \operatorname{sn}(\to), \ulcorner A \urcorner, \ulcorner B \urcorner \rangle, \\
\ulcorner \forall_{x_i} A \urcorner &:= \langle \operatorname{sn}(\forall), i, \ulcorner A \urcorner \rangle.
\end{aligned}$$

Assume that 0 is a constant and S is a unary function symbol in $\mathcal{L}$. For every $a \in \mathbb{N}$ the *numeral* $\underline{a} \in \mathrm{Ter}_{\mathcal{L}}$ is defined by $\underline{0} := 0$ and $\underline{n+1} := S\underline{n}$. We can define an elementary function s such that for every formula $C = C(z)$ with $z := x_0$,

$$s(\ulcorner C \urcorner, k) = \ulcorner C(\underline{k}) \urcorner;$$

the proof is an exercise.

We define symbol numbers for the names of the natural deduction rules: $\operatorname{sn}(\mathrm{AssVar}) := \langle 4, 0 \rangle$, $\operatorname{sn}(\to^+) := \langle 4, 1 \rangle$, $\operatorname{sn}(\to^-) := \langle 4, 2 \rangle$, $\operatorname{sn}(\forall^+) := \langle 4, 3 \rangle$, $\operatorname{sn}(\forall^-) := \langle 4, 4 \rangle$. For a derivation M we define its Gödel number $\ulcorner M \urcorner$ by

$$\begin{aligned}
\ulcorner u_i^A \urcorner &:= \langle \operatorname{sn}(\mathrm{AssVar}), i, \ulcorner A \urcorner \rangle, \\
\ulcorner \lambda_{u_i^A} M \urcorner &:= \langle \operatorname{sn}(\to^+), i, \ulcorner A \urcorner, \ulcorner M \urcorner \rangle, \\
\ulcorner M N \urcorner &:= \langle \operatorname{sn}(\to^-), \ulcorner M \urcorner, \ulcorner N \urcorner \rangle, \\
\ulcorner \lambda_{x_i} M \urcorner &:= \langle \operatorname{sn}(\forall^+), i, \ulcorner M \urcorner \rangle, \\
\ulcorner M t \urcorner &:= \langle \operatorname{sn}(\forall^-), \ulcorner M \urcorner, \ulcorner t \urcorner \rangle.
\end{aligned}$$

Let T be an $\mathcal{L}$-theory determined by an elementary axiom system Ax_T (containing $\mathrm{Stab}_{\mathcal{L}}$) plus the equality axioms $\mathrm{Eq}_{\mathcal{L}}$:

$$\begin{aligned}
& x = x \quad \text{(Reflexivity)}, \\
& x = y \to y = x \quad \text{(Symmetry)}, \\
& x = y \to y = z \to x = z \quad \text{(Transitivity)}, \\
& x_1 = y_1 \to \dots \to x_n = y_n \to f(x_1, \dots, x_n) = f(y_1, \dots, y_n), \\
& x_1 = y_1 \to \dots \to x_n = y_n \to R(x_1, \dots, x_n) \to R(y_1, \dots, y_n),
\end{aligned}$$

for all n-ary function symbols f and relation symbols R of the language $\mathcal{L}$. For such axiomatized theories we can define an elementary binary relation Prf_T such that $\mathrm{Prf}_T(d, a)$ holds if and only if d is the GN of a derivation with a closed end formula with GN a from a context composed of equality axioms and formulas from Ax_T.

3.1.2. Representable relations and functions. In this section we assume that $\mathcal{L}$ is an elementarily presented language with 0, S and $=$ in $\mathcal{L}$, and T an $\mathcal{L}$-theory containing the equality axioms $\mathrm{Eq}_{\mathcal{L}}$.

Definition. A relation $R \subseteq \mathbb{N}^n$ is *representable* in T if there is a formula $A(x_1, \dots, x_n)$ such that

$$\begin{aligned} &T \vdash A(\underline{a_1}, \dots, \underline{a_n}) && \text{if } (a_1, \dots, a_n) \in R, \\ &T \vdash \neg A(\underline{a_1}, \dots, \underline{a_n}) && \text{if } (a_1, \dots, a_n) \notin R. \end{aligned}$$

A function $f\colon \mathbb{N}^n \to \mathbb{N}$ is called *representable* in T if there is a formula $A(x_1, \dots, x_n, y)$ representing the graph $G_f \subseteq \mathbb{N}^{n+1}$ of f, i.e., such that

$$T \vdash A(\underline{a_1}, \dots, \underline{a_n}, \underline{f(a_1, \dots, a_n)}), \tag{20}$$

$$T \vdash \neg A(\underline{a_1}, \dots, \underline{a_n}, \underline{c}) \qquad \text{if } c \neq f(a_1, \dots, a_n) \tag{21}$$

and such that in addition

$$T \vdash A(\underline{a_1}, \dots, \underline{a_n}, y) \wedge A(\underline{a_1}, \dots, \underline{a_n}, z) \to y{=}z \text{ for all } a_1, \dots, a_n \in \mathbb{N}. \tag{22}$$

Note that in case $T \vdash \underline{b} \neq \underline{c}$ for $b < c$ condition (21) follows from (20) and (22).

Lemma. *If the characteristic function c_R of a relation $R \subseteq \mathbb{N}^n$ is representable in T, then so is the relation R itself.*

Proof. For simplicity assume $n = 1$. Let $A(x, y)$ be a formula representing c_R. We show that $A(x, \underline{1})$ represents the relation R. Assume $a \in R$. Then $c_R(a) = 1$, hence $(a, 1) \in G_{c_R}$, hence $T \vdash A(\underline{a}, \underline{1})$. Conversely, assume $a \notin R$. Then $c_R(a) = 0$, hence $(a, 1) \notin G_{c_R}$, hence $T \vdash \neg A(\underline{a}, \underline{1})$. □

3.1.3. Undefinability of the notion of truth in formal theories.

Lemma (Fixed point lemma). *Assume that all elementary functions are representable in T. Then for every formula $B(z)$ we can find a closed formula A such that*

$$T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner}).$$

Proof. Let s be the elementary function introduced in Section 3.1.1 and $A_s(x_1, x_2, x_3)$ a formula representing s in T. Let

$$C(z) := \forall_x (A_s(z, z, x) \to B(x)), \quad A := C(\underline{\ulcorner C \urcorner}),$$

and therefore

$$A = \forall_x (A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x) \to B(x)).$$

Because of $s(\ulcorner C \urcorner, \ulcorner C \urcorner) = \ulcorner C(\underline{\ulcorner C \urcorner}) \urcorner = \ulcorner A \urcorner$ we can prove in T

$$A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x) \leftrightarrow x = \underline{\ulcorner A \urcorner},$$

hence by definition of A also

$$A \leftrightarrow \forall_x (x = \underline{\ulcorner A \urcorner} \to B(x))$$

and therefore

$$A \leftrightarrow B(\underline{\ulcorner A \urcorner}). \qquad \square$$

Theorem. *Let T be a consistent theory such that all elementary functions are representable in T. Then there cannot exist a formula $B(z)$ defining the notion of truth, i.e., such that for all closed formulas A*

$$T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner}).$$

Proof. Assume we would have such a $B(z)$. Consider the formula $\neg B(z)$ and choose by the fixed point lemma a closed formula A such that

$$T \vdash A \leftrightarrow \neg B(\underline{\ulcorner A \urcorner}).$$

For this A we obtain $T \vdash A \leftrightarrow \neg A$, contradicting the consistency of T. $\square$

3.2. Undecidability and incompleteness

Consider a consistent formal theory T with the property that all recursive functions are representable in T. This is a very weak assumption, as we shall show in the next section: it is always satisfied if the theory allows to develop a certain minimum of arithmetic. We shall show that such a theory necessarily is undecidable. Then we prove Gödel's (first) incompleteness theorem saying that every axiomatized such theory must be incomplete. In fact, we prove a sharpened form of this theorem due to Gödel and then Rosser, which explicitly provides a closed formula A such that neither A nor $\neg A$ is provable in the theory T.

In this section let $\mathcal{L}$ be an elementarily presented language with 0, S, $=$ in $\mathcal{L}$ and T a theory containing the equality axioms $\mathrm{Eq}_{\mathcal{L}}$. Call a relation *recursive* if its (total) characteristic function is recursive. A set S of formulas is called *recursive* (*elementarily enumerable*), if $\ulcorner S \urcorner := \{ \ulcorner A \urcorner \mid A \in S \}$ is recursive (elementarily enumerable).

Theorem (Undecidability). *Assume that T is a consistent theory such that all recursive functions are representable in T. Then T is not recursive.*

Proof. Assume that T is recursive. By assumption there exists a formula $B(z)$ representing $\ulcorner T \urcorner$ in T. Choose by the fixed point lemma a closed formula A such that

$$T \vdash A \leftrightarrow \neg B(\underline{\ulcorner A \urcorner}).$$

We shall prove $(*)$ $T \nvdash A$ and $(**)$ $T \vdash A$; this is the desired contradiction.

Ad ($*$). Assume $T \vdash A$. Then $A \in T$, hence $\ulcorner A \urcorner \in \ulcorner T \urcorner$, hence $T \vdash B(\underline{\ulcorner A \urcorner})$ (because $B(z)$ represents in T the set $\ulcorner T \urcorner$). By the choice of A it follows that $T \vdash \neg A$, which contradicts the consistency of T.

Ad ($**$). By ($*$) we know $T \nvdash A$. Therefore $A \notin T$, hence $\ulcorner A \urcorner \notin \ulcorner T \urcorner$ and therefore $T \vdash \neg B(\underline{\ulcorner A \urcorner})$. By the choice of A it follows that $T \vdash A$. □

Theorem (Gödel-Rosser). *Let T be axiomatized and consistent. Assume that there is a formula $L(x, y)$ – written $x < y$ – such that*

$$T \vdash \forall_{x<\underline{n}}(x = \underline{0} \mathrel{\tilde{\vee}} \cdots \mathrel{\tilde{\vee}} x = \underline{n-1}), \tag{23}$$

$$T \vdash \forall_x(x = \underline{0} \mathrel{\tilde{\vee}} \cdots \mathrel{\tilde{\vee}} x = \underline{n} \mathrel{\tilde{\vee}} \underline{n} < x). \tag{24}$$

Assume also that every elementary function is representable in T. Then we can find a closed formula A such that neither A nor $\neg A$ is provable in T.

Proof. We first define $\mathrm{Refut}_T \subseteq \mathbb{N} \times \mathbb{N}$ by

$$\mathrm{Refut}_T(d, a) := \mathrm{Prf}_T(d, \dot{\neg} a)$$

with $\dot{\neg} a := \langle \mathrm{sn}(\to), a, \mathrm{sn}(\bot) \rangle$. Then Refut_T is elementary and $\mathrm{Refut}_T(d, a)$ holds if and only if d is the GN of a derivation of the negation of a formula with GN a from a context composed of equality axioms and formulas from Ax_T. Let $B_{\mathrm{Prf}_T}(x_1, x_2)$ and $B_{\mathrm{Refut}_T}(x_1, x_2)$ be formulas representing Prf_T and Refut_T, respectively. Choose by the fixed point lemma a closed formula A such that

$$T \vdash A \leftrightarrow \forall_x(B_{\mathrm{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \to \tilde{\exists}_{y<x} B_{\mathrm{Refut}_T}(y, \underline{\ulcorner A \urcorner})).$$

A expresses its own underivability, in the form (due to Rosser): "For every proof of me there is a shorter proof of my negation".

We shall show ($*$) $T \nvdash A$ and ($**$) $T \nvdash \neg A$.

Ad ($*$). Assume $T \vdash A$. Choose n such that

$$\mathrm{Prf}_T(n, \ulcorner A \urcorner).$$

Then we also have

$$\text{not } \mathrm{Refut}_T(m, \ulcorner A \urcorner) \qquad \text{for all } m,$$

since T is consistent. Hence

$$T \vdash B_{\mathrm{Prf}_T}(\underline{n}, \underline{\ulcorner A \urcorner}),$$
$$T \vdash \neg B_{\mathrm{Refut}_T}(\underline{m}, \underline{\ulcorner A \urcorner}) \qquad \text{for all } m.$$

By (23) we can conclude

$$T \vdash B_{\mathrm{Prf}_T}(\underline{n}, \underline{\ulcorner A \urcorner}) \wedge \forall_{y<\underline{n}} \neg B_{\mathrm{Refut}_T}(y, \underline{\ulcorner A \urcorner}).$$

Hence

$$
\begin{aligned}
&T \vdash \tilde{\exists}_x(B_{\mathrm{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \wedge \forall_{y<x} \neg B_{\mathrm{Refut}_T}(y, \underline{\ulcorner A \urcorner})),\\
&T \vdash \neg A.
\end{aligned}
$$

This contradicts the assumed consistency of T.

Ad $(**)$. Assume $T \vdash \neg A$. Choose n such that

$$\mathrm{Refut}_T(n, \ulcorner A \urcorner).$$

Then we also have

$$\text{not } \mathrm{Prf}_T(m, \ulcorner A \urcorner) \qquad \text{for all } m,$$

since T is consistent. Hence

$$
\begin{aligned}
&T \vdash B_{\mathrm{Refut}_T}(\underline{n}, \underline{\ulcorner A \urcorner}),\\
&T \vdash \neg B_{\mathrm{Prf}_T}(\underline{m}, \underline{\ulcorner A \urcorner}) \qquad \text{for all } m.
\end{aligned}
$$

This implies

$$T \vdash \forall_x(B_{\mathrm{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \to \tilde{\exists}_{y<x} B_{\mathrm{Refut}_T}(y, \underline{\ulcorner A \urcorner})),$$

as can be seen easily by cases on x, using (24). Hence $T \vdash A$. But this again contradicts the assumed consistency of T. □

Finally we formulate a variant of this theorem which does not assume that the theory T talks about numbers only. Call T a *theory with defined natural numbers* if there is a formula $N(x)$ – written Nx – such that $T \vdash N0$ and $T \vdash \forall_{x \in N} N(Sx)$ where $\forall_{x \in N} A$ is short for $\forall_x(Nx \to A)$. Representing a function in such a theory of course means that the free variables in (22) are relativized to N:

$$T \vdash \forall_{y,z \in N}(A(\underline{a_1}, \dots, \underline{a_n}, y) \to A(\underline{a_1}, \dots, \underline{a_n}, z) \to y{=}z) \text{ for all } \vec{a} \in \mathbb{N}.$$

Theorem (Gödel-Rosser). *Assume that T is an axiomatized consistent theory with defined natural numbers, and that there is a formula $L(x, y)$ – written $x < y$ – such that*

$$
\begin{aligned}
&T \vdash \forall_{x \in N}(x < \underline{n} \to x = \underline{0} \,\tilde{\vee}\, \cdots \,\tilde{\vee}\, x = \underline{n-1}),\\
&T \vdash \forall_{x \in N}(x = \underline{0} \,\tilde{\vee}\, \cdots \,\tilde{\vee}\, x = \underline{n} \,\tilde{\vee}\, \underline{n} < x).
\end{aligned}
$$

Assume also that every elementary function is representable in T. Then one can find a closed formula A such that neither A nor $\neg A$ is provable in T.

Proof. As for the Gödel-Rosser theorem above; just relativize all quantifiers to N. □

3.3. Representability of recursive functions

We show in this section that already very simple theories have the property that all recursive functions are representable in them; an example is a finitely axiomatized arithmetical theory Q due to Robinson (1950). A consequence will be the (even "essential") undecidability of Q.

3.3.1. Weak arithmetical theories.

THEOREM. *Let $\mathcal{L}$ be an elementarily presented language with 0, S, $=$ in $\mathcal{L}$ and T a consistent theory with defined natural numbers containing the equality axioms $\mathrm{Eq}_{\mathcal{L}}$. Assume that*

(25) $\quad T \vdash S\underline{a} \neq 0 \qquad$ *for all* $a \in \mathbb{N}$,

(26) $\quad T \vdash S\underline{a} = S\underline{b} \to \underline{a} = \underline{b} \qquad$ *for all* $a, b \in \mathbb{N}$,

(27) $\quad$ *the functions $+$ and $\cdot$ are representable in T*

and that there is a formula $L(x, y)$ – written $x < y$ – such that

(28) $\quad T \vdash \forall_{x \in N}(x \not< 0)$,

(29) $\quad T \vdash \forall_{x \in N}(x < S\underline{b} \to x < \underline{b} \mathbin{\tilde{\vee}} x = \underline{b}) \quad$ *for all* $b \in \mathbb{N}$,

(30) $\quad T \vdash \forall_{x \in N}(x < \underline{b} \mathbin{\tilde{\vee}} x = \underline{b} \mathbin{\tilde{\vee}} \underline{b} < x) \qquad$ *for all* $b \in \mathbb{N}$.

Then every recursive function is representable in T.

PROOF. First note that the formulas $x = y$ and $x < y$ actually do represent in T the equality and the less-than relations, respectively. From (25) and (26) we can see immediately that $T \vdash \underline{a} \neq \underline{b}$ when $a \neq b$. Assume $a \not< b$. We show $T \vdash \underline{a} \not< \underline{b}$ by induction on b. $T \vdash \underline{a} \not< 0$ follows from (28). In the step we have $a \not< Sb$, hence $a \not< b$ and $a \neq b$, hence by induction hypothesis and the representability (above) of the equality relation, $T \vdash \underline{a} \not< \underline{b}$ and $T \vdash \underline{a} \neq \underline{b}$, hence by (29) $T \vdash \underline{a} \not< S\underline{b}$. Now assume $a < b$. Then $T \vdash \underline{a} \neq \underline{b}$ and $T \vdash \underline{b} \not< \underline{a}$, hence by (30) $T \vdash \underline{a} < \underline{b}$.

We now show by induction on the definition of μ-recursive functions that every recursive function is representable in T. Recall (from Section 3.1.2) that the second condition (21) in the definition of representability of a function automatically follows from the other two (and hence need not be checked further). This is because $T \vdash \underline{a} \neq \underline{b}$ for $a \neq b$.

The *initial functions* constant 0, successor and projection (onto the i-th coordinate) are trivially represented by the formulas $0 = y$, $Sx = y$ and $x_i = y$ respectively. Addition and multiplication are represented in T by assumption. Recall that the one remaining initial function of μ-recursiveness is $\dot{-}$, but this is definable from the characteristic function of $<$ by $a \dot{-} b = \mu_i(b + i \geq a) = \mu_i(c_<(b + i, a) = 0)$. We now show that the characteristic function of $<$ is representable in T. (It will then follow that

$\dot{-}$ is representable, once we have shown that the representable functions are closed under μ.) We show that

$$A(x_1, x_2, y) := (x_1 < x_2 \land y = 1) \mathbin{\tilde{\lor}} (x_1 \not< x_2 \land y = 0)$$

represents $c_<$. First notice that $\forall_{y,z \in N}(A(\underline{a_1}, \underline{a_2}, y) \to A(\underline{a_1}, \underline{a_2}, z) \to y = z)$ already follows logically from the equality axiom (by cases on the alternatives of A). Assume $a_1 < a_2$. Then $T \vdash \underline{a_1} < \underline{a_2}$, hence $T \vdash A(\underline{a_1}, \underline{a_2}, 1)$. Now assume $a_1 \not< a_2$. Then $T \vdash \underline{a_1} \not< \underline{a_2}$, hence $T \vdash A(\underline{a_1}, \underline{a_2}, 0)$.

For the *composition* case, suppose f is defined from $h, g_1, \dots, g_m$ by

$$f(\vec{a}\,) = h(g_1(\vec{a}\,), \dots, g_m(\vec{a}\,)).$$

By induction hypothesis we already have representing formulas $A_{g_i}(\vec{x}, y_i)$ and $A_h(\vec{y}, z)$. As representing formula for f we take

$$A_f := \tilde{\exists}_{\vec{y} \in N}(A_{g_1}(\vec{x}, y_1) \mathbin{\tilde{\land}} \dots \mathbin{\tilde{\land}} A_{g_m}(\vec{x}, y_m) \mathbin{\tilde{\land}} A_h(\vec{y}, z)).$$

Assume $f(\vec{a}\,) = c$. Then there are $b_1, \dots, b_m$ such that $T \vdash A_{g_i}(\underline{\vec{a}}, \underline{b_i})$ for each i, and $T \vdash A_h(\underline{\vec{b}}, \underline{c})$ so by logic $T \vdash A_f(\underline{\vec{a}}, \underline{c})$. It remains to show uniqueness $T \vdash \forall_{z_1, z_2 \in N}(A_f(\underline{\vec{a}}, z_1) \to A_f(\underline{\vec{a}}, z_2) \to z_1 = z_2)$. But this follows by logic from the induction hypothesis for g_i, which gives

$$T \vdash \forall_{y_{1i}, y_{2i} \in N}(A_{g_i}(\underline{\vec{a}}, y_{1i}) \to A_{g_i}(\underline{\vec{a}}, y_{2i}) \to y_{1i} = y_{2i} = \underline{g_i(\vec{a}\,)})$$

and the induction hypothesis for h, which gives

$$T \vdash \forall_{z_1, z_2 \in N}(A_h(\underline{\vec{b}}, z_1) \to A_h(\underline{\vec{b}}, z_2) \to z_1 = z_2) \quad \text{with } b_i = g_i(\vec{a}\,).$$

For the μ case, suppose f is defined from g (taken here to be binary for notational convenience) by $f(a) = \mu_i(g(i, a) = 0)$, assuming $\forall_a \tilde{\exists}_i(g(i, a) = 0)$. By induction hypothesis we have a formula $A_g(y, x, z)$ representing g. In this case we represent f by the formula

$$A_f(x, y) := Ny \land A_g(y, x, 0) \land \forall_{v \in N}(v < y \to \tilde{\exists}_{u \in N; u \neq 0} A_g(v, x, u)).$$

We first show the representability condition (20), that is $T \vdash A_f(\underline{a}, \underline{b})$ when $f(a) = b$. Because of the form of A_f this follows from the assumed representability of g together with $T \vdash \forall_{v \in N}(v < \underline{b} \to v = \underline{0} \mathbin{\tilde{\lor}} \cdots \mathbin{\tilde{\lor}} v = \underline{b - 1})$.

We now tackle the uniqueness condition (22). Given a, let $b := f(a)$ (thus $g(b, a) = 0$ and b is the least such). It suffices to show

$$T \vdash \forall_{y \in N}(A_f(\underline{a}, y) \to y = \underline{b}).$$

We prove $T \vdash \forall_{y \in N}(y < \underline{b} \to \neg A_f(\underline{a}, y))$ and $T \vdash \forall_{y \in N}(\underline{b} < y \to \neg A_f(\underline{a}, y))$, and then appeal to the trichotomy law.

We first show $T \vdash \forall_{y \in N}(y < \underline{b} \to \neg A_f(\underline{a}, y))$. Now since, for any $i < b$, $T \vdash \neg A_g(\underline{i}, \underline{a}, 0)$ by the assumed representability of g, we obtain immediately $T \vdash \neg A_f(\underline{a}, \underline{i})$. Hence because of $T \vdash \forall_{y \in N}(y < \underline{b} \to y = \underline{0} \mathbin{\tilde{\lor}} \cdots \mathbin{\tilde{\lor}} y = \underline{b - 1})$ the claim follows.

Secondly, $T \vdash \forall_{y \in N}(\underline{b} < y \to \neg A_f(\underline{a}, y))$ follows almost immediately from $T \vdash \forall_{y \in N}(\underline{b} < y \to A_f(\underline{a}, y) \to \tilde{\exists}_{u \in N; u \neq 0} A_g(\underline{b}, \underline{a}, u))$ and the uniqueness for g, $T \vdash \forall_{u \in N}(A_g(\underline{b}, \underline{a}, u) \to u = 0)$. □

3.3.2. Robinson's theory Q. We conclude this section by considering a special and particularly simple arithmetical theory due originally to Robinson (1950). Let $\mathcal{L}_1$ be the language given by 0, S, $+$, $\cdot$ and $=$, and let Q be the theory determined by the axioms $\mathrm{Eq}_{\mathcal{L}_1}$ and

$$Sx \neq 0, \tag{31}$$

$$Sx = Sy \to x = y, \tag{32}$$

$$x + 0 = x, \tag{33}$$

$$x + Sy = S(x + y), \tag{34}$$

$$x \cdot 0 = 0, \tag{35}$$

$$x \cdot Sy = x \cdot y + x, \tag{36}$$

$$\tilde{\exists}_z(x + Sz = y) \mathbin{\tilde{\vee}} x = y \mathbin{\tilde{\vee}} \tilde{\exists}_z(y + Sz = x). \tag{37}$$

Theorem (Robinson's Q). *Every consistent theory $T \supseteq Q$ fulfills the assumptions of the previous theorem w.r.t. the definition $L(x, y) := \tilde{\exists}_z(x + Sz = y)$ of the $<$-relation. Hence every recursive function is representable in T.*

Proof. We show that T satisfies the conditions of the previous theorem. For (25) and (26) this is clear. For (27) we can take $x + y = z$ and $x \cdot y = z$ as representing formulas. For (28) we have to show $\neg\tilde{\exists}_z(x + Sz = 0)$; this follows from (34) and (31). For the proof of (29) we need the auxiliary proposition

$$x = 0 \mathbin{\tilde{\vee}} \tilde{\exists}_y(x = 0 + Sy), \tag{38}$$

which will be attended to below. Assume $x + Sz = S\underline{b}$, hence also $S(x + z) = S\underline{b}$ and therefore $x + z = \underline{b}$. We must show $\tilde{\exists}_{y'}(x + Sy' = \underline{b}) \mathbin{\tilde{\vee}} x = \underline{b}$. But this follows from (38) for z. In case $z = 0$ we obtain $x = \underline{b}$, and in case $\tilde{\exists}_y(z = 0 + Sy)$ we have $\tilde{\exists}_{y'}(x + Sy' = \underline{b})$, since $0 + Sy = S(0 + y)$. Thus (29) is proved. (30) follows immediately from (37). For the proof of (38) we use (37) with $y = 0$. It clearly suffices to exclude the first case $\tilde{\exists}_z(x + Sz = 0)$. But this means $S(x + z) = 0$, contradicting (31). □

Corollary (Essential undecidability of Q). *Every consistent theory $T \supseteq Q$ in an elementarily presented language is non-recursive.*

Proof. This follows from the theorem above and the undecidability theorem in Section 3.2. □

Corollary (Undecidability of logic). *The set of formulas derivable in minimal logic is non-recursive.*

Proof. Otherwise Q would be recursive, because a formula A is derivable in Q if and only if the implication $B \to A$ is derivable, where B is the conjunction of the finitely many axioms and equality axioms of Q. □

Remark. Note that it suffices that the underlying language contains one binary relation symbol (for $=$), one constant symbol (for 0), one unary function symbol (for S) and two binary functions symbols (for $+$ and $\cdot$). The study of decidable fragments of first-order logic is one of the oldest research areas of mathematical logic. For more information see Börger et al. (1997).

3.3.3. Σ_1-formulas. Reading the above proof of representability, one can see that the representing formulas used are of a restricted form, having no unbounded universal quantifiers and therefore defining Σ^0_1-relations. This will be of crucial importance for our proof of Gödel's second incompleteness theorem to follow, but in addition we need to make a syntactically precise definition of the class of formulas involved, more specific and apparently more restrictive than the notion of Σ_1-formula used earlier. However, as proved in the corollary below, we can still represent all recursive functions even in the weak theory Q by means of Σ_1-formulas in this more restrictive sense. Consequently provable Σ_1-ness will be the same whichever definition we take.

Definition. For the remainder of this chapter, the Σ_1-formulas of the language $\mathcal{L}_1$ will be those generated inductively by the following clauses:

(a) Only atomic formulas of the restricted forms $x = y$, $x \neq y$, $0 = x$, $Sx = y$, $x + y = z$ and $x \cdot y = z$ are allowed as Σ_1-formulas.
(b) If A and B are Σ_1-formulas, then so are $A \wedge B$ and $A \tilde{\vee} B$.
(c) If A is a Σ_1-formula, then so is $\forall_{x<y} A$, which is an abbreviation for $\forall_x(\tilde{\exists}_z(x + Sz = y) \to A)$.
(d) If A is a Σ_1-formula, then so is $\tilde{\exists}_x A$.

Corollary. *Every recursive function is representable in Q by a Σ_1-formula in the language $\mathcal{L}_1$.*

Proof. This can be seen immediately by inspecting the proof of the theorem above on weak arithmetical theories. Only notice that because of the equality axioms $\tilde{\exists}_z(x+Sz = y)$ is equivalent to $\tilde{\exists}_z\tilde{\exists}_w(Sz = w \wedge x+w = y)$ and $A(0)$ is equivalent to $\tilde{\exists}_x(0 = x \wedge A(x))$. □

3.4. Unprovability of consistency

We have seen in the theorem of Gödel-Rosser how, for every axiomatized consistent theory T safisfying certain weak assumptions, we can construct an undecidable sentence A meaning "For every proof of me there is a shorter proof of my negation". Because A is unprovable, it is clearly true.

Gödel's second incompleteness theorem provides a particularly interesting alternative to A, namely a formula Con_T expressing the consistency of T. Again it turns out to be unprovable and therefore true. We shall prove this theorem in a sharpened form due to Löb.

3.4.1. Σ_1-completeness. We prove an auxiliary proposition, expressing the completeness of Q with respect to Σ_1-formulas.

Lemma (Σ_1-completeness). *Let $A(x_1, \dots, x_n)$ be a Σ_1-formula of the language $\mathcal{L}_1$. Assume that $\mathcal{N}_1 \models A(\underline{a_1}, \dots, \underline{a_n})$ where $\mathcal{N}_1$ is the standard model of $\mathcal{L}_1$. Then $Q \vdash A(\underline{a_1}, \dots, \underline{a_n})$.*

Proof. By induction on the Σ_1-formulas of the language $\mathcal{L}_1$. For atomic formulas, the cases have been dealt with either in the earlier parts of the proof of the theorem above on weak arithmetical theories, or (for $x + y = z$ and $x \cdot y = z$) they follow from the recursion equations (33) - (36).

Cases $A \wedge B$, $A \tilde{\vee} B$. The claim follows immediately from the induction hypothesis.

Case $\forall_{x<y} A(x, y, z_1, \dots, z_n)$; for simplicity assume $n = 1$. Suppose $\mathcal{N}_1 \models (\forall_{x<y} A)(\underline{b}, \underline{c})$. Then also $\mathcal{N}_1 \models A(\underline{i}, \underline{b}, \underline{c})$ for each $i < b$ and hence by induction hypothesis $Q \vdash A(\underline{i}, \underline{b}, \underline{c})$. Now by the theorem above on Robinson's Q

$$Q \vdash \forall_{x<\underline{b}}(x = \underline{0} \,\tilde{\vee} \cdots \tilde{\vee}\, x = \underline{b-1}),$$

hence

$$Q \vdash (\forall_{x<y} A)(\underline{b}, \underline{c}).$$

Case $\tilde{\exists}_x A(x, y_1, \dots, y_n)$; for simplicity again take $n = 1$. Assume $\mathcal{N}_1 \models (\tilde{\exists}_x A)(\underline{b})$. Then $\mathcal{N}_1 \models A(\underline{a}, \underline{b})$ for some $a \in \mathbb{N}$, hence by induction hypothesis $Q \vdash A(\underline{a}, \underline{b})$ and therefore $Q \vdash (\tilde{\exists}_x A)(\underline{b})$. □

3.4.2. Derivability conditions. Let T be an axiomatized consistent theory with $T \supseteq Q$, and let $\mathrm{Prf}_T(p, z)$ be a Σ_1-formula of the language $\mathcal{L}_1$ which represents in Robinson's theory Q the recursive relation "a is the Gödel number of a proof in T of the formula with Gödel number b". Consider the following $\mathcal{L}_1$-formulas:

$$\begin{aligned} \mathrm{Thm}_T(x) &:= \tilde{\exists}_y \mathrm{Prf}_T(y, x), \\ \mathrm{Con}_T \quad &:= \neg \tilde{\exists}_y \mathrm{Prf}_T(y, \underline{\ulcorner \bot \urcorner}). \end{aligned}$$

Then $\mathrm{Thm}_T(x)$ defines in $\mathcal{N}_1$ the set of formulas provable in T, and we have $\mathcal{N}_1 \models \mathrm{Con}_T$ if and only if T is consistent. We write $\Box A$ for $\mathrm{Thm}_T(\underline{\ulcorner A \urcorner})$;

hence Con_T can be written $\neg\Box\bot$. Now suppose, in addition, that T satisfies the following two *derivability conditions*, due to Hilbert and Bernays (1939):

$$T \vdash \Box A \to \Box\Box A, \tag{39}$$

$$T \vdash \Box(A \to B) \to \Box A \to \Box B. \tag{40}$$

(39) formalizes Σ_1-completeness of the theory T for closed formulas, and (40) is a formalization of its closure under modus ponens (i.e., $\to^-$). The derivability conditions place further restrictions on the theory T and its proof predicate Prf_T. We check them under the assumption that Prf_T is as defined earlier. (There are non-standard ways of coding proofs which lead to various "pathologies" - see, e.g., Feferman (1960)).

The formalized version of modus ponens is easy to see, assuming that T can be conservatively extended to include a "proof-term" $t(y, y')$ such that one may prove

$$\mathrm{Prf}_T(y, \underline{\ulcorner A \to B \urcorner}) \to \mathrm{Prf}_T(y', \underline{\ulcorner A \urcorner}) \to \mathrm{Prf}_T(t(y, y'), \underline{\ulcorner B \urcorner})$$

for then (40) follows immediately by quantifier rules.

(39) is harder. A detailed proof requires a great deal of syntactic machinery to do with the construction of proof terms, as above, acting on Gödel numbers so as to mimic the various rules inside T. We merely content ourselves here with a short indication of why (39) holds; this should be sufficient to convince the reader of its validity.

As we have seen at the beginning of this chapter, the elementary functions are provably recursive and so we may take their definitions as having been added conservatively. Working informally "inside" T one shows, by induction on y, that

$$\mathrm{Prf}_T(y, \underline{\ulcorner A \urcorner}) \to \mathrm{Prf}_T(f(y), \underline{\ulcorner \Box A \urcorner})$$

where f is elementary. Then (39) follows by the quantifier rules.

If y is the Gödel number of a derivation (in T) consisting of an axiom A then there will be a term t, elementarily computable from y, such that $\mathrm{Prf}_T(t, \underline{\ulcorner A \urcorner})$ and hence $\Box A$ are derivable in T. This derivation may be syntactically complex, but it will essentially consist of checking that t, as a Gödel number, encodes the right thing. Thus the derivation of $\Box A$ has a fixed Gödel number (depending on t and hence y) and this is what we take as the value of $f(y)$.

If y is the Gödel number of a derivation of A in which one of the rules is finally applied, say to premises A' and A'', then there will be $y', y'' < y$ such that $\mathrm{Prf}_T(y', \underline{\ulcorner A' \urcorner})$ and $\mathrm{Prf}_T(y'', \underline{\ulcorner A'' \urcorner})$. By the induction hypothesis, $f(y')$ and $f(y'')$ will be the Gödel numbers of T-derivations of $\Box A'$ and $\Box A''$, and as in the modus-ponens case above, there will be a fixed derivation which combines these two into a new derivation of $\Box A$. We take, as the value $f(y)$, the Gödel number of this final derivation, computable from $f(y')$ and $f(y'')$ by applying some additional (sub-elementary) coding corresponding to the additional steps from $\Box A'$ and $\Box A''$ to $\Box A$.

The function f will be definable from elementary functions by a course-of-values recursion in which the recursion steps are in fact computed sub-elementarily. Therefore it will be a limited course-of-values recursion and, by a result in Chapter 2, f will therefore be elementary as required.

THEOREM (Gödel's second incompleteness theorem). *Let T be an axiomatized consistent extension of Q, satisfying the derivability conditions* (39) *und* (40)*. Then $T \nvdash \mathrm{Con}_T$.*

PROOF. Let $C := \bot$ in Löb's theorem below, which is a generalization of Gödel's original result. □

THEOREM (Löb). *Let T be an axiomatized consistent extension of Q satisfying the derivability conditions* (39) *and* (40)*. Then for any closed $\mathcal{L}_1$-formula C, if $T \vdash \Box C \to C$, then already $T \vdash C$.*

PROOF. Assume $T \vdash \Box C \to C$. We must show $T \vdash C$. Choose A by the fixed point lemma such that

$$Q \vdash A \leftrightarrow (\Box A \to C). \tag{41}$$

First we show $T \vdash \Box A \to C$. We obtain

$$\begin{array}{ll} T \vdash A \to \Box A \to C & \text{by (41)} \\ T \vdash \Box(A \to \Box A \to C) & \text{by } \Sigma_1\text{-completeness} \\ T \vdash \Box A \to \Box(\Box A \to C) & \text{by (40)} \\ T \vdash \Box A \to \Box\Box A \to \Box C & \text{again by (40)} \\ T \vdash \Box A \to \Box C & \text{since } T \vdash \Box A \to \Box\Box A \text{ by (39).} \end{array}$$

Therefore the assumption $T \vdash \Box C \to C$ implies $T \vdash \Box A \to C$. Hence $T \vdash A$ by (41), and then $T \vdash \Box A$ by Σ_1-completeness. But $T \vdash \Box A \to C$ as we have just shown, therefore $T \vdash C$. □

REMARK. It follows that if T is any axiomatized consistent extension of Q satisfying the derivability conditions (39) und (40), then the reflection schema

$$\Box C \to C \quad \text{for closed } \mathcal{L}_1\text{-formulas } C$$

is not derivable in T. For by Löb's theorem, it cannot be derivable when C is underivable.