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

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

Assign numbers to the logical symbols by $\mathrm{sn}(\to) := \langle 3, 0 \rangle$ and $\mathrm{sn}(\forall) := \langle 3, 1 \rangle$. 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

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

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: $\mathrm{sn}(\mathrm{AssVar}) := \langle 4, 0 \rangle$, $\mathrm{sn}(\to^+) := \langle 4, 1 \rangle$, $\mathrm{sn}(\to^-) := \langle 4, 2 \rangle$, $\mathrm{sn}(\forall^+) := \langle 4, 3 \rangle$, $\mathrm{sn}(\forall^-) := \langle 4, 4 \rangle$. For a derivation M we define its Gödel number $\ulcorner M \urcorner$ by

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

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

$$T \vdash A(\underline{a_1}, \dots, \underline{a_n}, \underline{f(a_1, \dots, a_n)}), \tag{15}$$

$$T \vdash \neg A(\underline{a_1}, \dots, \underline{a_n}, \underline{c}) \qquad \text{if } c \neq f(a_1, \dots, a_n) \tag{16}$$

and such that in addition

$$T \vdash A(\underline{a_1}, \dots, \underline{a_n}, y) \to A(\underline{a_1}, \dots, \underline{a_n}, z) \to y{=}z \text{ for all } a_1, \dots, a_n \in \mathbb{N}. \tag{17}$$

Note that in case $T \vdash \underline{b} \neq \underline{c}$ for $b < c$ condition (16) follows from (15) and (17).

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*

$$T \vdash \forall_{x<\underline{n}}(x = \underline{0} \lor \cdots \lor x = \underline{n-1}), \tag{18}$$

$$T \vdash \forall_x(x = \underline{0} \lor \cdots \lor x = \underline{n} \lor \underline{n} < x). \tag{19}$$

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

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

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 (18) we can conclude

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