CHAPTER 4

Gödel's Theorems

4.1. 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.

4.1.1. Gödel numbers of terms, formulas and derivations. 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 r we define recursively its Gödel number $\ulcorner r \urcorner$ by

$$\begin{aligned} \ulcorner x_i \urcorner &:= \langle \mathrm{SN}(\mathrm{Var}), i\rangle, \\ \ulcorner f r_1 \dots r_n \urcorner &:= \langle \mathrm{SN}(f), \ulcorner r_1 \urcorner, \dots, \ulcorner r_n \urcorner\rangle. \end{aligned}$$

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

$$\begin{aligned} \ulcorner R r_1 \dots r_n \urcorner &:= \langle \mathrm{SN}(R), \ulcorner r_1 \urcorner, \dots, \ulcorner r_n \urcorner\rangle, \\ \ulcorner A \to B \urcorner &:= \langle \mathrm{SN}(\to), \ulcorner A \urcorner, \ulcorner B \urcorner\rangle, \\ \ulcorner \forall_{x_i} A \urcorner &:= \langle \mathrm{SN}(\forall), i, \ulcorner A \urcorner\rangle. \end{aligned}$$

We define symbol numbers for the names of the natural deduction rules: $\mathrm{SN}(\mathrm{AVar}) := \langle 4, 0\rangle$, $\mathrm{SN}(\to^+) := \langle 4, 1\rangle$, $\mathrm{SN}(\to^-) := \langle 4, 2\rangle$, $\mathrm{SN}(\forall^+) := \langle 4, 3\rangle$,

$\mathrm{SN}(\forall^-) := \langle 4, 4\rangle$. For a derivation M we define its Gödel number $\ulcorner M \urcorner$ by

$$\begin{aligned}
\ulcorner u_i^A \urcorner &:= \langle \mathrm{SN}(\mathrm{AVar}), i, \ulcorner A \urcorner\rangle,\\
\ulcorner \lambda_{u_i^A} M \urcorner &:= \langle \mathrm{SN}(\to^+), i, \ulcorner A \urcorner, \ulcorner M \urcorner\rangle,\\
\ulcorner MN \urcorner &:= \langle \mathrm{SN}(\to^-), \ulcorner M \urcorner, \ulcorner N \urcorner\rangle,\\
\ulcorner \lambda_{x_i} M \urcorner &:= \langle \mathrm{SN}(\forall^+), i, \ulcorner M \urcorner\rangle,\\
\ulcorner Mr \urcorner &:= \langle \mathrm{SN}(\forall^-), \ulcorner M \urcorner, \ulcorner r \urcorner\rangle.
\end{aligned}$$

4.1.2. Elementary functions on Gödel numbers. We shall define an elementary predicate Deriv such that $\mathrm{Deriv}(d)$ if and only if d is the Gödel number of a derivation. To this end we need a number of auxiliary functions and relations, which will all be elementary and have the properties formulated below. First we need to provide some notions.

$\mathrm{Ter}(t)$	t is GN of a term,
$\mathrm{For}(a)$	a is GN of a formula,
$\mathrm{FV}(i, y)$	the variable x_i is free in the term or formula with GN y,
$\mathrm{fla}(d)$	GN of the formula derived by the derivation with GN d.

By the *context* of a derivation M we mean the set $\{u_{i_0}^{A_0}, \dots, u_{i_{n-1}}^{A_{n-1}}\}$ of its free assumption variables, where $i_0 < \dots < i_{n-1}$. Its Gödel number is defined to be the least number c such that $\forall_{\nu<n}((c)_{i_\nu} = \ulcorner A_i \urcorner)$.

$\mathrm{ctx}(d)$	GN of the context of the derivation with GN d,
$\mathrm{Cons}(c_1, c_2)$	the contexts with GN c_1, c_2 are consistent.

Then Deriv can be defined by course-of-values recursion:

$$\begin{aligned}
\mathrm{Deriv}(d) := {} & ((d)_0 = \mathrm{SN}(\mathrm{AVar}) \wedge \mathrm{lh}(d) = 3 \wedge \mathrm{For}((d)_2)) \vee \\
& ((d)_0 = \mathrm{SN}(\to^+) \wedge \mathrm{lh}(d) = 4 \wedge \mathrm{For}((d)_2) \wedge \mathrm{Deriv}((d)_3) \wedge \\
& \quad ((\mathrm{ctx}((d)_3))_{(d)_1} \neq 0 \to (\mathrm{ctx}((d)_3))_{(d)_1} = (d)_2)) \vee \\
& ((d)_0 = \mathrm{SN}(\to^-) \wedge \mathrm{lh}(d) = 3 \wedge \mathrm{Deriv}((d)_1) \wedge \mathrm{Deriv}((d)_2) \wedge \\
& \quad \mathrm{Cons}(\mathrm{ctx}((d)_1), \mathrm{ctx}((d)_2)) \wedge \\
& \quad (\mathrm{fla}((d)_1))_0 = \mathrm{SN}(\to) \wedge (\mathrm{fla}((d)_1))_1 = \mathrm{fla}((d)_2)) \vee \\
& ((d)_0 = \mathrm{SN}(\forall^+) \wedge \mathrm{lh}(d) = 3 \wedge \mathrm{Deriv}((d)_2) \wedge \forall_{i<\mathrm{lh}(\mathrm{ctx}((d)_2))}(\\
& \quad (\mathrm{ctx}((d)_2))_i \neq 0 \to \neg\mathrm{FV}((d)_1, (\mathrm{ctx}((d)_2))_i))) \vee \\
& ((d)_0 = \mathrm{SN}(\forall^-) \wedge \mathrm{lh}(d) = 3 \wedge \mathrm{Deriv}((d)_1) \wedge \mathrm{Ter}((d)_2) \wedge \\
& \quad (\mathrm{fla}((d)_1)_0 = \mathrm{SN}(\forall)).
\end{aligned}$$

For to define the auxiliary functions and relations we need further notions. A *substitution* is a map $x_{i_0} \mapsto r_0, \dots, x_{i_{n-1}} \mapsto r_{n-1}$ with $i_0 < \cdots < i_{n-1}$ from variables to terms; its Gödel number is the least number s such that $\forall_{\nu<n}((s)_{i_\nu} = \ulcorner r_i \urcorner)$.

$\mathrm{union}(c_1, c_2)$	GN of the union of the consistent contexts with GN c_1, c_2,
$\mathrm{remove}(c, i)$	GN of result of removing u_i from the context with GN c,
$\mathrm{Sub}(x, s, y)$	y is the GN of the result of applying the substitution with GN s to the term or formula with GN x,
$\mathrm{update}(s, i, t)$	GN of the result of updating the substitution with GN s by changing its entry at i to the term with GN t,

We now give definitions of all these; from the form of the definitions it will be clear that they have the required properties, and that the relations and functions thus defined are elementary.

$$\mathrm{update}(s,i,t) :=$$
$$\mu_{x<h(\max(s,t),\max(\mathrm{lh}(s),i))}((x)_i = t \land \forall_{k<\max(\mathrm{lh}(s),i)}(k \neq i \to (x)_i = (s)_i)).$$

Here $h(n,k) := (n+1)^{2^k}$ is the elementary function defined earlier with the property

$$\langle \underbrace{n, \dots, n}_{k} \rangle \leq h(n,k).$$

We define substitution by

$$\begin{aligned}
&\mathrm{Sub}(x,s,y) := \\
&\quad ((x)_0 = \mathrm{SN}(\mathrm{Var}) \land (s)_{(x)_1} = 0 \land y = x) \lor \\
&\quad ((x)_0 = \mathrm{SN}(\mathrm{Var}) \land (s)_{(x)_1} \neq 0 \land y = (s)_{(x)_1}) \lor \\
&\quad (((x)_{0,0} = 1 \lor (x)_{0,0} = 2 \lor (x)_0 = \mathrm{SN}(\to)) \land \\
&\quad\; \mathrm{lh}(x) = \mathrm{lh}(y) \land (x)_0 = (y)_0 \land \forall_{k<\mathrm{lh}(x) \dot{-} 1}((x)_{i+1} = (y)_{i+1})) \lor \\
&\quad ((x)_0 = \mathrm{SN}(\forall) \land \forall_{i<\mathrm{lh}(s)}((s)_i \neq 0 \to \neg\mathrm{FV}((x)_1, (s)_i)) \land \\
&\quad\; \mathrm{lh}(x) = \mathrm{lh}(y) \land (y)_0 = \mathrm{SN}(\forall) \land (y)_1 = (x)_1 \land \mathrm{Sub}((x)_2, s, (y)_2)) \lor \\
&\quad ((x)_0 = \mathrm{SN}(\forall) \land \exists_{i<\mathrm{lh}(s)}((s)_i \neq 0 \land \mathrm{FV}((x)_1, (s)_i)) \land \\
&\quad\; \mathrm{lh}(x) = \mathrm{lh}(y) \land (y)_0 = \mathrm{SN}(\forall) \land (y)_1 = x \land \\
&\quad\; \mathrm{Sub}((x)_2, \mathrm{update}(s, (x)_1, \langle \mathrm{SN}(\mathrm{Var}), x \rangle), (y)_2)).
\end{aligned}$$

Using the lemma in 4.1.3 below one can see that, if x is GN of a term or a formula and s is GN of a substitution, $\mathrm{Sub}(x,s,y) \to y \leq (x+s)^{2^{x+s}}$.

Removal of an assumption variable from a context is defined by

$$\mathrm{remove}(c,i) := \mu_{x \leq c}((x)_i = 0 \land \forall_{j<\mathrm{lh}(c)}(j \neq i \to (x)_j = (c)_j)).$$

The union of two consistent contexts can again be defined by the bounded μ-operator:

$$\mathrm{union}(c_1, c_2) := \mu_{c \le c_1 * c_2} \forall_{i < \max(\mathrm{lh}(c_1), \mathrm{lh}(c_2))} ((c)_i = \max((c_1)_i, (c_2)_i)).$$

Consistency of two contexts is defined by

$$\mathrm{Cons}(c_1, c_2) := \forall_{i < \max(\mathrm{lh}(c_1), \mathrm{lh}(c_2))} ((c_1)_i \ne 0 \to (c_2)_i \ne 0 \to (c_1)_i = (c_2)_i).$$

The context of a derivation is defined by

$$\begin{aligned}\mathrm{ctx}(d) := \mu_{c \le d}(((d)_0 = \mathrm{SN}(\mathrm{AVar}) \wedge (c)_{(d)_1} = (d)_2) \vee \\ ((d)_0 = \mathrm{SN}(\to^+) \wedge c = \mathrm{remove}(\mathrm{ctx}((d)_1), i)) \vee \\ ((d)_0 = \mathrm{SN}(\to^-) \wedge c = \mathrm{union}(\mathrm{ctx}((d)_1), \mathrm{ctx}((d)_2))) \vee \\ ((d)_0 = \mathrm{SN}(\forall^+) \wedge c = \mathrm{ctx}((d)_2)) \vee \\ ((d)_0 = \mathrm{SN}(\forall^-) \wedge c = \mathrm{ctx}((d)_1))).\end{aligned}$$

The formula of a derivation is defined by

$$\begin{aligned}\mathrm{fla}(d) := \mu_{a \le d^{2^d}}(((d)_0 = \mathrm{SN}(\mathrm{AVar}) \wedge a = (d)_2) \vee \\ ((d)_0 = \mathrm{SN}(\to^+) \wedge a = \langle \mathrm{SN}(\to), (d)_2, \mathrm{fla}((d)_3)\rangle) \vee \\ ((d)_0 = \mathrm{SN}(\to^-) \wedge a = (\mathrm{fla}((d)_1))_2) \vee \\ ((d)_0 = \mathrm{SN}(\forall^+) \wedge a = \langle \mathrm{SN}(\forall), (d)_1, \mathrm{fla}((d)_2)\rangle) \vee \\ ((d)_0 = \mathrm{SN}(\forall^-) \wedge \\ \mathrm{Sub}((\mathrm{fla}((d)_1))_2, \mu_{s \le d}((s)_{(\mathrm{fla}((d)_1))_1} = (d)_2), a))).\end{aligned}$$

Freeness of a variable x_i in a term or formula is defined by

$$\begin{aligned}\mathrm{FV}(i, y) := ((y)_0 = \mathrm{SN}(\mathrm{Var}) \wedge (y)_1 = i) \vee \\ ((y)_{0,0} = 1 \wedge \exists_{j < \mathrm{lh}(y) \dot- 1} \mathrm{FV}(i, (y)_{j+1})) \vee \\ ((y)_{0,0} = 2 \wedge \exists_{j < \mathrm{lh}(y) \dot- 1} \mathrm{FV}(i, (y)_{j+1})) \vee \\ ((y)_0 = \mathrm{SN}(\to) \wedge (\mathrm{FV}(i, (y)_1) \vee \mathrm{FV}(i, (y)_2))) \vee \\ ((y)_0 = \mathrm{SN}(\forall) \wedge i \ne (y)_1 \wedge \mathrm{FV}(i, (y)_2)).\end{aligned}$$

The sets of formulas and terms are defined by

$$\begin{aligned}&\mathrm{For}(a) := \\ &((a)_{0,0} = 1 \wedge \mathrm{Symb}_{\mathcal{L}}((a)_0) \wedge \mathrm{lh}(a) = (a)_{0,1} + 1 \wedge \forall_{j < (a)_{0,1}} \mathrm{Ter}((a)_{j+1})) \vee \\ &((a)_0 = \mathrm{SN}(\to) \wedge \mathrm{lh}(a) = 3 \wedge \mathrm{For}((a)_1) \wedge \mathrm{For}((a)_2)) \vee \\ &((a)_0 = \mathrm{SN}(\forall) \wedge \mathrm{lh}(a) = 3 \wedge \mathrm{For}((a)_2)),\end{aligned}$$

$$\begin{aligned}\mathrm{Ter}(t) := ((t)_0 = \mathrm{SN}(\mathrm{Var}) \wedge \mathrm{lh}(t) = 2) \vee \\ ((t)_{0,0} = 2 \wedge \mathrm{Symb}_{\mathcal{L}}((t)_0) \wedge \mathrm{lh}(t) = (t)_{0,1} + 1 \wedge \forall_{j < (t)_{0,1}} \mathrm{Ter}((t)_{j+1})).\end{aligned}$$

Recall that for simplicity we have left out the logical connectives $\wedge$, $\vee$ and $\exists$. They could be added easily, including an extension of the notion of a derivation to also allow their axioms as listed in 1.2.3.

4.1.3. Substitution of numerals. In the sequel the elementary relation Sub representing substitution will not suffice. We shall need an elementary *function* s such that $s(\ulcorner C(z) \urcorner, k) = \ulcorner C(\underline{k}) \urcorner$ where $\underline{k}$ denotes the *numeral* $\underline{k}$ for the number k, defined by $\underline{0} := 0$ and $\underline{k+1} := \mathrm{S}\underline{k}$ (in any language with the function symbols 0 and S). It is easy to write down the course-of-values recursion equations for s; however, we also need to find an elementary bounding function. We shall do this for terms $r(z)$; the argument can easily be extended to formulas $C(z)$.

For a term r we define its *sum of maximal sequence lengths* $\|r\|$ by

$$\|x_i\| := 2 \quad \text{and} \quad \|fr_0 \dots r_{k-1}\| := k + 1 + \max_{i<k}(\|r_i\|)$$

and its *symbol bound* $\mathrm{Symb}(r)$ by

$$\mathrm{Symb}(x_i) := \max(\mathrm{SN}(\mathrm{Var}), i) + 1$$
$$\mathrm{Symb}(fr_0 \dots r_{k-1}) := \max(\mathrm{SN}(f), \max_{i<k}(\mathrm{Symb}(r_i))).$$

LEMMA. $\|r\| \leq \ulcorner r \urcorner < \mathrm{Symb}(r)^{2^{\|r\|}}$.

PROOF. We prove the first inequality by induction on r. The case of a variable x_i is easy:

$$\|x_i\| = 2 \leq \langle \mathrm{SN}(\mathrm{Var}), i \rangle = \ulcorner x_i \urcorner,$$

and for a term $fr_0 \dots r_{k-1}$ we have by the estimate in 3.2.5

$$\begin{aligned} \ulcorner fr_0 \dots r_{k-1} \urcorner &= \langle \mathrm{SN}(f), \ulcorner r_0 \urcorner, \dots, \ulcorner r_{k-1} \urcorner \rangle \\ &\geq (k+1)(\max(\mathrm{SN}(f), \ulcorner r_0 \urcorner, \dots, \ulcorner r_{k-1} \urcorner \rangle) + 1) \\ &> k + 1 + \max_{i<k}(\ulcorner r_i \urcorner) \\ &\geq k + 1 + \max_{i<k}(\|r_i\|) \quad \text{by induction hypothesis} \\ &= \|fr_0 \dots r_{k-1}\|. \end{aligned}$$

For the second inequality we again use induction on k. For a variable x_i we obtain again by the estimate in 3.2.5

$$\ulcorner x_i \urcorner = \langle \mathrm{SN}(\mathrm{Var}), i \rangle < \mathrm{Symb}(x_i)^{2^2} = \mathrm{Symb}(x_i)^{2^{\|x_i\|}}$$

and for a term built with a function symbol f we have

$$\begin{aligned} &\ulcorner fr_0 \dots r_{k-1} \urcorner \\ &= \langle \mathrm{SN}(f), \ulcorner r_0 \urcorner, \dots, \ulcorner r_{k-1} \urcorner \rangle \end{aligned}$$

$$\le \langle \underbrace{n \dot- 1, n \dot- 1, \dots, n \dot- 1}_{k+1} \rangle \quad \text{with } n := \mathrm{Symb}(r)^{2^{\max_{i<k} \|r_i\|}}, \text{ by ind. hyp.}$$
$$< n^{2^{k+1}} \quad \text{by the estimate in 3.2.5}$$
$$= \mathrm{Symb}(r)^{2^{k+1+\max_{i<k}\|r_i\|}} = \mathrm{Symb}(r)^{2^{\|fr_0\dots r_{k-1}\|}}. \qquad \square$$

As a consequence we obtain the desired estimates. First notice that $\ulcorner \underline{k} \urcorner < c^{2^{2k+1}}$ with $c := \max(\mathrm{SN(S)}, \mathrm{SN}(0))$, since by the previous lemma $\ulcorner \underline{k} \urcorner < c^{2^{\|\underline{k}\|}}$, and $\|\underline{k}\| = 2k+1$ can be proved easily by induction on k: $\|0\| = 1$, and $\|\mathrm{S}\underline{k}\| = 2 + \|k\|$. Now we obtain for $z := x_0$, again by the previous lemma

$$\begin{aligned} s(\ulcorner r(z) \urcorner, k) = \ulcorner r(\underline{k}) \urcorner &< \max(\mathrm{Symb}(r), \mathrm{SN(S)}, \mathrm{SN}(0))^{2^{\|r(\underline{k})\|}} \\ &\le \max(\ulcorner r(z) \urcorner, \mathrm{SN(S)}, \mathrm{SN}(0))^{2^{\|r(z)\|+2k+1}} \\ &\le \max(\ulcorner r(z) \urcorner, \mathrm{SN(S)}, \mathrm{SN}(0))^{2^{\ulcorner r(z) \urcorner+2k+1}}. \end{aligned}$$

Therefore we can bound the substitution function s for numerals by

$$s(y,k) \le \max(y, \mathrm{SN(S)}, \mathrm{SN}(0))^{2^{y+2k+1}}.$$

4.1.4. Axiomatized theories. Call a relation *recursive* if its (total) characteristic function is recursive. A set S of formulas is called *recursive* (*elementary*, *recursively enumerable*), if $\ulcorner S \urcorner := \{ \ulcorner A \urcorner \mid A \in S \}$ is recursive (elementary, recursively enumerable). Clearly the sets $\mathrm{Stab}_{\mathcal{L}}$ of stability axioms and $\mathrm{Eq}_{\mathcal{L}}$ of $\mathcal{L}$-equality axioms are elementary. Now let $\mathcal{L}$ be an elementarily presented language with $=$ in $\mathcal{L}$. A theory T with $L(T) \subseteq \mathcal{L}$ is *recursively (elementarily) axiomatizable*, if there is a recursive (elementary) set S of closed $\mathcal{L}$-formulas such that $T = \{ A \in \overline{\mathcal{L}} \mid S \cup \mathrm{Eq}_{\mathcal{L}} \vdash A \}$.

THEOREM. *For theories T with $L(T) \subseteq \mathcal{L}$ the following are equivalent.*

(a) *T is recursively axiomatizable.*
(b) *T is elementarily axiomatizable.*
(c) *T is recursively enumerable.*

PROOF. (c) → (b). Let $\ulcorner T \urcorner$ be recursively enumerable. Then there is an elementary f such that $\ulcorner T \urcorner = \mathrm{ran}(f)$. Let $f(n) = \ulcorner A_n \urcorner$. We define an elementary function g with the property $g(n) = \ulcorner A_0 \land \dots \land A_n \urcorner$ by

$$\begin{aligned} g(0) &:= f(0), \\ g(n+1) &:= g(n) \mathbin{\dot\land} f(n+1), \\ g(n) &\le \max(\mathrm{SN}(\land), m)^{2^{4n+m}} \quad \text{where } m := \max\nolimits_{i<n} f(i) \end{aligned}$$

with $a \mathbin{\dot\wedge} b := \langle \mathrm{SN}(\wedge), a, b\rangle$. For $S := \{\, A_0 \wedge \dots \wedge A_n \mid n \in \mathbb{N}\,\}$ we have $\ulcorner S \urcorner = \mathrm{ran}(g)$, and this set is elementary because of $a \in \mathrm{ran}(g) \leftrightarrow \exists_{n<a}(a = g(n))$. T is elementarily axiomatizable, since $T = \{\, A \in \overline{\mathcal{L}} \mid S \cup \mathrm{Eq}_{\mathcal{L}} \vdash A\,\}$.

(b) → (a) is clear.

(a) → (c). Let T be axiomatized by S with $\ulcorner S \urcorner$ recursive. Then

$$a \in \ulcorner T \urcorner \leftrightarrow \exists_d(\mathrm{Deriv}(d) \wedge \mathrm{fla}(d) = a \wedge \forall_{i<a}\neg\mathrm{FV}(i,a) \wedge \forall_{i<\mathrm{lh}(\mathrm{ctx}(d))}((\mathrm{ctx}(d))_i \in \ulcorner \mathrm{Eq} \urcorner \cup \ulcorner S \urcorner)).$$

Hence $\ulcorner T \urcorner$ is recursively enumerable. □

Call a theory T in our elementarily presented language $\mathcal{L}$ *axiomatized* if it is given by a recursively enumerable axiom system Ax_T. By the theorem just proved we can even assume that Ax_T is elementary. For such axiomatized theories we define a binary relation Prf_T by

$$\mathrm{Prf}_T(d,a) := \mathrm{Deriv}(d) \wedge \mathrm{fla}(d) = a \wedge \forall_{i<\mathrm{lh}(\mathrm{ctx}(d))}((\mathrm{ctx}(d))_i \in \ulcorner \mathrm{Eq} \urcorner \cup \ulcorner S \urcorner).$$

Clearly Prf_T is elementary and $\mathrm{Prf}_T(d,a)$ if and only if d is the GN of a derivation of the formula with GN a from a context composed of equality axioms and formulas from Ax_T. A theory T is *consistent* if $\bot \notin T$; otherwise T is *inconsistent*. A theory T is *complete* if for every closed formula A we have $A \in T$ or $A \notin T$, and *incomplete* otherwise.

Corollary. *Let T be a consistent theory. If T is axiomatized and complete, then T is recursive.*

Proof. We define the characteristic function $c_{\ulcorner T \urcorner}$ of $\ulcorner T \urcorner$ as follows. $c_{\ulcorner T \urcorner}(a)$ is 0 if $\neg\mathrm{For}(a)$ or $\exists_{i<a}\mathrm{FV}(i,a)$. Otherwise it is defined by

$$c_{\ulcorner T \urcorner}(a) = (\mu_x((\mathrm{Prf}_T((x)_0, a) \wedge (x)_1 = 1) \vee (\mathrm{Prf}_T((x)_0, \dot\neg a) \wedge (x)_1 = 0)))_1$$

with $\dot\neg a := \langle \mathrm{SN}(\to), a, \mathrm{SN}(\bot)\rangle$. Completeness of T implies that $c_{\ulcorner T \urcorner}$ is total, and consistency that it indeed is the characteristic function of $\ulcorner T \urcorner$. □

4.1.5. Undefinability of the notion of truth. Let $\mathcal{M}$ be an $\mathcal{L}$-structure. A relation $R \subseteq |\mathcal{M}|^n$ is called *definable* in $\mathcal{M}$ if there is an $\mathcal{L}$-formula $A(x_1,\dots,x_n)$ such that

$$R = \{\, (a_1,\dots,a_n) \in |\mathcal{M}|^n \mid \mathcal{M} \models A(x_1,\dots x_n)[x_1 := a_1,\dots,x_n := a_n]\,\}.$$

We assume in this section that $|\mathcal{M}| = \mathbb{N}$, 0 is a constant in $\mathcal{L}$ and S is a unary function symbol in $\mathcal{L}$ with $0^{\mathcal{M}} = 0$ and $\mathrm{S}^{\mathcal{M}}(a) = a+1$. Recall that for every $a \in \mathbb{N}$ the *numeral* $\underline{a} \in \mathrm{Ter}_{\mathcal{L}}$ is defined by $\underline{0} := 0$ and $\underline{n+1} := \mathrm{S}\underline{n}$. Observe that in this case the definability of $R \subseteq \mathbb{N}^n$ by $A(x_1,\dots,x_n)$ is equivalent to

$$R = \{\, (a_1,\dots,a_n) \in \mathbb{N}^n \mid \mathcal{M} \models A(\underline{a_1},\dots,\underline{a_n})\,\}.$$

Furthermore let $\mathcal{L}$ be an elementarily presented language. We will always assume in this chapter that every elementary relation is definable in $\mathcal{M}$. A set S of formulas is called *definable* in $\mathcal{M}$ if $\ulcorner S \urcorner := \{ \ulcorner A \urcorner \mid A \in S \}$ is.

We shall show that already from these assumptions it follows that the notion of truth for $\mathcal{M}$, more precisely the set $\mathrm{Th}(\mathcal{M})$ of all closed formulas valid in $\mathcal{M}$, is undefinable in $\mathcal{M}$. From this it will follow that the notion of truth is in fact undecidable, for otherwise the set $\mathrm{Th}(\mathcal{M})$ would be recursive (Church's Thesis), hence recursively enumerable, and hence definable, because we have assumed already that all elementary relations are definable in $\mathcal{M}$ and so their projections are definable also. For the proof we shall need the following Fixed Point Lemma, which will be generalized in 4.2.2.

Lemma (Semantical Fixed Point Lemma). *If every elementary relation is definable in $\mathcal{M}$, then for every $\mathcal{L}$-formula $B(z)$ we can find a closed $\mathcal{L}$-formula A such that*

$$\mathcal{M} \models A \quad \textit{if and only if} \quad \mathcal{M} \models B(\underline{\ulcorner A \urcorner}).$$

Proof. Let s be the elementary function from 4.1.3 satisfying for every formula $C = C(z)$ (with $z := x_0$)

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

hence in particular

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

By assumption the graph G_s of s is definable in $\mathcal{M}$, by $A_s(x_1, x_2, x_3)$ say. Let

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

and therefore

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

Hence $\mathcal{M} \models A$ if and only if $\exists_{a \in \mathbb{N}}((\mathcal{M} \models B(\underline{a})) \wedge a = \ulcorner C(\underline{\ulcorner C \urcorner}) \urcorner)$, which is the same as $\mathcal{M} \models B(\underline{\ulcorner A \urcorner})$. □

Theorem (Tarski's Undefinability Theorem). *Assume that every elementary relation is definable in $\mathcal{M}$. Then* $\mathrm{Th}(\mathcal{M})$ *is undefinable in $\mathcal{M}$, hence in particular not recursively enumerable.*

Proof. Assume that $\ulcorner \mathrm{Th}(\mathcal{M}) \urcorner$ is definable by $B_W(z)$. Then for all closed formulas A

$$\mathcal{M} \models A \quad \text{if and only if} \quad \mathcal{M} \models B_W(\underline{\ulcorner A \urcorner}).$$

Now consider the formula $\neg B_W(z)$ and choose by the Fixed Point Lemma a closed $\mathcal{L}$-formula A such that

$$\mathcal{M} \models A \quad \text{if and only if} \quad \mathcal{M} \models \neg B_W(\underline{\ulcorner A \urcorner}).$$

This contradicts the equivalence above.

We already have noticed that all recursively enumerable relations are definable in $\mathcal{M}$. Hence it follows that $\ulcorner\mathrm{Th}(\mathcal{M})\urcorner$ cannot be recursively enumerable. □

4.2. The Notion of Truth in Formal Theories

We now want to generalize the arguments of the previous section. There we have made essential use of the notion of truth in a structure $\mathcal{M}$, i.e., of the relation $\mathcal{M} \models A$. The set of all closed fromulas A such that $\mathcal{M} \models A$ has been called the theory of $\mathcal{M}$, denoted $\mathrm{Th}(\mathcal{M})$.

Now instead of $\mathrm{Th}(\mathcal{M})$ we shall start more generally from an arbitrary theory T. We consider the question as to whether in T there is a *notion of truth* (in the form of a *truth formula* $B(z)$), such that $B(z)$ "means" that z is "true". A consequence is that we have to explain all the notions used without referring to semantical concepts at all.

- z ranges over closed formulas (or sentences) A, or more precisely over their Gödel numbers $\ulcorner A\urcorner$.
- A "true" is to be replaced by $T \vdash A$.
- C "equivalent" to D is to be replaced by $T \vdash C \leftrightarrow D$.

Hence the question now is whether there is a truth formula $B(z)$ such that $T \vdash A \leftrightarrow B(\underline{\ulcorner A\urcorner})$ for all sentences A. The result will be that this is impossible, under rather weak assumptions on the theory T. Technically, the issue will be to replace the notion of definability by the notion of "representability" within a formal theory. We begin with a discussion of this notion.

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 $\mathrm{Eq}_{\mathcal{L}}$.

4.2.1. Representable relations and functions.

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

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

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

$$T \vdash A(\underline{a_1}, \dots, \underline{a_n}, \underline{f(a_1, \dots, a_n)}), \tag{4.1}$$
$$T \vdash \neg A(\underline{a_1}, \dots, \underline{a_n}, \underline{c}) \qquad \text{if } c \neq f(a_1, \dots, a_n) \tag{4.2}$$

and such that in addition

$$T \vdash A(\underline{a_1}, \dots, \underline{a_n}, y) \wedge A(\underline{a_1}, \dots, \underline{a_n}, z) \to y{=}z \text{ for all } a_1, \dots, a_n \in \mathbb{N}. \tag{4.3}$$

Note that in case $T \vdash \underline{b} \neq \underline{c}$ for $b < c$ condition (4.2) follows from (4.1) and (4.3).

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 $T \vdash \neg A(\underline{a}, \underline{1})$. □

4.2.2. 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. The proof is very similar to the proof of the Semantical Fixed Point Lemma. Let s be the elementary function from 4.1.3 and $A_s(x_1, x_2, x_3)$ a formula representing s in T. Let

$$C := \tilde{\exists}_x (B(x) \wedge A_s(z, z, x)), \quad A := C(\underline{\ulcorner C \urcorner}),$$

and therefore

$$A = \tilde{\exists}_x (B(x) \wedge A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x)).$$

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

$$A \leftrightarrow \tilde{\exists}_x (B(x) \wedge x = \underline{\ulcorner A \urcorner})$$

and therefore

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

Note that for $T = \mathrm{Th}(\mathcal{M})$ we obtain the Semantical Fixed Point Lemma above as a special case.

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. □

With $T := \mathrm{Th}(\mathcal{M})$ Tarski's Undefinability Theorem is a special case.

4.3. 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. Moreover we shall prove Gödel's First Incompleteness Theorem, which says that every axiomatized such theory must be incomplete. We will also prove a sharpened form of this theorem due to 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 $\mathcal{L}$ again be an elementarily presented language with 0, S, = in $\mathcal{L}$ and T a theory containing the equality axioms $\mathrm{Eq}_{\mathcal{L}}$.

4.3.1. Undecidability.

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 $B(z)$ in 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 $T \vdash A$. $\square$

4.3.2. Incompleteness.

Theorem (Gödel's First Incompleteness Theorem). *Assume that T is an axiomatized consistent theory with the property that all recursive functions are representable in T. Then T is incomplete.*

Proof. This is an immediate consequence of the fact that every axiomatized consistent theory which is complete is also recursive (a corollary in 4.1.4), and the Undecidability Theorem above. $\square$

As already mentioned, we now sharpen the Incompleteness Theorem in the sense that we actually produce a formula A such that neither A nor $\neg A$ is provable. The original idea for this sharpening is due to Rosser.

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

$$T \vdash \forall_{x<\underline{n}}(x=\underline{0} \,\tilde{\vee}\, \dots \,\tilde{\vee}\, x=\underline{n-1}), \tag{4.4}$$

$$T \vdash \forall_{x}(x=\underline{0} \,\tilde{\vee}\, \dots \,\tilde{\vee}\, x=\underline{n} \,\tilde{\vee}\, \underline{n}<x). \tag{4.5}$$

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

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

Then Refut_T is elementary and $\mathrm{Refut}_T(d,a)$ 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 we have

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

By (4.4) 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 we have

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

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 we have

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

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 (4.5). 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 $T \vdash \forall_{x \in N} N(\mathrm{S}x)$ 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 (4.3) are relativized to N:

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

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*

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

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. □

4.4. Representability

We show in this section that already very simple theories have the property that all recursive functions are representable in them.

4.4.1. Weak arithmetical theories.

Theorem (Weak arithmetical theories). *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 $\mathrm{Eq}_{\mathcal{L}}$ and the stablity axiom $\forall_{x,y \in N}(\neg\neg x = y \to x = y)$. Assume that there is a formula $L(x, y)$ – written $x < y$ – such that*

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

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

(4.8) *the functions* $+$ *and* $\cdot$ *are representable in* T,

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

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

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

Then T fulfills the assumptions of the Gödel-Rosser Theorem relativized to N, i.e.,

(4.12) $T \vdash \forall_{x \in N}(x < \underline{a} \to x = \underline{0} \mathbin{\tilde{\vee}} \dots \mathbin{\tilde{\vee}} x = \underline{a-1})$ *for all* $a \in \mathbb{N}$,

(4.13) $T \vdash \forall_{x \in N}(x = \underline{0} \mathbin{\tilde{\vee}} \dots \mathbin{\tilde{\vee}} x = \underline{a} \mathbin{\tilde{\vee}} \underline{a} < x)$ *for all* $a \in \mathbb{N}$,

and every recursive function is representable in T.

Proof. (4.12) can be proved easily by induction on a. The base case follows from (4.9), and the step from the induction hypothesis and (4.10). (4.13) immediately follows from the trichotomy law (4.11), using (4.12).

For the representability of recursive functions, first note that the formulas $x = y$ and $x < y$ actually do represent in T the equality and the less-than relations, respectively. From (4.6) and (4.7) 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 (4.9). In the step we have $a \not< b + 1$, 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 (4.10) $T \vdash \underline{a} \not< \mathrm{S}\underline{b}$. Now assume $a < b$. Then $T \vdash \underline{a} \neq \underline{b}$ and $T \vdash \underline{b} \not< \underline{a}$, hence by (4.11) $T \vdash \underline{a} < \underline{b}$.

We now show by induction on the definition of μ-recursive functions, that every recursive function is representable in T. Recall (from 4.2.1) that the second condition (4.2) 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 $0 = y$, $\mathrm{S}x = y$ 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.) So define

$$A := (x_1 < x_2 \wedge y = 1) \tilde{\vee} (x_1 \not< x_2 \wedge y = 0).$$

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 $T \vdash A(\underline{a_1}, \underline{a_2}, 0)$. Furthermore notice that $\forall_{y,z \in N}(A(\underline{a_1}, \underline{a_2}, y) \wedge A(\underline{a_1}, \underline{a_2}, z) \to y = z)$ already follows logically from the equality axioms (by cases on $\underline{a_1} < \underline{a_2}$, using stablity of equality).

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

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

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 $T \vdash \forall_{z_1,z_2 \in N}(A_f(\underline{\vec{a}}, z_1) \wedge A_f(\underline{\vec{a}}, z_2) \to z_1 = z_2)$. But this follows by logic from the induction hypothesis for g_i, which gives

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

and the induction hypothesis for h, which gives

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

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

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

We first show the representability condition (4.1), 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 $T \vdash \forall_{v \in N}(v < \underline{b} \to v = \underline{0} \tilde{\vee} \dots \tilde{\vee} v = \underline{b-1})$.

We now tackle the uniqueness condition (4.3). 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 and stablity of equality.

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 $T \vdash \forall_{y \in N}(y < \underline{b} \to y = \underline{0} \tilde{\vee} \dots \tilde{\vee} y = \underline{b-1})$ 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, $T \vdash \forall_{u \in N; u \neq 0}(A_g(\underline{b}, \underline{a}, u) \to u = 0)$. This completes the proof. □

4.4.2. Robinson's theory Q. We conclude this section by considering a special and particularly simple arithmetical theory due originally to Robinson. Let $\mathcal{L}_1$ be the language given by 0, S, +, · and =, and let Q be the theory determined by the axioms $\mathrm{Eq}_{\mathcal{L}_1}$, stability of equality $\neg\neg x = y \to x = y$ and

$$\mathrm{S}x \neq 0, \tag{4.14}$$

$$\mathrm{S}x = \mathrm{S}y \to x = y, \tag{4.15}$$

$$x + 0 = x, \tag{4.16}$$

$$x + \mathrm{S}y = \mathrm{S}(x + y), \tag{4.17}$$

$$x \cdot 0 = 0, \tag{4.18}$$

$$x \cdot \mathrm{S}y = x \cdot y + x, \tag{4.19}$$

$$\tilde{\exists}_z(x + \mathrm{S}z = y) \tilde{\vee} x = y \tilde{\vee} \tilde{\exists}_z(y + \mathrm{S}z = x). \tag{4.20}$$

THEOREM (Robinson's Q). *Every consistent theory $T \supseteq Q$ fulfills the assumptions of the Gödel-Rosser Theorem w.r.t. the definition $L(x, y) := \tilde{\exists}_z(x + \mathrm{S}z = y)$ of the $<$-relation. In particular, every recursive function is representable in T.*

PROOF. We show that T satisfies the conditions of the previous theorem. For (4.6) and (4.7) this is clear. For (4.8) we can take $x + y = z$ and $x \cdot y = z$ as representing formulas. For (4.9) we have to show $\neg\tilde{\exists}_z(x + \mathrm{S}z = 0)$; this follows from (4.17) and (4.14). For the proof of (4.10) we need the auxiliary proposition

$$x = 0 \tilde{\vee} \tilde{\exists}_y(x = 0 + \mathrm{S}y), \tag{4.21}$$

which will be attended to below. Assume $x + \mathrm{S}z = \mathrm{S}\underline{b}$, hence also $\mathrm{S}(x + z) = \mathrm{S}\underline{b}$ and therefore $x + z = \underline{b}$. We must show $\tilde{\exists}_{y'}(x + \mathrm{S}y' = \underline{b}) \tilde{\vee} x = \underline{b}$. But this follows from (4.21) for z. In case $z = 0$ we obtain $x = \underline{b}$, and in case $\tilde{\exists}_y(z = 0 + \mathrm{S}y)$ we have $\tilde{\exists}_{y'}(x + \mathrm{S}y' = \underline{b})$, since $0 + \mathrm{S}y = \mathrm{S}(0 + y)$. Thus (4.10) is proved. (4.11) follows immediately from (4.20). For the proof of (4.21) we use (4.20) with $y = 0$. It clearly suffices to exclude the first case $\tilde{\exists}_z(x + \mathrm{S}z = 0)$. But this means $\mathrm{S}(x + z) = 0$, contradicting (4.14). □

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

PROOF. This follows from the theorem above and the Undecidability Theorem in 4.3.1. □

COROLLARY (Undecidability of logic). *The set of formulas derivable in the classical fragment of 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).