5. Peano arithmetic and Gödel's incompleteness theorem

In this chapter we give the proof of Gödel's incompleteness theorem, modulo technical details treated in subsequent chapters. The incompleteness theorem is formulated and proved for decidable extensions of Peano arithmetic. Peano arithmetic is a natural collection of sentences concerning natural numbers.

We deal throughout with the language described in Chapter 2 appropriate for the structure $(\omega, S, 0, +, \cdot)$. Recall that this language has the following non-logical symbols:

$+$, a binary function symbol, taken to be 7.

$\bullet$, a binary function symbol, taken to be 9.

$\mathbf{S}$, a unary function symbol, taken to be 6.

$\mathbf{0}$, an individual constant, taken to be 8.

Also recall from Chapter 2 that the logical symbols are negation, implication, equality, and universal quantifier, taken to be the integers 1,2,3,4 respectively, and variables $v_0, v_1, \ldots$, taken to be the integers $5, 10, 15, \ldots$. Thus a symbol in our language is one of the integers $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, \ldots$.

If φ is a formula and $\sigma_0, \ldots, \sigma_{m-1}$ are terms, then by $\varphi(\sigma_0, \ldots, \sigma_{m-1})$ we mean the formula

$$\mathrm{Subf}^{v_0}_{\sigma_0} \cdots \mathrm{Subf}^{v_{m-1}}_{\sigma_{m-1}} \varphi.$$

Peano arithmetic consists of the following formulas in this language:

(P1) $\forall v_0 \forall v_1 [\mathbf{S} v_0 = \mathbf{S} v_1 \rightarrow v_0 = v_1]$.
(P2) $\forall v_0 (\neg(\mathbf{S} v_0 = \mathbf{0}))$.
(P3) $\forall v_0 [v_0 + \mathbf{0} = v_0]$.
(P4) $\forall v_0 \forall v_1 [v_0 + \mathbf{S} v_1 = \mathbf{S}(v_0 + v_1)]$.
(P5) $\forall v_0 [v_0 \bullet \mathbf{0} = \mathbf{0}]$.
(P6) $\forall v_0 \forall v_1 [v_0 \bullet \mathbf{S} v_1 = (v_0 \bullet v_1) + v_0]$.
(P7) $\varphi(\mathbf{0}) \wedge \forall v_0 (\varphi \rightarrow \varphi(\mathbf{S} v_0)) \rightarrow \forall v_0 \varphi$ for any formula φ.

Let $\mathbf{P}$ consist of all of these formulas.

The simplest form of the incompleteness theorem is that $\mathbf{P}$ is incomplete. The theorem actually applies much more generally, and our formulation gives a fairly general version. The steps in the proof of the theorem are as follows:

(1) Assign numbers to formulas and proofs. This is straightforward, and we carry it out fully in this chapter.

(2) Show that certain relations connected with the notion of proof can be represented in a certain sense within $\mathbf{P}$. In this chapter we formulate this precisely, but proofs are left to a later chapter.

(3) The actual proof of incompleteness. We do this in the present chapter, based on results about representation.

The symbols of our language are certain positive integers. A formula is a finite sequence of symbols, and a proof is a finite sequence of formulas. Thus to assign numbers to formulas and proofs we just need to assign numbers to finite sequences of positive integers. To do this we make use of the unique decomposition of a positive integer into a product of primes. Recall that a prime is a natural number greater than 1 which cannot be written as a product of smaller positive integers. Thus the primes are $2, 3, 5, 7, \ldots$. We let $p_0, p_1, \ldots$ enumerate all the primes. So $p_0 = 2$, $p_1 = 3$, $p_2 = 5$, etc. There are infinitely many primes, and so this list is infinite. The unique decomposition theorem is that every positive integer can be written as a product of primes, and this decomposition is unique except for the order of the primes. We can formulate this theorem as follows:

For every integer $m > 1$ there is a positive integer k and nonnegative integers $i(0), \ldots, i(k)$ such that

$$m = p_0^{i(0)} \cdot p_1^{i(1)} \cdot \ldots \cdot p_k^{i(k)},$$

with $i(k) \neq 0$, and if we also have

$$m = p_0^{j(0)} \cdot p_1^{j(1)} \cdot \ldots \cdot p_l^{j(l)},$$

with $j(l) \neq 0$, then $k = l$ and $i(0) = j(0), i(1) = j(1), \ldots, i(k) = j(k)$.

Now if $\varphi \stackrel{\text{def}}{=} \langle \varphi_0, \ldots, \varphi_k \rangle$ is a sequence of positive integers, we define its *Gödel number* to be

$$gn(\varphi) = p_0^{\varphi_0} \cdot p_1^{\varphi_1} \cdot \ldots \cdot p_k^{\varphi_k}.$$

If $\langle \varphi_0, \ldots, \varphi_m \rangle$ is a sequence of sequences of positive integers (for example, a proof), we define its Gödel number to be

$$gn_1(\langle \varphi_0, \ldots, \varphi_m \rangle) = p_0^{gn(\varphi_0)} \cdot \ldots p_m^{gn(\varphi_m)}.$$

Thus if Γ is a set of formulas containing $\mathbf{P}$, then $gn[\Gamma]$, which by definition is $\{gn(\varphi) : \varphi \in \Gamma\}$, is a set of natural numbers. If $\Phi \stackrel{\text{def}}{=} \langle \varphi_0, \ldots, \varphi_k \rangle$ is a Γ-proof, then $gn_1(\Phi)$ is a natural number. In general, we can use the functions gn and gn_1 to translate syntactic notions about our language into numbers or sets of numbers or relations between numbers.

Gödel numbers are large, even for simple syntactic notions, although this is not really significant for the incompleteness proof. Here are some examples. The simple formula $v_0 = v_0$ is actually the sequence $\langle 3, 5, 5 \rangle$, and its Gödel number is

$$p_0^3 \cdot p_1^5 \cdot p_2^5 = 2^3 \cdot 3^5 \cdot 5^5 = 6,075,000.$$

The Peano postulate (P2), which is the formula $\forall v_0(\neg(\mathbf{S}v_0 = \mathbf{0}))$, is actually the sequence $\langle 4, 5, 1, 3, 6, 5, 8 \rangle$, and its Gödel number is

$$p_0^4 \cdot p_1^5 \cdot p_2^1 \cdot p_3^3 \cdot p_4^6 \cdot p_5^5 \cdot p_6^8 = 2^4 \cdot 3^5 \cdot 5 \cdot 7^3 \cdot 11^6 \cdot 13^5 \cdot 17^8 \approx 3 \cdot 10^{28}.$$

The one-termed sequence $\langle\exists v_0(v_0 = v_1)\rangle$ is a very short proof, involving one instance of logical axiom (L4). The formula $\exists v_0(v_0 = v_1)$ is actually $\neg\forall v_0\neg(v_0 = v_1)$, or as a sequence, $\langle 1, 4, 5, 1, 3, 5, 10\rangle$. Hence

$$\begin{aligned} gn_1(\langle\exists v_0(v_0 = v_1)\rangle) &= 2^{gn(\langle 1,4,5,1,3,5,10\rangle)} \\ &= 2^{p_0^1\cdot p_1^4\cdot p_2^5\cdot p_3^1\cdot p_4^3\cdot p_5^5\cdot p_6^{10}} \\ &= 2^{2\cdot 3^4\cdot 5^5\cdot 7\cdot 11^3\cdot 13^5\cdot 17^{10}} \\ &\approx 2^{3.5\cdot 10^{27}}. \end{aligned}$$

Proposition 5.1. *gn is a one-one function.*

Proof. Note that (implicitly) the domain of gn is the set of all nonempty sequences of positive integers. Suppose that two such sequences $\langle m_0, \ldots, m_k\rangle$ and $\langle n_0, \ldots, n_l\rangle$ are given, and $gn\langle m_0, \ldots, m_k\rangle) = gn(\langle n_0, \ldots, n_l\rangle$. Thus

$$p_0^{m_0+1} \cdot p_1^{m_1+1} \cdot \ldots \cdot p_k^{m_k+1} = p_0^{n_0+1} \cdot p_1^{n_1+1} \cdot \ldots \cdot p_l^{n_l+1}.$$

By the uniqueness of prime decompositions, $k = l$ and $m_i = n_i$ for all $i \leq k$. □

We can use Gödel numbers to introduce a notation which will be technically convenient. If φ is a formula, then $v_{gn(\varphi)}$ is a variable which does not occur in φ. In fact, with $\varphi = \langle\varphi_0, \ldots, \varphi_m\rangle$, for each $i \leq m$ we have $\varphi_i < gn(\varphi)$, and so we cannot have $\varphi_i = 5(gn(\varphi) + 1) = v_{gn(\varphi)}$. Also note that $gn(\varphi) \neq 0$. Now we define for any term σ involving at most the variable v_0 and any formula φ,

$$\mathrm{Subff}^{v_0}_\sigma\varphi = \forall v_{gn(\varphi)}[v_{gn(\varphi)} = \sigma \to \forall v_0[v_0 = v_{gn(\varphi)} \to \varphi]].$$

Proposition 5.2. *If σ is a term involving at most the variable v_0, then $\vdash \mathrm{Subff}^{v_0}_\sigma\varphi \leftrightarrow \mathrm{Subf}^{v_0}_\sigma\varphi$.*

Proof. Again we argue model-theoretically, showing that $\models \mathrm{Subff}^{v_0}_\sigma\varphi \leftrightarrow \mathrm{Subf}^{v_0}_\sigma\varphi$, so that the proposition follows by the completeness theorem.

Suppose that $\overline{A}$ is a structure for our language, and suppose that $a : \omega \to A$.

First suppose that $\overline{A} \models \mathrm{Subff}^{v_0}_\sigma\varphi[a]$. Let $x = \sigma^{\overline{A}}(a)$. Since $gn(\varphi) > 0$ it follows that $v_{gn(\varphi)}$ does not occur in σ. Hence $\sigma^{\overline{A}}(a) = \sigma^{\overline{A}}(a^{gn(\varphi)}_x)$ by Proposition 2.4. Thus $\overline{A} \models (v_{gn(\varphi)} = \sigma)[a^{gn(\varphi)}_x]$. Now $\overline{A} \models (v_0 = v_{gn(\varphi)})[a^{0\ gn(\varphi)}_{x\ x}]$ so, since $\overline{A} \models \mathrm{Subff}^{v_0}_\sigma\varphi[a]$, we get $\overline{A} \models \varphi[a^{0\ gn(\varphi)}_{x\ x}]$. Since $v_{gn(\varphi)}$ does not occur in φ, it follows from Lemma 4.4 that $\overline{A} \models \varphi[a^0_x]$. Now no free occurrence of v_0 in φ is within a subformula $\forall v_k\mu$ with v_k occurring in σ, since only v_0 possibly occurs in σ. So by Lemma 4.6 it follows that $\overline{A} \models \mathrm{Subf}^{v_0}_\sigma\varphi[a]$.

Conversely, suppose that $\overline{A} \models \mathrm{Subf}^{v_0}_\sigma\varphi[a]$. Let $x = \sigma^{\overline{A}}(a)$. Again by Lemma 4.6 we get

$$(*) \qquad \overline{A} \models \varphi[a^0_x].$$

Now suppose that $y \in A$; we want to show that $(v_{gn(\varphi)} = \sigma \to \forall v_0[v_0 = v_{gn(\varphi)} \to \varphi])[a^{gn(\varphi)}_y]$. To this end, suppose that $\overline{A} \models (v_{gn(\varphi)} = \sigma)[a^{gn(\varphi)}_y]$; we want to show that $\overline{A} \models \forall v_0[v_0 = v_{gn(\varphi)} \to \varphi][a^{gn(\varphi)}_y]$, and to do this we take any $x \in A$, assume that $\overline{A} \models (v_0 = v_{gn(\varphi)})[a^{0\ gn(\varphi)}_{x\ y}]$, and prove that $\overline{A} \models \varphi[a^{0\ gn(\varphi)}_{x\ y}]$. Since $\overline{A} \models ((v_{gn(\varphi)} = \sigma)[a^{gn(\varphi)}_y]$, we have $y = \sigma^{\overline{A}}(a^{gn(\varphi)}_y)$, and so by Proposition 2.4 $y = \sigma^{\overline{A}}(a)$ since $v_{gn(\varphi)}$ does not occur in σ. Also, since $\overline{A} \models (v_0 = v_{gn(\varphi)})[a^{0\ gn(\varphi)}_{x\ y}]$, we have $x = y$. Now by $(*)$, since $v_{gn(\varphi)}$ does not occur in φ we have $\overline{A} \models \varphi[a^{0\ gn(\varphi)}_{x\ y}]$, as desired. □

A particular use of this notation is the following replacement of the Peano Postulate (P7):

(P7′) $\mathrm{Subff}^{v_0}_{\mathbf{0}}\varphi \wedge \forall v_0(\varphi \to \mathrm{Subff}^{v_0}_{\mathbf{S}v_0}\varphi) \to \forall v_0\varphi$ for any formula φ.

Let $\mathbf{P}'$ be $\mathbf{P}$ with (P7) replaced by (P7′).

We turn to the second step in the proof of the incompleteness theorem: representation of number-theoretic relations in $\mathbf{P}$. To do this we need an auxiliary important notion: recursive functions. Recursive functions are functions which are computable in a general sense. In fact, there are many equivalent definitions of the set of recursive functions; for example, Turing-computable functions, functions computable by an abstract computer, or functions representable in $\mathbf{P}$.

Essentially the set of recursive functions is defined as the closure of a simple set of functions under some operations on functions. The starting set of functions is as follows. First we have the successor function $\mathbf{s}$, which assigns to each natural number m its successor $m+1$. Second, if $i, m \in \omega$ and $i < m$, we have the *projection function* $\mathbf{I}^m_i : {}^m\omega \to \omega$ defined by setting

$$\mathbf{I}^m_i(a_0, \ldots, a_{m-1}) = a_i$$

for any $a_0, \ldots, a_{m-1} \in \omega$. Those are all of the starting functions.

There are three kinds of operations on functions. First we have *composition operations.* If m, n are positive integers, then $\mathbf{C}^m_n$ acts upon an $(m+1)$-tuple $\langle f, g_0, \ldots, g_{m-1}\rangle$ such that f is an m-ary operation on ω and each g_i is an n-ary operation on ω; $\mathbf{C}^m_n(f, g_0, \ldots, g_{m-1})$ itself is an n-ary operation on ω defined like this: let $a_0, \ldots, a_{n-1} \in \omega$; then

$$\mathbf{C}^m_n(f, g_0, \ldots, g_{m-1})(a_0, \ldots, a_{n-1}) = \\ f(g_0(a_0, \ldots, a_{n-1}), \ldots, g_{m-1}(a_0, \ldots, a_{n-1})).$$

Second, we have operations of *primitive recursion.* There are two kinds of primitive recursion, without and with parameters. The operation $\mathbf{Q}_0$ acts on a pair (a, f) consisting of a natural number a and a binary operation f on ω to produce a function from ω into ω; it is defined, by recursion, as follows: for any $m \in \omega$,

$$\mathbf{Q}_0(a, f)(0) = a;$$
$$\mathbf{Q}_0(a, f)(m+1) = f(m, \mathbf{Q}_0(a, f)(m)).$$

With parameters, for each positive integer n we have an operation $\mathbf{Q}_n$ acting on a pair (f, g) consisting of an n-ary operation f on ω and an $(n+2)$-ary operation on ω to produce an

$(n+1)$-ary operation on ω, defined, by recursion, as follows. For any $a_0, \ldots, a_{n-1}, m \in \omega$,

$$\mathbf{Q}_n(f,g)(a_0,\ldots,a_{n-1},0) = f(a_0,\ldots,a_{n-1});$$
$$\mathbf{Q}_n(f,g)(a_0,\ldots,a_{n-1},m+1) = g(a_0,\ldots,a_{n-1},m,\mathbf{Q}_n(f,g)(a_0,\ldots,a_{n-1},m)).$$

The third kind of operations on functions is *minimalization.* For each positive integer m we have such an operation, $\mathbf{M}_m$, which acts only on certain *special* $(m+1)$-ary functions f on ω. Namely, an $(m+1)$-ary function f is *special* provided that for all $a_0, \ldots, a_{m-1} \in \omega$ there is a $b \in \omega$ such that $f(a_0, \ldots, a_{m-1}, b) = 0$. Then $\mathbf{M}_m(f)$ is defined like this: for any $a_0, \ldots, a_{m-1} \in \omega$, $\mathbf{M}_m(f)(a_0, \ldots, a_{m-1})$ is the least $b \in \omega$ such that $f(a_0, \ldots, a_{m-1}, b) = 0$.

Now a *recursive function construction sequence* is a sequence $\langle f_0, \ldots, f_m \rangle$ of functions such that for each $i \leq m$ one of the following conditions holds:

(1) $f_i = \mathbf{s}$;

(2) There exist $i, m \in \omega$ with $i < m$ such that $f = \mathbf{I}^m_i$.

(3) There exist positive integers m, n and $j, k_0, \ldots, k_{m-1} < i$ such that each f_{k_s} is an n-ary operation on ω, f_j is an m-ary operation on ω, and $f_i = \mathbf{C}^m_n(f_j, f_{k_0}, \ldots, f_{k_{m-1}})$.

(4) There exist $a \in \omega$ and $j < i$ such that f_j is a binary operation on ω and $f = \mathbf{Q}_0(a, f_j)$.

(5) There exist a positive integer n and $j, k < i$ such that f_j is an n-ary operation on ω, f_k is an $(n+2)$-ary operation on ω, and $f_i = \mathbf{Q}_n(f_j, f_k)$.

(6) There exist a positive integer m and a $j < i$ such that f_j is a special $(m+1)$-ary operation on ω and $f_i = \mathbf{M}_m(f_j)$.

Now a *recursive function* is a function that appears in some recursive function construction sequence.

Furthermore, for any positive integer m we say that a set B of m-termed sequences of members of ω is *recursive* if its characteristic function χ_B is recursive. Recall that $\chi_B(a_0, \ldots, a_{m-1}) = 1$ if $\langle a_0, \ldots, a_{m-1} \rangle \in B$, and $\chi_B(a_0, \ldots, a_{m-1}) = 0$ if $\langle a_0, \ldots, a_{m-1} \rangle \notin B$.

We need the following notation. With each natural number m we associate a term $\overline{m}$ of our language, by recursion:

$$\overline{0} = \mathbf{0}; \quad \overline{m+1} = \mathbf{S}\overline{m}.$$

Thus

$$\overline{m} = \overbrace{\mathbf{SS}\cdots\mathbf{S}}^{m \text{ times}}\mathbf{0}.$$

An m-ary function f is *representable* provided that there is a formula φ with free variables among $v_0, \ldots, v_m$ such that for all $a_0, \ldots, a_{m-1} \in \omega$ the following conditions hold:

$$\mathbf{P} \vdash \varphi(\overline{a_0}, \ldots, \overline{a_{m-1}}, \overline{f(a_0, \ldots, a_{m-1})});$$
$$\mathbf{P} \vdash \forall v_m[\varphi(\overline{a_0}, \ldots, \overline{a_{m-1}}, v_m) \rightarrow v_m = \overline{f(a_0, \ldots, a_{m-1})}].$$

An m-ary relation is a set of m-tuples of natural numbers. Let R be an m-ary relation. It is *representable* if there is a formula φ with free variables among $v_0, \ldots, v_{m-1}$ such that for all $a_0, \ldots, a_{m-1} \in \omega$ the following conditions hold:

$$\langle a_0, \ldots, a_{m-1} \rangle \in R \quad \text{implies that} \quad \mathbf{P} \vdash \varphi(\overline{a_0}, \ldots, \overline{a_{m-1}});$$
$$\langle a_0, \ldots, a_{m-1} \rangle \notin R \quad \text{implies that} \quad \mathbf{P} \vdash \neg\varphi(\overline{a_0}, \ldots, \overline{a_{m-1}}).$$

The first technical result we need is:

Theorem A. *All recursive functions and all recursive relations are representable.*

This will be proved in a later chapter Also, later we prove:

Theorem B. *The set $gn[\mathbf{P}']$ is recursive. Moreover, if Δ is a finite set of formulas, then $gn[\mathbf{P}' \cup \Delta]$ is recursive.*

The proof of the incompleteness theorem depends upon the following function $G : \omega \to \omega$ which involves a crucial trick. For any $m \in \omega$,

$$G(m) = \begin{cases} gn(\text{Subff}^{v_0}_{\overline{m}}\varphi) & \text{if } m = gn(\varphi) \text{ for some formula } \varphi, \\ 0 & \text{otherwise.} \end{cases}$$

Thus G is defined like this. Given any natural number m, if m is not the Gödel number of a formula, then $G(m) = 0$. If m is the Gödel number of a formula φ, then that formula φ is uniquely determined by Theorem 5.1, and we let $G(m)$ be the formula $\text{Subff}^{v_0}_{\overline{m}}\varphi$, which by Theorem 5.2 is provably equivalent to $\varphi(\overline{m})$. This is a kind of diagonal procedure. Now in a later chapter we prove:

Theorem C. *G is recursive.*

Now we can prove the following fundamental theorem, based on Theorems A and C.

Theorem 5.3. (Fixed point theorem) *For any formula φ there is a formula ψ such that $\mathbf{P} \vdash \psi \leftrightarrow \varphi(\overline{gn(\psi)})$.*

Proof. Let χ be a formula with at most v_0 and v_1 free which represents G; χ exists by Theorems A and C. Let v_i be a variable not occurring in φ or χ. Let θ be the formula $\exists v_i[\varphi(v_i) \wedge \chi(v_0, v_i)]$, and for brevity let $m = gn(\theta)$. Let $\psi = \text{Subff}^{v_0}_{\overline{m}}\theta$. Thus

(1) $G(m) = gn(\psi)$.

Now by definition of "representable", $\mathbf{P} \vdash \chi(\overline{m}, \overline{G(m)})$. Hence a tautology gives

(2) $\mathbf{P} \vdash \varphi(\overline{G(m)}) \to [\varphi(\overline{G(m)}) \wedge \chi(\overline{m}, \overline{G(m)})]$.

By Theorem 3.33 we have $\mathbf{P} \vdash [\varphi(\overline{G(m)}) \wedge \chi(\overline{m}, \overline{G(m)})] \to \exists v_i[\varphi(v_i) \wedge \chi(\overline{m}, v_i)]$; hence by (2), $\mathbf{P} \vdash \varphi(\overline{G(m)}) \to \exists v_i[\varphi(v_i) \wedge \chi(\overline{m}, v_i)]$. Since $\theta(\overline{m})$ is $\exists v_i[\varphi(v_i) \wedge \chi(\overline{m}, v_i)]$, it follows from the definition of ψ and Theorem 5.2 that $\mathbf{P} \vdash \varphi(\overline{G(m)}) \to \psi$. Hence by (1) we have

(3) $\mathbf{P} \vdash \varphi(\overline{gn(\psi)}) \to \psi$.

Now by the second condition on representability we have $\mathbf{P} \vdash \forall v_1[\chi(\overline{m}, v_1) \to v_1 = \overline{G(m)}]$. Hence by Theorem 3.27, since v_i does not occur in χ, we get

(4) $\mathbf{P} \vdash \chi(\overline{m}, v_i) \to v_i = \overline{G(m)}$.

By Theorem 3.18, $\mathbf{P} \vdash v_i = \overline{G(m)} \to [\varphi(v_i) \leftrightarrow \varphi(\overline{G(m)})]$. Hence by (4) and a tautology, $\mathbf{P} \vdash \varphi(v_i) \wedge \chi(\overline{m}, v_i) \to \varphi(\overline{G(m)})$. Hence by generalization and Proposition 3.39, since v_i does not occur in φ or χ we get $\mathbf{P} \vdash \exists v_i[\varphi(v_i) \wedge \chi(\overline{m}, v_i)] \to \varphi(\overline{G(m)})$. By Theorem 5.2 and the definition of ψ, $\mathbf{P} \vdash \psi \to \varphi(\overline{G(m)})$. Together with (1) and (3) this finishes the proof. □

Note from the proof of Theorem 5.3 that if the free variables of φ are among v_0, then the formula ψ which is defined is a sentence.

The following result will be used several times.

Theorem 5.4. *Let $\overline{M}$ be the structure $(\omega, S, 0, +, \cdot)$. Then $\overline{M} \models \mathbf{P}$.*

Proof. (P1)–(P6) are straightforward: (P1): If $a, b \in \omega$ and $S(a) = S(b)$, then $a + 1 = b + 1$, and hence $a = b$. (P2): For any $a \in \omega$, $S(a) = a + 1 \neq 0$. (P3): For any $a \in \omega$, $a + 0 = a$. (P4): For any $a, b \in \omega$, $a + S(b) = a + b + 1 = S(a + b)$. (P5): For any $a \in \omega$, $a \cdot 0 = 0$. (P6): For any $a, b \in \omega$, $a \cdot S(b) = a \cdot (b + 1) = a \cdot b + a$.

(P7) requires more work. Let φ be any formula, and let $a : \omega \to \omega$ be any assignment. Assume that

$$\overline{M} \models [\varphi(\mathbf{0}) \wedge \forall v_0(\varphi \to \varphi(\mathbf{S}v_0))][a].$$

We prove by induction that for every $u \in \omega$, $\overline{M} \models \varphi[a^0_u]$. For $u = 0$ we have $\overline{A} \models \varphi(\mathbf{0})[a]$. That is, $\overline{A} \models \mathrm{Subf}^{v_0}_{\mathbf{0}}[a]$. By Lemma 4.6 it follows that $\overline{A} \models \varphi[a^0_{\mathbf{0}(a)}]$. Now $\mathbf{0}(a) = 0$, so $\overline{A} \models \varphi[a^0_0]$. So our statement holds for $u = 0$. Now assume that $\overline{M} \models \varphi[a^0_u]$. Since $\overline{M} \models \forall v_0(\varphi \to \varphi(\mathbf{S}v_0))][a]$, it follows that $\overline{M} \models (\varphi \to \varphi(\mathbf{S}v_0))][a^0_u]$. Since $\overline{A} \models \varphi[a^0_u]$, we then have $\overline{A} \models \varphi(\mathbf{S}v_0))][a^0_u]$. That is, $\overline{M} \models (\mathrm{Subf}^{v_0}_{\mathbf{S}v_0}\varphi)[a^0_u]$. By Lemma 4.6 we then get $\overline{A} \models \varphi[(a^0_u)^0_x]$, where $x = (\mathbf{S}v_0)^{\overline{A}}(a^0_u) = u + 1$. Since $(a^0_u)^0_x = a^0_x$, this completes the inductive proof.

Now it follows that $\overline{A} \models \forall v_0\varphi[a]$. □

Our first important theorem concerning incompleteness is as follows.

Theorem 5.5. (Tarski's undefinability of truth theorem) *There is no formula φ with only v_0 free such that for every sentence ψ, $\overline{M} \models \psi \leftrightarrow \varphi(\overline{gn(\psi)})$, where $\overline{M} = (\omega, S, 0, +, \cdot)$.*

Proof. Suppose there is such a formula φ. By the comment following the proof of Theorem 5.3 let ψ be a sentence such that $\mathbf{P} \vdash \psi \leftrightarrow \neg\varphi(\overline{gn(\psi)})$. By Theorem 5.4 and Theorem 3.2 we then get $\overline{M} \models \psi \leftrightarrow \neg\varphi(\overline{gn(\psi)})$. But by hypothesis, $\overline{M} \models \psi \leftrightarrow \varphi(\overline{gn(\psi)})$. However, a structure cannot model both a sentence and its negation. □

A formula φ as described in Theorem 5.5 would have provided a truth definition for $\overline{M}$: to check whether a sentence ψ held in $\overline{M}$ one could look at its Gödel number m and see whether $\overline{M} \models \varphi(\overline{m})$ or $\overline{M} \models \neg\varphi(\overline{m})$, giving $\overline{M} \models \psi$ or $\overline{M} \models \neg\psi$ respectively.

For the proof of the incompleteness theorem we need another result proved in a later chapter.

Theorem D. *If* Γ *is a set of formulas and* $gn[\Gamma]$ *is recursive, then the following binary relation is also recursive:*

$$\begin{aligned}\mathrm{Prf}_\Gamma \overset{\text{def}}{=} \{(n,m) : & \textit{ there is a } \Gamma\textit{-proof } \Phi \textit{ with last entry } \varphi \\ & \textit{such that } m = gn_1(\Phi) \textit{ and } n = gn(\varphi)\}\end{aligned}$$

Theorem 5.6. (Gödel's incompleteness theorem) *If* Γ *is a set of sentences containing* $\mathbf{P}'$, $(\omega, S, 0, +, \cdot)$ *is a model of* Γ, *and* $gn[\Gamma]$ *is recursive, then* Γ *is incomplete.*

The hypotheses say intuitively that Γ is a set of true sentences, and there is an effective procedure for deciding whether a given sentence is in Γ or not.

Proof. By Theorems A and D, the relation Prf_Γ is representable, say by a formula χ that has at most v_0 and v_1 free. Thus

(1) For all $m, n \in \omega$, if $(m, n) \in \mathrm{Prf}_\Gamma$ then $\mathbf{P} \vdash \chi(\overline{m}, \overline{n})$.

(2) For all $m, n \in \omega$, if $(m, n) \notin \mathrm{Prf}_\Gamma$ then $\mathbf{P} \vdash \neg\chi(\overline{m}, \overline{n})$.

Let φ be the formula $\forall v_1(\neg\chi)$. Thus φ has at most v_0 free. For any sentence σ, $\varphi(\overline{gn(\sigma)})$ intuitively says that σ is not provable. Now we apply the comment after the fixed point theorem 5.3 to get a sentence ψ such that

(3) $\mathbf{P} \vdash \psi \leftrightarrow \varphi(\overline{gn(\psi)})$.

Note that

(4) $\varphi(\overline{gn(\psi)})$ is the formula $\forall v_1 \neg\chi(\overline{gn(\psi)})$.

So ψ says intuitively that it itself is not provable. This is the essential trick of the proof. It is derived from the liar paradox, concerning a man who says "I am lying": it true, it false, and if false, it is true.

We claim that ψ shows that Γ is incomplete: $\Gamma \not\vdash \psi$ and $\Gamma \not\vdash \neg\psi$.

Suppose that $\Gamma \vdash \psi$. Let Φ be a Γ-proof with last entry ψ. Then by (1), $\mathbf{P} \vdash \chi(\overline{gn(\psi)}, \overline{gn_1(\Phi)})$. Hence by Theorem 3.33, $\mathbf{P} \vdash \exists v_1 \chi(\overline{gn(\psi)})$. Thus $\mathbf{P} \vdash \neg\forall v_1 \neg\chi(\overline{gn(\psi)})$ by the definition of $\exists$. By (4) this says that $\mathbf{P} \vdash \neg\varphi(\overline{gn(\psi)})$. Hence by (3) we get $\mathbf{P} \vdash \neg\psi$. But $(\omega, S, 0, +, \cdot)$ is a model of both Γ and $\mathbf{P}$, and $\Gamma \vdash \psi$ by hypothesis. This contradicts Theorem 3.2.

It follows that not$(\Gamma \vdash \psi)$. Hence for every natural number n we have $(gn(\psi), n) \notin \mathrm{Prf}_\Gamma$, and hence by (2), $\mathbf{P} \vdash \neg\chi(\overline{gn(\psi)}, \overline{n})$. It follows that

(5) not$(\Gamma \vdash \exists v_1 \chi(\overline{gn(\psi)}))$.

In fact, otherwise by Theorem 3.2, since $(\omega, S, 0, +, \cdot)$ is a model of Γ, we would get $(\omega, S, 0, +, \cdot) \models \exists v_1 \chi(\overline{g(\psi)})$, and hence there is some $n \in \omega$ such that $(\omega, S, 0, +, \cdot) \models$

$\chi(\overline{g(\psi)}, \overline{n})$. This contradicts $\mathbf{P} \vdash \neg\chi(\overline{g(\psi)}, \overline{n})$ in view of Theorems 5.4 and 3.2. Thus (5) holds.

Since $\exists v_1 \chi(\overline{g(\psi)})$ is $\neg\forall v_1 \neg\chi(\overline{gn(\psi)})$, i.e. by (4) it is $\neg\varphi(\overline{gn(\psi)})$, it follows from (3) and (5) that not$(\Gamma \vdash \neg\psi)$. □

It is important from an intuitive point of view that the sentence produced in this proof is actually true. This is expressed in the following corollary.

Corollary 5.7. *If Γ is a set of sentences containing $\mathbf{P}'$, $(\omega, S, 0, +, \cdot)$ is a model of Γ, and $g[\Gamma]$ is recursive, then the sentence ψ defined in the proof of Theorem 5.6 holds in the structure $(\omega, S, 0, +, \cdot)$.*

Proof. Since ψ is not provable from Γ, it follows that for every $n \in \omega$, $(gn(\psi), n) \notin \mathrm{Prf}_\Gamma$, and hence $\mathbf{P} \vdash \neg\chi(\overline{gn(\psi)}, \overline{n})$. Hence by Theorems 3.2 and 5.4, $(\omega, S, 0, +, \cdot) \models \neg\chi(\overline{gn(\psi)}, \overline{n})$. Since this is true for every $n \in \omega$, it follows that

$$(\omega, S, 0, +, \cdot) \models \forall v_1 \neg\chi(\overline{gn(\psi)}, v_1).$$

By (3) and (4) in the proof of Theorem 5.6 it follows that $(\omega, S, 0, +, \cdot) \models \psi$. □

EXERCISES

E5.1. Give an estimate for the size of $gn(\varphi)$, where φ is the Peano Postulate (P1).

E5.2. Describe $G(gn(v_0 = v_0))$ and express it as a product of primes.

Suppose that Γ is a set of sentences containing $\mathbf{P}'$. A formula ρ with at most v_0 free is a Γ-*provability condition* iff for any sentence φ, $\Gamma \vdash \varphi$ iff $\Gamma \vdash \rho(\overline{gn(\varphi)})$.

E5.3. Suppose that Γ is a set of sentences containing $\mathbf{P}'$, and $\overline{M} \overset{\text{def}}{=} (\omega, S, 0, +, \cdot)$ is a model of Γ. Let χ be as in the proof of Gödel's incompleteness theorem, and let π be the formula $\exists v_1 \chi$. Prove that π is a Γ-provability condition.

E5.4. Suppose that Γ is a set of sentences containing $\mathbf{P}'$, and $\overline{M} \overset{\text{def}}{=} (\omega, S, 0, +, \cdot)$ is a model of Γ. Assume that ρ is a Γ-provability condition. Apply the fixed point theorem to get a sentence ψ such that $\mathbf{P} \vdash \psi \leftrightarrow \neg\rho(\overline{gn(\psi)})$, as in the proof of Gödel's incompleteness theorem. Prove that not$(\Gamma \vdash \psi)$ and not$(\Gamma \vdash \neg\psi)$.

E5.5. Let χ be as in the proof of Gödel's incompleteness theorem, and let π be the formula $\exists v_1 \chi$. The following can be shown for χ.

(i) For any sentences φ and ψ,

$$\Gamma \vdash \pi(\overline{gn(\varphi \to \psi)}) \to (\pi(\overline{gn(\varphi)}) \to \pi(\overline{gn(\psi)})).$$

("Γ proves that if $\varphi \to \psi$ is provable, then from the provability of φ it follows that ψ is provable")

(ii) For any sentence φ,

$$\Gamma \vdash \pi(\overline{gn(\varphi)}) \to \pi(\overline{gn(\pi(\overline{gn(\varphi)}))}).$$

("Γ proves that if φ is provable, then it is provable that φ is provable."

By the fixed point theorem, let ψ be a sentence such that $\Gamma \vdash \psi \leftrightarrow \pi(\overline{gn(\psi)})$. Note that ψ says "I am provable". By the fixed point theorem again, let θ be a sentence such that $\Gamma \vdash \theta \leftrightarrow (\pi(\overline{gn(\theta)}) \rightarrow \psi)$. Thus θ says "If I am provable, then ψ holds."

Show that if $\Gamma \vdash \theta$, then $\Gamma \vdash \psi$.

E5.6. (Continuing E5.5.) Show that $\Gamma \vdash \pi(\overline{gn(\theta)}) \rightarrow \pi(\overline{gn(\psi)})$.

E5.7. (Continuing E5.5.) Show that $\Gamma \vdash \psi$.

E5.8. Assume that $\Gamma \vdash \neg(\overline{m} = \overline{n})$ for all distinct $m, n \in \omega$. Let χ be as in the proof of Gödel's incompleteness theorem, and let θ be the sentence $\forall v_0(v_0 = v_0)$. Prove that the following formula $\rho(v_0)$ is a provabilty condition:

$$v_0 = \overline{gn(\theta)} \vee \exists v_1 \chi(v_0, v_1)$$

E5.9. (Continuing E5.8) Prove that $\Gamma \vdash \theta \leftrightarrow \rho(\overline{gn(\theta)})$ (so that θ asserts its own provability with respect to this condition).

E5.10. (Continuing E5.8) Prove that $\Gamma \vdash \theta$.

E5.11. Assume that $\Gamma \vdash \neg(\overline{m} = \overline{n})$ for all distinct $m, n \in \omega$. Let χ be an in the proof of Gödel's incompleteness theorem, and let θ be the sentence $\neg\forall v_0(v_0 = v_0)$. Prove that the following formula $\rho(v_0)$ is a provabilty condition:

$$\neg(v_0 = \overline{gn(\theta)}) \wedge \exists v_1 \chi(v_0, v_1)$$

E5.12. (Continuing E5.11) Show that $\Gamma \vdash \theta \leftrightarrow \rho(\overline{gn(\theta)})$, so that θ asserts its own provability.

E5.13. (Continuing E5.11) Show that $\Gamma \vdash \neg\theta$.