2. INCOMPLETENESS

The methods of arithmetization and self–reference were originally used to prove incompleteness theorems for arithmetical theories. In this chapter we present the most important theorems of this type.

A sentence φ (in the language of S) is *undecidable* in S if $S \nvdash \varphi$ and $S \nvdash \neg\varphi$. S is *complete* if no sentence is undecidable in S, otherwise *incomplete*.

§1. Incompleteness. We begin with the first and most important result of the whole subject, Gödel's incompleteness theorem (for theories in L_A).

Theorem 1. Let φ be a Π_1 sentence such that

(G) $\quad Q \vdash \varphi \leftrightarrow \neg Pr_T(\varphi)$.

Then φ is true and $T \nvdash \varphi$. Thus, if T is Σ_1–sound, then also $T \nvdash \neg\varphi$.

Proof. Suppose $T \vdash \varphi$. Then, by Fact 7 (b), $Q \vdash Pr_T(\varphi)$. But then, by (G), $Q \vdash \neg\varphi$ and so $T \vdash \neg\varphi$. It follows that T is inconsistent, contrary to Convention 2. Thus, $T \nvdash \varphi$. By (G), φ is true. Thus, $\neg\varphi$ is a false Σ_1 sentence and so $T \nvdash \neg\varphi$ if T is Σ_1–sound. ■

Notice the close similarity between the proofs of Theorem 1, Lemma 1.2, and Theorem 1.3 (the liar paradox).

To derive the conclusion that $T \nvdash \neg\varphi$ in Theorem 1, we needed the assumption that T is Σ_1–sound. We can now see that this is stronger than mere consistency: $T + \neg\varphi$ is consistent but not Σ_1–sound. (Note that it does not follow from Theorem 1 that $T + \neg\varphi$ is incomplete.) Thus, the question arises if, assuming consistency only, there is a (Π_1) sentence which is undecidable in T. Our next result, known as Rosser's theorem, shows that the answer is affirmative.

Theorem 2. Let θ be a Π_1 sentence such that

(R) $\quad Q \vdash \theta \leftrightarrow \forall z(Prf_T(\theta,z) \rightarrow \exists u{\le}z Prf_T(\neg\theta,u))$.

Then θ is undecidable in T.

Proof. We first prove that $T \nvdash \theta$. Suppose, for *reductio ad absurdum*, $T \vdash \theta$ and let p be a proof of θ in T. Then, by Fact 7 (a),

(1) $\quad Q \vdash Prf_T(\theta,p)$.

Since T is consistent, we have $T \nvdash \neg\theta$. By Fact 7 (d), $Q \vdash \neg Prf_T(\neg\theta,q)$ for every q. But then, by Fact 1 (iv),

$$Q \vdash u \le p \rightarrow \neg Prf_T(\neg\theta,u).$$

Combining this with (1) we get

$$Q \vdash \exists z(Prf_T(\theta,z) \wedge \forall u{\le}z \neg Prf_T(\neg\theta,u)).$$

But then, by (R), $Q \vdash \neg\theta$ and so $T \vdash \neg\theta$, a contradiction. Thus, $T \nvdash \theta$ as desired.

Next we prove that $T \nvdash \neg\theta$. Suppose $T \vdash \neg\theta$ and let p be a proof of $\neg\theta$ in T. Then $T \nvdash \theta$ and so, by Fact 7 (d), $Q \vdash \neg Prf_T(\theta,q)$ for every q. By Fact 1 (iv), it follows that

$$Q \vdash z < p \rightarrow \neg Prf_T(\theta,z),$$

whence, by Fact 1 (v),

(2) $\quad Q \vdash Prf_T(\theta,z) \rightarrow p \leq z.$

By Fact 7 (a), $Q \vdash Prf_T(\neg\theta,p)$. Hence, trivially,

$$Q \vdash p \leq z \rightarrow \exists u {\leq} z Prf_T(\neg\theta,u).$$

Combining this with (2) and (R) we get $Q \vdash \theta$ and so $T \vdash \theta$, again a contradiction. It follows that $T \nvdash \neg\theta$, as desired. ■

Arguments similar to the above proof will occur time and again in the following pages.

Theorem 2 can also be proved by considering a Σ_1 sentence ψ such that

(R′) $\quad Q \vdash \psi \leftrightarrow \exists z(Prf_T(\neg\psi,z) \wedge \forall u {\leq} z \neg Prf_T(\psi,u)),$

a condition that is, of course, (almost) satisfied by $\neg\theta$, where θ is as in (R). A sentence satisfying (R) or (R′) is called a *Rosser sentence* for T.

The difference between (the proofs of) Theorems 1 and 2 can be described in the following way. The formula $\xi(x) := Pr_T(x)$ used in the former has the properties: (i) if $T \vdash \varphi$, then $T \vdash \xi(\varphi)$, and (ii) if $T \vdash \neg\varphi$, then ($T \nvdash \varphi$ and so) $\xi(\varphi)$ is false. The corresponding formula which is (almost) used in the latter,

$$\xi(x) := \exists z(Prf_T(x,z) \wedge \forall u {\leq} z \neg Prf_T(\neg x,u)),$$

satisfies (i) and (iii): if $T \vdash \neg\varphi$, then $T \vdash \neg\xi(\varphi)$. From (i) and (iii) it follows at once that if

$$T \vdash \psi \leftrightarrow \neg\xi(\psi)$$

(or $T \vdash \psi \leftrightarrow \xi(\neg\psi)$), then ψ is undecidable in T.

If $PA \dashv T$, the above proof of Theorem 2 can be replaced by the following argument. Suppose $T \vdash \theta$. Then $T \nvdash \neg\theta$. By (R), it follows that $\neg\theta$ is true and so, $\neg\theta$ being Σ_1, $T \vdash \neg\theta$, by Fact 9 (a), a contradiction. (This part does not use the assumption that $PA \dashv T$.) Next suppose $T \vdash \neg\theta$. Then $T \nvdash \theta$. But then, by Corollary 1.10 (a) and (R), $T \vdash \theta$, again a contradiction.

That T is incomplete also follows from Theorem 1.2, since every complete r.e. theory is decidable. This proof, however, does not (directly) yield an example of a sentence undecidable in T. Furthermore, the present proof of Theorem 1 is needed in the proof of Theorem 4, below.

That every complete r.e. theory U is decidable is seen as follows: If U is inconsistent, decidability is trivial; thus, suppose U is consistent. Let φ be any sentence of U. To decide whether or not $\varphi \in Th(U)$, generate, in some effective way, all proofs in U. If a proof of φ is found, conclude that $\varphi \in Th(U)$; if a proof of $\neg\varphi$ is found, conclude that $\varphi \notin Th(U)$.

Conversely, Theorem 1.2 follows from Theorem 2. Indeed, suppose U is a consistent, decidable extension of Q. There is then a complete, recursive, consistent extension U′ of U. U′ is an extension of Q. Hence, by Craig's theorem (Theorem 1.1), there is a complete, consistent primitive recursive extension of Q. This, however, contradicts Theorem 2.

That any consistent, decidable theory U has a complete, consistent, decidable extension can be seen as follows: Let $\varphi_0, \varphi_1, \ldots$ be an effective enumeration of all sentences of the language of U. Define U_n by: $U_0 = U$, $U_{n+1} = U_n + \varphi_n$ if $U_n \nvdash \neg\varphi_n$, $U_{n+1} = U_n + \neg\varphi_n$ otherwise. Let $U' = \bigcup\{U_n: n\in N\}$. Then U' is complete and consistent. By assumption, it can be effectively decided whether $U_n \vdash \neg\varphi_n$ or not. It follows that U' is decidable.

Theorem 2 can be strengthened as follows. A family $\{T_k: k\in N\}$ of theories is *r.e.* if the binary relation $\varphi\in T_k$ is r.e.

Theorem 3. If $\{T_k: k\in N\}$ is an r.e. family of theories, there is a Π_1 sentence which is simultaneously undecidable in all the theories T_k.

We derive Theorem 3 from the following slight improvement of Theorem 2.

Let us say that a set X of sentences is *monoconsistent with* T if T + φ is consistent for every $\varphi\in X$. Thus, for example, if φ is undecidable in T, then $\{\varphi, \neg\varphi\}$ is monoconsistent with T. Also, if X and Y are monoconsistent with T, so is $X \cup Y$. Let $\varphi^0 := \varphi$ and $\varphi^1 := \neg\varphi$.

Lemma 1. If X is r.e. and monoconsistent with Q, then there is a Π_1 sentence θ such that $\theta^i \notin X$, i = 0, 1.

Proof. The proof is almost the same as the proof of Rosser's theorem. Let R(k,m) be a primitive recursive relation such that $X = \{k: \exists m R(k,m)\}$ and let $\rho(x,y)$ be a PR binumeration of R(k,m). Let θ be such that

(1) $\quad Q\vdash \theta \leftrightarrow \forall z(\rho(\theta,z) \rightarrow \exists u{\le}z\rho(\neg\theta,u))$.

Suppose either $\theta\in X$ or $\neg\theta\in X$. Let m be the smallest number such that either $R(\theta,m)$ or $R(\neg\theta,m)$. Suppose first $R(\neg\theta,m)$. Then $\neg\theta\in X$. Also not $R(\theta,n)$ and so $Q\vdash \neg\rho(\theta,n)$ for n < m. It follows, by Fact 1 (v), that $Q\vdash \rho(\theta,z) \rightarrow m \le z$. Now $Q\vdash \rho(\neg\theta,m)$ and so

$$Q\vdash \forall z(\rho(\theta,z) \rightarrow \exists u{\le}z\rho(\neg\theta,u)).$$

But then, by (1), $Q\vdash \theta$ which is impossible, since $\neg\theta\in X$.

Thus, not $R(\neg\theta,m)$ and so $R(\theta,m)$ whence $\theta\in X$. Also not $R(\neg\theta,n)$ for $n \le m$. It follows that $Q\vdash \rho(\theta,m)$ and, by Fact 1 (iv), $Q\vdash u \le m \rightarrow \neg\rho(\neg\theta,u)$. But then

$$Q\vdash \exists z(\rho(\theta,z) \wedge \forall u{\le}z\neg\rho(\neg\theta,u))$$

and so, by (1), $Q\vdash \neg\theta$, which is impossible, since $\theta\in X$. Thus, we have derived the desired contradiction and the proof is complete. ■

Proof of Theorem 3. The set $\bigcup\{Th(T_k): k\in N\}$ is r.e. and monoconsistent with Q. Now use Lemma 1. ■

§2. Consistency statements. Most arguments carried out in this book can be formalized in PA. In particular this is true of the proof of Theorem 1. This leads to a proof of the following very important result, Gödel's second incompleteness theorem (for theories in L_A). (Recall that a *numeration* of a set X is a formula numerating X in PA.)

Theorem 4. (a) Suppose PA⊣ T. Let φ be as in (G). Then $PA \vdash Con_T \to \varphi$ and consequently $T \nvdash Con_T$.

(b) If $\tau(x)$ is any Σ_1 numeration of T, then $T \nvdash Con_\tau$.

Proof. (a) We follow closely the proof of Theorem 1 (a). By (BLiii),

(1) $\quad PA \vdash Pr_T(\varphi) \to Pr_T(Pr_T(\varphi))$.

By (G) and (BLi), $PA \vdash Pr_T(Pr_T(\varphi) \to \neg\varphi)$ and so, by (BLii),

$$PA \vdash Pr_T(Pr_T(\varphi)) \to Pr_T(\neg\varphi).$$

But then, by (1),

$$PA \vdash Pr_T(\varphi) \to Pr_T(\neg\varphi),$$

whence, by Corollary 1.5 (iii), $PA \vdash Pr_T(\varphi) \to \neg Con_T$ and so, by (G),

$$PA \vdash Con_T \to \varphi.$$

But then, assuming that $T \vdash Con_T$, we get $T \vdash \varphi$, contradicting Theorem 1 (a). It follows that $T \nvdash Con_T$. ♦

The proof of (b) is obtained from the above by replacing $Pr_T(x)$ by $Pr_\tau(x)$. ■

In Theorem 4 (b) it is sufficient to assume that $\tau(x)$ is Σ_1 and numerates T in some theory S such that PA⊣ S⊣ T; but the assumption that $\tau(x)$ is Σ_1 cannot be omitted; see Theorem 7, below.

In applying Theorem 4 to an extension S of PA, we often show that there is a PR binumeration (Σ_1 numeration) $\sigma(x)$ of S such that $S \vdash Con_\sigma$ and conclude that S is inconsistent.

A somewhat shorter proof of Theorem 4 (a) is as follows. By (G),

$$PA \vdash \neg\varphi \to Pr_T(\varphi).$$

By provable Σ_1-completeness (Fact 9 (b)),

$$PA \vdash \neg\varphi \to Pr_T(\neg\varphi).$$

But then, by Corollary 1.5 (iii), $PA \vdash \neg\varphi \to \neg Con_T$ and so

$$PA \vdash Con_T \to \varphi.$$

A similar proof yields Theorem 4 (b).

Combining Theorem 4 and Corollary 1.8, we get.

Corollary 1. If PA⊣ T, then T is not finitely axiomatizable.

Proof. Suppose T is finitely axiomatizable. Then there is a k such that T⊣ T | k. Also, by Corollary 1.8, $T \vdash Con_{T|k}$, whence $T|k \vdash Con_{T|k}$. But, since PA⊣ T | k, this contradicts Theorem 4. ■

Corollary 1 will be strengthened in Chapter 4 (Corollary 4.1) and Chapter 6 (Theorem 6.3).

The proof of Theorem 4 can also be formalized in PA yielding:

Corollary 2. If PA$\dashv$ T, then PA + $Con_T \vdash Con_{T+\neg ConT}$.

Proof. Let φ be as in (G). By Theorem 4 (a),

(1) $PA \vdash Con_T \to \varphi$.

But then, by (BLi) and (BLii), $PA \vdash Pr_T(Con_T) \to Pr_T(\varphi)$ and so, by (G)

(2) $PA \vdash Pr_T(Con_T) \to \neg\varphi$.

From (1) and (2) we get $PA \vdash Pr_T(Con_T) \to \neg Con_T$ which, by Corollary 1.5 (iv), yields the desired conclusion. ■

The proof of our next result is another exercise in formalization, in this case of the proof of Theorem 2.

Theorem 5. Let θ be a Rosser sentence for T. Then

$$PA + Con_T \vdash \neg Pr_T(\theta) \wedge \neg Pr_T(\neg\theta).$$

Proof. We follow closely the above proof of Theorem 2. By Corollary 1.5 (iii),

(1) $PA + Con_T \vdash Pr_T(\theta) \to \neg Pr_T(\neg\theta)$.

It follows that

$$PA + Con_T \vdash Prf_T(\theta,z) \to \neg Prf_T(\neg\theta,u)$$

and so

(2) $PA + Con_T \vdash Prf_T(\theta,z) \to \forall u{\le}z\, \neg Prf_T(\neg\theta,u))$.

Let

$$\gamma(z) := Prf_T(\theta,z) \wedge \forall u{\le}z \neg Prf_T(\neg\theta,u)).$$

Then, by (2),

(3) $PA + Con_T \vdash Prf_T(\theta,z) \to \gamma(z)$.

By Fact 9 (b), we have, $PA \vdash \gamma(z) \to Pr_T(\gamma(\dot{z}))$. Combining this with (3) yields

$$PA + Con_T \vdash Prf_T(\theta,z) \to Pr_T(\gamma(\dot{z})),$$

whence, by Corollary 1.5 (i),

(4) $PA + Con_T \vdash Pr_T(\theta) \to Pr_T(\exists z\gamma(z))$.

By (R), $T \vdash \exists z\gamma(z) \to \neg\theta$. But then, by (BLi) and (BLii),

$$PA \vdash Pr_T(\exists z\gamma(z)) \to Pr_T(\neg\theta).$$

Combining this with (4), we get $PA + Con_T \vdash Pr_T(\theta) \to Pr_T(\neg\theta)$. But then, by (1),

(5) $PA + Con_T \vdash \neg Pr_T(\theta)$,

as desired.

Next we prove that

(6) $PA + Con_T \vdash \neg Pr_T(\neg\theta)$.

From (1), we get

$$PA + Con_T \vdash Prf_T(\neg\theta,u) \to \neg Prf_T(\theta,z)$$

and so

$$PA + Con_T \vdash Prf_T(\neg\theta,u) \to \forall z{<}u \neg Prf_T(\theta,z)).$$

Let

$$\delta(u) := Prf_T(\neg\theta,u) \wedge \forall z{<}u \neg Prf_T(\theta,z)).$$

By an argument similar to the proof of (4), we get

$$PA + Con_T \vdash Pr_T(\neg\theta) \rightarrow Pr_T(\exists u\delta(u)).$$

(R) easily implies that $T \vdash \exists u\delta(u) \rightarrow \theta$. But then (6) follows, by an argument almost the same as the proof of (5). ■

If $PA \dashv T$, this proof of Theorem 5 can be replaced by the formalization of the above short proof of Theorem 2. By (R),

$$PA \vdash Pr_T(\theta) \wedge \neg Pr_T(\neg\theta) \rightarrow \neg\theta.$$

Since $\neg\theta$ is Σ_1, $PA \vdash \neg\theta \rightarrow Pr_T(\neg\theta)$. It follows that $PA \vdash Pr_T(\theta) \rightarrow Pr_T(\neg\theta)$ and so, by Corollary 1.5 (iii),

$$PA \vdash Con_T \rightarrow \neg Pr_T(\theta).$$

Next, by Corollary 1.10 (b), (R), (BLi), and (BLii), $PA \vdash Pr_T(\neg\theta) \wedge \neg Pr_T(\theta) \rightarrow Pr_T(\theta)$, whence $PA \vdash Pr_T(\neg\theta) \rightarrow Pr_T(\theta)$ and so, by Corollary 1.5 (iii),

$$PA \vdash Con_T \rightarrow \neg Pr_T(\neg\theta).$$

Combining Theorem 5 and Corollary 1.5 (iv) we get:

Corollary 3. Let θ be as in (R). Then

$$PA + Con_T \vdash Con_{T+\theta} \wedge Con_{T+\neg\theta}.$$

The sentence φ in (G) above says of itself that it is not provable in T. Let us now consider a sentence χ saying of itself that it *is* provable in T, i.e. such that

$$Q \vdash \chi \leftrightarrow Pr_T(\chi).$$

Is χ provable in T? In this case no simple argument in terms truth will yield an answer, not even if T is true. Nevertheless, it turns out that $T \vdash \chi$ provided that $PA \dashv T$. This follows from our next result, known as Löb's theorem.

Theorem 6. Suppose $PA \dashv T$ and let φ be any sentence such that $T \vdash Pr_T(\varphi) \rightarrow \varphi$. Then $T \vdash \varphi$.

Proof. Let θ be such that

(1) $PA \vdash \theta \leftrightarrow (Pr_T(\theta) \rightarrow \varphi)$.

From this, (BLi), and (BLii), we get

(2) $PA \vdash Pr_T(\theta) \rightarrow (Pr_T(Pr_T(\theta)) \rightarrow Pr_T(\varphi))$.

By (BLiii),

(3) $PA \vdash Pr_T(\theta) \rightarrow Pr_T(Pr_T(\theta))$.

From (2) and (3) it follows that

(4) $PA \vdash Pr_T(\theta) \rightarrow Pr_T(\varphi)$.

Since, by hypothesis, $T \vdash Pr_T(\varphi) \rightarrow \varphi$, this implies that

(5) $T \vdash Pr_T(\theta) \rightarrow \varphi$.

But then, by (1), $T \vdash \theta$, whence, by (BLi), $PA \vdash Pr_T(\theta)$. Finally, this together with (5) yields $T \vdash \varphi$, as desired. ■

There is a semantic paradox related to the above proof in somewhat the same way as the liar paradox is related to the proof of Theorem 1. Let

(**) If (**) is true, the earth is flat.

"Prove", by considering (**), that the earth is flat.

Theorem 6 is a strengthening of Theorem 4: let $\varphi := \bot$. But Theorem 6 can also be derived from Theorem 4 as follows. Suppose $T \vdash Pr_T(\varphi) \to \varphi$. Then $T + \neg\varphi \vdash \neg Pr_T(\varphi)$, whence, by Corollary 1.5 (iv), $T + \neg\varphi \vdash Con_{T+\neg\varphi}$. But then, by Theorem 4, $T + \neg\varphi$ is inconsistent and so $T \vdash \varphi$.

By slightly modifying the proof of Theorem 6 we can derive the stronger result that for every sentence φ,

(L) $\quad PA \vdash Pr_T(Pr_T(\varphi) \to \varphi) \to Pr_T(\varphi)$.

In fact, from (4) we get

$$PA \vdash (Pr_T(\varphi) \to \varphi) \to (Pr_T(\theta) \to \varphi).$$

But then, by (1), $PA \vdash (Pr_T(\varphi) \to \varphi) \to \theta$, whence, by (BLi) and (BLii),

$$PA \vdash Pr_T(Pr_T(\varphi) \to \varphi) \to Pr_T(\theta).$$

Finally, (L) follows from this and (4).

Theorem 4 is sometimes informally expressed by saying that if T is as assumed, then T does not prove that T is consistent. That this must be interpreted with some care is clear from the following result.

Theorem 7. Suppose $PA \dashv T$. Let $\tau(x)$ be any formula binumerating T in T and let

$$\tau^*(x) := \tau(x) \wedge Con_{\tau|x}.$$

Then (i) $\tau^*(x)$ binumerates T in T and (ii) $PA \vdash Con_{\tau^*}$.

The following intuitive proof of Theorem 7 (ii) (formalizable in PA) is probably easier to understand than the formal argument below, but its formalization would be somewhat longer: "Any proof p from the set X defined by $\tau(x) \wedge Con_{\tau|x}$ contains a greatest sentence $\varphi \in X$. Since φ satisfies $Con_{\tau|x}$, it follows that the set of members of X occurring in p is consistent. Thus, p cannot be a proof of $\bot$."

Proof of Theorem 7. Note that x is free in $Con_{\tau|x}$.

(i) If $k \in T$, then $T \vdash \tau(k)$. By Corollary 1.9 (a), $T \vdash Con_{\tau|k}$. Thus, $T \vdash \tau^*(k)$. If, on the other hand, $k \notin T$, then $T \vdash \neg\tau(k)$ and so $T \vdash \neg\tau^*(k)$.

(ii) Trivially $\vdash \tau^*(x) \to \tau(x)$. Hence, by Fact 6,

(1) $\quad \vdash Con_\tau \to Con_{\tau^*}$.

Since PA is reflexive, we have $PA \vdash Con_{\tau|0}$. (We assume that 0 is not a formula.) Also, by Fact 8 (iii), $PA \vdash \forall z Con_{\tau|z} \to Con_\tau$. By the least number principle, it follows that

(2) $\quad PA \vdash \neg Con_\tau \to \exists z(\neg Con_{\tau|z+1} \wedge Con_{\tau|z})$.

By Fact 6,

$$PA \vdash \neg Con_{\tau|z+1} \to (Con_{\tau|x} \to x \leq z).$$

Hence, by the definition of $\tau^*(x)$,

$$PA \vdash \neg Con_{\tau|z+1} \to (\tau^*(x) \to \tau(x) \wedge x \leq z).$$

Hence, again by Fact 6,

$$PA \vdash \neg Con_{\tau|z+1} \wedge Con_{\tau|z} \to Con_{\tau^*}.$$

But then, by (2),

$$\mathrm{PA}\vdash \neg\mathrm{Con}_\tau \to \mathrm{Con}_{\tau*}$$

and so, by (1), PA$\vdash$ $\mathrm{Con}_{\tau*}$, as desired. ■

If $\tau(x)$ is PR, then $\tau^*(x)$ is Π_1. By Theorems 4 and 7, $\tau^*(x)$ is not provably in T equivalent to a Σ_1 formula.

The formula $\tau^*(x)$ may seem like a mere curiosity, but certain closely related formulas are actually of crucial importance in connection with interpretability (see the proof of Lemma 6.2.).

By Theorems 4 and 7, there are formulas $\tau_0(x)$ and $\tau_1(x)$ binumerating T in T such that Con_{τ_0} and Con_{τ_1} are not provably equivalent in T. We now show that this is so even if we restrict ourselves to PR formulas.

Theorem 8. Suppose PA$\dashv$ T. Let $\tau(x)$ be any PR binumeration of T.
(a) There is a PR binumeration $\tau'(x)$ of T such that
(i) T$\vdash$ $\mathrm{Con}_\tau \to \mathrm{Con}_{\tau'}$,
(ii) T$\nvdash$ $\mathrm{Con}_{\tau'} \to \mathrm{Con}_\tau$.
(b) Let π be a true Π_1 sentence such that T$\vdash$ $\pi \to \mathrm{Con}_\tau$. There is then a PR binumeration $\tau'(x)$ of T such that T$\vdash$ $\pi \leftrightarrow \mathrm{Con}_{\tau'}$.

Proof. (a) Let $\tau'(x)$ be such that

$$\mathrm{PA}\vdash \tau'(x) \leftrightarrow \tau(x) \wedge \forall y{\le}x\neg\mathrm{Prf}_T(\mathrm{Con}_{\tau'} \to \mathrm{Con}_\tau, y).$$

By Fact 6, (i) holds. Suppose (ii) is false, i.e.
(1) T$\vdash$ $\mathrm{Con}_{\tau'} \to \mathrm{Con}_\tau$.
Let p be a proof of $\mathrm{Con}_{\tau'} \to \mathrm{Con}_\tau$ in T. Then, by Fact 7 (a) and Fact 1 (v),

$$\mathrm{PA}\vdash \forall y{\le}x\neg\mathrm{Prf}_T(\mathrm{Con}_{\tau'}\to\mathrm{Con}_\tau, y) \to x < p$$

and so PA$\vdash$ $\tau'(x) \to \tau(x) \wedge x < p$. By Fact 6, it follows that,
(2) PA$\vdash$ $\mathrm{Con}_{\tau|p} \to \mathrm{Con}_{\tau'}$.
But T$\vdash$ $\mathrm{Con}_{\tau|p}$, by Corollary 1.9 (a). Hence, by (1) and (2), T$\vdash$ Con_τ, contradicting Theorem 4 (a). This proves (ii). Finally, by (ii), Fact 1 (iv), and Fact 7 (d), $\tau'(x)$ is a PR binumeration of T. ♦

(b) By Fact 5 (b), we may assume that $\pi := \forall x\delta(x)$, where $\delta(x)$ is PR. Let $\tau'(x) := \tau(x) \vee \exists y{\le}x\neg\delta(y)$. Since π is true, $\tau'(x)$ is a PR binumeration of T. Clearly

$$\mathrm{T} + \pi\vdash \tau'(x) \to \tau(x).$$

Thus, by Fact 6, T + $\pi\vdash$ $\mathrm{Con}_\tau \to \mathrm{Con}_{\tau'}$ and so T$\vdash$ $\pi \to \mathrm{Con}_{\tau'}$.

To show that the converse implication is provable in T we use the fact that evidently

$$\mathrm{PA}\vdash \exists y\neg\delta(y) \to \neg\mathrm{Con}_{\tau'}.$$

But then, by Fact 6, T + $\neg\pi\vdash$ $\neg\mathrm{Con}_{\tau'}$ and so T$\vdash$ $\mathrm{Con}_{\tau'} \to \pi$. ■

Suppose $\tau(x)$ is a PR binumeration of T. Then, by Theorem 4, it may be true that T$\vdash$ $\neg\mathrm{Con}_\tau$. However, from Theorem 8 (a) it follows that we can always choose $\tau(x)$ so that this does not hold:

Corollary 4. If PA$\dashv$ T, there is a PR binumeration $\tau(x)$ of T such that T$\nvdash$ $\neg\mathrm{Con}_\tau$.

§3. Independent formulas. A formula $\xi(x)$ is *independent over* T if the only propositional combinations of sentences of the form $\xi(k)$ provable in T are the tautologies. This, of course, is the same as saying that $T + \{\xi(k)^{f(k)}: k \in N\}$ is consistent for any $f \in 2^N$.

The following result is a strengthening of Theorem 2.

Theorem 9. There is a Π_1 formula which is independent over T.

Proof. Let $R(k,i,\gamma,p)$ be the primitive recursive relation:

there is a binary sequence s such that $s_k = i$ (so $i = 0$ or $i = 1$) and p is a proof in T of $\neg(\gamma(0)^{s_0} \wedge ... \wedge \gamma(k)^{s_k})$.

Let $\rho(x,y,z,u)$ be a PR binumeration of $R(k,i,\gamma,p)$. Let $\mu(x)$ be such that

$$Q \vdash \mu(x) \leftrightarrow \forall z(\rho(x,1,\mu,z) \rightarrow \exists u \leq z \rho(x,0,\mu,u)).$$

Suppose, for *reductio ad absurdum*, that $\mu(x)$ is not independent over T. There is then a smallest n for which there is a sequence s such that

(1) $\neg(\mu(0)^{s_0} \wedge ... \wedge \mu(n)^{s_n})$

is provable in T. Let s be the sequence for which the shortest proof p of (1) in T is minimal. There are then two cases. (We assume that $n > 0$ and leave the case $n = 0$ to the reader.)

Case 1. $s_n = 0$. Then

(2) $T \vdash \mu(0)^{s_0} \wedge ... \wedge \mu(n\text{-}1)^{s_{n-1}} \rightarrow \neg\mu(n)$,

(3) $T \vdash \rho(n,0,\mu,p)$,

(4) $T \vdash \neg\rho(n,1,\mu,q)$ for $q \leq p$.

From (3) and (4) we get $T \vdash \mu(n)$ as in the proof of Rosser's theorem. But then, by (2),

(5) $T \vdash \neg(\mu(0)^{s_0} \wedge ... \wedge \mu(n\text{-}1)^{s_{n-1}})$,

contrary to the fact that n is minimal.

Case 2. $s_n = 1$. Then

(6) $T \vdash \mu(0)^{s_0} \wedge ... \wedge \mu(n\text{-}1)^{s_{n-1}} \rightarrow \mu(n)$,

(7) $T \vdash \rho(n,1,\mu,p)$,

(8) $T \vdash \neg\rho(n,0,\mu,q)$ for $q < p$.

From (7) and (8) we get $T \vdash \neg\mu(n)$ and so, by (6), we again get (5), again contrary to the minimality of n. ■

Theorem 9 can be improved as follows; Theorem 10 will be used in Chapter 6 (proof of Lemma 6.8).

Theorem 10. For any Σ_n formula $\delta(x)$, there is a Σ_{n+1} formula $\eta(x)$ such that for any $f, g \in 2^N$, if $T_f = T + \{\delta(k)^{f(k)}: k \in N\}$ is consistent, so is $T_f + \{\eta(k)^{g(k)}: k \in N\}$.

Proof. For every $f \in 2^N$, let $R^f(k,i,\gamma,p)$ be the relation:

there is a binary sequence s such that $(s)_k = i$ and p is a proof in T_f of $\neg(\gamma(0)^{(s)_0} \wedge ... \wedge \gamma(k)^{(s)_k})$

(compare the relation $R(k,i,\gamma,p)$ defined in the proof of Theorem 9). Using the formula $\delta(x)$, we are going to define a formula $\rho^*(x,y,z,w)$ such that for every f,

(1) $\rho^*(x,y,z,w)$ binumerates $R^f(k,i,\gamma,p)$ in T_f.

(Thus, $\rho^*(x,y,z,w)$ behaves in relation to T_f, in the same way as the formula $\rho(x,y,z,u)$ in the proof of Theorem 9 behaves in relation to T.) Let $R^+(k,i,\gamma,t,n,p)$ be the following primitive recursive relation, where t is a binary sequence:

there is a binary sequence s such that $(s)_k = i$ and p is a proof of $\neg(\gamma(0)^{(s)_0} \wedge \ldots \wedge \gamma(k)^{(s)_k})$ in $T + \delta(0)^{(t)_0} + \ldots + \delta(n)^{(t)_n}$.

Then

(2) $R^f(k,i,\gamma,p)$ iff $\exists nt \leq p\big(\forall m \leq n((t)_m = f(m))\ \&\ R^+(k,i,\gamma,t,n,p)\big)$.

This is trivial except that it isn't clear that assuming that $R^f(k,i,\gamma,p)$, we can choose $t \leq p$. But this holds if we assume, as we may, that if $\delta(n)^{f(n)}$ occurs in p, then $p \geq 2\times3\times\ldots\times p_n$, where p_n is the n^{th} prime mumber. Let $\rho^+(x,y,z,u,v,w)$ be a PR binumeration of $R^+(k,i,\gamma,t,n,p)$.

By Fact 2, there is a PR formula $\sigma(x,z,u)$ such that

$$Q \vdash \sigma(k,m,u) \leftrightarrow u = (k)_m.$$

Let

$$\beta(x,y) := \forall z \leq y((\delta(z) \rightarrow \sigma(x,z,0)) \wedge (\neg\delta(z) \rightarrow \sigma(x,z,1))).$$

Then for every n and every t,

(3) $T_f \vdash \beta(t,n) \leftrightarrow \bigwedge\{(t)_m = f(m) : m \leq n\}$.

In view of (2) and (3), the obvious definition of $\rho^*(x,y,z,w)$ is now:

$$\rho^*(x,y,z,w) := \exists uv \leq w(\beta(u,v) \wedge \rho^+(x,y,z,u,v,w)).$$

To prove (1), suppose first $R^f(k,i,\gamma,p)$. By (2), there are then n, $t \leq p$ such that $(t)_m = f(m)$ for all $m \leq n$ and $R^+(k,i,\gamma,t,n,p)$. But then $T \vdash \rho^+(k,i,\gamma,t,n,p)$. By (3), it follows that $T_f \vdash \beta(t,n)$ and so that $T_f \vdash \rho^*(k,i,\gamma,p)$.

Next suppose $\neg R^f(k,i,\gamma,p)$. Then, by (2), $\neg R^+(k,i,\gamma,t,n,p)$ for every $n \leq p$ and every $t \leq p$ such that $(t)_m = f(m)$ for all $m \leq n$. It follows that $T \vdash \neg\rho^+(k,i,\gamma,t,n,p)$ for all such n and t. Also, by (3), $T_f \vdash \neg\beta(t,n)$ for all t such that $(t)_m \neq f(m)$ for some $m \leq n$. It follows that $T_f \vdash \neg\rho^*(k,i,\gamma,p)$. This proves (1).

Let $\eta(x)$ be such that

$$Q \vdash \eta(x) \leftrightarrow \exists z(\rho^*(x,0,\eta,z) \wedge \forall u \leq z \neg\rho^*(x,1,\eta,u)).$$

The proof that $\eta(x)$ is as desired is now the same as the proof of Theorem 9 except that T is replaced by any consistent theory T_f, and the fact that $\rho(x,y,z,u)$ is decidable in T is replaced by (1). We leave this part of the proof to the reader.

Finally, if $\delta(x)$ is Σ_n, then $\beta(x,y)$ is Δ_{n+1}, whence the same is true of $\rho^*(x,y,z,w)$ and so $\eta(x)$ is Σ_{n+1}, as desired. ■

The proof of the final theorem of this § is quite different from the proofs of Theorems 9 and 10; instead of a Rosser type construction it uses the formulas $Sat_\Phi(x,y)$ and so does not apply to Q (and its extensions).

In the proof of Theorem 11 we assume, as we may, that

(+) $PA \vdash \langle x,y \rangle = z \leftrightarrow (x = (z)_0 \wedge y = (z)_1)$.

Theorem 11. Suppose PA⊣ T. Then there is a Γ (Δ_{n+1}) formula $\gamma(x)$ such that T +

$\forall x(\gamma(x) \leftrightarrow \delta(x))$ is consistent for every Γ (B_n) formula $\delta(x)$.

Proof. Suppose first $\Gamma = \Sigma_n$; the case $\Gamma = \Pi_n$ follows by taking negations. Let $S(k,m,n)$ be a primitive recursive relation such that

$$\delta(x) \in \Sigma_n \ \&\ T \vdash \neg\forall x(\eta(x) \leftrightarrow \delta(x)) \text{ iff } \exists n S(\eta,\delta,n).$$

Let $\sigma(x,y,z)$ be a PR binumeration of $S(k,m,n)$ and let $\sigma^*(x,y,z) :=$

$$\sigma(x,y,z) \wedge \forall y'z'(<y',z'> < <y,z> \rightarrow \neg\sigma(x,y',z')).$$

Finally, let $\gamma(x)$ be such that

(1) $\quad PA \vdash \gamma(x) \leftrightarrow \exists yz(\sigma^*(\gamma,y,z) \wedge Sat_{\Sigma_n}(x,y))$.

Suppose there is a Σ_n formula $\delta(x)$ such that

(2) $\quad T \vdash \neg\forall x(\gamma(x) \leftrightarrow \delta(x))$.

For each such formula, there is an n such that $S(\gamma,\delta,n)$. Now pick $\delta(x)$ and n so that $<\delta,n>$ is minimal. Then, by (+),

$$PA \vdash \sigma^*(\gamma,y,z) \leftrightarrow y = \delta \wedge z = n.$$

Hence, by (1) and Fact 10 (a) (i), $PA \vdash \forall x(\gamma(x) \leftrightarrow \delta(x))$, contradicting (2). Thus, (2) is false for all Σ_n formulas $\delta(x)$, as desired.

To obtain a Δ_{n+1} formula as desired, replace $Sat_{\Sigma_n}(x,y)$ by $Sat_{B_n}(x,y)$ in (1). ■

For extensions T of PA, Theorem 9 follows at once from Theorem 11.

Theorem 11 has the following:

Corollary 5. Suppose $PA \dashv T$. There is a Γ (Δ_{n+1}) sentence not in $\Gamma^{d,T}$ (B_n^T).

Proof. Let $\gamma(x)$ be as in Theorem 11 and let $\varphi := \gamma(0)$. ■

§4. The length of proofs. We begin by showing that the length of proofs of (Π_1) sentences φ is not bounded by any recursive function of φ.

Theorem 12. Let f(k) be any recursive function. There is then a Π_1 sentence φ such that $T \vdash \varphi$ and the least proof of φ in T is $> f(\varphi)$.

Proof. Let $\delta_f(x,y)$ be a Σ_1 formula defining f in Q (cf. Fact 3 (b)). Let φ be such that

$$Q \vdash \varphi \leftrightarrow \forall y(\delta_f(\varphi,y) \rightarrow \forall z \le y \neg Prf_T(\varphi,z)).$$

Suppose φ has a proof $p \le f(\varphi)$ in T. Since

$$Q \vdash \delta_f(\varphi,y) \leftrightarrow y = f(\varphi)$$

and, by Fact 7 (a), $Q \vdash Prf_T(\varphi,p)$, it follows that $Q \vdash \neg\varphi$ and so $T \vdash \neg\varphi$, a contradiction. Thus, φ has no proof $p \le f(\varphi)$ in T. But then, by Fact 1 (iv) and Fact 7 (d), $Q \vdash \forall z \le f(\varphi) \neg Prf_T(\varphi,z))$, whence $Q \vdash \varphi$ and so $T \vdash \varphi$. ■

In Theorem 12 and in Theorems 13 and 14, below, we use (the Gödel number of) the proof as a measure of its "length". We could also have used the number of (occurrences of) symbols as a (more natural) measure of "length" and proved the same results.

Suppose $T \nvdash \varphi$. Then $T + \varphi$ is stronger than T not only in the sense that it proves more theorems but also in the sense that there are infinitely many theorems of T which have "much shorter" proofs in $T + \varphi$; more exactly:

Theorem 13. Suppose $T \nvdash \varphi$. Let f be any recursive function. There is then a sentence θ such that $T \vdash \theta$ and there is a proof q of θ in $T + \varphi$ such that θ has no proof $\leq f(q)$ in T.

Proof. We may assume that f is increasing. Let $\delta_f(x,y)$ be a formula defining f in Q (cf. Fact 3 (a)). Let ψ be such that

$$Q \vdash \psi \leftrightarrow \exists yz(Prf_{T+\varphi}(\varphi \vee \psi, y) \wedge \delta_f(y,z) \wedge \forall u \leq y+z \neg Prf_T(\varphi \vee \psi, u)).$$

Let $\theta := \varphi \vee \psi$. Suppose $T \nvdash \theta$. Since, trivially, $T + \varphi \vdash \theta$, it follows, by Fact 1 (iv) and Fact 7 (a) and (d), that $T \vdash \psi$ and so $T \vdash \theta$, a contradiction. Thus, $T \vdash \theta$.

Let q be the least proof of θ in $T + \varphi$. Suppose there is a proof $\leq f(q)$ of θ in T. Then, again by Fact 1 (iv) and Fact 7 (a) and (d), $T \vdash \neg\psi$ and so $T \vdash \varphi$, contrary to hypothesis. It follows that θ has no proof $\leq f(q)$ in T. ■

Another way of obtaining "much shorter proofs", in this case without getting any new theorems, is to add new (nonlogical but correct) rules of inference: for example, if T is Σ_1–sound, the rule

R: from $Pr_T(\varphi)$ derive φ,

is *correct for* T in the sense that every sentence which can be derived (from the axioms of T) using this rule can be proved without it, i.e. is a theorem of T. That R occasionally leads to "much shorter proofs" follows from our next:

Theorem 14. Suppose $PA \dashv T$ and T is Σ_1–sound. Let g(k,m) be any primitive recursive function. There are then a (Σ_1, Π_1) sentence φ such that $T \vdash \varphi$ and a proof q of $Pr_T(\varphi)$ in T such that φ has no proof $\leq g(\varphi,q)$ in T.

Proof. We may assume that g(k,m) is increasing in m. Let φ be such that

$$T \vdash \varphi \leftrightarrow \exists y(Prf_T(Pr_T(\varphi),y) \wedge \forall z \leq g(\varphi,y) \neg Prf_T(\varphi,z)).$$

Clearly

$$T + Pr_T(Pr_T(\varphi)) + \neg Pr_T(\varphi) \vdash \varphi.$$

Since φ is Σ_1, we have, by provable Σ_1–completeness, $T + \varphi \vdash Pr_T(\varphi)$. It follows that

$$T + Pr_T(Pr_T(\varphi)) \vdash Pr_T(\varphi),$$

and so, by Theorem 6, $T \vdash Pr_T(\varphi)$. Since T is Σ_1–sound, this implies that $T \vdash \varphi$ and that φ is true.

Let q be the least proof of $Pr_T(\varphi)$ in T. Since φ is true and g(k,m) is increasing in m, it follows that φ has no proof $\leq g(\varphi,q)$.

To obtain a Π_1 sentence as desired, let φ be such that

$$T \vdash \varphi \leftrightarrow \forall z(Prf_T(\varphi,z) \rightarrow \exists y \leq z(g(\varphi,y) < z \wedge Prf_T(Pr_T(\varphi),y)))$$

and set

$$\varphi^* := \exists y(Prf_T(Pr_T(\varphi),y) \wedge \forall z \leq g(\varphi,y) \neg Prf_T(\varphi,z)).$$

Then

$$T + Pr_T(Pr_T(\varphi)) + \neg Pr_T(\varphi) \vdash \varphi^*.$$

Clearly, $T \vdash \varphi^* \to \varphi$ and so, by (BLi) and (BLii), $T \vdash Pr_T(\varphi^*) \to Pr_T(\varphi)$. Since φ^* is Σ_1, we have $T + \varphi^* \vdash Pr_T(\varphi^*)$. It follows that

$$T + Pr_T(Pr_T(\varphi)) \vdash Pr_T(\varphi).$$

The rest of the proof is now the same as above, except that we observe that, since φ is Π_1 and $T \vdash \varphi$, φ must be true (Fact 9 (a)). ■

For any sequence p of formulas and any formula θ, let $p^\wedge\theta$ be p followed by θ. If p is a proof of $Pr_T(\theta)$ in T we may think of $p^\wedge\theta$ as an R–proof of θ in T, i.e. a proof in T in which we are allowed to use the rule R. Now, let h be any primitive recursive function and let $g(\theta,p) = h(p^\wedge\theta)$. Then g is primitive recursive. Let φ and q be as in Theorem 14 and let $r = q^\wedge\varphi$. Then r is an R–proof of φ in T and φ has no proof $\leq h(r)$ in T.

Exercises for Chapter 2.

In the following Exercises we write Prf(x,y), Pr(x), Con for $Prf_T(x,y)$, $Pr_T(x)$, Con_T, respectively.

1. Suppose T is true. Show that T is not complete by using the fact that Th(T), being r.e., is definable in **N** together with Corollary 1.7.

2. Let U be a (not necessarily r.e. or true) consistent extension of Q. Suppose there is a formula $\upsilon(x)$ binumerating U in U. Show that U is not complete.

3. Let $Ref(T) = \{\varphi: T \vdash \neg\varphi\}$. Let X be any set such that $Th(T) \subseteq X$ and $Ref(T) \cap X = \emptyset$. Show that there is no formula binumerating X in T. (This improves Lemma 1.2.) Conclude that Th(T) and Ref(T) are *recursively inseparable*, i.e. there is no recursive set Y such that $Th(T) \subseteq Y$ and $Ref(T) \cap Y = \emptyset$. (This implies Theorem 1.2.)

4. (a) Suppose T is Σ_1–sound. Use the fact that there is an r.e. nonrecursive set to show that there is a (true) Π_1 sentence not provable in T.

(b) Let X_0 and X_1 be disjoint r.e. sets. Let $\rho_i(x,y)$ be a PR formula such that $X_i = \{k: \exists m Q \vdash \rho_i(k,m)\}$, $i = 0, 1$. Let

$$\xi(x) := \exists y(\rho_0(x,y) \wedge \forall z{\leq}y \neg\rho_1(x,z)).$$

Show that if $k \in X_0$, then $Q \vdash \xi(k)$, and if $k \in X_1$, then $Q \vdash \neg\xi(k)$ (compare Theorem 3.2).

(c) Show that the sets of Π_1 and Σ_1 sentences provable in T are not recursive and, therefore, there is a true Π_1 sentence which is unprovable in T (compare Theorem 2). [Hint: There are disjoint r.e. recursively inseparable sets (see Exercise 3).]

(d) Suppose $PA \dashv T$. Show that the set of Δ_n^T sentences is not recursive. [Hint: Let

σ be a Σ_n formula which is not Δ^T_n. Let X_i and $\rho_i(x,y)$ be as in (b). Suppose X_0 and X_1 are recursively inseparable. Let $\eta(x) :=$

$$\exists y(\rho_0(x,y) \wedge \forall z{\le}y \neg\rho_1(x,z)) \vee (\exists z(\rho_1(x,z) \wedge \forall y{\le}z \neg\rho_0(x,y)) \wedge \sigma).$$

Let $Y = \{k: \eta(k) \text{ is } \Delta^T_n\}$. Then $X_0 \subseteq Y$ and $X_1 \cap Y = \emptyset$.]

5. Suppose $Q \dashv S$. Show that there is a Π_1 sentence θ such that $S \nvdash \theta$, $S \nvdash \neg\theta$, $T \nvdash \neg Pr_S(\theta)$, $T \nvdash \neg Pr_S(\neg\theta)$.

6. Let φ be as in Theorem 1.

(a) Show that $PA \vdash \varphi \rightarrow Con$. Conclude that $PA \vdash \varphi \leftrightarrow Con$ and so $PA \vdash Con \leftrightarrow \neg Pr(Con)$. (Thus, there is a sentence, Con, satisfying (G) not constructed using self-reference.) We also have $PA \vdash Pr(\neg\varphi) \rightarrow Pr(\neg Con)$ (compare the last part of Theorem 1).

(b) Suppose $PA \dashv T$ and T is Σ_1-sound. Show that $T \nvdash Con \rightarrow \neg Pr(\neg\varphi)$ (compare Theorem 5).

7. T is *ω-consistent* iff for every formula $\alpha(x)$, if $T \vdash \neg\alpha(k)$ for every k, then $T \nvdash \exists x\alpha(x)$.

(c) Show that if T is ω-consistent, then T is Π_3-sound.

(d) Suppose T is true. Show that there is a false Σ_3 sentence φ such that $T + \varphi$ is ω-consistent. Conclude that ω-consistency does not imply Σ_3-soundness. [Hint: Let φ be a sentence "saying" that $T + \varphi$ is not ω-consistent.]

(e) Suppose $PA \dashv T$ and T is true. Show that for every n, there is an extension S of T which is Σ_n-sound but not ω-consistent. [Hint: Let $\delta(x)$ be a Π_n formula such that $PA \vdash \varphi \leftrightarrow \exists x\delta(x)$, where φ as in Exercise 1.6 (b). Let $S = T + \exists x\delta(x) + \{\neg\delta(k): k \in N\}$.]

8. Suppose $PA \dashv T$. Let θ be a Π_1 Rosser sentence for T and let $\psi :=$

$$\forall u(Prf(\neg\theta,u) \rightarrow \exists z{<}u Prf(\theta,z)).$$

Show that $T \nvdash \theta \rightarrow Con$, $T \nvdash \psi \rightarrow Con$, and $PA \vdash \theta \wedge \psi \rightarrow Con$. Conclude that $T \nvdash \psi$. [Hint: Use Theorem 4 and Corollary 3.]

9. Suppose $PA \dashv T$. Suppose φ is undecidable in T. Show that there is a (Σ_1, Π_1) sentence ψ such that

$$T \nvdash \varphi \rightarrow \psi,$$
$$PA \vdash Pr(\varphi) \rightarrow Pr(\psi).$$

[Hint: Construct ψ is such a way that $PA \vdash Pr(\varphi) \wedge \neg Pr(\varphi \rightarrow \psi) \rightarrow Pr(\psi)$.]

10. Strengthen Lemma 1 in the following way. Suppose X is r.e. and monoconsistent with Q. Show that there is a Π_1 formula $\eta(x)$ such that the only propositional combinations of sentences of the form $\eta(k)$ which are members of X are the tautologies.

11. Suppose PA$\dashv$ T. Show that there is a Π_1 formula $\kappa(x)$ such that T$\nvdash$ $\kappa(k)$ for every k, but T$\vdash$ $\kappa(k) \vee \kappa(m)$ whenever $k \neq m$. (This can also be obtained as a special case of Theorem 3.5.). [Hint: Let $\kappa(x)$ be such that

$$PA\vdash \kappa(x) \leftrightarrow \forall y(Prf(\kappa(\dot{x}),y) \rightarrow \exists zu(<z,u><<x,y> \wedge Prf(\kappa(\dot{z}),u))).]$$

12. Suppose PA$\dashv$ T. Let f(k) be any recursive function. Show that there is a Π_1 sentence θ such that T$\vdash$ θ and T$\restriction$f(θ)$\nvdash$ θ. (This improves Corollary 1; also compare Exercise 4.5.) [Hint: Let $\delta_f(x,y)$ be a Σ_1 formula defining f in Q (cf. Fact 3 (b)) and let θ be such that

$$Q\vdash \theta \leftrightarrow \forall y(\delta_f(\theta,y) \rightarrow \neg Pr_{T\restriction y}(\theta).]$$

13. Prove Löb's theorem by considering a sentence θ such that

$$PA\vdash \theta \leftrightarrow Pr(\theta\rightarrow\varphi).$$

(This is essentially the proof of Theorem 6 using Theorem 4 mentioned in the text.) Show that this proof can be formalized in PA.

14. Suppose PA$\dashv$ T. Show that there is a PR formula $\delta(x)$ such that T$\vdash$ $\forall x Pr(\delta(\dot{x}))$ and T$\nvdash$ $\forall x\delta(x)$. [Hint: Let φ be as in the proof of Theorem 1 and let $\delta(x) :=$ $\neg Prf(\varphi,x)$.]

15. Suppose PA$\dashv$ T. Let $\tau(x)$ be a PR binumeration of T and let $\tau^*(x)$ be as in Theorem 7.

(a) Let ψ be such that PA$\vdash$ $\psi \leftrightarrow \neg Pr_{\tau^*}(\psi)$. Show that ψ is undecidable in T.

(b) Show that T$\vdash$ $Con_{\tau^*+\neg Con\tau}$. [Hint: Let φ be as in Theorem 1. Then T$\vdash$ $\neg\varphi \rightarrow Pr_{\tau^*}(\neg\varphi)$ and so T$\vdash$ $\neg\varphi \rightarrow \neg Pr_{\tau^*}(\varphi)$. It follows that T$\vdash$ $Pr_\tau(\varphi) \rightarrow \neg Pr_{\tau^*}(\varphi)$. Also T$\vdash$ $\neg Pr_\tau(\varphi) \rightarrow \neg Pr_{\tau^*}(\varphi)$ and so T$\vdash$ $\neg Pr_{\tau^*}(\varphi)$. Now use the fact that T$\vdash$ $Con_\tau \rightarrow \varphi$.]

16. Suppose PA$\dashv$ T. Prove the following strengthening of Corollary 4. Suppose X is r.e. and monoconsistent with T. There is then a PR binumeration $\tau(x)$ of T such that $\neg Con_\tau \notin X$ (see Exercise 6.6 (b)). [Hint: Let $\tau'(x)$ be a PR binumeration of T, let $\rho(x,y)$ be a PR binumeration of a relation R(k,m) such that X = {k: $\exists$mR(k,m)}, let φ be such that

$$PA\vdash \varphi \leftrightarrow Con_{\tau'(x)\wedge\forall y\leq x\neg\rho(\neg\varphi,y)},$$

and set $\tau(x) := \tau'(x)\wedge\forall y\leq x\neg\rho(\neg\varphi,y)$.]

17. Suppose PA$\dashv$ T. Let $\tau_0(x)$ and $\tau_1(x)$ be PR binumerations of T.

(a) Show that there is a PR binumeration $\tau(x)$ of T such that

$$T\vdash Con_\tau \leftrightarrow Con_{\tau_0} \wedge Con_{\tau_1}.$$

[Hint: Let $\tau(x) := \tau_0(x) \vee \exists y\leq x Prf_{\tau_1}(\bot,y)$. See also Theorem 8 (b).]

(b) Show that there is a PR binumeration $\tau(x)$ of T such that

$$T\vdash Con_\tau \leftrightarrow Con_{\tau_0} \vee Con_{\tau_1}.$$

[Hint: Let $\tau(x) := (\tau_0(x) \wedge \tau_1(x)) \vee (\exists y\leq x Prf_{\tau_0}(\bot,y) \wedge \exists y\leq x Prf_{\tau_1}(\bot,y))$.]

(c) Suppose $T \nvdash \varphi$. Show that there is a PR binumeration $\tau(x)$ of T such that $T \nvdash \mathrm{Con}_\tau \to \varphi$ (compare Theorem 8 (a)).

(d) Suppose $T \nvdash \varphi$ and $T \nvdash \neg\psi$. Show that there is a PR binumeration $\tau(x)$ of T such that $T \nvdash \mathrm{Con}_\tau \to \varphi$ and $T \nvdash \psi \to \mathrm{Con}_\tau$. [Hint: Use Lemma 1 and Theorem 8 (b).]

18. Suppose $PA \dashv T$. Let $\alpha(x)$, $\beta(x)$ be PR formulas and let $\alpha \leq \beta$ mean that there is a primitive recursive function g such that

$$PA \vdash \forall x(\mathrm{Prf}_\alpha(\bot,x) \to \mathrm{Prf}_\beta(\bot,g(x))).$$

($\alpha \leq \beta$ implies $PA \vdash \mathrm{Con}_\beta \to \mathrm{Con}_\alpha$, bur not conversely.) Let $\alpha \equiv \beta$ mean that $\alpha \leq \beta \leq \alpha$. Let $\tau(x)$ be a PR binumeration of T and let $\alpha(x)$ be such that $PA \vdash \tau(x) \to \alpha(x)$. Show that there is a Π_1 (Σ_1) sentence φ such that $\alpha \equiv \tau + \varphi$. [Hint: In the Π_1 case let φ be such that

$$PA \vdash \varphi \leftrightarrow \forall x(\mathrm{Prf}_\alpha(\bot,x) \to \exists y{\leq}x\mathrm{Prf}_{\tau+\varphi}(\bot,y)).$$

Use the fact that to every PR formula $\delta(x)$, there is a primitive recursive function h such that

$$PA \vdash \delta(x) \to \mathrm{Prf}_\tau(\delta(\dot{x}),h(x)).]$$

19. Prove the following strengthening of Theorems 3 and 9. If $\{T_k: k \in N\}$ is an r.e. family of theories, there is a Π_1 formula which is simultaneously independent over all the theories T_k. Strengthen Theorems 10 and 11 in the same way.

20. (a) Derive Theorem 9 for extensions of PA from Theorem 11.

(b) Formulate and prove a generalization of Theorem 11 which implies Theorem 10 for extensions of PA.

21. Suppose $PA \dashv T$. Let σ be any Σ_1 sentence. Show that there is a Σ_1 sentence χ such that

$$PA \vdash (\sigma \vee \Pr(\bot)) \leftrightarrow \Pr(\chi).$$

Conclude that (i) for every Σ_1 sentence σ such that $T \vdash \Pr(\bot) \to \sigma$, there is a Σ_1 sentence χ such that $T \vdash \sigma \leftrightarrow \Pr(\chi)$ and so for any sentences φ, ψ, there is a Σ_1 sentence χ such that $T \vdash \Pr(\chi) \leftrightarrow \Pr(\varphi) \vee \Pr(\psi)$, and (ii) for every Π_1 sentence π such that $T \vdash \pi \to \mathrm{Con}$, there is a Π_1 sentence θ such that $T \vdash \pi \leftrightarrow \mathrm{Con}_{T+\theta}$ (compare Theorem 8 (b)). [Hint: Let $\delta(y)$ be a PR formula such that $\sigma := \exists y\delta(y)$. Let χ be such that

$$PA \vdash \chi \leftrightarrow \exists y(\delta(y) \wedge \forall z{\leq}y\neg\mathrm{Prf}(\chi,z)).$$

Then $PA \vdash \Pr(\chi) \wedge \neg\sigma \to \Pr(\neg\chi)$.]

22. Suppose $PA \dashv T$. Show that the following conditions are equivalent:

(i) T is Σ_1–sound.

(ii) For any two Σ_1 sentences σ_0, σ_1, if $T \vdash \sigma_0 \vee \sigma_1$, then either $T \vdash \sigma_0$ or $T \vdash \sigma_1$.

(iii) If σ is Δ_1^T, then either $T \vdash \sigma$ or $T \vdash \neg\sigma$ (compare Exercise 3.6 (a)).

(iv) $\Pr(x)$ numerates Th(T) in T (compare Exercise 6.18).

[Hint: (iii) implies (i). Let $\delta(z)$ be a PR formula such that $\exists z\delta(z)$ is false and provable in T. Let σ be such that

$$Q\vdash \sigma \leftrightarrow \exists z((Prf(\neg\sigma,z) \vee \delta(z)) \wedge \forall u{\le}z\neg Prf(\sigma,u)).$$

(iv) implies (i). Let $\delta(z)$ be as above. Let φ be such that

$$Q\vdash \varphi \leftrightarrow \exists z(\delta(z) \wedge \forall u{\le}z\neg Prf(\varphi,u)).$$

Then $T\vdash Pr(\varphi)$ and $T\nvdash \varphi$.]

23. Suppose PA⊣ T.

(a) Let φ be any Γ sentence. Show that there is a formula $\xi(x)$ such that if ψ is Γ, then ψ is a fixed point of $\xi(x)$ in T iff $\psi := \varphi$.

(b) Suppose $\gamma(x)$ is Γ. Show that $\gamma(x)$ has infinitely many Γ fixed points in T. Conclude that the formula $\xi(x)$ mentioned in (a) cannot be Γ.

(c) Let X be any r.e. set of sentences. Show that there is a formula $\xi(x)$ such that if $\varphi\in X \cap \Gamma$, then φ is a fixed point of $\xi(x)$ in T and if $\varphi\in \Gamma - X$, then φ is not a fixed point of $\xi(x)$ in T. [Hint: Let $\rho(x,y)$ be a PR formula such that $X = \{k: \exists m\ PA\vdash \rho(k,m)\}$. Let $\xi(x)$ be such that

$$PA\vdash \xi(\varphi) \leftrightarrow \big(Tr_\Gamma(\varphi) \wedge \exists y(\rho(\varphi,y) \wedge \forall z{\le}y\neg Prf(\varphi{\leftrightarrow}\xi(\varphi),z))\big) \vee$$
$$\big(\neg Tr_\Gamma(\varphi) \wedge \exists z(Prf(\varphi{\leftrightarrow}\xi(\varphi),z) \wedge \forall y{\le}z\neg\rho(\varphi,y))\big).]$$

In Exercises 24 – 28 "proof" means "proof in T".

24. Let f(k) be any recursive function.

(a) Show that there is a Σ_1 sentence φ such that $T\vdash \varphi$ and the least proof of φ is $> f(\varphi)$ (compare Theorem 12).

(b) Show that there is a Π_1 formula $\xi(x)$ such that for every n, $T\vdash \xi(n)$ and the least proof of $\xi(n)$ is $> f(n)$.

25. Suppose PA⊣ T and let g(k) be any recursive function. Show that there are Π_1 sentences ψ_0, ψ_1 provable and a proof p of $\psi_0 \vee \psi_1$ such that neither ψ_0 nor ψ_1 has a proof $\le g(p)$. [Hint: Let $\delta_g(x,y)$ be a Σ_1 formula defining g(k) in T. Let $Prf'(x,y) := Prf(x,y) \wedge \forall z{<}y\neg Prf(x,z)$ and let ψ_i be such that

$$T\vdash \psi_i \leftrightarrow \forall yz\big(Prf'(\psi_0{\vee}\psi_1,y) \wedge \exists v(\delta_g(y,v) \wedge z \le v) \wedge Prf(\psi_i,z) \rightarrow$$
$$\exists u{<}z{+}i\,Prf(\psi_{1-i},u)\big).]$$

26. Suppose PA⊣ T and T is Σ_1–sound. There is then a recursive function g(k) which given a proof p of a sentence $Pr(\varphi_0) \vee Pr(\varphi_1)$ picks out a φ_i such that $Pr(\varphi_i)$ is true; in other words, g(p) = 0 or g(p) = 1, if g(p) = 0, then $Pr(\varphi_0)$ is true, and if g(p) = 1, then $Pr(\varphi_1)$ is true. Show that g(k) is not provably recursive in T even if we restrict ourselves to Σ_1 sentences φ_0, φ_1. [Hint: Suppose not. Assume that $T\vdash g(y) = 0 \vee g(y) = 1$. Let ψ_i be such that

$$T\vdash \psi_i \leftrightarrow \exists y(Prf'(Pr(\psi_0){\vee}Pr(\psi_1),y) \wedge g(y) = 1{-}i),$$

where $Prf'(x,y) := Prf(x,y) \wedge \forall z{<}y\neg Prf(x,z)$.]

27. Suppose PA$\dashv$ T, T is Σ_1–sound, and g is primitive recursive.

(a) Show that there are true Σ_1 sentences σ_0, σ_1 and a proof p of $Pr(\sigma_0) \vee Pr(\sigma_1)$ such that neither $Pr(\sigma_0)$ nor $Pr(\sigma_1)$ has a proof $\leq g(p)$. [Hint: Show that there is a primitive recursive function h such that h is provably increasing in T and if

$$T \vdash \sigma \leftrightarrow \exists z(Prf(Pr(\chi),z) \wedge \forall u \leq z \neg Prf(Pr(\sigma),u)),$$

(*) $$T \vdash \chi \leftrightarrow \exists z(Prf(Pr(\sigma) \vee Pr(\chi),z) \wedge \forall u \leq h(g(z)) \neg Prf(\sigma,u)),$$

r is a proof of $Pr(\chi)$, and $Pr(\sigma)$ has no proof $\leq r$, then there is a proof $\leq h(r)$ of σ. (Analyze the proof of Lemma 1.1 (c).) Let $\sigma_0 := \sigma$ and $\sigma_1 := \chi$. Use Löb's theorem to show that $T \vdash Pr(\sigma_0) \vee Pr(\sigma_1)$.]

(b) Show that there are Σ_1 sentences χ_0, χ_1 such that χ_0 is true, χ_1 is false (in fact, $T \vdash \neg\chi_1$) and $Pr(\chi_0) \vee Pr(\chi_1)$ has a proof p such that $Pr(\chi_0)$ has no proof $\leq g(p)$. [Hint: In (*) replace $Pr(\chi)$ by $Pr(\sigma^*)$, where $\sigma^* :=$

$$\exists u(Prf(Pr(\sigma),u) \wedge \forall z{<}u \neg Prf(Pr(\chi),z)).$$

Let $\chi_0 := \sigma$ and $\chi_1 := \sigma^*$.]

28. Suppose PA$\dashv$ T and T is Σ_1–sound.

(a) Show that Theorem 14 and Exercises 26, 27 hold with "primitive recursive" replaced by "provably recursive in T".

(b) There is a recursive function f such that if p is a proof of the sentence $Pr(\varphi)$, then f(p) is a proof of φ. Show that f is not provably recursive in T.

Notes for Chapter 2.

Theorem 1 is due to Gödel (1931). (However, Gödel assumed that T is ω–consistent (see Exercise 7) but then applied this assumption only to the formula (corresponding to) $Prf_T(\varphi,x)$.) For a quick proof of what is the essential content of Gödel's theorem, namely: truth and provability in arithmetic are not equivalent (or: the set of true sentences of L_A is not r.e.), see Exercise 1; this also follows from each of the Exercises 1.2 (a), 1.3 (a), and 1.6 (b). Theorem 2 is due to Rosser (1936). Theorems 1 and 2 can be strengthened and generalized in a number of different directions as indicated in Exercises 1, 2, 3, 4 (see also Chapter 8). However, these "directions" lead away from the central theme of this book and so will not be pursued further; but see, for example, Kleene (1952a), Mostowski (1952b), (1961), and Kreisel and Lévy (1968). Lemma 1 is due to Lindström (1979). Theorem 3 is due to Mostowski (1961); for a stronger result also due to Mostowski (1961), see Exercise 19.

Theorem 4 is essentially due to Gödel (1931); the present general formulation is due to Feferman (1960). Corollary 1 is due to Mostowski (1952a) and Ryll-Nardzewski (1952); this result is strengthened in Chapter 4 (Corollary 4.1) and Chapter 6 (Theorem 6.3). Theorem 6 is due to Löb (1955). Löb's theorem or, more exactly, (L), is one of the keys to the modal logic of provability (cf. Boolos (1979), (1993), Smoryński (1985), Lindström (199?)). Theorem 7 is due to Feferman (1960).

Theorem 8 (a) (with a different proof) is due to Feferman (1960); Theorem 8 (b) is due to Orey (see Feferman (1960)).

Theorem 9 is due to Mostowski (1961); for a stronger result also due to Mostowski (1961), see Exercise 19. Theorem 10 is due to Scott (1962). Theorem 11 (with a different proof) is due to Montagna (1982).

For Theorem 12 with Π_1 replaced by Σ_1, see Exercise 24 (a). A result similar to Theorem 13 was first obtained by Gödel (1936) (cf. also Mostowski (1952b)); for a stronger result, see Exercise 3.3. Theorems 12 and 13 can also be derived from the fact that the set of (Π_1) sentences provable in T (T + φ) is not recursive (cf. Exercise 4 (c) and Theorem 1.2). Theorem 14, improved as in Exercise 28 (a), is due to Parikh (1971); the present proof was pointed out to me by Christian Bennet; see also de Jongh and Montagna (1989); a more general result has been proved by Montagna (1992); cf. also Hájek, Montagna, Pudlák (1992); for related results, see Exercise 5.15.

Exercise 1 is implicit in Tarski (1933) (see Gödel (1934) and Mostowski (1952b)). Exercise 6 (a) is a special case of a general result, the fixed point theorem of provability logic due to Dick de Jongh (unpublished) and Sambin (1976) (cf. also Boolos (1979), (1993), Smoryński (1985), Lindström (199?)). Exercise 11 (with a different proof) is due to Kripke (1963). Exercise 13 is due to Kreisel (see Smoryński (1985)). Exercise 15 (b) is due to Feferman (1960); it was used by him to prove Theorem 6.8. Exercise 17 is due to Hájkova (1971); her papers contain many related results. Exercise 18 is due to Bennet (1986). Exercise 19 is due to Mostowski (1961). Exercise 21 is due to Warren Goldfarb. The equivalence of (i), (ii), (iii) in Exercise 22 is due to Jensen and Ehrenfeucht (1976) and Guaspari (1979); for similar results, see Exercise 5.2. Exercise 27, improved as in Exercise 28 (a), is due to Shavrukov (1993) (with different proofs).