5. PARTIAL CONSERVATIVITY

A sentence φ is Γ*–conservative over* T if for every Γ sentence θ, if $T + \varphi \vdash \theta$, then $T \vdash \theta$. In this chapter we study this phenomenon for its own sake. Results on Γ–conservativity are, however, also very useful in many contexts, in particular in connection with interpretability (see Chapters 6 and 7).

Our task in this chapter is to develop general methods for constructing partially conservative sentences satisfying additional conditions such as being nonprovable in a given theory.

We assume throughout that $PA \dashv T$. The results of this chapter do not depend on the assumption that T is reflexive.

A first example of a Π_1–conservative sentence is given in the following:

Theorem 1. $\neg Con_T$ is Π_1–conservative over T.

Proof. Suppose θ is Π_1 and

(1) $T + \neg Con_T \vdash \theta$.

From (1) we get $PA \vdash Pr_T(\neg\theta) \rightarrow Pr_T(Con_T)$, whence

(2) $PA \vdash Pr_T(\neg\theta) \rightarrow \neg Con_{T+\neg Con_T}$.

By provable Σ_1–completeness,

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

By Corollary 2.2,

(4) $PA + Con_T \vdash Con_{T+\neg Con_T}$.

Combining (2), (3), (4) we get $PA \vdash \neg\theta \rightarrow \neg Con_T$ and so by (1), $T \vdash \theta$. ■

By Corollary 2.4, Theorem 1 provides us with an example of a (Σ_1) sentence φ which is Π_1–conservative over T and nontrivially so, i.e. such that $T \nvdash \varphi$, even if T is not Σ_1–sound.

If φ is Γ–conservative over T and ψ is Γ^d, then clearly φ is Γ–conservative over $T + \psi$. Also note that if T is Σ_1–sound and π is Π_1, then π is Σ_1–conservative over T iff π is true iff $T + \pi$ is consistent.

Let us now try to construct a sentence φ which is nontrivially Γ–conservative over T. Thus, given that

(1) $T + \varphi \vdash \theta$,

where θ is Γ, we want to be able to conclude that $T \vdash \theta$. This follows if (1) implies that

(2) $T + \neg\theta \vdash \varphi$.

The natural way to ensure that (1) implies (2) is to let φ be a sentence saying of itself that there is a false Γ sentence (namely θ) which φ implies in T. Thus, let φ be such that

(3) $PA \vdash \varphi \leftrightarrow \exists u(\Gamma(u) \wedge Pr_{T+\varphi}(u) \wedge \neg Tr_\Gamma(u))$,

where $\Gamma(x)$ is a PR binumeration of the set of Γ sentences. Then (1) implies (2).

It is, however, not generally true that $T \nvdash \varphi$. This holds if T is true, since φ is then false. But, for example, $T + \neg Con_T \vdash \varphi$, and so if $T \vdash \neg Con_T$, then $T \vdash \varphi$. To prevent this from happening, we redefine φ as follows: let φ be such that

$$PA \vdash \varphi \leftrightarrow \exists y \exists uv \le y(\Gamma(u) \wedge Prf_{T+\varphi}(u,v) \wedge \neg Tr_\Gamma(u) \wedge \forall z \le y \neg Prf_T(\varphi,z)).$$

Then $T \nvdash \varphi$ and φ is Γ–conservative over T. Also, if $\Gamma = \Pi_n$, then φ is Σ_n which is optimal; in fact, this is the sentence used in the proof of Theorem 2 (a), below, for $\Gamma = \Pi_n$.

From our present point of view the proof of Theorem 4.2 with S = T can be understood as follows (see the remarks following Corollary 4.1). Let ψ be as in that proof. It is sufficient to show that $\neg\psi$ is Γ^d–conservative over T; in fact, that is exactly what is done in the proof of Theorem 4.2. This also follows from the fact that (3) with φ replaced by $\neg\psi$ and Γ by Γ^d is true.

Let $[\Gamma]_S(x,y) :=$

$$\forall uv \le y(\Gamma(u) \wedge Prf_{S+x}(u,v) \rightarrow Tr_\Gamma(u)).$$

This formula is constructed to yield the following:

Lemma 1. $[\Gamma]_T(x,y)$ is a Γ formula such that

(i) $PA \vdash [\Gamma]_T(x,y) \wedge z \le y \rightarrow [\Gamma]_T(x,z)$,

(ii) $T + \varphi \vdash [\Gamma]_T(\varphi,m)$ for all φ and m,

(iii) if ψ is Γ and $T + \varphi \vdash \psi$, there is a q such that $PA + [\Gamma]_T(\varphi,q) \vdash \psi$.

Proof. (i) is clear. (ii) Let $\theta_0,\dots,\theta_k$ be all Γ sentences $\le$ m provable in $T + \varphi$ and whose proofs are $\le$ m. Then

$$PA \vdash \forall uv \le m(\Gamma(u) \wedge Prf_{T+\varphi}(u,v) \rightarrow u = \theta_0 \vee \dots \vee u = \theta_k).$$

Also clearly, by Fact 10 (a) (ii),

$$T + \varphi \vdash u = \theta_0 \vee \dots \vee u = \theta_k \rightarrow Tr_\Gamma(u).$$

It follows that $T + \varphi \vdash [\Gamma]_T(\varphi, m)$.

(iii) Suppose ψ is Γ and $T + \varphi \vdash \psi$. Let p be a proof of ψ in $T + \varphi$ and let $q = \max\{\psi,p\}$. Then $PA + [\Gamma]_T(\varphi,q) \vdash Tr_\Gamma(\psi)$ and so $PA + [\Gamma]_T(\varphi,q) \vdash \psi$. ■

S is a Γ*-subtheory* of T, $S \dashv_\Gamma T$, if every Γ sentence provable in S is provable in T. We write $[\Gamma](x,y)$ for $[\Gamma]_T(x,y)$.

Lemma 2. Suppose $\chi(x,y)$ is Γ^d. There is then a Γ^d formula $\xi(x)$ such that for all k and m,

(i) $T + \xi(k) \vdash \chi(k,m)$,

(ii) $T + \xi(k) \dashv_\Gamma T + \{\chi(k,q) : q \in N\}$.

Proof. *Case 1.* $\Gamma = \Sigma_n$. Let $\xi(x)$ be such that

(1) $PA \vdash \xi(k) \leftrightarrow \forall y([\Sigma_n](\xi(k),y) \rightarrow \chi(k,y))$.

Then (i) follows from Lemma 1 (ii) and (1). To prove (ii), suppose ψ is Σ_n and $T + \xi(k) \vdash \psi$. By Lemma 1 (iii), there is a q such that

$$PA + [\Sigma_n](\xi(k),q) \vdash \psi.$$

Hence, by Lemma 1 (i),

$$PA + \forall y\leq q\chi(k,y) + \neg\psi \vdash \forall y([\Sigma_n](\xi(k),y) \rightarrow \chi(k,y))$$

and so, by (1),

$$PA + \forall y\leq q\chi(k,y) + \neg\psi \vdash \xi(k).$$

But then, since $T + \xi(k) \vdash \psi$, it follows that $T + \{\chi(k,q): q\in N\} \vdash \psi$, as desired.

Case 2. $\Gamma = \Pi_n$. Let $\xi(x)$ be such that

$$PA \vdash \xi(k) \leftrightarrow \exists y(\neg[\Pi_n](\xi(k),y) \wedge \forall z\leq y\chi(k,z)).$$

The proof that $\xi(x)$ is as claimed is then almost the same as in Case 1. ■

From Lemma 2 we derive the following result on numerations of r.e. sets.

Lemma 3. Let X be an r.e. set. There is then a Γ^d formula $\xi(x)$ such that

(i) if $k\in X$, then $T \vdash \neg\xi(k)$,

(ii) if $k\notin X$, then $\xi(k)$ is Γ–conservative over T.

Proof. Let $\rho(x,y)$ be a PR formula such that $X = \{k: \exists m PA \vdash \rho(k,m)\}$ and let $\xi(x)$ be as in Lemma 2 with $\chi(x,y) := \neg\rho(x,y)$. ■

For extensions of PA Lemma 3 implies Theorem 3.1.

We can now prove our first general theorem on the existence of nontrivially partially conservative sentences.

Theorem 2. (a) There is a Γ^d sentence φ such that $T \nvdash \varphi$ and φ is Γ–conservative over T.

(b) If X is r.e. and monoconsistent with T, there is a Γ^d sentence φ such that $\varphi\notin X$ and φ is Γ–conservative over T.

Proof. (a) is the special case of (b) where X = Th(T). ♦

(b) Let $\xi(x)$ be as in Lemma 3 and let φ be such that $PA \vdash \varphi \leftrightarrow \xi(\varphi)$. If $\varphi\in X$, then, by Lemma 3 (i), $T \vdash \neg\xi(\varphi)$ and so $T \vdash \neg\varphi$, which is impossible. Thus, $\varphi\notin X$ and so, by Lemma 3 (ii), φ is Γ–conservative over T. ■

Of course, the Γ^d sentence mentioned in Theorem 2 (a) is not Γ^T (compare Corollary 2.5).

The following result is a natural strengthening of Theorem 2.

Theorem 3. (a) There is a Γ sentence φ such that φ is Γ^d–conservative over T and $\neg\varphi$ is Γ–conservative over T.

(b) If X is r.e. and monoconsistent with T, there is a Γ sentence φ such that φ is Γ^d–conservative over T, $\neg\varphi$ is Γ–conservative over T, and $\varphi, \neg\varphi\notin X$.

We derive Theorem 3 from:

Lemma 4. Suppose $\chi_0(x,y)$ is Γ^d and $\chi_1(x,y)$ is Γ. Then there is a Γ formula $\xi(x)$ such that for i = 0, 1,

(i) $T + \xi^i(k) \vdash \forall y \le m \chi_i(k,y) \to \chi_{1-i}(k,m)$,

(ii) if ψ is Γ^d and $T + \xi^i(k) \vdash \psi^i$, then $T + \{\chi_{1-i}(k,q): q \in N\} \vdash \psi^i$.

Proof. We need only prove this for $\Gamma = \Sigma_n$. Let $\xi(x)$ be such that

(1) $PA \vdash \xi(k) \leftrightarrow \exists y((\neg[\Pi_n](\xi(k),y) \vee \neg\chi_0(k,y)) \wedge \forall z<y([\Sigma_n](\neg\xi(k),z) \wedge \chi_1(k,z)))$.

We verify (i) and (ii) for $i = 0$ and leave the case $i = 1$ to the reader.

(i) By Lemma 1 (ii),

$$T + \xi(k) \vdash \neg[\Pi_n](\xi(k),y) \to y > m.$$

It follows that

$$T + \xi(k) + \forall y \le m \chi_0(k,y) \vdash (\neg[\Pi_n](\xi(k),y) \vee \neg\chi_0(k,y)) \to y > m.$$

But then, by (1),

$$T + \xi(k) + \forall y \le m \chi_0(k,y) \vdash \chi_1(k,m),$$

as desired.

(ii) Suppose ψ is Π_n and

(2) $T + \xi(k) \vdash \psi$.

By Lemma 1 (iii), there is a q such that $T + [\Pi_n](\xi(k),q) \vdash \psi$ and so

(3) $T + \neg\psi \vdash \neg[\Pi_n](\xi(k),q)$.

By Lemma 1 (ii), for every m,

(4) $T + \neg\xi(k) \vdash [\Sigma_n](\neg\xi(k),m)$.

By (3), (4), Lemma 1 (i), and (1), it follows that

$$T + \neg\psi + \neg\xi(k) + \forall y \le q \chi_1(k,y) \vdash \xi(k)$$

and so

$$T + \neg\psi + \forall y \le q \chi_1(k,y) \vdash \xi(k).$$

But then, by (2), $T + \forall y \le q \chi_1(k,y) \vdash \psi$, as desired. ■

Proof of Theorem 3. (a) is a special case of (b). ♦

(b) Let $\rho_i(x,y)$, $i = 0, 1$, be PR binumerations of relations $R_i(k,m)$ such that $X = \{k: \exists m R_0(k,m)\}$ and $\{\varphi: \neg\varphi \in X\} = \{k: \exists m R_1(k,m)\}$. Let $\xi(x)$ be as in Lemma 4 with $\chi_i(x,y) := \neg\rho_{1-i}(x,y)$. Let φ be such that $PA \vdash \varphi \leftrightarrow \xi(\varphi)$. Suppose $\varphi \in X$ or $\neg\varphi \in X$. Let m be the least number such that $R_0(\varphi,m)$ or $R_1(\varphi,m)$. Suppose $R_i(\varphi,m)$. Then not $R_{1-i}(\varphi,n)$ for $n \le m$. (We may assume that $R_0(k,n)$ implies not $R_1(k,n)$.) But then, by Lemma 4 (i), $T \vdash \neg\xi^i(\varphi)$, whence $T \vdash \neg\varphi^i$. But this is impossible, since $\varphi^i \in X$. It follows that φ, $\neg\varphi \notin X$. But then, by Lemma 4 (ii), φ is Γ^d–conservative over T and $\neg\varphi$ is Γ–conservative over T. ■

Let $Prf'_{T,\Gamma}(x,y) :=$

$$\exists uv \le y(\Gamma(u) \wedge Tr_\Gamma(u) \wedge Prf_{T+u}(x,v));$$

a slight modification of the formula $Prf_{T,\Gamma}(x,y)$ defined in Chapter 4. In the proofs of Lemmas 2 and 4 $[\Gamma](x,y)$ can be replaced by $\neg Prf'_{T,\Gamma^d}(\neg x,y)$. Then, for example, formula (1) in the proof of Lemma 4 becomes:

(Sm) $PA \vdash \xi(k) \leftrightarrow \exists y((Prf'_{T,\Sigma_n}(\neg\xi(k),y) \vee \neg\chi_0(k,y)) \wedge \forall z<y(\neg Prf'_{T,\Pi_n}(\xi(k),y) \wedge \chi_1(k,z)))$.

This formula may be compared with formula (1) in the proof of Theorem 3.2 and (R′) following the proof of Theorem 2.2.

Our next result is related to Theorem 4.3; it will be used several times, in some cases indirectly, in Chapters 6 and 7.

S is a *Γ–conservative extension* of T if $T \dashv S \dashv_\Gamma T$. By Theorems 4.4 (a) and 4.5, T + Rfn_T is a Π_1–conservative extension of PA + Con_T^ω.

Theorem 4. (a) Let X be an r.e. set of Γ sentences. There is then a Γ sentence θ such that T + θ is a Γ^d–conservative extension of T + X.

(b) Let $\gamma(x,y)$ be any Γ formula. There is then a Γ formula $\eta(x)$ such that for every k, T + $\eta(k)$ is a Γ^d–conservative extension of T + $\{\gamma(k,m): m \in N\}$.

Proof. (a) By Craig's theorem, we may assume that X is primitive recursive. Let $\eta(x)$ be a PR binumeration of X. Then for every q,

(1) $PA + X \vdash \eta(q) \to Tr_\Gamma(q)$.

By Lemma 2 with (Γ replaced by Γ^d and) $\chi(x,y) := \eta(y) \to Tr_\Gamma(y)$, there is a Γ sentence θ such that for all φ,

(2) $T + \theta \vdash \eta(\varphi) \to Tr_\Gamma(\varphi)$,

(3) $T + \theta \dashv_{\Gamma^d} T + \{\eta(q) \to Tr_\Gamma(q): q \in N\}$.

From (2) it follows that $T + \theta \vdash X$ and from (1) and (3) it follows that $T + \theta \dashv_{\Gamma^d}$ T + X. ♦

(b) Left to the reader. ■

So far there has been no indication that the properties of Σ_n and Π_n, n > 1, in terms of partial conservativity may be different, but we shall now show that they are.

Let ψ_0 and ψ_1 be Γ sentences. If

(1) $T \vdash \psi_0 \vee \psi_1$,

then, trivially,

(2) ψ_i is Γ^d–conservative over $T + \neg\psi_{1-i}$, i = 0, 1.

If $\Gamma = \Pi_n$, the converse of this is true. This follows from our next:

Lemma 5. Let ψ_0 and ψ_1 be any Π_n sentences. There are then Π_n sentences θ_0 and θ_1 such that

(i) $T \vdash \theta_0 \vee \theta_1$,

(ii) $T \vdash \psi_i \to \theta_i$, i = 0, 1,

(iii) $T \vdash \theta_0 \wedge \theta_1 \to \psi_0 \wedge \psi_1$.

Proof. By Fact 5, we may assume that $\psi_i := \forall x \delta_i(x)$, where $\delta_i(x)$ is Σ_{n-1}. Let $\theta_i :=$
$\forall x(\neg\delta_i(x) \to \exists y{<}x{+}i \neg\delta_{1-i}(y))$.

Then (i), (ii), (iii) are easily verified (cf. Lemma 1.3). ■

From (ii) and (iii) of Lemma 5 it follows that $T + \neg\psi_i + \psi_{1-i} \vdash \neg\theta_i$. Hence, assuming (2), $T + \neg\psi_i \vdash \neg\theta_i$. It follows that $T \vdash \theta_0 \vee \theta_1 \to \psi_0 \vee \psi_1$ and so, by Lemma 5 (i), we get (1).

We now prove that if $\Gamma = \Sigma_n$, then (2) does not imply (1).

Theorem 5. (a) There are Σ_n sentences ψ_0, ψ_1 such that

(i) $T\vdash \neg(\psi_0 \wedge \psi_1)$,

(ii) $T\nvdash \psi_0 \vee \psi_1$,

(iii) ψ_i is Π_n–conservative over $T + \neg\psi_{1-i}$, $i = 0, 1$.

(b) Suppose X is r.e. and monoconsistent with T. Then there are Σ_n sentences ψ_0, ψ_1 such that (i) and (iii) hold and

(iv) $\psi_0 \vee \psi_1 \notin X$.

We derive this theorem from:

Lemma 6. Let X be an r.e. set. There are then Σ_n formulas $\xi_0(x)$ and $\xi_1(x)$ such that for $i = 0, 1$,

(i) $T\vdash \neg(\xi_0(x) \wedge \xi_1(x))$,

(ii) if $k\in X$, then $T\vdash \neg\xi_i(k)$,

(iii) if $k\notin X$, then $\xi_i(k)$ is Π_n–conservative over $T + \neg\xi_{1-i}(k)$.

Proof. Let $\rho(x,y)$ be a PR formula such that $X = \{k: \exists m PA\vdash \rho(k,m)\}$. For $i = 0, 1$, let $\xi_i(x)$, $\chi_i(x)$, $\delta_i(x,y)$ be, respectively, Σ_n, Σ_n, and Π_{n-1} formulas such that

(1) $PA\vdash \chi_i(k) \leftrightarrow \exists y(\neg[\Pi_n](\xi_i(k),y) \wedge \forall z{\le}y\neg\rho(k,z))$,

(2) $PA\vdash \chi_i(x) \leftrightarrow \exists y\delta_i(x,y)$,

$\xi_i(x) := \exists y(\delta_i(x,y) \wedge \forall z{<}y{+}i\neg\delta_{1-i}(x,z))$.

This application of (double) self–reference is more complicated than any we have encountered so far and it requires some thought to see that it is admissible. But in view of Fact 5 it is.

(i) is then clear. To prove (ii), suppose $k\in X$. Let m be such that $PA\vdash \rho(k,m)$. By Lemma 1 (ii),

$T + \xi_i(k)\vdash \neg[\Pi_n](\xi_i(k),y) \rightarrow m < y$.

So, by (1),

(3) $T + \xi_i(k)\vdash \neg\chi_i(k)$.

Also, by (2), $PA\vdash \xi_i(x) \rightarrow \chi_i(x)$. Now (ii) follows from this and (3).

Finally, to prove (iii), suppose $k\notin X$. Now suppose ψ is Π_n and

$T + \neg\xi_{1-i}(k) + \xi_i(k)\vdash \psi$.

By (i), it follows that

(4) $T + \xi_i(k)\vdash \psi$.

But then, by Lemma 1 (iii), there is a q such that $T + [\Pi_n](\xi_i(k),q)\vdash \psi$. Also $T\vdash \neg\rho(k,m)$ for all m. By (1), it now follows that $T + \neg\psi\vdash \chi_i(k)$. Thus, by (2), $T + \neg\psi\vdash \exists y\delta_i(k,y)$. But then

$T + \neg\psi + \neg\xi_{1-i}(k)\vdash \xi_i(k)$.

Combining this with (4) we get $T + \neg\xi_{1-i}(k)\vdash \psi$. This proves (iii). ■

Proof of Theorem 5. (a) follows from (b). ♦

(b) We may assume that if $\psi\in X$ and $T\vdash \psi \rightarrow \theta$, then $\theta\in X$. Let $\xi_i(x)$ be as in Lemma 6. Let φ be such that

$PA\vdash \varphi \leftrightarrow \xi_0(\varphi) \vee \xi_1(\varphi)$.

Set $\psi_i := \xi_i(\varphi)$. If $\varphi\in X$, then, by Lemma 6 (ii), $T\vdash \neg\xi_i(\varphi)$ for i = 0, 1, and so $T\vdash \neg\varphi$, impossible. Thus, $\varphi\notin X$ and so (iv) holds. (i) and (iii) follow from Lemma 6 (i) and (iii), respectively. ■

Theorem 5 (b) will be used in the proof of Theorem 7.7 (b), below. Note that, by Theorem 5, Lemma 5 with Π_n replaced by Σ_n is false.

We can now partially improve Corollary 2.5 as follows:

Corollary 1. There are Σ_n sentences ψ_0, ψ_1, such that $T\vdash \psi_0 \rightarrow \neg\psi_1$ and there is no Δ_n sentence φ for which $T\vdash \psi_0 \rightarrow \varphi$ and $T\vdash \varphi \rightarrow \neg\psi_1$.

Proof. Let ψ_0, ψ_1 be as in Theorem 5 (a). Suppose φ is Δ_n, $T\vdash \psi_0 \rightarrow \varphi$, and $T\vdash \varphi \rightarrow \neg\psi_1$. Then $T\vdash \neg\psi_1 \rightarrow \varphi$ and $T\vdash \neg\psi_0 \rightarrow \neg\varphi$ and so $T\vdash \psi_0 \vee \psi_1$, a contradiction. ■

Let Cons(Γ,T) be the set of sentences Γ–conservative over T. It is clear from the definition of Cons(Γ,T) that it is a Π^0_2 set. We now show that this classification is correct.

Our next lemma follows at once from Lemma 3.2 (b) but has a simpler direct proof which we leave to the reader.

Lemma 7. Let R(k,m) be any r.e. relation. There are then formulas $\rho_0(x,y)$ and $\rho_1(x,y)$ such that $\rho_0(x,y)$ is Σ_1, $\rho_1(x,y)$ is Π_1, $\rho_0(x,y)$ numerates R(k,m) in T, $PA\vdash \rho_0(k,m) \rightarrow \rho_1(k,m)$, and if not R(k,m), then $T\nvdash \rho_1(k,m)$.

Theorem 6. (a) Cons(Γ,T) is a complete Π^0_2 set.

(b) If $\Gamma \neq \Sigma_1$, then $\Gamma^d \cap$ Cons(Γ,T) is a complete Π^0_2 set.

Proof. Let X be any Π^0_2 set and let R(k,m) be an r.e. relation such that X = {k: $\forall$mR(k,m)}. Let $\rho(x,y)$ be a formula numerating R(k,m) in T, which is Σ_1 if $\Gamma = \Sigma_n$ and Π_1 if $\Gamma = \Pi_n$. Let $\xi(x)$ be as in (the proof of) Lemma 2 with $\chi(x,y) := \rho(x,y)$. To prove (a) it is now sufficient to show that

(1) $k\in X$ iff $\xi(k)\in$ Cons(Γ,T).

By Lemma 2,

(2) $T + \xi(k)\vdash \rho(k,m)$,

(3) $T + \xi(k)\dashv_\Gamma T + \{\rho(k,q): q\in N\}$.

If $k\in X$, then $T\vdash \rho(k,q)$ for every q and so, by (3), $\xi(k)\in$ Cons(Γ,T). If $k\notin X$, there is an m such that $T\nvdash \rho(k,m)$ and so, by (2), $\xi(k)\notin$ Cons(Γ,T) (in fact, $\xi(k)$ is not Σ_1– or not Π_1–conservative over T, as the case may be). Thus, (1) holds.This proves (a).

If Γ is Σ_n or Π_n with $n \geq 2$, then $\xi(x)$ is Γ^d as claimed in (b). Finally, suppose $\Gamma = \Pi_1$. Let $\rho_0(x,y)$ and $\rho_1(x,y)$ be as in Lemma 7. Let $\rho(x,y) := \rho_0(x,y)$. Then $\xi(x)$ is Σ_1. By Lemma 7, $\xi(k)\notin$ Cons(Π_1,T) if $k\notin X$. Thus, (b) holds in this case, too. ■

Suppose T is Σ_1–sound and θ is Π_1. Then θ is Σ_1–conservative over T iff θ is true. Thus, $\Pi_1 \cap$ Cons(Σ_1,T) is Π^0_1.

We conclude this chapter with a proof of Theorem 4.8. We derive this result from the following lemma; a refinement of this lemma (for n = 1) will be proved in Chapter 7 (Lemma 7.22).

Lemma 8. There is a Π_n formula $\xi(x)$ such that for every k,
(i) $T \nvdash \xi(k)$,
(ii) $T \vdash \xi(k+1) \rightarrow \xi(k)$,
(iii) $\xi(k)$ is Σ_n–conservative over $T + \neg\xi(k+1)$.

Proof. In a first attempt to prove Lemma 8 it is natural to let $\xi(x)$ be such that

$$PA \vdash \xi(k) \leftrightarrow \xi(k+1) \vee \forall v([\Sigma_n](\neg\xi(k+1) \wedge \xi(k),v) \rightarrow \neg Prf_T(\xi(k),v)).$$

But then (i) does not follow and so we have to proceed in a more indirect way.

Let $\delta(u)$ be any formula. Let $\kappa(z,x,y)$ be a Π_n formula such that
(1) $PA \vdash \neg\kappa(z,x,0)$,
(2) $PA \vdash \kappa(\delta,k,y+1) \leftrightarrow \kappa(\delta,k+1,y) \vee \forall v([\Sigma_n](\neg\eta_\delta(k) \wedge \xi_\delta(k),v) \rightarrow \neg Prf_T(\xi_\delta(k),v))$,
where

$$\xi_\delta(x) := \forall u(\delta(u) \rightarrow \kappa(\delta,x,(u \dot{-} x) + 1)),$$
$$\eta_\delta(x) := \forall u(\delta(u) \rightarrow \kappa(\delta,x+1,u \dot{-} x)).$$

($\dot{-}$ is the function such that $k \dot{-} m = k - m$ if $k \geq m$ and $= 0$ otherwise.) In (2) set $y = u \dot{-} k$. Then, since neither y nor u is free in the second disjunct of (2), by predicate logic, we get
(3) $PA \vdash \xi_\delta(k) \leftrightarrow \eta_\delta(k) \vee \forall v([\Sigma_n](\neg\eta_\delta(k) \wedge \xi_\delta(k),v) \rightarrow \neg Prf_T(\xi_\delta(k),v))$.
It follows that
(4) if $T \vdash \xi_\delta(k)$, then $T \vdash \eta_\delta(k)$.
For let p be a proof of $\xi_\delta(k)$ in T. By Lemma 1 (ii),

$$T + \neg\eta_\delta(k) \wedge \xi_\delta(k) \vdash \neg Prf_T(\xi_\delta(k),p),$$

whence $T + \xi_\delta(k) \vdash \eta_\delta(k)$ and so $T \vdash \eta_\delta(k)$.

Clearly
(5) if $T \vdash \delta(u) \rightarrow u > k$, then $T \vdash \eta_\delta(k) \leftrightarrow \xi_\delta(k+1)$.
Suppose now $\delta(u)$ is PR. Then
(6) if $\exists u\delta(u)$ is true, then $T \nvdash \xi_\delta(0)$.
Suppose $\exists u\delta(u)$ is true and $T \vdash \xi_\delta(0)$. Let m be the least number such that $\delta(m)$ is true. Then $T \vdash \delta(u) \rightarrow u \geq m$. By (4) and (5), it follows that $T \vdash \eta_\delta(m)$. But also $T \vdash \delta(m)$ and so, by (1), $T \vdash \neg\eta_\delta(m)$, a contradiction. Thus, (6) is proved.

Now let $\delta'(u)$ be a PR formula such that
(7) $PA \vdash \exists u\delta'(u) \leftrightarrow Pr_T(\xi_{\delta'}(0))$.
If $\exists u\delta'(u)$ is true, then, by (6), $Pr_T(\xi_{\delta'}(0))$ is false and, by (7), it is true. Thus, $\exists u\delta'(u)$ is false, whence, by (7), $Pr_T(\xi_{\delta'}(0))$ is false and so $T \nvdash \xi_{\delta'}(0)$.

Let $\xi(x) := \xi_{\delta'}(x)$ and $\eta(x) := \eta_{\delta'}(x)$. Then $T \nvdash \xi(0)$. Hence, by (3) and (5) with $\delta(u) := \delta'(u)$, we get (i) and (ii).

(iii) can be verified as follows. Suppose
(8) $T + \neg\xi(k+1) + \xi(k) \vdash \sigma$,

where σ is Σ_n. Then, by (5), $T + \neg\eta(k) + \xi(k) \vdash \sigma$. Hence, by Lemma 1 (iii), there is a q such that

$$T + [\Sigma_n](\neg\eta(k) \wedge \xi(k), q) \vdash \sigma.$$

But then, by (i), (3), and Lemma 1 (i), $T + \neg\sigma \vdash \xi(k)$, whence $T + \neg\xi(k) \vdash \sigma$ and so, by (8), $T + \neg\xi(k+1) \vdash \sigma$, proving (iii). ■

Proof of Theorem 4.8. Let $\xi(x)$ be as in Lemma 8. By Lemma 8 (i) and (iii), $T \nvdash \xi(k) \rightarrow \xi(k+1)$. It follows that $T + \xi(0) + \{\xi(k) \rightarrow \xi(k+1): k \in N\}$ is an axiomatization of $T + \{\xi(k): k \in N\}$ which is irredundant over T. Let π_k, $k \in N$, be Π_n sentences such that $T + \{\pi_k: k \in N\}$ is an axiomatization of $T + \{\xi(k): k \in N\}$. Let r be arbitrary. By Lemma 8 (ii), there is an m such that $T + \xi(m) \vdash \pi_r$. Let s be such that $T + \pi_0 \wedge ... \wedge \pi_s \vdash \xi(m+1)$. We may assume that $s > r$. It follows that

$$T + \xi(m) \wedge \neg\xi(m+1) \vdash \neg(\pi_0 \wedge ... \wedge \pi_{r-1} \wedge \pi_{r+1} \wedge ... \wedge \pi_s).$$

But then, by Lemma 8 (iii),

$$T + \pi_0 \wedge ... \wedge \pi_{r-1} \wedge \pi_{r+1} \wedge ... \wedge \pi_s \vdash \xi(m+1).$$

It follows, by Lemma 8 (ii), that $T + \{\pi_k: k \neq r\} \vdash \pi_r$. Thus, $T + \{\pi_k: k \in N\}$ is not irredundant over T. ■

We have actually proved more than is stated in Theorem 4.8. First of all, for every r, $T + \{\pi_k: k \neq r\} \vdash \pi_r$; in fact, for every m, $T + \{\pi_k: k > m\} \vdash \pi_r$. Secondly, this holds for all, not necessarily r.e., sets $\{\pi_k: k \in N\}$ of Π_n sentences such that $T + \{\pi_k: k \in N\} \dashv\vdash T + \{\xi(k): k \in N\}$. The theory $T + \{\eta(k): k \in N\}$ constructed in the proof of Theorem 4.7, on the other hand, is deductively equivalent to $T + \{\eta(k): k \notin H\}$ and $\{\eta(k): k \notin H\}$ is irredundant over T. (The set $\{\eta(k): k \notin H\}$ is not r.e. (cf. Lemma 4.6).)

Exercises for Chapter 5.

In the following exercises we assume that $PA \dashv T$.

1. Let θ be a Π_1 Rosser sentence for T. Show that $\neg\theta$ is not Π_1–conservative over T (compare Exercise 2 (c)).

2. Suppose T is not Σ_1–sound.

(a) Show that Con_T is not Σ_1–conservative over T. [Hint: Let $\delta(y)$ be a PR formula such that $\exists y\delta(y)$ is false and provable in T. Let χ be as in Exercise 2.21. Then $T \nvdash \chi$ and $T + \neg\chi \vdash Pr_T(\chi) \wedge Pr_T(\neg\chi)$.]

(b) Improve (a) by showing that if $T \nvdash \neg Con_T$, there is a Σ_1 sentence σ such that $T + Con_T \vdash Pr_T(\sigma)$ and $T \nvdash Pr_T(\sigma)$.

(c) Improve (a) by showing that if θ is a Π_1 Rosser sentence for T, θ is not Σ_1–conservative over T. [Hint: Let $\psi := \exists u(Prf_T(\neg\theta,u) \wedge \forall z \leq u \neg Prf_T(\theta,z))$. $T + \neg\psi$ is consistent. $T + \neg\psi + \theta \vdash Con_{T+\neg\theta}$ and $T + \neg\theta \vdash \neg\psi$. Thus, $T + \neg\psi + \theta \vdash Con_{T+\neg\psi}$. Apply (a) to $T + \neg\psi$.]

3. Show that the result of replacing Σ_n by Π_n in Corollary 1 is false.

4. φ is a *self-prover in* T if $T \vdash \varphi \rightarrow Pr_T(\varphi)$. Every Σ_1 sentence is a self-prover.

(a) Show that φ is a self-prover in T iff there is a sentence θ such that $T \vdash \varphi \leftrightarrow (\theta \wedge Pr_T(\theta))$.

(b) Show that for every $n > 0$, there is a Σ_n (Π_{n+1}) self-prover in T which is not Π_n^T (Σ_{n+1}^T).

5. (a) Show that Lemma 2 (ii) can be replaced by

if $PA \dashv S \dashv T$, then $S + \xi(k) \dashv_\Gamma S + \{\chi(k,q)\colon q \in N\}$.

(b) φ is *hereditarily* Γ*-conservative* over T if φ is Γ-conservative over S for every S such that $PA \dashv S \dashv T$. Show that in Lemma 3 and Theorem 2 we can replace "Γ^d-conservative over T" by "hereditarily Γ^d-conservative over T".

(c) Show that in Theorem 3 we cannot in general replace "Γ- (Γ^d-) conservative" by "hereditarily Γ- (Γ^d-) conservative". [Hint: Let φ be a Σ_1 sentence and ψ a Π_1 sentence such that $PA + \varphi \wedge \psi$ is consistent and $PA \nvdash \varphi \vee \psi$. Let $T = PA + \varphi \wedge \psi$.]

6. (a) Show that there are sentences φ and ψ such that, $T + \varphi \nvdash \psi$, $T + \psi \nvdash \varphi$, φ is Π_n-conservative over $T + \psi$, and ψ is Σ_n-conservative over $T + \varphi$.

(b) Improve (a) by showing that there are sentences φ and ψ as in (a) such that φ is Σ_n and ψ is Π_n. [Hint: Let

$T \vdash \varphi \leftrightarrow \exists z(\neg[\Pi_n]_{T+\psi}(\varphi,z) \wedge \forall u{\le}z \neg Prf_T(\varphi,u))$,

$T \vdash \psi \leftrightarrow \forall z([\Sigma_n]_{T+\varphi}(\psi,z) \rightarrow \neg Prf_T(\psi,z))$.

Use Exercise 5 (b).]

7. Show that there are Σ_n sentences ψ_0, ψ_1 as in Theorem 5 satisfying the additional condition that $\neg\psi_i$ is Σ_n-conservative over T, $i = 0, 1$.

8. (a) S is a *proper* Γ-subtheory of T if $S \vdash_\Gamma T \not\dashv_\Gamma S$. Suppose $A \dashv B \not\dashv_{\Pi_1} A$. Show that there is a sentence χ such that A is a proper Π_1-subtheory of $A + \chi^i$ and $A + \chi^i \dashv_\Gamma B$, $i = 0, 1$.

(b) Show that there are sentences φ_0, φ_1 such that φ_0, φ_1, $\neg\varphi_0 \vee \neg\varphi_1$ are Γ-conservative over T and $\neg\varphi_0$, $\neg\varphi_1$, $\varphi_0 \wedge \varphi_1$ are not Π_1-conservative over T. [Hint: Use Lemma 4.]

9. (a) Show that there is a Δ_{n+1} sentence φ such that φ and $\neg\varphi$ are Π_n-conservative over T. [Hint: Let φ be such that

$PA \vdash \varphi \leftrightarrow \exists y(\neg[\Pi_n](\varphi,y) \wedge \forall z{<}y[\Pi_n](\neg\varphi,z))$.]

(b) Show that if T is Σ_n-sound, there is no Δ_{n+1} sentence φ such that φ and $\neg\varphi$ are Σ_n-conservative over T.

(c) Show that there is no B_n sentence φ such that φ and $\neg\varphi$ are Π_n- (Σ_n-) conservative over T. Conclude that there is a Δ_{n+1} sentence which is not B_n^T (compare Corollary 2.5). [Hint: Suppose not. Let $\varphi := (\pi_0 \wedge \sigma_0) \vee \ldots \vee (\pi_n \wedge \sigma_n)$. In the Π_n case, for $k \le n+1$, show that

$T\vdash \bigvee\{\bigwedge_{j\in X}\neg\sigma_j: X\subseteq\{0,\dots,n\}$ & X has exactly k elements}.]

10. Let X_0 and X_1 be disjoint r.e. sets.

(a) Show that there is a Σ_n formula $\xi(x)$, such that $\xi^i(x)$ numerates X_i in T, i = 0, 1, and if $k\notin X_0\cup X_1$, then $\xi(k)$ is Π_n–conservative over T and $\neg\xi(k)$ is Σ_n–conservative over T.

(b) Show that there is a formula $\xi(x)$ such that (i) if $k\in X_0$, then $T\vdash \xi(k)$, (ii) if $k\in X_1$, then $T\vdash \neg\xi(k)$, (iii) if Y_0 and Y_1 are any disjoint finite subsets of $(X_0\cup X_1)^c$, then $\bigwedge\{\xi(k): k\in Y_0\}\wedge\bigwedge\{\neg\xi(k): k\in Y_1\}$ is Γ–conservative over T. [Hint: First define a formula $\eta(k)$ such that all the sentences $(\neg)\eta(0)\wedge\dots\wedge(\neg)\eta(k)$ are Γ–conservative over T. Then let $\xi(x) := (\xi_0(x)\vee\eta(x))\wedge\neg\xi_1(x)$ for suitable $\xi_0(x)$, $\xi_1(x)$.]

11. (a) Let X and Y be r.e. sets of Γ and Γ^d sentences, respectively, such that if $\varphi\in X$ and $\psi\in Y$, then $T\vdash\varphi\vee\psi$. Show that there is a Γ sentence θ such that $T+\theta$ is a Γ^d–conservative extension of T + X and $T+\neg\theta$ is a Γ–conservative extension of T + Y.

(b) Let $\theta_0, \theta_1, \theta_2, \dots$ be a recursive sequence of Γ sentences such that $T\vdash \neg(\theta_k\wedge\theta_m)$ for $k\neq m$. Let X_0 and X_1 be disjoint r.e. sets. Show that there is a sentence φ such that $X_0 = \{k: T\vdash \theta_k\to\varphi\}$ and $X_1 = \{k: T\vdash\theta_k\to\neg\varphi\}$.

12. Suppose T is not Σ_1–sound. Show that $\Pi_1\cap \mathrm{Cons}(\Sigma_1,T)$ is a complete Π^0_2set. [Hint: Let R(k,m) and S(k,m,n) be an r.e. and a primitive recursive relation such that $X = \{k: \forall m R(k,m)\}$ and R(k,m) iff $\exists n S(k,m,n)$. Let $\sigma(x,y,z)$ be a PR binumeration of S(k,m,n). Let $\gamma(x)$ be a PR formula such that $\exists x\gamma(x)$ is false and provable in T. Let $\rho_0(x,y)$, $\rho_1(x,y)$, and $\delta(x,y,z)$ be such that

$$PA\vdash \rho_0(x,y)\leftrightarrow\forall z(\mathrm{Prf}_T(\rho_1(\dot{x},\dot{y}),z)\to\exists u\le z\sigma(x,y,u)),$$
$$\rho_0(x,y) := \forall z\delta(x,y,z),$$
$$\rho_1(x,y) := \exists z(\gamma(z)\wedge\forall u\le z\delta(x,y,z)).$$

Then

$$T\vdash \rho_0(x,y)\to\rho_1(x,y),$$
if R(k,m), then $T\vdash\rho_0(k,m)$,
if not R(k,m), then $T\nvdash\rho_1(k,m)$.]

13. (a) Let HCons(Γ,T) be the set of sentences hereditarily Γ–conservative over T. Suppose $\Gamma\neq\Sigma_1$. Show that $\Gamma^d\cap \mathrm{HCons}(\Gamma,T)$ is a complete Π^0_2 set.

(b) Show that

$$\Gamma^d\cap\mathrm{Cons}(\Gamma,T)\cap\{\varphi: \neg\varphi\in\mathrm{Cons}(\Gamma^d,T)\}$$
is a complete Π^0_2 set.

(c) Show that

$$\Sigma_n\times\Sigma_n\cap\{<\varphi_0,\varphi_1>: \varphi_i\in\mathrm{Cons}(\Pi_n, T+\neg\varphi_{1-i}), i=0,1\}$$
is a complete Π^0_2 set.

14. (a) Suppose φ is Σ_n, and Π_n–conservative over T. Let ψ be any Π_n sentence which is Σ_n–conservative over T + φ. Show that T + $\neg\varphi \vdash \psi$. Conclude that no Π_n sentence is nontrivially Σ_n–conservative over T + φ and T + $\neg\varphi$. [Hint: Let $\varphi := \exists x\gamma(x)$ and $\psi := \forall x\delta(x)$, where $\gamma(x)$ and $\delta(x)$ are Π_{n-1} and Σ_{n-1}, respectively. Then T + φ + $\psi \vdash \exists x(\gamma(x) \wedge \forall y{\le}x\delta(y))$.]

(b) Show that there is an r.e. family of consistent extensions of PA such that for no Γ does there exist a Γ sentence which is nontrivially Γ^d–conservative over every member of the family. [Hint: Let φ be a Π_1 sentence undecidable in PA. Then

$$\{PA + \neg\theta\colon PA \vdash \theta \to \varphi\} \cup \{PA + \theta\colon PA \vdash \varphi \to \theta\}$$

is an r.e. family of extensions of PA. Suppose θ is Π_n and nontrivially Σ_n–conservative over all members of this family. Then PA + $\varphi \nvdash \theta$. θ is Σ_n–conservative over T + $\neg(\theta \wedge \varphi)$. It follows that PA + $\varphi \vdash \theta$, a contradiction. The dual case is similar.]

15. This exercise may be compared with Theorems 2.13, 2.14.

(a) For each Γ, there is a primitive recursive function f such that for every Γ sentence φ, f(φ) is a proof in PA of $\varphi \leftrightarrow Tr_\Gamma(\varphi)$. Use this to show that there is a Γ sentence θ and a primitive recursive function g(k) such that θ is Γ^d–conservative over T and if ψ is any Γ^d sentence and q a proof of ψ in T + θ, then g(q) is a proof of ψ in T.

(b) Let f be any recursive function. Show that there are sentences φ, ψ such that φ is Γ–conservative over T, ψ is Γ, T$\vdash \psi$, and there is a proof p of ψ in T + φ such that q > f(p) for every proof q of ψ in T.

Notes for Chapter 5.

The general concept Γ–*conservative* is due to Guaspari (1979). Theorem 1 is due to Kreisel (1962). Lemma 2 is due to Lindström (1984a). Lemma 3 and Theorem 2 with X = Th(T) are due to Guaspari (1979); for somewhat stronger results, also due to Guaspari (1979), see Exercise 5 (b). The proofs of Lemma 3 and Theorem 2 are from Lindström (1984a). Lemma 4 is due to Lindström (1984a). (Lemmas 2 and 4 and their proofs are similar to and were inspired by results of Guaspari (1979), Solovay (cf. Guaspari (1979)), and Hájek (1971); for further applications, see e.g. Hájek and Pudlák (1993).) Theorem 3 less the references to the set X is due to Solovay (cf. Guaspari (1979); see also Jensen and Ehrenfeucht (1976); the full result is proved in Smoryński (1981a) and Lindström (1984a). The formula $Prf'_{T,\Gamma}(x,y)$ was introduced by Smoryński (1981a); (Sm) and the fixed point mentioned in Exercise 3.7 (a) are special cases of a very general construction due to Smoryński (1981a); however, in the proof of his main theorem Smoryński has to assume that the formulas $\chi_i(x,y)$ are PR. Theorem 4 is due to Lindström (1984a). Lemma 6 and Theorem 5 are due to Bennet (1986), (1986a). Corollary 1 with Σ_n replaced by Π_n is false (Exercise 3). Theorem 6 for $\Gamma = \Pi_1$ and for $\Gamma = \Pi_{n+1}$ are essentially due to Solovay (cf. Hájek (1979)) and Hájek (1979), respectively, (in both cases with different proofs);

Theorem 6 for $\Gamma = \Sigma_n$, $n > 1$, is due to Quinsey (1980), (1981) (with a different proof); the present proof is due to Lindström (1984a). For more information on Cons(Γ,T) and related sets, see Exercises 12 and 13. Lemma 8 is due to Lindström (1993); Lemma 8 with Π_n and Σ_n interchanged and restricted to Σ_n–sound theories is also true but the proof is quite different.

An alternative concept of *partial conservativity* has been introduced and studied by Hájek (1984).

Exercise 2 (a) is due to Smoryński (1980); Exercise 2 (c) is due to Švejdar (cf. Hájek and Pudlák (1993)). Exercise 4 is due to Kent (1973). Exercise 5 (b) is due to Guaspari (1979). Exercise 7 is due to Bennet (1986). Exercise 10 (a) is due to Smoryński (1981a). Exercise 12 is due to Quinsey (1981); the suggested proof is due to Bennet. Exercise 13 (c) is due to Bennet (1986). Exercise 14 is due to Misercque (1983).