# 5. PARTIAL CONSERVATIVITY

A sentence $\varphi$ is *$\Gamma$–conservative over* T if for every $\Gamma$ sentence $\theta$, if $T + \varphi \vdash \theta$, then $T \vdash \theta$. In this chapter we study this phenomenon for its own sake. Results on $\Gamma$–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 $\Pi_1$–conservative sentence is given in the following:

**Theorem 1.** $\neg Con_T$ is $\Pi_1$–conservative over T.

**Proof.** Suppose $\theta$ is $\Pi_1$ 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 $\Sigma_1$–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 ($\Sigma_1$) sentence $\varphi$ which is $\Pi_1$–conservative over T and nontrivially so, i.e. such that $T \nvdash \varphi$, even if T is not $\Sigma_1$–sound.

If $\varphi$ is $\Gamma$–conservative over T and $\psi$ is $\Gamma^d$, then clearly $\varphi$ is $\Gamma$–conservative over $T + \psi$. Also note that if T is $\Sigma_1$–sound and $\pi$ is $\Pi_1$, then $\pi$ is $\Sigma_1$–conservative over T iff $\pi$ is true iff $T + \pi$ is consistent.

Let us now try to construct a sentence $\varphi$ which is nontrivially $\Gamma$–conservative over T. Thus, given that

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

where $\theta$ is $\Gamma$, 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 $\varphi$ be a sentence saying of itself that there is a false $\Gamma$ sentence (namely $\theta$) which $\varphi$ implies in T. Thus, let $\varphi$ 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 $\Gamma$ sentences. Then (1) implies (2).