10.5 Jech's Proof of Gödel's Second Incompleteness Theorem for Set Theory

This section is dedicated to a very short proof of Gödel's Second Incompleteness Theorem in the framework of set theory due to Tom Jech [15]. This proof is short only if one does not count how many pages are needed for the whole coding process that will only be exposed in a few lines.

Theorem 258 (Jech's version of Gödel's second incompleteness theorem)**.** *It is unprovable in set theory — unless it is inconsistent — that there exists a model of set theory.*

Proof of Theorem 258: Towards a contradiction, we assume that set theory is consistent and that it proves that there exists some model of set theory. We also let $\mathcal{T}$ be any finite set of axioms that is large enough to be able to formulate the concepts of "model of a finite theory", "satisfiability", and also to prove the existence of a model of set theory.

From now on, we focus on the finite sub-theory $\mathcal{T}$, and "model" means "model of $\mathcal{T}$".
Given $\mathscr{M} = \langle M, \in_M \rangle, \mathscr{N} = \langle N, \in_N \rangle$ two models of $\mathcal{T}$, we write $\mathscr{M} \prec \mathscr{N}$ if there exists some set $m \in N$ and some E_N such that $\mathscr{N} \models$ "E_N is a binary relation" and $E_N \upharpoonright m \times m = \in_M$.

Notice that if $\mathscr{M} \prec \mathscr{N}$ holds, then for each closed formula[5] φ, we have

$$\mathscr{M} \models \varphi \text{ if and only if } \mathscr{N} \models \text{“} m \models \varphi \text{”}. \tag{10.1}$$

In particular, we have

$$\text{if there exists } \mathscr{M} \prec \mathscr{N}, \text{ then } \mathscr{N} \models \text{“} m \models \mathcal{T} \text{”};$$

and also

$$\text{if there exists } m \in N \text{ s.t. } \mathscr{N} \models \text{“} m \models \mathcal{T} \text{”}, \text{ then there exists some } \mathscr{M} \prec \mathscr{N}.$$

It follows that the relation $\prec$ is transitive. i.e.

$$\mathscr{M}_1 \prec \mathscr{M}_2 \prec \mathscr{M}_3 \Longrightarrow \mathscr{M}_1 \prec \mathscr{M}_3. \tag{10.2}$$

We now consider $(\varphi_n)_{n \in \omega}$ some Gödel numbering of the formulas with one free variable x, and let

$$S_n = \{k \in \omega \mid \varphi_n(k)\}.$$

We set

$$S = \{n \in \omega \mid \text{“there exists a model } \mathscr{M} \text{ such that } \mathscr{M} \models n \notin S_n \text{”}\} \tag{10.3}$$

[5] From the language of set theory.

Since the sentence "there exists a model $\mathscr{M}$ such that $\mathscr{M} \models x \notin S_x$" stands for a formula with one free variable x, there exists some integer k such that

$$\varphi_k(x) := \text{"there exists a model } \mathscr{M} \text{ such that } \mathscr{M} \models x \notin S_x\text{"},$$

hence

$$S = S_k$$

We see that $\mathcal{T}$ proves the following sentence:

$$\text{"}k \in S\text{"} \longleftrightarrow \text{"}\exists \mathscr{M} \quad \mathscr{M} \models k \notin S\text{"} \tag{10.4}$$

If $\mathscr{M}$ is any model (of $\mathcal{T}$), then we have

$$\mathscr{M} \models \text{"}k \in S\text{"} \Longleftrightarrow \exists \mathscr{N} \prec \mathscr{M} \quad \mathscr{N} \models k \notin S \tag{10.5}$$

.

Let us say that $\mathscr{M}$ is *positive* if $\mathscr{M} \models$ "$k \in S$", and *negative* if $\mathscr{M} \not\models$ "$k \in S$".

By 10.5, we have:

$$\text{if } \mathscr{M} \text{ is } \textit{positive}\text{, then there exists some } \textit{negative } \mathscr{N} \prec \mathscr{M}. \tag{10.6}$$

By 10.3, we have:

$$\text{If } \mathscr{M} \text{ is } \textit{negative}\text{, then all } \mathscr{N} \prec \mathscr{M} \text{ are } \textit{positive}. \tag{10.7}$$

Since $\mathcal{T}$ proves that there exists a model of $\mathcal{T}$, for every model $\mathscr{M}$ (of $\mathcal{T}$), we have

- if $\mathscr{M}$ is *negative*, by 10.5 there exists some *positive* $\mathscr{N} \prec \mathscr{M}$, and by 10.5 again, there exists some *negative* $\mathscr{O} \prec \mathscr{N}$. Hence, $\mathscr{O} \prec \mathscr{M}$ holds by 10.2, which contradicts 10.7.
- In case $\mathscr{M}$ is *positive*, by 10.5 there exists some *negative* $\mathscr{N} \prec \mathscr{M}$, which leads to the same contradiction as the previous case.

□ 258