PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 121, Number 1, May 1994

ON GÖDEL'S SECOND INCOMPLETENESS THEOREM

THOMAS JECH

(Communicated by Andreas R. Blass)

ABSTRACT. In this note we give a short proof of Gödel's Second Incompleteness Theorem.

Gödel's Second Incompleteness Theorem states that no sufficiently strong consistent mathematical theory can prove its own consistency [1]. In this note we give a short proof of the theorem.

Theorem. *It is unprovable in set theory (unless it is inconsistent) that there exists a model of set theory.*

Proof. Assume that set theory is consistent and that it proves that a model of set theory exists. Let Σ be a finite set of axioms sufficiently strong to formulate the concepts "model" and "satisfies" and to prove the existence of a model of set theory. For the rest of the proof, a *model* means a model of Σ and letters M and N denote models. If m is a set with a binary relation, $\in^m$ denotes that relation. If $N \models (E \text{ is a relation})$, then E^* is the relation consisting of all pairs (x, y) such that $N \models xEy$.

If M and N are models we define

$$M < N \quad \text{if there exists some } m \in N \text{ such that } \in^M = (\in^m)^*.$$

(Informally, $M < N$ means that M is, in the real world, the structure that N thinks m is.) If $M < N$ then for every sentence σ

$$(1) \qquad M \models \sigma \quad \text{if and only if} \quad N \models (m \models \sigma).$$

In particular, $N \models (m \text{ is a model})$. Also, if $N \models (m \text{ is a model})$, then $(\in^m)^*$ is the $\in$-relation of some model $M < N$. It follows that

$$(2) \qquad \text{if } M_1 < M_2 \text{ and } M_2 < M_3 \text{ then } M_1 < M_3.$$

Let us consider some fixed coding of formulas by numbers (Gödel numbering), and let S_n be the name for the nth definable set of numbers.

Received by the editors September 29, 1992 and, in revised form, November 30, 1992.
1991 *Mathematics Subject Classification.* Primary 03B99, 03C99, 03E99, 03F99.
Supported in part by NSF grant DMS–8918299.

Definition. S is the set of all numbers n with the property that there is a model M such that $M \models n \notin S_n$.

Let k be the Gödel number of S, and let A be the sentence "$k \in S$". Then the following equivalence is provable in Σ:

$$A \leftrightarrow \exists M (M \models \neg A). \tag{3}$$

By (1) if M is any model then

$$M \models A \leftrightarrow \exists N < M (N \models \neg A). \tag{4}$$

We say that M is *positive* if $M \models A$ and *negative* otherwise. As a consequence of (4), if M is negative then all $N < M$ are positive.

Since Σ is consistent and proves that a model exists, we have

(5) there exists a model

and also (using (1))

(6) for every model M there exists a model $N < M$.

Toward a contradiction, let M_1 be a model by (5). If M_1 is positive, there is, by (4), a negative model $M_2 < M_1$; otherwise let $M_2 = M_1$. By (6) there is a model $M_3 < M_2$, and since M_2 is negative, M_3 is positive. By (4) there is a negative $M_4 < M_3$, and we have $M_4 < M_2$ by (2), a contradiction.

Remark 1. The sentence A in (3) is the analog of Gödel's "I am unprovable." Another way to obtain A is as follows: A *property* is a formula of the language of set theory with one free variable. Let p be the property (of properties q) $\exists M\, M \models \neg q(q)$, and let $A = p(p)$. Then (3) holds.

Remark 2. Even though our proof of Gödel's Theorem uses the Completeness Theorem, it can be modified to apply to weaker theories such as Peano Arithmetic (PA). To prove that PA does not prove its own consistency (unless it is inconsistent), we argue as follows:

Assume that PA is consistent and that "PA is consistent" is provable in PA. There is a conservative extension Γ of PA in which the Completeness Theorem is provable [2, Theorem 5.5, p. 443], and moreover, PA $\vdash$ (Γ is a conservative extension of PA). Therefore, $\Gamma \vdash$ (Γ is a conservative extension of a consistent theory) and thus proves its own consistency. Consequently, Γ proves that Γ has a model.

Now let Σ be a sufficiently strong finite subset of Γ that proves that Σ has a model; the proof above leads to a contradiction.

Remark 3. In [3] P. Vopěnka proved Gödel's Theorem for set theory using Richard's Paradox ("the least number not definable in 1000 words").

Acknowledgment

I am grateful to Andreas Blass for comments and improvements.

References

1. Kurt Gödel, *Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme* I, Monatsh. Math. **38** (1931), 173–198.
2. Gaisi Takeuti, *Proof theory*, 2nd ed., Appendix by S. Simpson, North-Holland, Amsterdam, 1987.
3. Petr Vopěnka, *A new proof of Gödel's result on non-provability of consistency*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **14** (1966), 111–116.

Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania 16802

E-mail address: jech@math.psu.edu