Notre Dame Journal of Formal Logic
Volume 35, Number 3, Summer 1994

On Formalization of Model-Theoretic Proofs of Gödel's Theorems

MAKOTO KIKUCHI and KAZUYUKI TANAKA

Abstract Within a weak subsystem of second-order arithmetic WKL_0, that is Π^0_2-conservative over PRA, we reformulate Kreisel's proof of the Second Incompleteness Theorem and Boolos' proof of the First Incompleteness Theorem.

1 Introduction To bypass the pedagogic obstacles in Gödel's original proof of the Incomplete Theorems, model-theoretic methods have been invented by several people (cf. Smoryński [7], Boolos [1], and Kreisel [4]). However, such semantic proofs usually need such strong assumptions in metamathematics that they do not readily yield the formalized versions. For instance, as Smoryński [7] points out, Kreisel's proof of the Second Incompleteness Theorem in [4] does not directly lead to the formalized Second Incompleteness Theorem: $PRA \vdash Con(PA) \rightarrow Con(PA + \neg Con(PA))$.

In this paper, we elucidate the mechanism of such semantic proofs and show that they can be carried out within a subsystem of second-order arithmetic WKL_0. Thus, we also obtain proofs of the Formalized Incompleteness Theorems in PRA, since any Π^0_2 theorem of WKL_0 is provable in PRA by a result of Friedman.

In most semantic proofs, the arithmetized version of the Completeness Theorem is repeatedly used. Although one can easily show in WKL_0 that any countable consistent theory has a model (see Theorem 2.2 below), what we really need for the proofs is not just the existence of such a model but a specific construction. Typically, a model can be constructed by taking the leftmost path through a certain binary branching tree, but, as shown in Section 3, we cannot prove the existence of such a path in WKL_0. Instead, our Theorem 5.5 and Corollary 5.6 jointly assert over RCA_0 that every model $\mathfrak{M}$ of $PA + Con(PA)$ has a definable end-extension $\mathfrak{N}$ which is also a model of PA (cf. Feferman [2], and Lemma 6.2.3 of Smoryński [6]).

In Section 2, we define the system WKL_0 and set up the basics of predicate calculus within WKL_0. In Section 3, we define the system ACA_0 and prove that the existence of the leftmost paths of binary branching trees is equivalent to ACA_0. In

Received June 11, 1993; revised November 14, 1994

Section 4, we introduce the derivability conditions, i.e., the three basic properties of the provability predicate, and give model-theoretic proofs of them in WKL_0. Section 5 is devoted to prove our version of the Arithmetized Completeness Theorem, i.e., Theorem 5.5 and Corollary 5.6. We apply these theorems to formalize Kreisel's proof in Section 6, and also to extract the First Incompleteness Theorem from Berry's paradox in Section 7, which strengthens the argument of Boolos [1]. See Kikuchi [3] for more information on the latter application.

2 The System WKL_0 As usual, first-order arithmetic is formalized in the language $\mathcal{L}_1$ with the symbols $+, \cdot, 0, 1, <$, and second order arithmetic is in the language $\mathcal{L}_2 = \mathcal{L}_1 \cup \{\in\}$. Variables $x, y, z, \ldots$ range over the natural numbers, and $X, Y, Z, \ldots$ over the sets of natural numbers. Terms and formulas are built up in the usual way. An $\mathcal{L}_2$-formula with no unbounded quantifiers is called a Σ^0_0 or Π^0_0 formula. If φ is a Σ^0_j formula, $(\forall x_1 \ldots \forall x_k)\varphi$ is a Π^0_{j+1} formula. If φ is a Π^0_j formula, $(\exists x_1 \ldots \exists x_k)\varphi$ is a Σ^0_{j+1} formula. Notice that a Σ^0_j or Π^0_j formula with no set parameters must be an $\mathcal{L}_1$-formula.

The system RCA_0, which stands for *Recursive Comprehension Axiom*, is an $\mathcal{L}_2$-theory consisting of the following axioms:

(I) Basic axioms of arithmetic, namely the axioms of discretely ordered semi-rings with the least positive element 1.

(II) Σ^0_1 induction: $\varphi(0) \wedge (\forall x)(\varphi(x) \to \varphi(x+1)) \to (\forall x)\varphi(x)$, where $\varphi(x)$ is a Σ^0_1 formula.

(III) Recursive comprehension: $(\forall n)(\varphi(n) \leftrightarrow \psi(n)) \to (\exists X)(\forall n)(\varphi(n) \leftrightarrow n \in X)$, where φ is a Σ^0_1 formula and ψ a Π^0_1 formula.

Within RCA_0, we can prove the existence and uniqueness of such a set X that $(\forall x)(x = x \leftrightarrow x \in X)$, and we write this set as $\tilde{\mathbf{N}}$.

By Seq_2, we denote the set of codes for finite sequences of zeros and ones. Weak König's Lemma (WKL) is an assertion that every infinite subtree of Seq_2 has an infinite path. The $\mathcal{L}_2$-theory WKL_0 is obtained from RCA_0 by adding WKL. Since WKL brings forth the compactness of the closed unit interval and other spaces, mathematics developed in WKL_0 is much richer than that in RCA_0. However, concerning the consistency issue, these two systems are equivalent to each other. In fact, Harrington shows that they have the same Σ^0_j theorems for any j (cf. Simpson [5]).

Friedman was the first to observe that if a Π^0_2 sentence is provable in WKL_0, it is also provable in PRA. Here, PRA is the system which has symbols for the primitive recursive functions and whose axioms contain all the defining equations for primitive recursive functions and induction for quantifier-free formulas.

Theorem 2.1 (Friedman's Conservation Theorem) $WKL_0 \vdash \varphi$ *implies* $PRA \vdash \varphi$ *for every* Π^0_2 *sentence* φ *in* $\mathcal{L}_1$.

Proof: As Simpson and Tanaka [6] point out, Friedman's Theorem itself can be proved within WKL_0, hence it is also provable in PRA by its own assertion.

Now, we define a basic notion of predicate calculus in RCA_0. We first fix a countable first-order language $\mathcal{L}$. We may assume that all the meaningful expressions such as terms and formulas in $\mathcal{L}$ are encoded by natural numbers under a standard Gödel numbering $\ulcorner\ \urcorner$ so that the basic operations on the expressions in $\mathcal{L}$ (e.g., $disj(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) = \ulcorner\varphi \vee \psi\urcorner$) are represented by primitive recursive functions. Let $\mathcal{L}\text{-}trm(x)$, $\mathcal{L}\text{-}fml(x)$, $\mathcal{L}\text{-}snt(x)$ be formulas in $\mathcal{L}_1$ which express that x is the Gödel number of an $\mathcal{L}$-term, of an $\mathcal{L}$-formula, or of an $\mathcal{L}$-sentence, respectively. A Σ^0_1-formula $T(x)$ is called an $\mathcal{L}$-*theory* if RCA_0 proves $(\forall x)(T(x) \to \mathcal{L}\text{-}snt(x))$. For an $\mathcal{L}$-theory T, we can define an $\mathcal{L}_1$-formula $Proof_T(x, y)$ which means that y encodes a proof of the formula with code x in the theory T. Then we define two $\mathcal{L}_1$-formulas $Pr_T(x)$ and $Con(T)$ by

$$Pr_T(x) \leftrightarrow (\exists y)Proof_T(x, y),$$
$$Con(T) \leftrightarrow \neg Pr_T(\ulcorner\bot\urcorner),$$

where $\bot$ denote a contradiction such as $\neg(\forall x)(x = x)$. Notice that $Pr_T(x)$ is a Σ^0_1 formula, and $Con(T)$ is a Π^0_1 sentence.

To define models, take a set $C = \{c_n\}$ of new constants and let $\mathcal{L}' = \mathcal{L} \cup C$. We can safely assume that $\mathcal{L}'$ has a Gödel numbering $\ulcorner\urcorner$ which coincides with the Gödel numbering of $\mathcal{L}$ on $\mathcal{L}$-terms and $\mathcal{L}$-formulas. Let T be an $\mathcal{L}$-theory. A set $\mathfrak{M}$ of (the Gödel numbers of) $\mathcal{L}'$-sentences is called a *model* of T if it satisfies $(\forall x)(T(x) \to x \in \mathfrak{M})$ and the following Tarski clauses:

$$\neg\varphi \in \mathfrak{M} \leftrightarrow \varphi \notin \mathfrak{M},$$
$$(\varphi \wedge \psi) \in \mathfrak{M} \leftrightarrow (\varphi \in \mathfrak{M} \wedge \psi \in \mathfrak{M}),$$
$$(\varphi \vee \psi) \in \mathfrak{M} \leftrightarrow (\varphi \in \mathfrak{M} \vee \psi \in \mathfrak{M}),$$
$$(\varphi \to \psi) \in \mathfrak{M} \leftrightarrow (\varphi \in \mathfrak{M} \to \psi \in \mathfrak{M}),$$
$$(\exists x)\theta(x) \in \mathfrak{M} \leftrightarrow (\exists c \in C)(\theta(c) \in \mathfrak{M}),$$
$$(\forall x)\theta(x) \in \mathfrak{M} \leftrightarrow (\forall c \in C)(\theta(c) \in \mathfrak{M}).$$

where φ, ψ and $\theta(c)$ are $\mathcal{L}'$-sentences. In other words, a model of T is the elementary diagram of an $\mathcal{L}$-structure with universe C which satisfies T.

Theorem 2.2 (The Completeness Theorem (WKL_0)) *The following are equivalent:*

(i) $Con(T)$,

(ii) there is a model $\mathfrak{M}$ *of* T.

Proof: The theorem can be proved by the usual Henkin construction. See [5] for the details. It is also known that this theorem turns out to be equivalent to WKL_0 over RCA_0.

Using the derivability conditions explained in the next section, we can also easily obtain the following version of the Completeness Theorem.

Theorem 2.3 (The Completeness Theorem; revised (WKL_0)) *For any $\mathcal{L}$-sentence φ, T proves φ if and only if φ holds in every model of T.*

Proof: The left-to-right direction of the theorem is easily proved by induction on the length of proofs. We will prove the other direction by showing its contraposition. Assume that T does not prove φ. Then $T \cup \{\neg\varphi\}$ is consistent (by derivability condition D2), so there is a model $\mathfrak{M}$ of $T \cup \{\neg\varphi\}$.

***3 Binary Branching Trees and** ACA_0* In this section, we shall show that the existence of the leftmost paths of binary branching trees is equivalent to ACA_0. Except for the definition of a binary relation $\prec$ over Seq_2, the contents of this section will not be used in the succeeding sections.

The system ACA_0, which stands for *Arithmetical Comprehension Axiom*, is the theory in $\mathcal{L}_2$ obtained from RCA_0 by adding

Arithmetical Comprehension: $(\exists X)(\forall x)(\varphi(x) \leftrightarrow x \in X)$, where φ is any $\mathcal{L}_2$ formula which does not contain quantifiers for set variables.

It is known that, over RCA_0, ACA_0 is equivalent to König's Lemma, which asserts that *every finitely branching infinite tree has an infinite path*, and strictly stronger than WKL_0 (see Simpson [5]). The following lemma is often useful in showing that arithmetical comprehension is needed to prove various theorems.

Lemma 3.1 (RCA_0) *The following are equivalent:*

(i) ACA_0.
(ii) For any one-to-one function $f : \tilde{\mathbf{N}} \to \tilde{\mathbf{N}}$, *there exists a set* $X \subseteq \tilde{\mathbf{N}}$ *such that* $n \in X$ *if and only if* $n = f(m)$ *for some* $m \in \tilde{\mathbf{N}}$.

Proof: See Simpson [5].

Now, we introduce a binary relation $\prec$ on Seq_2 as follows:

$$x \prec y \quad \leftrightarrow \quad \text{there exists } n < \min\{lh(x), lh(y)\} \text{ such that } (\forall m < n)((x)_m = (y)_m) \wedge (x)_n = 0 \wedge (y)_n = 1,$$

where $lh(x)$ is the length of x and $(x)_n$ is the nth element of x. Namely, if $x \prec y$ then x occurs on the left of y in the tree Seq_2.

Let $T \subseteq Seq_2$ be an infinite tree. We say an infinite path $P \subseteq T$ is the leftmost infinite path of T if for any infinite path Q of T with $P \neq Q$, there exists $s \in P$ and $t \in Q$ such that $s \prec t$.

Theorem 3.2 (RCA_0) *The following are equivalent:*

(i) ACA_0.
(ii) For any infinite set $X \subseteq Seq_2$, *there exists the minimal tree* $T \subseteq Seq_2$ *such that* $X \subseteq T$.
(iii) Every infinite tree $T \subseteq Seq_2$ *has the leftmost infinite path.*

Proof: (i) $\Rightarrow$ (ii): Let X be an infinite subset of Seq_2. Then define a set $T \subseteq \tilde{\mathbf{N}}$ by $x \in T \Leftrightarrow x \in Seq_2 \wedge (\exists y)(y \in X \wedge y \upharpoonright lh(x) = x)$ for all $x \in \tilde{\mathbf{N}}$. This set exists by arithmetical comprehension. It is clear that T is a tree which satisfies the condition (ii).

(ii) $\Rightarrow$ (iii): Let $T \subseteq Seq_2$ be an infinite tree. Define a set X by $x \in X \Leftrightarrow x \in T \wedge (\forall y)(y \prec x \wedge lh(x) = lh(y) \to y \notin T)$ for all $x \in \tilde{\mathbf{N}}$. This set exists by the recursive comprehension. Then, by (ii), we can find the minimal tree U which contains X. Now, define a set P by $x \in P \Leftrightarrow x \in U \wedge (\forall y)(x \prec y \wedge lh(x) = lh(y) \to y \notin U)$ for all $x \in \tilde{\mathbf{N}}$. This set P is the leftmost infinite path of the tree T.

(iii) $\Rightarrow$ (i): Instead of proving ACA_0 directly, we prove the existence of images of one-to-one functions. Let $f : \tilde{\mathbf{N}} \to \tilde{\mathbf{N}}$ be a one-to-one function. Define a set T by $x \in T \Leftrightarrow x \in Seq_2 \wedge (\forall y < lh(x))\{(\exists z < lh(x))(f(z) = y) \to (x)_y = 1\}$ for all $x \in \tilde{\mathbf{N}}$. This set T exists by the recursive comprehension. Obviously, T forms an infinite tree. Let P be the leftmost infinite path of T. Then, x is an element of the image of f if and only if there is an element t of P such that $lh(t) = x + 1$ and $(t)_x = 1$. So the image of f exists by the recursive comprehension.

4 The Derivability Conditions It was Bernays who first exposed the machinery of the proof of the Second Incompleteness Theorem by listing the properties of the provability predicate, now called the Hilbert-Bernays Derivability Conditions. Then Löb simplified those conditions, and showed that the Second Incompleteness Theorem can be obtained from the following three assertions (with $S = T = PA$):

D1. $T \vdash \varphi$ implies $S \vdash Pr_T(\ulcorner\varphi\urcorner)$,
D2. $S \vdash Pr_T(\ulcorner\varphi\urcorner) \wedge Pr_T(\ulcorner\varphi \to \psi\urcorner) \to Pr_T(\ulcorner\psi\urcorner)$,
D3. $S \vdash Pr_T(\ulcorner\varphi\urcorner) \to Pr_T(\ulcorner Pr_T(\ulcorner\varphi\urcorner)\urcorner)$.

D1 is easily proved by induction on the length of a proof of φ in T. D2 is a formalization of modus ponens, and it is proved by showing that from proofs of φ and $\varphi \to \psi$, one can easily construct a proof of ψ, that is, the concatenation of the former two proofs plus modus ponens. D3 is a formalization of D1. But to formalize the syntactical proof of D1, one needs a lengthy argument. So, we will instead formalize the following semantic proof of D1 in the case $S = PRA$ (or a reasonably strong subsystem of PA) and T = any theory.

Proof Proof of D1: Assume $T \vdash \varphi$. Then, $Pr_T(\ulcorner\varphi\urcorner)$ holds in the standard model $\mathbb{N}$ of arithmetic. Let $\mathfrak{M}$ be any model of S. As $\mathfrak{M}$ is an end-extension of $\mathbb{N}$, every Σ^0_1 sentence true in $\mathbb{N}$ is also true in $\mathfrak{M}$. So $\mathfrak{M} \models Pr_T(\ulcorner\varphi\urcorner)$. By the Completeness Theorem, we have $S \vdash Pr_T(\ulcorner\varphi\urcorner)$.

In the rest of this section, we prove D3 in the case $S = PRA$ and $T = PA$ by formalizing the above proof in WKL_0.

Let $\mathfrak{M}$ be a model of $T = PA$ with universe $C = \{c_n\}$. Then, by the definition of models, there exist c_{n_0} and c_{n_1} in C such that $\mathfrak{M} \models 0 = c_{n_0}$ and $\mathfrak{M} \models 1 = c_{n_1}$. We denote such elements c_{n_0} and c_{n_1} in C by $0_{\mathfrak{M}}$ and $1_{\mathfrak{M}}$, respectively.

The next lemma claims that $\mathfrak{M}$ includes $\tilde{\mathbf{N}}$ as an initial segment.

Lemma 4.1 (RCA_0) *There exists a function $e_{\mathfrak{M}} : \tilde{\mathbf{N}} \to C$ which satisfies the following conditions; for each $i, j \in \tilde{\mathbf{N}}$,*

(i) $e_{\mathfrak{M}}(0) = 0_{\mathfrak{M}}, e_{\mathfrak{M}}(1) = 1_{\mathfrak{M}}$,
(ii) $\mathfrak{M} \models e_{\mathfrak{M}}(i + j) = e_{\mathfrak{M}}(i) + e_{\mathfrak{M}}(j), \mathfrak{M} \models e_{\mathfrak{M}}(i \cdot j) = e_{\mathfrak{M}}(i) \cdot e_{\mathfrak{M}}(j)$,
(iii) $i \neq j$ *implies* $\mathfrak{M} \models e_{\mathfrak{M}}(i) \neq e_{\mathfrak{M}}(j)$,
(iv) *if* $\mathfrak{M} \models c_n < e_{\mathfrak{M}}(i)$, *then there exists* $k \in \tilde{\mathbf{N}}$ *such that* $\mathfrak{M} \models c_n = e_{\mathfrak{M}}(k)$.

Proof: First, we define the successor function $s \subseteq C^2$ by $(c_m, c_n) \in s$ if and only if $(\mathfrak{M} \models c_m + 1 = c_n) \wedge (\forall l < n)(\mathfrak{M} \models c_m + 1 \neq c_l)$. The existence of s is obvious from recursive comprehension, and clearly s can be viewed as a function $s : C \to C$ such that $\mathfrak{M} \models s(c) = c + 1$ for all $c \in C$.

Then, we define a desired function $e_{\mathfrak{M}} : \tilde{\mathbf{N}} \to C$ by primitive recursion:

$$e_{\mathfrak{M}}(0) = 0_{\mathfrak{M}}, \quad e_{\mathfrak{M}}(i+1) = s(e_{\mathfrak{M}}(i)).$$

Clearly, this function $e_{\mathfrak{M}}(x)$ satisfies the condition (i). The condition (ii) can be easily proved by induction on j. The condition (iii) is an immediate consequence of the conditon (ii). The condition (iv) is proved by induction on i.

The next theorem is a semantical version of D3, and it also plays an important role in the following sections.

Theorem 4.2 (RCA_0) *For any model $\mathfrak{M}$ of PA, $(\forall x_1 \dots \forall x_n)(\varphi(x_1, \dots, x_n) \to (\mathfrak{M} \models \varphi(e_{\mathfrak{M}}(x_1), \dots, e_{\mathfrak{M}}(x_n))))$ for every Σ^0_1 formula $\varphi(x_1, \dots, x_n)$.*

Proof: It can be readily shown by induction on the complexity of formula φ, since $\mathfrak{M}$ obeys the Tarski clauses.

From the above theorem, we also have the following version of D3.

Corollary 4.3 *For any Σ^0_1 sentence σ, $PRA \vdash \sigma \to Pr_{PA}(\ulcorner \sigma \urcorner)$.*

Proof: Let σ be a Σ^0_1 sentence. By the above theorem, $RCA_0 \vdash (\sigma \to \mathfrak{M} \models \sigma)$ for any model $\mathfrak{M}$ of PA. So, $WKL_0 \vdash \sigma \to Pr_{PA}(\ulcorner \sigma \urcorner)$ by Theorem 2.3. As $\sigma \to Pr_{PA}(\ulcorner \sigma \urcorner)$ is a Π^0_2 sentence, we have $PRA \vdash \sigma \to Pr_{PA}(\ulcorner \sigma \urcorner)$ from Theorem 2.1.

In comparison with usual syntactic arguments for D3 as in 3.2.5 for D3 of [7], the mechanism of induction in our proof is considerably simpler because models satisfy the Tarski clauses whereas the provability predicate does not.

5 The Arithmetized Completeness Theorem Fix a countable first-order language $\mathcal{L}$. Let $T(x)$ be an $\mathcal{L}_1$-formula such that PA proves $(\forall x)(T(x) \to \mathcal{L}\text{-}snt(x))$. The Arithmetized Completeness Theorem asserts that over $PA + Con(T)$, we can construct an $\mathcal{L}_1$-formula $Tr_T(x)$ which expresses validity in a model of T. This version of the Completeness Theorem is due to Feferman [2].

To begin with, we will introduce a Henkin theory T_H of T. Take a set $C = \{c_n\}$ of new constants and put $\mathcal{L}' = \mathcal{L} \cup C$. The $\mathcal{L}'$-theory T_H is obtained from T by adding all the Henkin axioms. We here assume that the theory T_H is expressed by an $\mathcal{L}_1$-formula $T_H(x)$ in PA. As usual, we have $PA \vdash Con(T) \to Con(T_H)$, since we can easily rewrite a proof of inconsistency in T_H (if any) to a proof of it in T.

We now define an $\mathcal{L}_1$-formula $cns_T(x)$ as follows:

Definition 5.1 $cns_T(x) \leftrightarrow (x \in Seq_2) \wedge (\{\varphi \mid \mathcal{L}'\text{-}snt(\ulcorner\varphi\urcorner) \wedge \ulcorner\varphi\urcorner < lh(x) \wedge (x)_{\ulcorner\varphi\urcorner} = 0\}$ is consistent with $T_H)$.

Then $cns_T(x)$ can be seen as a certain subtree of Seq_2. Next, we set

$$lft_T(x) \leftrightarrow cns_T(x) \wedge (\forall y)(y \prec x \to \neg cns_T(y))$$

and

$$Tr_T(x) \leftrightarrow (\exists y)(lft_T(y) \wedge lh(y) = x + 1 \wedge (y)_x = 0).$$

We are going to show in $PA + Con(T)$ that $Tr_T(x)$ defines a model of T.

Lemma 5.2 $PA + Con(T) \vdash (\forall x)(\exists! y)(lh(y) = x \wedge lft_T(y))$.

Proof: We prove this lemma by induction on x. The case $x = 0$ is clear because $Con(T)$ guarantees that the null sequence satisfies lft_T. For $x = n$, assume that there exists a unique m such that $lh(m) = n$ and $lft_T(m)$. If n is the Gödel number of an $\mathcal{L}'$-sentence ψ and if the theory $\{\varphi \mid \mathcal{L}'\text{-}snt(\ulcorner\varphi\urcorner) \wedge \ulcorner\varphi\urcorner < n \wedge (m)_{\ulcorner\varphi\urcorner} = 0\} \cup \{\psi\}$ is shown in $PA + Con(T)$ to be consistent with T_H, then we put $m' = m^\frown\langle 0\rangle$. Otherwise, put $m' = m^\frown\langle 1\rangle$. Clearly, m' works for the case $x = n + 1$.

For an $\mathcal{L}_1$-formula $\varphi(x)$, we say that $\varphi(x)$ defines a model of T, if it satisfies $(\forall x)(T(x) \to \varphi(x))$ and the Tarski clauses, i.e., the definition of model $\mathfrak{M}$ (in Section 2) with $x \in \mathfrak{M}$ replaced by $\varphi(x)$. Then let $Mod_T(\varphi)$ be an $\mathcal{L}_1$-sentence expressing that φ defines a model of T.

Lemma 5.3 (The Arithmetized Completeness Theorem) $PA + Con(T) \vdash Mod_T(Tr_T)$.

Proof: From the definition of Tr_T, it is obvious that $(\forall x)(T(x) \to Tr_T(x))$ holds in $PA + Con(T)$. With the help of Lemma 5.1, it is also easy to prove in $PA + Con(T)$ that $Tr_T(x)$ obeys the Tarski clauses.

By applying D1 to the above theorem, we have

Lemma 5.4 $PRA \vdash Pr_{PA+Con(T)}(Mod_T(Tr_T))$.

From this lemma, we obtain

Theorem 5.5 (RCA_0) *For any model $\mathfrak{M}_0$ of $PA + Con(T)$, there exists a model $\mathfrak{M}_1$ of T such that $\mathfrak{M}_0 \models Tr_T(\ulcorner\varphi\urcorner)$ if and only if $\mathfrak{M}_1 \models \varphi$ for any $\mathcal{L}$-sentence φ.*

Proof: We define a set $\mathfrak{M}_1$ by $n \in \mathfrak{M}_1 \leftrightarrow \mathfrak{M}_0 \models Tr_T(e_{\mathfrak{M}_0}(n))$ for all $n \in \tilde{\mathbf{N}}$. Since $\mathfrak{M}_0$ is essentially the set of the Gödel numbers of sentences true in $\mathfrak{M}_0$ and the function which maps each n to the Gödel number of the formula $Tr_T(e_{\mathfrak{M}_0}(n))$ is recursive in $\mathfrak{M}_0$, $\mathfrak{M}_1$ is recursive in $\mathfrak{M}_0$. So the existence of $\mathfrak{M}_1$ is provable in RCA_0 although $\mathfrak{M}_1$ is defined by taking the leftmost path of a binary branching tree. We shall show in RCA_0 that $\mathfrak{M}_1$ forms a model of T.

Suppose we have partial functions neg, $disj$, etc., from $\tilde{\mathbf{N}}$ to $\tilde{\mathbf{N}}$ which are such that $neg(\ulcorner\varphi\urcorner) = \ulcorner\neg\varphi\urcorner$, $disj(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) = \ulcorner\varphi \vee \psi\urcorner$, etc. Since these functions can be defined by Σ^0_1 formulas, we have $(\forall m)(\forall n)(m = neg(n) \to \mathfrak{M}_0 \models e_{\mathfrak{M}_0}(m) = neg(e_{\mathfrak{M}_0}(n)))$, etc. by Theorem 4.2. Since $\mathfrak{M}_0$ is a model of $PA + Con(T)$, $\mathfrak{M}_0 \models Mod_T(Tr_T)$ by Lemma 5.4. So we have $(\forall x)\mathfrak{M}_0 \models (Tr_T(neg(x)) \leftrightarrow \neg Tr_T(x))$, etc., and hence $(\forall x)\mathfrak{M}_0 \models (Tr_T(e_{\mathfrak{M}_0}(neg(x))) \leftrightarrow \neg Tr_T(e_{\mathfrak{M}_0}(x)))$, etc., showing that $\mathfrak{M}_1$ obeys the Tarski clauses.

Since $(\forall x)(Pr_T(x) \to \mathfrak{M}_0 \models Pr_T(e_{\mathfrak{M}_0}(x)))$ holds by Theorem 3.2, we have, with Lemma 4.3 again, $(\forall x)(Pr_T(x) \to \mathfrak{M}_0 \models Tr_T(e_{\mathfrak{M}_0}(x)))$. So $(\forall x)(Pr_T(x) \to x \in \mathfrak{M}_1)$. This concludes that $\mathfrak{M}_1$ is a model of T.

Corollary 5.6 (RCA_0) *Let $\mathfrak{M}_0$ and $\mathfrak{M}_1$ be models defined in Theorem 5.5 with $T = PA$. Then $\mathfrak{M}_0 \models \varphi \to \mathfrak{M}_1 \models \varphi$ for any Σ^0_1 sentence φ.*

Proof: Suppose that φ is a Σ^0_1 sentence and $\mathfrak{M}_0 \models \varphi$. By Theorem 4.2 and Corollary 4.3, we have $Pr_{PA}(\ulcorner\varphi \to Pr_{PA}(\ulcorner\varphi\urcorner)\urcorner)$. Since $\mathfrak{M}_0$ is a model of PA, $\mathfrak{M}_0 \models (\varphi \to Pr_{PA}(\ulcorner\varphi\urcorner))$, thus $\mathfrak{M}_0 \models Pr_{PA}(\ulcorner\varphi\urcorner)$. So, by Lemma 5.4, $\mathfrak{M}_0 \models Tr_{PA}(\ulcorner\varphi\urcorner)$, hence $\mathfrak{M}_1 \models \varphi$.

6 Kreisel's Proof of the Second Incompleteness Theorem Now we are ready to carry out Kreisel's model-theoretic proof of the Second Incompleteness Theorem in WKL_0. For a concise account of Kreisel's proof, see Smoryński [7].

Theorem 6.1 (The Diagonalization Lemma) *For any $\mathcal{L}_1$-formula $\varphi(x)$ with at most one free variable x, there is an $\mathcal{L}_1$-sentence σ such that $PA \vdash \sigma \leftrightarrow \varphi(\ulcorner \sigma \urcorner)$.*

Theorem 6.2 (The Second Incompleteness Theorem (WKL_0)) *$Con(PA)$ implies $\neg Pr_{PA}(\ulcorner Con(PA) \urcorner)$.*

Proof: First of all, from Theorem 6.1, we obtain an $\mathcal{L}_1$-sentence σ_0 such that $PA \vdash \sigma_0 \leftrightarrow \neg Tr_{PA}(\ulcorner \sigma_0 \urcorner)$. Fix such a σ_0 and let n_0 be the numeral denoting $\ulcorner \sigma_0 \urcorner$. Now reasoning within WKL_0, assume $Con(PA)$ and $Pr_{PA}(\ulcorner Con(PA) \urcorner)$ by way of contradiction. By Theorem 2.2, there is a model $\mathfrak{M}$ of PA. Since $Con(PA)$ holds in any model of PA, by the repeated use of Theorem 5.5, we can construct a sequence of models of PA, $\mathfrak{M} = \mathfrak{M}_0, \mathfrak{M}_1, \dots, \mathfrak{M}_{2^{n_0+1}}$ such that, for each $i < 2^{n_0+1}$, $\mathfrak{M}_i \models Tr_T(\ulcorner \varphi \urcorner)$ if and only if $\mathfrak{M}_{i+1} \models \varphi$ for any $\mathcal{L}_1$-sentence φ. Then, for each $i \leq 2^{n_0+1}$, we take a numeral m_i such that $\mathfrak{M}_i \models lft_{PA}(m_i) \wedge lh(m_i) = n_0 + 1$. The existence of such m_i's follows from Lemma 5.2 and a simple fact in WKL_0 that there are no more than 2^{n_0+1} sequences from $\{0, 1\}$ with length $n_0 + 1$.

Next we want to show that $m_i \prec m_{i+1}$. Since $\mathfrak{M}_j \models \sigma_0 \leftrightarrow \mathfrak{M}_{j+1} \models \neg\sigma_0$ by the definition of σ_0, clearly we have $m_i \neq m_{i+1}$. So it suffices to show $\neg(m_{i+1} \prec m_i)$. By way of contradition, assume that $m_{i+1} \prec m_i$. Then by Theorem 4.5, $\mathfrak{M}_i \models m_{i+1} \prec m_i$. Hence $\mathfrak{M}_i \models \neg cns_{PA}(m_{i+1})$ by the choice of m_i. Since $\neg cns_{PA}(m_{i+1})$ is a Σ^0_1 sentence, it would follow from Corollary 5.6 that $\mathfrak{M}_{i+1} \models \neg cns_{PA}(m_{i+1})$, which contradicts the choice of m_{i+1}. Thus, we have $m_i \prec m_{i+1}$ within WKL_0.

Now, we obtain a chain $m_0 \prec m_1 \prec \cdots \prec m_{2^{n_0+1}}$, which contradicts the fact that the number of finite sequences from $\{0, 1\}$ with length $n_0 + 1$ is 2^{n_0+1}.

Finally, we have the following theorem by virtue of Theorem 2.1.

Theorem 6.3 (Formalized Second Incompleteness Theorem) *$PRA \vdash Con(PA) \to \neg Pr_{PA}(\ulcorner Con(PA) \urcorner)$.*

7 Berry's Paradox and the First Incompleteness Theorem *The least integer not nameable in fewer than nineteen syllables* is named in eighteen syllables. This is a paradox due to Berry. In [1], Boolos obtained an interesting proof of the First Incompleteness Theorem based on this paradox. Inspired by his proof, we extract the First Incompleteness Theorem from Berry's paradox in WKL_0, hence in PRA by Theorem 2.1.

An $\mathcal{L}_1$-formula is called an *m-formula* ($m \in \tilde{\mathbf{N}}$) if it consists of fewer than m symbols, and if its variables are taken from $v_0, v_1, \dots, v_{m-1}$ and only v_0 may occur freely in it. Let $F(x, y)$ denote an $\mathcal{L}_1$-formula expressing that x is the Gödel number of a y-formula. We say that an m-formula $\varphi(v_0)$ names a number $n \in \tilde{\mathbf{N}}$ in a model $\mathfrak{M}$ of PA if

$$\mathfrak{M} \models (\forall v_0)(v_0 = \overbrace{1 + \cdots + 1}^{n \text{ times}} \leftrightarrow \varphi(v_0)).$$

For the sake of convenience, we write *name* (g, n) for the formula

$$(\forall v_0)(v_0 = \overbrace{1 + \cdots + 1}^{n \text{ times}} \leftrightarrow \varphi(v_0))$$

if $g = \ulcorner \varphi(v_0) \urcorner$. Since "$\mathfrak{M} \models name\ (x, z)$" is a predicate recursive in $\mathfrak{M}$, it can be shown within RCA_0 that for each $m \in \tilde{\mathbf{N}}$, there exists the least number in $\tilde{\mathbf{N}}$ which cannot be named by any m-formula in $\mathfrak{M}$.

We define $P(x, y)$ to be an $\mathcal{L}_1$-formula expressing that some y-formula names x in the model defined by Tr_{PA}, that is, $(\exists z)(F(z, y) \wedge Tr_{PA}(\ulcorner name(z, x) \urcorner))$. Finally, let $Q(x, y)$ be $\neg P(x, y) \wedge (\forall z < x) P(z, y)$. Thus, $Q(x, y)$ means that x is the smallest number that cannot be named by any y-formula in the model defined by Tr_{PA}. Now, let k be the number of symbols occurring in $Q(x, y)$, and t be the closed term

$$\overbrace{(1 + \cdots + 1)}^{10 \text{ times}} \cdot \overbrace{(1 + \cdots + 1)}^{k \text{ times}}.$$

Without loss of generality, we may assume that $Q(v_0, t)$ is a $(10 \cdot k)$-formula. Then we have

Lemma 7.1 (RCA_0) *Suppose $\mathfrak{M}_0$ is a model of $PA + Con(PA)$, and $\mathfrak{M}_1$ is the model of PA defined by $Tr_{PA}(x)$ in $\mathfrak{M}_0$ in the same way as in Theorem 4.4. Let s be the numeral denoting the least number $n \in \tilde{\mathbf{N}}$ such that $(\forall x)(F(x, t) \to \mathfrak{M}_1 \models \neg name(x, n))$. Then,*

(i) $\mathfrak{M}_0 \models Q(s, t)$,
(ii) $\mathfrak{M}_1 \models \neg Q(s, t)$.

Proof: (i): By way of contradiction, we assume that $\mathfrak{M}_0 \models \neg Q(s, t)$. Then $\mathfrak{M}_0 \models P(s, t) \vee \neg(\forall z < s) P(z, t)$ by the definition of $Q(x, y)$. Thus, $(\forall z < s)(\exists x)(F(x, t) \wedge \mathfrak{M}_0 \models Tr_{PA}(\ulcorner name(x, z) \urcorner))$ by the definition of s, so $(\forall z < s)(\exists x)(\mathfrak{M}_0 \models F(x, t) \wedge Tr_{PA}(\ulcorner name(x, z) \urcorner))$ since $F(x, y)$ is recursive. Hence, by Lemma 4.1, $\mathfrak{M}_0 \models (\forall z < s)\ (\exists x)(F(x, t) \wedge Tr_{PA}(\ulcorner name(x, z) \urcorner))$, namely, $\mathfrak{M}_0 \models (\forall z < s)\ P(z, t)$. Therefore, we must have $\mathfrak{M}_0 \models P(s, t)$, that is, $\mathfrak{M}_0 \models (\exists x)(F(x, t) \wedge Tr_{PA}(name(x, s)))$. By using Lemma 4.1 and the recursiveness of $F(x, y)$ again, we also obtain $(\exists x)(F(x, t) \wedge \mathfrak{M}_0 \models Tr_{PA}(name(x, s)))$, that is, $(\exists x)(F(x, t) \wedge \mathfrak{M}_1 \models name(x, s))$. This contradicts the definition of s.

(ii): If we had $\mathfrak{M}_1 \models Q(s, t)$, then we would have $\mathfrak{M}_1 \models name(\ulcorner Q(v_0, t) \urcorner, s)$, which contradicts the definition of s since $Q(v_0, t)$ is a $(10 \cdot k)$-formula.

This lemma leads to the following version of the First Incompleteness Theorem.

Theorem 7.2 (The First Incompleteness Theorem (WKL_0)) *If $PA + Con(PA)$ is consistent, PA is incomplete.*

Proof: Assume that $PA + Con(PA)$ is consistent. By Theorem 2.2, we have a model $\mathfrak{M}_0$ of $PA + Con(PA)$. Then define a model $\mathfrak{M}_1$ of PA and the numeral s as in the above lemma. Since both $\mathfrak{M}_0$ and $\mathfrak{M}_1$ are models of PA, we have $PA \nvdash Q(s, t)$ and $PA \nvdash \neg Q(s, t)$, which means that PA is incomplete.

We remark that the consistency of $PA + Con(PA)$ is easily deduced from the 1-consistency of PA, hence also from the ω-consistency of PA.

Corollary 7.3 (The Formalized First Incompleteness Theorem) $PRA \vdash Con(PA + Con(PA)) \rightarrow \neg Pr_{PA}(\ulcorner Q(s,t) \urcorner) \wedge \neg Pr_{PA}(\ulcorner \neg Q(s,t) \urcorner)$.

Proof: Apply Theorem 2.1 to Theorem 7.2.

REFERENCES

[1] Boolos, G., "A new proof of the Gödel Incompleteness Theorem," *Notices of the American Mathematical Society*, vol. 36 (1989), pp. 388–390. Zbl 0972.03544

[2] Feferman, S., "Arithmetization of metamathematics in a general setting," *Fundamenta Mathematicae*, vol. 49 (1960), pp. 35–92. Zbl 0095.24301 MR 26:4913

[3] Kikuchi, M., "A note on Boolos' proof of the Incompleteness Theorem," *Mathematical Logic Quarterly*, vol. 40 (1994), pp. 528–532. Zbl 0805.03052 MR 95j:03095

[4] Kreisel, G., "Notes on arithmetical models for consistent formulae of the predicate calculus," *Fundamenta Mathematicae*, vol. 37 (1950), pp. 265–285. Zbl 0040.00302 MR 12,790a

[5] Simpson, S., *Subsystems of Second Order Arithmetic*, forthcoming. Zbl 0909.03048 MR 2001i:03126

[6] Simpson, S., and K. Tanaka, "On the strong soundness of the theory of real closed fields," *Proceedings of the Fourth Asian Logic Conference*, (1990), pp. 7–10.

[7] Smoryński, C., "The Incompleteness Theorems," pp. 821–865 in *Handbook of Mathematical Logic*, edited by J. Barwise, North Holland, Amsterdam, 1977.

Mathematical Institute
Tohoku University
Sendai 980, Japan
email: makoto@math.tohoku.ac.jp

Mathematical Institute
Tohoku University
Sendai 980, Japan
email: tanaka@math.tohoku.ac.jp