Notre Dame Journal of Formal Logic
Volume 31, Number 4, Fall 1990

Derivability Conditions on Rosser's Provability Predicates

TOSHIYASU ARAI*

Abstract This paper is complementary to a paper by Guaspari and Solovay. Let $\mathrm{Th}(x)$ denote a Σ_1 provability predicate $\exists y\theta(y,x)$ for PRA (Primitive Recursive Arithmetic). We assume that formulas are in negation normal form, and hence $\neg\neg\phi$ is literally equal to ϕ for a formula ϕ. The symmetric form of Rosser's provability predicate Th^{R} for Th is defined by $\mathrm{Th}^{\mathrm{R}}(x) :\Leftrightarrow \exists y[\theta(y,x)\ \&\ \forall z \leq y\ \neg\theta(z,\dot{\neg}x)]$, where $\dot{\neg}$ denotes a function such that $\dot{\neg}\ulcorner\phi\urcorner = \ulcorner\neg\phi\urcorner$ with the Gödel number $\ulcorner\phi\urcorner$ of ϕ. For a 'canonical' provability predicate P for PRA, we construct Σ_1 formulas Th_2 and Th_3 such that PRA proves $\forall x,y[F(x \dot{\to} y) \to (F(x) \to F(y))]$, $\forall x[G(x) \to G(\ulcorner G(\dot{x})\urcorner)]$, and $\forall x[P(x) \leftrightarrow \mathrm{Th}_2(x) \leftrightarrow \mathrm{Th}_3(x)]$, where $\ulcorner\phi\urcorner \dot{\to} \ulcorner\psi\urcorner = \ulcorner\phi \to \psi\urcorner$, $F(x) :\Leftrightarrow \mathrm{Th}_2^{\mathrm{R}}(x)$, and $G(x) :\Leftrightarrow \mathrm{Th}_3^{\mathrm{R}}(x)$.

Let PRA (Primitive Recursive Arithmetic) denote the theory obtainable from PA (Peano Arithmetic formulated in a language containing function symbols for all primitive recursive functions) by restricting induction axioms to quantifier-free formulas. All results in this paper hold for any 1-consistent r.e. extension of PRA, but for the sake of definiteness we state results only for PRA.

We will consider derivability conditions on the symmetric form of Rosser's provability predicates. Let P be a Σ_1^0-formula, $\exists y\theta(y,x)$, with θ quantifier-free. P is said to be a *provability predicate* (for PRA) if P numerates the theorems of PRA in PRA, i.e., P satisfies the following:

D1 $\quad \vdash\varphi \Leftrightarrow \vdash P(\ulcorner\varphi\urcorner)$ for every formula φ,

where $\vdash\varphi$ means that φ is derivable in PRA and $\ulcorner\varphi\urcorner$ is the Gödel number of φ. Then the so-called symmetric form of Rosser's provability predicate P^{R} for P is defined by:

$$P^{\mathrm{R}}(x) :\Leftrightarrow \exists y[\theta(y,x) \wedge \forall v \forall z \leq y(v = \dot{\neg}x \vee x = \dot{\neg}v \to \neg\theta(z,v))]$$

*I would like to express my thanks to the referee for some valuable suggestions.

Received September 28, 1988; revised March 27, 1989

where $\dot{\neg}$ is a primitive recursive function such that $\dot{\neg}\ulcorner\varphi\urcorner = \ulcorner\neg\varphi\urcorner$ for every formula φ. P^{R} is a provability predicate, since PRA is 1-consistent. Obviously $\mathrm{Con}_{P^{\mathrm{R}}}$, which is defined by $\forall \mathit{formula}\ x \neg [P^{\mathrm{R}}(x) \wedge P^{\mathrm{R}}(\dot{\neg}x)]$, is derivable in PRA. Thus Gödel's Second Incompleteness Theorem does not hold with this provability predicate P^{R}. Therefore P^{R} does not satisfy both of the following two derivability conditions D2 and D3: For every formula φ and ψ,

D2 $\quad \vdash P^{\mathrm{R}}(\ulcorner\varphi \to \psi\urcorner) \to (P^{\mathrm{R}}(\ulcorner\varphi\urcorner) \to P^{\mathrm{R}}(\ulcorner\psi\urcorner))$

D3 $\quad \vdash P^{\mathrm{R}}(\ulcorner\varphi\urcorner) \to P^{\mathrm{R}}(\ulcorner P^{\mathrm{R}}(\ulcorner\varphi\urcorner)\urcorner)$.

To state results simply we will assume that formulas in PRA are in negation normal form, i.e., formulas are built up from atomic formulas and negated atomic formulas by applying the propositional connectives $\wedge$ and $\vee$, and the quantifiers $\forall$ and $\exists$. For a formula φ, the formula $\neg\varphi$ is defined recursively as follows: for atomic φ, $\neg(\neg\varphi) :\Leftrightarrow \varphi$; $\neg(\varphi \wedge \psi) :\Leftrightarrow \neg\varphi \vee \neg\psi$; $\neg(\varphi \vee \psi) :\Leftrightarrow \neg\varphi \wedge \neg\psi$; $\neg(\forall x\varphi) :\Leftrightarrow \exists x\neg\varphi$; $\neg(\exists x\varphi) :\Leftrightarrow \forall x\neg\varphi$. Therefore $\neg\neg\varphi$ is literally equal to φ. For formulas φ and ψ, the formula $\neg\varphi \vee \psi$ is denoted by $\varphi \to \psi$.

Now P^{R} can be written simply as follows:

$$P^{\mathrm{R}}(x) :\Leftrightarrow P(x) < P(\dot{\neg}x) \qquad \text{(in Guaspari's witness comparison notation)}$$
$$:\Leftrightarrow \exists y[\theta(y,x) \wedge \forall z \leq y\ \neg\theta(z,\dot{\neg}x)].$$

Kreisel [4] asked if there was a canonical P_0 (cf. Kreisel [3] for 'canonical') such that P_0^{R} satisfies D2 or D3. In [1], Guaspari and Solovay partly answered a modified version of Kreisel's question. Let D4 denote the following derivability condition (*demonstrable* Σ_1^0-*completeness*):

D4 $\quad \vdash \varphi \to P(\ulcorner\varphi\urcorner) \qquad$ for every Σ_1^0-sentence φ.

Then for any given provability predicate P satisfying D2 and D4, there exist Σ_1^0 P_2 and P_3 such that P_2^{R} (P_3^{R}) does not satisfy D2 (D3), respectively, and both P_2 and P_3 are *demonstrably extensionally equal to P*, i.e., there exist (Σ_1^0) φ, ψ, and σ so that,

$$\nvdash P_2^{\mathrm{R}}(\ulcorner\varphi \to \psi\urcorner) \to (P_2^{\mathrm{R}}(\ulcorner\varphi\urcorner) \to P_2^{\mathrm{R}}(\ulcorner\psi\urcorner))$$
$$\nvdash P_3^{\mathrm{R}}(\ulcorner\sigma\urcorner) \to P_3^{\mathrm{R}}(\ulcorner P_3^{\mathrm{R}}(\ulcorner\sigma\urcorner)\urcorner)$$
$$\vdash \forall \mathit{formula}\ x[P(x) \leftrightarrow P_2(x) \leftrightarrow P_3(x)].$$

In this paper, we will give the following complements to Guaspari–Solovay's result: Let D2′, D3′, and D4′ denote the uniform version of D2, D3, and D4, respectively:

D2′ $\quad \vdash \forall \mathit{formulas}\ x,y[P(x \dot{\to} y) \to P(x) \to P(y)]$

where $\dot{\to}$ is a function such that

$$\ulcorner\varphi\urcorner \dot{\to} \ulcorner\psi\urcorner = \ulcorner\varphi \to \psi\urcorner \qquad \text{for formulas } \varphi \text{ and } \psi.$$

D3′ $\quad \vdash \forall \mathit{formula}\ x[P(x) \to P(\ulcorner P(\dot{x})\urcorner)]$

where $\ulcorner P(\dot{x}) \urcorner$ denotes a term $t(x)$ such that if the nth numeral $\bar{n}$ ($\equiv$ S...S0 with n applications of the successor function S) is substituted for the variable x in $t(x)$, then the value of the result $t(\bar{n})$ is equal to $\ulcorner P(\bar{n}) \urcorner$.

D4′ $\quad \vdash \varphi(x) \rightarrow P(\ulcorner \varphi(\dot{x}) \urcorner) \qquad$ for every Σ_1^0 φ.

Definition 1 Let Fml denote the set of formulas. A function V: Fml $\rightarrow \{0,1\}$ is said to be a *truth valuation* if V satisfies the following conditions (1 for truth, 0 for falsehood):

(1) $V(\neg\varphi) = 1 \dot{-} V(\varphi)$ for an atomic or existential formula φ, i.e., a formula of the form $\exists x \psi$.
(2) $V(\varphi \wedge \psi) = \min(V(\varphi), V(\psi))$,
$V(\varphi \vee \psi) = \max(V(\varphi), V(\psi))$.

Definition 2 A formula φ is said to be a *tautology* if $V(\varphi) = 1$ for any truth valuation V.

Definition 3 A formula φ is said to be a *tautological consequence* of a set Γ of formulas if $V(\varphi) = 1$ for any truth valuation V such that $V(\psi) = 1$ for any $\psi \in \Gamma$.

Definition 4 A set Γ of formulas is said to be *satisfiable* if there exists a truth valuation V such that $V(\varphi) = 1$ for any $\varphi \in \Gamma$.

Moreover, let Taut denote the following derivability condition:

Taut $\quad \vdash x$ is a tautology $\rightarrow P(x)$.

Note that if P satisfies D2′ and Taut, then P satisfies the following two conditions:

(*) $\vdash \forall$*finite set of formulas* Γ, $\forall$*formula* x [x is a tautological consequence of $\Gamma \rightarrow \forall y \in \Gamma (P(y) \rightarrow P(x))$]
(**) $\vdash \exists$*formula* $y[P(y) \wedge P(\dot{\neg} y)] \rightarrow \forall$*formula* $xP(x)$.

Then our result runs as follows: For any given provability predicate P satisfying D2′, D4′, and Taut, there exist Σ_1 Th_2 and Th_3 such that $\mathrm{Th}_2^{\mathrm{R}}$ ($\mathrm{Th}_3^{\mathrm{R}}$) satisfies D2′ [D3′], respectively, and both Th_2 and Th_3 are demonstrably extensionally equal to P:

$$\vdash \forall \textit{formulas } x, y[\mathrm{Th}_2^{\mathrm{R}}(x \dot{\rightarrow} y) \rightarrow \mathrm{Th}_2^{\mathrm{R}}(x) \rightarrow \mathrm{Th}_2^{\mathrm{R}}(y))]$$

$$\vdash \forall \textit{formula } x[\mathrm{Th}_3^{\mathrm{R}}(x) \rightarrow \mathrm{Th}_3^{\mathrm{R}}(\ulcorner \mathrm{Th}_3^{\mathrm{R}}(\dot{x}) \urcorner)]$$

$$\vdash \forall \textit{formula } x[P(x) \leftrightarrow \mathrm{Th}_2(x) \leftrightarrow Th_3(x)].$$

Therefore, to answer Kreisel's original question some properties about the order of theorems of PRA under a canonical proof predicate must be used.

Remarks.
(1) Clearly $\mathrm{Th}_2^{\mathrm{R}}$ ($\mathrm{Th}_3^{\mathrm{R}}$) does not satisfy D3 (D2).
(2) Jeroslow [2] showed that if a provability predicate P satisfies D3″:

D3″ $\quad \vdash P(t) \rightarrow P(\ulcorner P(t) \urcorner)$

for every closed term t whose value is the Gödel number of a formula, then Gödel's Second Incompleteness Theorem holds with P:

$$\nvdash \mathrm{Con}_p (\equiv \forall \textit{formula } x \neg [P(x) \wedge P(\dot{\neg} x)]).$$

$\mathrm{Th}_3^{\mathrm{R}}$ shows that to have Gödel's Second Incompleteness Theorem with a provability predicate it is not sufficient that the predicate satisfies D3 (and the stronger D3′); D3 is weaker than D3″ since in D3 a closed term t is restricted to a numeral.

Next we will set forth the constructions of Th_2 and Th_3. We consider a provability predicate as an enumeration of theorems of PRA with infinite repetitions. Without loss of generally we can assume that a provability predicate P is of the form $\exists y(x = f(y))$, for some primitive recursive function f, such that

$$\vdash \forall y \exists z > y[f(y) \text{ is a formula} \wedge f(z) = f(y)].$$

To construct a provability predicate, i.e., an enumeration of theorems, we need to arrange the order of theorems.

Example (Rosser sentences) Let us first give some definitions.

(1) A provability predicate is said to be *standard* if it satisfies D2 and D4.
(2) A sentence φ is said to be a *Rosser sentence* if for a standard P

$$\vdash \varphi \leftrightarrow P(\ulcorner \neg \varphi \urcorner) < P(\ulcorner \varphi \urcorner).$$

(3) Let σ_+ and σ_- be Σ_1^0 sentences. The pair $\langle \sigma_+, \sigma_- \rangle$ is said to be a *Rosser pair* if for a standard P

$$\vdash P(\ulcorner \sigma_\pm \urcorner) \leftrightarrow P(\ulcorner \neg \sigma_\pm \urcorner),$$

$$\vdash \neg (\sigma_+ \wedge \sigma_-), \quad \text{and}$$

$$\vdash \sigma_+ \vee \sigma_- \leftrightarrow P(\ulcorner \bot \urcorner) \qquad (\bot \text{ is a false formula, e.g., } 0 = 1).$$

Then we have the following theorem (cf. the construction of nonequivalent Rosser sentences in Guaspari–Solovay [1]):

Theorem. *For a sentence φ, φ is a Rosser sentence iff there exists a Rosser pair $\langle \sigma_+, \sigma_- \rangle$ such that $\vdash \varphi \leftrightarrow \sigma_+$.*

Corollary *For sentences φ and ψ, if φ is a Rosser sentence and $\varphi \leftrightarrow \psi$ is derivable, then ψ is also a Rosser sentence.*

Proof of theorem: ($\Rightarrow$) Assume that φ is a Rosser sentence with a standard P. Define Σ_1^0-sentences σ_+ and σ_- as

$$\begin{aligned} \sigma_+ &:\Leftrightarrow P(\ulcorner \neg \varphi \urcorner) < P(\ulcorner \varphi \urcorner) \\ &:\Leftrightarrow \exists y [\theta(y, \ulcorner \neg \varphi \urcorner) \wedge \forall z \le y \, \neg \theta(z, \ulcorner \varphi \urcorner)] \qquad \text{with } P \equiv \exists y \theta(y, x); \\ \sigma_- &:\Leftrightarrow P(\ulcorner \varphi \urcorner) \le P(\ulcorner \neg \varphi \urcorner) \\ &:\Leftrightarrow \exists z [\theta(z, \ulcorner \varphi \urcorner) \wedge \forall y < z \, \neg \theta(y, \ulcorner \neg \varphi \urcorner)]. \end{aligned}$$

Claim *$\langle \sigma_+, \sigma_- \rangle$ is a Rosser pair with P.*

Proof of Claim: For readability we write $\Box \psi$ for $P(\ulcorner \psi \urcorner)$. Then D1, D2, and D4 can be written as;

D1 $\vdash \psi \Leftrightarrow \vdash \Box \psi$

D2 $\vdash \Box(\varphi \to \psi) \to \Box\varphi \to \Box\psi$

D4 $\vdash \varphi \to \Box\varphi$ for $\Sigma_1^0\ \varphi$,

and we have

$\vdash \varphi \leftrightarrow (\Box\neg\varphi \prec \Box\varphi)$

$\vdash \sigma_+ \leftrightarrow \varphi$

$\vdash \sigma_- \leftrightarrow (\Box\varphi \preceq \Box\neg\varphi)$.

We now prove that $\vdash \Box\sigma_\pm \leftrightarrow \Box\neg\sigma_\pm$.

(1) $\vdash \Box\neg\sigma_+ \to \Box\sigma_+$

$\sigma_+ \to \Box\sigma_+$ by D4.

On the other hand

$$\begin{aligned} \Box\neg\sigma_+ \wedge \neg\sigma_+ &\to \Box\neg\varphi \wedge \neg\sigma_+ && \text{by } \varphi \leftrightarrow \sigma_+ \\ &\to \Box\varphi && \text{by the definition of } \sigma_+ \\ &\to \Box\sigma_+. \end{aligned}$$

From these we get the assertion

(2) $\vdash \Box\neg\sigma_- \to \Box\sigma_-$

$$\begin{aligned} \neg\sigma_- \wedge \Box\varphi &\to \sigma_+ \\ &\to \varphi && \text{by } \varphi \leftrightarrow \sigma_+. \end{aligned}$$

Hence by D1 and D2, $\Box\neg\sigma_- \to \Box(\Box\varphi \to \varphi)$.

So by the formalized Löb's Theorem we have:

(+) $\Box\neg\sigma_- \to \Box\varphi$.

From (+) and D4 we have

(a) $\Box\neg\sigma_- \to \Box\Box\varphi$

$$\begin{aligned} \neg\sigma_- \wedge \Box\varphi &\to \Box\neg\varphi && \text{by the definition of } \sigma_- \\ &\to \Box\neg\sigma_+ && \text{by } \varphi \leftrightarrow \sigma_+. \end{aligned}$$

From this and (+) we have:

(b) $\Box\neg\sigma_- \wedge \neg\sigma_- \to \Box\neg\sigma_+$

$\neg\sigma_+ \wedge \Box\varphi \to \sigma_-$ by the definition of σ_-.

Hence we get the following by D1, D2, (a), and (b): $\Box\neg\sigma_- \wedge \neg\sigma_- \to \Box\sigma_-$. On the other hand, we have by D4 that $\sigma_- \to \Box\sigma_-$. From these we get the assertion:

(3) $\vdash \Box\sigma_+ \to \Box\neg\sigma_+$ and $\vdash \Box\sigma_- \to \Box\neg\sigma_-$.

From (1), (2), and $\vdash \neg(\sigma_+ \wedge \sigma_-)$ we have

$$\begin{aligned} \Box\sigma_+ &\to \Box\neg\sigma_- \qquad \text{and} \qquad & \Box\sigma_- &\to \Box\neg\sigma_+ \\ &\to \Box\sigma_- & &\to \Box\sigma_+ \\ &\to \Box\neg\sigma_+ & &\to \Box\neg\sigma_-. \end{aligned}$$

($\Leftarrow$) Let $\langle \sigma_+, \sigma_- \rangle$ be a Rosser pair with a standard P so that $\vdash \varphi \leftrightarrow \sigma_+$. σ_+ and σ_- are of the forms $\exists x \tau_+(x)$ and $\exists x \tau_-(x)$ with quantifier-free τ_+ and τ_-, respectively. Let P be of the form $\exists y(x = f(y))$ with a primitive recursive f. We will define a primitive recursive function g and a Σ_1^0-formula Th so that:

- $\mathrm{Th}(x) :\Leftrightarrow \exists y(x = g(y))$;
- Th is demonstrably extensionally equal to P and hence Th is a standard provability predicate;
- φ is a Rosser sentence with Th, i.e., $\vdash \varphi \leftrightarrow (\mathrm{Th}(\ulcorner \neg\varphi \urcorner) \prec \mathrm{Th}(\ulcorner \varphi \urcorner))$.

Let $*$ denote a number that is not the Gödel number of any formula. Then define g by recursion, as follows:

(a) $g(m) = f(m)$ if $f(m) \neq \ulcorner \neg\varphi \urcorner$ and $\ulcorner \varphi \urcorner$.
(b) If $f(m) = \ulcorner \neg\varphi \urcorner$, then we put

$$g(m) = \begin{cases} \ulcorner \neg\varphi \urcorner, & \text{if } \exists x \leq m\ \tau_+(x) \vee \exists n < m(g(n) = \ulcorner \varphi \urcorner) \\ *, & \text{otherwise.} \end{cases}$$

(c) If $f(m) = \ulcorner \varphi \urcorner$, then we put

$$g(m) = \begin{cases} \ulcorner \varphi \urcorner, & \text{if } \exists x \leq m\ \tau_-(x) \vee \exists n < m(g(n) = \ulcorner \neg\varphi \urcorner) \\ *, & \text{otherwise.} \end{cases}$$

Now note that the following hold by assumption:

$$\vdash \neg(\sigma_+ \wedge \sigma_-)$$

$$\vdash P(\ulcorner \varphi \urcorner) \leftrightarrow P(\ulcorner \neg\varphi \urcorner) \leftrightarrow P(\ulcorner \perp \urcorner) \leftrightarrow \sigma_+ \vee \sigma_-.$$

Using these we can see what g looks like:

- If $\neg P(\ulcorner \perp \urcorner)$, then clearly $\forall m(f(m) = g(m))$.
- Suppose $P(\ulcorner \perp \urcorner)$. Then g outputs formulas except φ and $\neg\varphi$ as f does. Let m denote the smallest number such that $\tau_+(m) \vee \tau_-(m)$. This m exists because $\sigma_+ \vee \sigma_-$. Put $\varphi_+ :\Leftrightarrow \varphi$, $\varphi_- :\Leftrightarrow \neg\varphi$. Let n denote the number defined as follows:

$$\begin{aligned} n = \min(n \mid \exists x(m \leq x < n \text{ and } (f(x), f(n)) \\ = (\ulcorner \varphi_- \urcorner, \ulcorner \varphi_+ \urcorner), (\ulcorner \varphi_+ \urcorner, \ulcorner \varphi_- \urcorner))). \end{aligned}$$

  This n exists because $P(\ulcorner \perp \urcorner)$ and because every theorem occurs cofinally for f. If $f(x) \in \{\ulcorner \varphi_+ \urcorner, \ulcorner \varphi_- \urcorner\}$, then

  (1) $g(x) = *$, if $x < m$.
  (2) If $m \leq x < n$ and $\sigma_+ [\sigma_-]$ holds, then

$$g(x) = \begin{cases} f(x), & \text{if } f(x) = \ulcorner \varphi_- \urcorner, [\ulcorner \varphi_+ \urcorner] \\ *, & \text{otherwise} \end{cases}$$

  respectively.

  (3) $g(x) = f(x)$, if $n \leq x$.

Thus we see that the following hold:

$\vdash \forall formula\ x[P(x) \leftrightarrow (\mathrm{Th}(x))]$

$\vdash \varphi \leftrightarrow (\mathrm{Th}(\ulcorner \neg\varphi \urcorner) < \mathrm{Th}(\ulcorner \varphi \urcorner))$.

Let P be a provability predicate satisfying D2′, D4′, and Taut of the form $\exists y(x = f(y))$ with a primitive recursive f. We will define primitive recursive functions g and h, and put

$\mathrm{Th}_2(x) :\Leftrightarrow \exists y(x = g(y))$

$\mathrm{Th}_3(x) :\Leftrightarrow \exists y(x = h(y))$

so that $\mathrm{Th}_2^{\mathrm{R}}$ ($\mathrm{Th}_3^{\mathrm{R}}$) satisfies D2′ (D3′), respectively, and both Th_2 and Th_3 are demonstrably extensionally equal to P.

A construction of g Let $\mathrm{Sat}(n)$ denote a formula such that

$$\vdash \mathrm{Sat}(n) \leftrightarrow \{\ulcorner \varphi \urcorner : \exists k \leq n(\ulcorner \varphi \urcorner = f(k))\} \text{ is satisfiable.}$$

Then g is defined as follows:

$$g(m) = f(m), \text{ if } \mathrm{Sat}(m).$$

Suppose that there exists an m such that $\neg\mathrm{Sat}(m)$, i.e., $\exists formula\ y[P(y) \wedge P(\dot\neg y)]$. Choose the minimal such m; so $\neg\mathrm{Sat}(m) \wedge [\mathrm{Sat}(m \dot- 1) \vee m = 0]$. Let Γ be the finite set of formulas $\{\varphi : \exists n < m(\ulcorner \varphi \urcorner = f(n))\}$, and let V be a truth valuation such that:

$$V(\varphi) = 1 \qquad \text{for all } \varphi \in \Gamma.$$

Let $\{\theta_i\}_{i<\omega}$ be an enumeration of all formulas. Then put

$$g(m + 2i) = \begin{cases} \ulcorner \theta_i \urcorner, & \text{if } V(\theta_i) = 1 \\ *, & \text{if } V(\theta_i) = 0 \end{cases}$$

(where $*$ is a number that is not the Gödel number of any formula, i.e., $* \neq \ulcorner \theta_i \urcorner$ for $\forall i < \omega$).

$$g(m + 2i + 1) = \begin{cases} \ulcorner \neg\theta_i \urcorner, & \text{if } V(\theta_i) = 1 \\ *, & \text{if } V(\theta_i) = 0. \end{cases}$$

Then clearly Th_2 is demonstrably extensionally equal to P.

Assertion *Th_2^R satisfies* D2′ *and* Taut.

Proof: (i) Suppose that $\neg\exists formula\ y[P(y) \wedge P(\dot\neg y)]$. Then for some formula x

$$\mathrm{Th}^{\mathrm{R}}(x) \leftrightarrow \mathrm{Th}(x) \leftrightarrow P(x).$$

Hence the assertion is trivial by our assumption.

(ii) Suppose now that $\exists formula\ y[P(y) \wedge P(\dot\neg y)]$. Let V be the truth valuation mentioned above. Then we see easily the following: for any i

$$\mathrm{Th}_2^{\mathrm{R}}(\ulcorner \theta_i \urcorner) \leftrightarrow V(\theta_i) = 1.$$

From this we get the assertion.

A construction of h. Here we assume that the coding of (formal) expressions (i.e., finite sequences of alphabets) satisfies the following condition (Assumption on coding):

$$\vdash e_1 \text{ is a proper subexpression of } e_2 \to (\ulcorner e_1 \urcorner < \ulcorner e_2 \urcorner).$$

Then we have that $\vdash \forall x(x \le \ulcorner \dot{x} \urcorner)$, where $\ulcorner \dot{x} \urcorner$ is a term $t(x)$ such that the value of $t(\bar{n})$ is equal to $\ulcorner \bar{n} \urcorner$ for every n. Therefore for every formula $\psi(v)$ in which a variable v occurs we have that $\vdash \forall x(x < \ulcorner \psi(\dot{x}) \urcorner)$; in particular, $\vdash \ulcorner \varphi \urcorner < \ulcorner \psi(\ulcorner \varphi \urcorner) \urcorner$ for every formula φ.

We will define h by using the primitive recursion theorem. For readability, we write Th for Th_3, i.e.,

$$\mathrm{Th}(x) :\Leftrightarrow \exists y(x \mathrel{\dot{\in}} h(y)),$$

where $x \mathrel{\dot{\in}} y$ means that y is (the code of) a finite set Γ of numbers and x belongs to Γ.

Definition 5

(1) $x \mathrel{\dot{\in}} \bar{f}(m) :\Leftrightarrow \exists n < m(x = f(n))$.

(2) *contradictory at m* $:\Leftrightarrow$ $\exists$*formula* $y(y = f(m) \wedge \dot{\neg} y \mathrel{\dot{\in}} \bar{f}(m))$.

(3) For each formula x, we define a number $Nx \le x + 1$ and a finite sequence $\{x_i\}_{i<Nx}$ of formulas, as follows.

(3.1) $x_0 := x$.

(3.2) Assume x_i is defined. If for a formula y, x_i is of the form $\ulcorner \mathrm{Th}^{\mathrm{R}}(y) \urcorner$, then x_{i+1} is defined to be the formula $y : x_i = \ulcorner \mathrm{Th}^{\mathrm{R}}(\dot{x}_{i+1}) \urcorner > x_{i+1}$. Otherwise put $Nx := i + 1$.

In what follows, we argue in PRA.

Proposition 1 $\forall$*formula* $x\, \forall i < Nx[Nx = i + Nx_i \wedge \forall y < Nx_i((x_i)_j = x_{i+j})]$.

Definition 6

(1) We say that *bell 1 rings at m* if m is the minimal m such that $\forall n \le m$ $\neg$*contradictory at n* and

$$\exists\textit{formula } x[\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner = f(m) \wedge \neg \exists i < N\dot{\neg}x(\dot{\neg}x)_i \mathrel{\dot{\in}} \bar{f}(m)].$$

(2) We say that *bell 2 rings at m* if m is the minimal m such that bell 1 has not rung before m (i.e., $\forall n < m \neg$ (*bell 1 rings at n*)) $\wedge$ *contradictory at m.*

(3) *bell rings at m* $:\Leftrightarrow$ either bell 1 or 2 rings at m

(4) *bell rings* $:\Leftrightarrow \exists m$ (bell rings at m).

Then we define h as follows: if $\forall n \le m$ (bell does not ring at n), then put $h(m) := \{f(m)\}$. In the following we assume that bell rings at m.

Case 1: Bell 1 rings at m. Let x be the formula such that

$$\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner = f(m).$$

Then put

$$h(m) = \{\dot{\neg}(\dot{\neg}x)_i : i < N\dot{\neg}x\}.$$

Case 2: Bell 2 rings at m. Then put

$$h(m) = \varnothing(\text{empty set}).$$

For $m' > m$ the definition of $h(m')$ is independent of which bell rings:

$$h(m+1) = \{\dot{\neg} y : y \text{ is a formula} \wedge \ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner \mathrel{\dot{\in}} \bar{f}(m) \wedge N\dot{\neg} y > 1\}.$$

Let $\{\ulcorner \theta_n \urcorner\}_{n<\omega}$ denote the enumeration of all the formulas in increasing order:

$$\vdash \forall n, n'(n < n' \to \ulcorner \theta_n \urcorner < \ulcorner \theta_{n'} \urcorner).$$

Then for each n, put

$$h(m+2+n) = \{\ulcorner \theta_n \urcorner\} \cup \{\ulcorner \mathrm{Th}^{\mathrm{R}}(\ulcorner \theta_n \urcorner) \urcorner\}.$$

This completes the definition of h.

Lemma 1
(1) *bell rings* $\leftrightarrow \exists$*formula* $y[P(y) \wedge P(\dot{\neg} y)]$
(2) $\forall$*formula* x $[\mathrm{Th}(x) \leftrightarrow P(x)]$.

Proof: (1) Suppose bell 1 rings at m and x is the formula such that $\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner = f(m)$. Then by definition we have $\neg(\dot{\neg} x \mathrel{\dot{\in}} \bar{f}(m))$ and $x = \dot{\neg}\dot{\neg} x \mathrel{\dot{\in}} h(m)$. Therefore $\mathrm{Th}^{\mathrm{R}}(\dot{x})$ is true. By D4′ we have $P(\ulcorner \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner)$. Thus $P(y) \wedge P(\dot{\neg} y)$, with $y = \ulcorner \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner$. The other case is easy.

(2) If bell does not ring, then $\forall m(h(m) = \{f(m)\})$. If bell rings, then by (1) and (**), $\forall$*formula* $x\, P(x)$. By the definition of h we have $\forall$*formula* $x\, \mathrm{Th}(x)$.

Lemma 2 *Assume that bell rings at m, y is a formula, and* $\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner \mathrel{\dot{\in}} \bar{f}(m)$. Then
(1) $N\dot{\neg} y = 1 \to \dot{\neg} y \mathrel{\dot{\in}} \bar{f}(m)$.
(2) $\forall i < Ny - 1[\dot{\neg} y_i \mathrel{\dot{\in}} \bar{f}(m)]$, i.e., $\forall i < Ny[\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}_i) \urcorner \mathrel{\dot{\in}} \bar{f}(m)]$. *In fact, if* $n < m$ *is a number such that* $\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner = f(n)$, *then there exist* $n = n_0 > n_1 > \ldots > n_{Ny-1}$ *such that* $\dot{\neg} y_{i-1} = \ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}_i) \urcorner = f(n_i)$ *for all* $i < Ny$, $(y_{-1} := \ulcorner \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner)$.
(3) $\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner = f(m) \wedge$ *bell 1 rings at* $m \to \forall i < N\dot{\neg} x(y \neq \dot{\neg}(\dot{\neg} x)_i)$.
(4) $\neg(y \mathrel{\dot{\in}} \bar{f}(m))$.

Proof: Let $n < m$ be a number such that $\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner = f(n)$.

(1) Since bell has not rung until n, we have

$$\exists i < N\dot{\neg} y((\dot{\neg} y)_i \mathrel{\dot{\in}} \bar{f}(n)).$$

By the assumption that $N\dot{\neg} y = 1$ we have that $\dot{\neg} y \mathrel{\dot{\in}} \bar{f}(n)$.

(2) By induction on i using (1). Note that if $i + 1 < Ny$ then $y_i = \ulcorner \mathrm{Th}^{\mathrm{R}}(\dot{y}_{i+1}) \urcorner$, and so $N\dot{\neg} y_i = 1$.

(3) Assume that

$$\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner = f(m) \wedge \text{bell 1 rings at } m.$$

Suppose that $y = \dot{\neg}(\dot{\neg} x)_i$, for some $i < N\dot{\neg} x$. From the hypothesis of the lemma we have that $(\dot{\neg} y)_j \mathrel{\dot{\in}} \bar{f}(m)$, for some $j < N\dot{\neg} y$. But then by Proposition 1 $(\dot{\neg} y)_j = (\dot{\neg} x)_{i+j} \mathrel{\dot{\in}} \bar{f}(m)$. This contradicts our assumption, $\forall i < N\dot{\neg} x \neg((\dot{\neg} x)_i \mathrel{\dot{\in}} \bar{f}(m))$.

(4) By induction on y. Suppose $y \mathrel{\dot{\in}} \bar{f}(m)$. Then by (2)

(a) $\forall i < N\dot{\neg} y - 1[\dot{\neg}(\dot{\neg} y)_i \mathrel{\dot{\in}} \bar{f}(m)]$.

(If $N\dot{\neg}y = 1$, then (a) is trivial. Otherwise let z denote $(\dot{\neg}y)_1$. Then $\ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{z})\urcorner \mathrel{\dot{\in}} \bar{f}(m)$ and apply (2).) Since bell has not rung before m

$$\forall \mathit{formula}\ x \neg (x, \dot{\neg}x \mathrel{\dot{\in}} \bar{f}(m)),$$

and so

(b) $\forall i < N\dot{\neg}y - 1 \neg [(\dot{\neg}y)_i \mathrel{\dot{\in}} \bar{f}(m)]$.

Moreover, since $y \mathrel{\dot{\in}} \bar{f}(m)$, $\neg [\dot{\neg}y \mathrel{\dot{\in}} \bar{f}(m)]$; hence by (1)

(c) $N\dot{\neg}y > 1$.

Let z be the formula $(\dot{\neg}y)_{N\dot{\neg}y-1}$. Then by (a) and (c)

$$\ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{z})\urcorner = \dot{\neg}(\dot{\neg}y)_{N\dot{\neg}y-2} \mathrel{\dot{\in}} \bar{f}(m)$$

and

$$z = (\dot{\neg}y)_{N\dot{\neg}y-1} \leq u < \ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{u})\urcorner = \dot{\neg}\dot{\neg}y = y,$$

where $u = (\dot{\neg}y)_1$. Hence by the induction hypothesis we get

(d) $\neg [(\dot{\neg}y)_{N\dot{\neg}y-1} \mathrel{\dot{\in}} \bar{f}(m)]$.

From (b) and (d) we have

$$\forall i < N\dot{\neg}y \neg [(\dot{\neg}y)_i \mathrel{\dot{\in}} \bar{f}(m)].$$

But then bell 1 would ring at n, which is a contradiction.

Finally, we have:

Proposition 2 $\mathrm{Th}^{\mathrm{R}}(y) \to \mathrm{Th}^{\mathrm{R}}(\ulcorner\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner)$, *for any formula y.*

Proof: Assume that $\mathrm{Th}^{\mathrm{R}}(y)$. Then by D4′ and Lemma 1.2, $\mathrm{Th}(\ulcorner\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner)$.

Case 1: Bell does not ring. Then by Lemma 1, $\neg\mathrm{Th}(\ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner)$. Hence the assertion $\mathrm{Th}^{\mathrm{R}}(\ulcorner\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner)$ holds.

Case 2: Bell rings at m. Then there exists a k such that $\ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner \mathrel{\dot{\in}} h(k)$. Let k be the minimal such.

Case 2.1.: $k > m + 1$. Suppose that y is the nth formula $\ulcorner\theta_n\urcorner$ in the enumeration $\{\ulcorner\theta_i\urcorner\}_{i<\omega}$, and let n_0 denote the number $k - m - 2$. Then

$$\ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner = \ulcorner\theta_{n0}\urcorner$$

and

$$\ulcorner\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner \mathrel{\dot{\in}} h(m + 2 + n) \qquad (y = \ulcorner\theta_n\urcorner)$$
$$\ulcorner\theta_n\urcorner = y < \ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner = \ulcorner\theta_{n0}\urcorner,$$

therefore $n < n_0$ and $\mathrm{Th}^{\mathrm{R}}(\ulcorner\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner)$.

Case 2.2: $k = m + 1$.

$$\ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner \mathrel{\dot{\in}} h(m + 1)$$
$$= \{\dot{\neg}x : x \text{ is a formula} \wedge \ulcorner\neg\mathrm{Th}^{\mathrm{R}}(\dot{x})\urcorner \mathrel{\dot{\in}} \bar{f}(m) \wedge N\dot{\neg}x > 1\}.$$

Let x be the formula $\ulcorner\mathrm{Th}^{\mathrm{R}}(\dot{y})\urcorner$. Then $N\dot{\neg}x = 1$. Hence this is not the case.

Case 2.3: $k = m$. Then bell 1 rings at m. Let x be the formula such that $\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner = f(m)$. Then

$$\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner \dot{\in} h(m) = \{ \dot{\neg} (\dot{\neg} x)_i : i < N \dot{\neg} x \}$$

and so

$$\ulcorner \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner = (\dot{\neg} x)_i \text{ for some } i < N \dot{\neg} x,$$

$$y = (\dot{\neg} x)_{i+1}.$$

Hence $\neg [y \dot{\in} \bar{f}(m)]$.

On the other hand, $\dot{\neg} y = \dot{\neg} (\dot{\neg} x)_{i+1} \dot{\in} h(m)$. Therefore $\neg \mathrm{Th}^{\mathrm{R}}(y)$. Hence this is not the case.

Case 2.4: $k < m$. Then $\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner \dot{\in} \bar{f}(m)$. By Lemma 2.4, $\neg [y \dot{\in} \bar{f}(m)]$.

Next we show that $\neg [y \dot{\in} h(m)]$. If bell 2 rings, then the assertion is trivial. Assume that bell 1 rings and let x be the formula such that $\ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{x}) \urcorner = f(m)$. Then by Lemma 2.3, $\forall i < N \dot{\neg} x (y \neq \dot{\neg} (\dot{\neg} x)_i)$. But $h(m) = \{ \dot{\neg} (\dot{\neg} x)_i : i < N \dot{\neg} x \}$. Therefore we get the assertion, $\neg [y \dot{\in} h(m)]$. Hence we have $\neg \exists n \leq m [y \dot{\in} h(n)]$. On the other hand, by our assumption that $\mathrm{Th}^{\mathrm{R}}(y)$ and Lemma 2.1 we have $N \dot{\neg} y > 1$. And so

$$\dot{\neg} y \dot{\in} h(m+1) = \{ \dot{\neg} y : y \text{ is a formula} \wedge \ulcorner \neg \mathrm{Th}^{\mathrm{R}}(\dot{y}) \urcorner \dot{\in} \bar{f}(m) \wedge N \dot{\neg} y > 1 \}.$$

Therefore we have $\neg \mathrm{Th}^{\mathrm{R}}(y)$. Hence this is not the case.

REFERENCES

[1] Guaspari, D. and R. M. Solovay, "Rosser sentences," *Annals of Mathematical Logic*, vol. 16 (1979), pp. 81–99.

[2] Jeroslow, R., "Redundancies in the Hilbert-Bernays derivability conditions for Gödel's second incompleteness theorem," *The Journal of Symbolic Logic*, vol. 38 (1973), pp. 359–367.

[3] Kreisel, G., "Mathematical logic," pp. 95–195 in *Lectures on Modern Mathematics*, vol. 3, edited by T. L. Saaty, Wiley and Sons, New York, 1965.

[4] Kreisel, G. and G. Takeuti, "Formally self-referential propositions for cut-free classical analysis and related systems," *Dissertationes Mathematicae* 118, Warsaw, 1974.

Department of Mathematics
School of Science
Nagoya University
Chikusa-ku
Nagoya 464-01 Japan