THE JOURNAL OF SYMBOLIC LOGIC
Volume 82, Number 1, March 2017

# UNIVERSAL ROSSER PREDICATES

MAKOTO KIKUCHI AND TAISHI KURAHASHI

**Abstract.** Gödel introduced the original provability predicate in the proofs of Gödel's incompleteness theorems, and Rosser defined a new one. They are equivalent in the standard model $\mathbb{N}$ of arithmetic or any nonstandard model of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$, but the behavior of Rosser's provability predicate is different from the original one in nonstandard models of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$. In this paper, we investigate several properties of the derivability conditions for Rosser provability predicates, and prove the existence of a Rosser provability predicate with which we can define any consistent complete extension of PA in some nonstandard model of $\mathsf{PA}+\neg\mathsf{Con}_{\mathsf{PA}}$. We call it a universal Rosser predicate. It follows from the theorem that the true arithmetic TA can be defined as the set of theorems of PA in terms of a universal Rosser predicate in some nonstandard model of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$. By using this theorem, we also give a new proof of a theorem that there is a nonstandard model $M$ of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$ such that if $N$ is an initial segment of $M$ which is a model of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$ then every theorem of PA in $N$ is a theorem of PA in $\mathbb{N}$. In addition, we prove that there is a Rosser provability predicate such that the set of theorems of PA in terms of the Rosser provability predicate is inconsistent in any nonstandard model of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$.

**§1. Introduction.** Truth and provability are different. This is a well known fact which was revealed by Gödel's first incompleteness theorem. The notion of truth in a model of Peano arithmetic PA cannot coincide with the concept of PA-provability, since a sentence must be either true or false in the model while it is shown in the first incompleteness theorem that there is a sentence which cannot be proved nor disproved from PA. Furthermore, it is also well known that truth cannot be defined arithmetically, while the PA-provability can be formalized and represented by a formula $\mathsf{Pr}_{\mathsf{PA}}(x)$ in $\mathbb{N}$. That is, there is no formula $\varphi(x)$ which satisfies $\mathbb{N} \models \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$ for every sentence $\psi$, while the formula $\mathsf{Pr}_{\mathsf{PA}}(x)$ satisfies the condition that $\mathsf{PA} \vdash \psi$ if and only if $\mathbb{N} \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$ for every sentence $\psi$. The formula $\mathsf{Pr}_{\mathsf{PA}}(x)$ is called a provability predicate.

By using the provability predicate, we can consider the PA-provability in nonstandard models of PA. But things are unchanged, since the first incompleteness theorem can be formalized within PA. Truth and provability are different in any model of PA although there may be extra theorems of PA in some nonstandard model of PA. Also, truth in a model cannot be defined arithmetically in the model itself. It is worth mentioning that we can define a nonstandard model $M$ of PA

Received August 25, 2015.
2010 *Mathematics Subject Classification.* 03C62, 03F40, 03H15.
*Key words and phrases.* Rosser provability predicates, universal Rosser predicates, nonstandard models of arithmetic, the incompleteness theorems.


0022-4812/17/8201-0017
DOI:10.1017/jsl.2016.19

arithmetically in $\mathbb{N}$ or in other models of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$ by the arithmetized completeness theorem. Actually, we can define explicitly a formula $\varphi(x)$ such that for any $N \models \mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$ there is $M \models \mathsf{PA}$ which satisfies the condition that $M \models \psi$ if and only if $N \models \varphi(\ulcorner \psi \urcorner)$ for every sentence $\psi$. That is, the truth in $M$ is defined in $N$ by $\varphi(x)$. (This truth definability relation between models of PA is studied in Hájek and Pudlák [3].) However, such a formula $\varphi(x)$ cannot be a provability predicate. It may happen that truth in a model can be captured arithmetically in another model, but it is never specified in any model by using a provability predicate.

Rosser defined a new provability predicate $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ which is called a Rosser provability predicate by modifying the original provability predicate $\mathsf{Pr}_{\mathsf{PA}}(x)$, and Rosser improved the first incompleteness theorem to theories which are consistent but not $\omega$-consistent. There is no difference between $\mathsf{Pr}_{\mathsf{PA}}(x)$ and $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ in $\mathbb{N}$ since we can show that $\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}} \vdash \forall x(\mathsf{Pr}^R_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}_{\mathsf{PA}}(x))$ and $\mathbb{N} \models \mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}$. In this sense, Rosser provability predicates can be substituted for the original provability predicate in the arguments on the PA-provability in nonstandard models of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$. However, the situation is quite different in models of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$. There is an alternative way to consider the relationship between truth and provability when we use Rosser provability predicates.

According to Kikuchi and Kurahashi [5], we say that a model $M$ of PA is sane if $M \models \mathsf{Con}_{\mathsf{PA}}$ and insane if $M \models \neg\mathsf{Con}_{\mathsf{PA}}$. For any model $M$ of PA, we define two sets $\mathsf{Thm}_{\mathsf{PA}}(M)$ and $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ by $\mathsf{Thm}_{\mathsf{PA}}(M) = \{\varphi : \varphi$ is a sentence and $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)\}$ and $\mathsf{Thm}^R_{\mathsf{PA}}(M) = \{\varphi : \varphi$ is a sentence and $M \models \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner)\}$, respectively. $\mathsf{Thm}_{\mathsf{PA}}(M)$ is the set of all theorems of PA in $M$ and $\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N}) = \{\varphi : \varphi$ is a sentence and $\mathsf{PA} \vdash \varphi\}$. $\mathsf{Thm}_{\mathsf{PA}}(M) = \mathsf{Thm}^R_{\mathsf{PA}}(M)$ if $M$ is sane. If $M$ is insane, $\mathsf{Thm}_{\mathsf{PA}}(M) \neq \mathsf{Thm}^R_{\mathsf{PA}}(M)$ and $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ can be a consistent complete set of sentences while $\mathsf{Thm}_{\mathsf{PA}}(M)$ is always the set of all sentences of arithmetic. The structures of insane models of PA are investigated in detail in Kikuchi and Kurahashi [5]. In particular, it is shown in [5] that there is an insane model $M$ of PA such that $\mathsf{Thm}_{\mathsf{PA}}(N) = \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ for any initial segment $N$ of $M$ which is a sane model of PA.

This paper is devoted to analyzing $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ for insane models $M$ of PA. We investigate several properties of the derivability conditions for Rosser provability predicates, and then we show that there exists a Rosser provability predicate satisfying the condition that for any complete extension $T$ of PA there is an insane model $M$ of PA such that $T = \mathsf{Thm}^R_{\mathsf{PA}}(M)$. We call such a Rosser provability predicate a universal Rosser predicate. This theorem means that truth in any model of PA can be specified as the set of theorems of PA in some insane model of PA by using a universal Rosser predicate. The standard model $\mathbb{N}$ is not an exception and it follows that there is an insane model $M$ of PA such that $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ is equal to the true arithmetic $\mathsf{TA} = \{\varphi : \mathbb{N} \models \varphi\}$ when $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ is a universal Rosser predicate. We also give a new proof of the above-mentioned result of [5] by using this theorem. In addition, we prove that there is a Rosser provability predicate $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ such that $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ is inconsistent for any insane model $M$ of PA.

## §2. Preliminaries.

**2.1. Provability predicates.** In this paper, we assume that formulas are in negation normal form, namely the negations in formulas are allowed to be applied only to

https://doi.org/10.1017/jsl.2016.19 Published online by Cambridge University Press

atomic formulas. We use the symbol $\equiv$ to indicate the syntactical identity. For each formula $\varphi$, we define $\neg\varphi$ recursively as follows: $\neg\neg\varphi :\equiv \varphi$ for any atomic formula $\varphi$; $\neg(\varphi \wedge \psi) :\equiv \neg\varphi \vee \neg\psi$; $\neg\forall x\varphi :\equiv \exists x\neg\varphi$; and so on. Then we can prove $\neg\neg\varphi \equiv \varphi$ for every formula $\varphi$.

We call a set of sentences a *theory*. A theory $T$ is said to be sound if and only if the standard model $\mathbb{N}$ of arithmetic is a model of $T$. We assume that the theory Peano arithmetic PA is sound, in other words, the true arithmetic $\mathsf{TA} := \{\varphi : \mathbb{N} \models \varphi\}$ is an extension of PA. For each natural number $n$, $\overline{n}$ denotes the numeral for $n$. Also for each formula $\varphi$, $\ulcorner\varphi\urcorner$ denotes the numeral for $\varphi$ in some fixed Gödel numbering.

A formula is said to be $\Delta_1$ if and only if it is both $\Sigma_1$ and $\Pi_1$ in PA. We say that a formula $\mathsf{Prf}_{\mathsf{PA}}(x, y)$ is a *proof predicate* of PA if and only if it is $\Delta_1$ and satisfies the following conditions:

1. for any formula $\varphi$, $\mathsf{PA} \vdash \varphi$ if and only if $\mathsf{PA} \vdash \mathsf{Prf}_{\mathsf{PA}}(\ulcorner\varphi\urcorner, \overline{n})$ for some natural number $n$;
2. $\mathsf{PA} \vdash \forall x(\exists y\mathsf{Prf}_{\mathsf{PA}}(x, y) \rightarrow \forall z\exists w > z\mathsf{Prf}_{\mathsf{PA}}(x, w))$;
3. $\mathsf{PA} \vdash \forall y(\exists x\mathsf{Prf}_{\mathsf{PA}}(x, y) \rightarrow \exists! x\mathsf{Prf}_{\mathsf{PA}}(x, y))$.

We can say that each proof predicate $\mathsf{Prf}_{\mathsf{PA}}(x, y)$ represents the relation '$y$ is a code of a PA-proof of a sentence whose Gödel number is $x$'.

For each proof predicate $\mathsf{Prf}_{\mathsf{PA}}(x, y)$, we obtain a $\Sigma_1$ formula $\exists y\mathsf{Prf}_{\mathsf{PA}}(x, y)$, denoted by $\mathsf{Pr}_{\mathsf{PA}}(x)$, which is called a *provability predicate* of PA. Then each formula $\mathsf{Pr}_{\mathsf{PA}}(x)$ weakly represents the set of all Gödel numbers of theorems of PA, that is, for any formula $\varphi$, $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ if and only if $\varphi$ is provable in PA.

In this paper, we fix a proof predicate $\mathsf{Prf}_{\mathsf{PA}}(x, y)$ of PA whose provability predicate $\mathsf{Pr}_{\mathsf{PA}}(x)$ satisfies the following conditions: for any formulas $\varphi$ and $\psi$,

**D1**: if $\mathsf{PA} \vdash \varphi$, then $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$;
**D2**: $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi \rightarrow \psi\urcorner) \rightarrow (\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\psi\urcorner))$;
**D3**: $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)\urcorner)$;
**$\Sigma_1$-compl**: if $\varphi$ is $\Sigma_1$, then $\mathsf{PA} \vdash \varphi \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$.

Notice that the condition **D1** obviously holds for every provability predicate, and that the condition **D3** is an instance of the condition **$\Sigma_1$-compl** because the formula $\mathsf{Pr}_{\mathsf{PA}}(x)$ is $\Sigma_1$. The conditions **D1**, **D2**, and **D3** are called the *derivability conditions*. Also the last condition is referred to formalized $\Sigma_1$-completeness.

Let $\mathsf{Con}_{\mathsf{PA}}$ be the $\Pi_1$ sentence $\neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\overline{0} = \overline{1}\urcorner)$ which asserts the consistency of PA. Then Gödel's second incompleteness theorem holds for $\mathsf{Pr}_{\mathsf{PA}}(x)$, namely $\mathsf{PA} \nvdash \mathsf{Con}_{\mathsf{PA}}$. We say a model $M$ of PA is *sane* if $M \models \mathsf{Con}_{\mathsf{PA}}$, and *insane* if $M$ is not sane. By Gödel's second incompleteness theorem, there is an insane nonstandard model $M$ of PA. In such a model $M$, the sentence $\exists y\mathsf{Prf}_{\mathsf{PA}}(\ulcorner\overline{0} = \overline{1}\urcorner, y)$ is true, and thus $M$ contains a nonstandard proof of $\overline{0} = \overline{1}$. Moreover $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ holds for every formula $\varphi$ because $\mathsf{PA} \vdash \neg\mathsf{Con}_{\mathsf{PA}} \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$.

For each proof predicate $\mathsf{Prf}_{\mathsf{PA}}(x, y)$, we obtain a Rosser provability predicate $\mathsf{Pr}^{R}_{\mathsf{PA}}(x)$ of PA which is defined as the $\Sigma_1$ formula $\exists y(\mathsf{Prf}_{\mathsf{PA}}(x, y) \wedge \forall z \leq y$ $\neg\mathsf{Prf}_{\mathsf{PA}}(\dot{\neg}(x), z))$. Here $\dot{\neg}$ is a function symbol expressing a primitive recursive function calculating the code of $\neg\varphi$ from that of $\varphi$. The idea of using a Rosser provability predicate was introduced in Rosser [10] to improve Gödel's first incompleteness theorem. The following proposition is one of the most important properties of Rosser provability predicates.

https://doi.org/10.1017/jsl.2016.19 Published online by Cambridge University Press

PROPOSITION 2.1. *For any sentence* $\varphi$, *if* $\mathsf{PA} \vdash \neg\varphi$, *then* $\mathsf{PA} \vdash \neg\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$.

Since our formulas are formulated so that $\varphi$ and $\neg\neg\varphi$ are identical, we have $\mathsf{PA} \vdash \neg\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \vee \neg\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \neg\varphi \urcorner)$. Also since proof predicates $\mathsf{Prf}_{\mathsf{PA}}(x, y)$ satisfy $\mathsf{PA} \vdash \forall y(\exists x \mathsf{Prf}_{\mathsf{PA}}(x, y) \to \exists! x \mathsf{Prf}_{\mathsf{PA}}(x, y))$, we have $\mathsf{PA} \vdash \neg\mathsf{Con}_{\mathsf{PA}} \to \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \vee \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \neg\varphi \urcorner)$. By combining these observations, we obtain the following proposition.

PROPOSITION 2.2. $\mathsf{PA} \vdash \neg\mathsf{Con}_{\mathsf{PA}} \to (\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \leftrightarrow \neg\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \neg\varphi \urcorner))$ *for any sentence* $\varphi$.

**2.2. Definability of theories in models.** We say that a formula $\varphi(x)$ defines a theory $T$ in a model $M$ of PA if and only if $T = \{\psi : M \models \varphi(\ulcorner \psi \urcorner)\}$. It is clear that every recursive theory $T$ can be defined by some formula in any model. However, by the generalized version of Tarski's undefinability theorem, for any model $M$ of PA, there is no formula defining the full theory $\mathrm{Th}(M) := \{\varphi : M \models \varphi\}$ of $M$ in $M$.

On the other hand, the full theory $\mathrm{Th}(M)$ of each model $M$ can be defined in another model $N$. For a proof, we use the following fact due to Mostowski.

DEFINITION 2.3 (See Lindström [8]).

1. For each sentence $\varphi$, let $\varphi^0$ denote $\varphi$, and let $\varphi^1$ denote $\neg\varphi$.
2. A formula $\xi(x)$ is *independent* over PA if and only if $\mathsf{PA} + \{\xi(\overline{n})^{f(n)} : n \in \omega\}$ is consistent for every function $f : \omega \to \{0, 1\}$.

FACT 2.4 (Mostowski [9]). *There is a* $\Sigma_1$ *formula* $\xi(x)$ *which is independent over* PA.

PROPOSITION 2.5. *There exists a* $\Sigma_1$ *formula* $\xi(x)$ *such that for any model* $M$ *of* PA, *there exists a model* $N$ *of* PA *such that* $\xi(x)$ *defines* $\mathrm{Th}(M)$ *in* $N$.

PROOF. Let $\xi(x)$ be a $\Sigma_1$ formula independent over PA. Let $M$ be any model of PA. We define the function $f_M : \omega \to \{0, 1\}$ as follows:

$$f_M(n) = \begin{cases} 0 & \text{if } n \text{ is the Gödel number of some sentence } \varphi \text{ such that } M \models \varphi; \\ 1 & \text{otherwise.} \end{cases}$$

Then $\mathsf{PA} + \{\xi(\overline{i})^{f_M(i)} : i \in \omega\}$ is consistent since $\xi(x)$ is independent over PA, and hence it has a model $N$. Let $\varphi$ be any sentence with the Gödel number $n$. Then $M \models \varphi$ if and only if $f_M(n) = 0$. Also $f_M(n) = 0$ if and only if $N \models \xi(\ulcorner \varphi \urcorner)$. Therefore, we have $M \models \varphi$ if and only if $N \models \xi(\ulcorner \varphi \urcorner)$. This means that $\xi(x)$ defines $\mathrm{Th}(M)$ in $N$. ⊣

In this paper, we prove that a $\Sigma_1$ formula $\xi(x)$ and a model $N$ in the statement of Proposition 2.5 can be taken as a Rosser provability predicate of PA and a nonstandard insane model of PA, respectively. Here, we introduce $\mathsf{Thm}_{\mathsf{PA}}(M)$ and $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ which are sets defined in a model $M$ by the provability predicate $\mathsf{Pr}_{\mathsf{PA}}(x)$ and the Rosser provability predicate $\mathsf{Pr}^R_{\mathsf{PA}}(x)$, respectively.

DEFINITION 2.6. Let $M$ be any model of PA.

1. $\mathsf{Thm}_{\mathsf{PA}}(M) := \{\varphi : \varphi \text{ is a sentence and } M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)\}$.
2. $\mathsf{Thm}^R_{\mathsf{PA}}(M) := \{\varphi : \varphi \text{ is a sentence and } M \models \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner)\}$.

We can say that $\mathsf{Thm}_{\mathsf{PA}}(M)$ is the set of all sentences that the model $M$ considers provable in PA. In Kikuchi and Kurahashi [5], it is shown that $\mathsf{Thm}_{\mathsf{PA}}(M)$ is not a complete theory for any model $M$ of PA.

https://doi.org/10.1017/jsl.2016.19 Published online by Cambridge University Press

For every sane model $M$ of PA, it is easy to check $\mathsf{Thm}_{\mathsf{PA}}(M) = \mathsf{Thm}^R_{\mathsf{PA}}(M)$. Thus as far as we consider sane models, Rosser provability predicate $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ gives no new information.

Rosser provability predicate plays an important role in investigating insane models. The usual provability predicate $\mathsf{Pr}_{\mathsf{PA}}(x)$ defines the set of all sentences in any insane model, however $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ defines different theories in such models. In fact, $\mathsf{PA} \vdash \neg\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \overline{0} = \overline{1} \urcorner)$ by Proposition 2.1, and thus $\overline{0} = \overline{1} \notin \mathsf{Thm}^R_{\mathsf{PA}}(M)$ for every model $M$ of PA. Moreover, $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ is a complete extension of PA for any insane model $M$ because $M \models \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \leftrightarrow \neg\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \neg\varphi \urcorner)$ holds for any sentence $\varphi$ by Proposition 2.2.

**§3. The derivability conditions for Rosser predicates.** It is known that some properties of Rosser provability predicates are dependent on the choice of a proof predicate. In particular, Guaspari and Solovay [2] and Arai [1] proved that whether $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ satisfies each of the derivability conditions is dependent on the choice of a proof predicate. We describe their results more precisely. We consider the following conditions for Rosser provability predicates.

Definition 3.1. For all formulas $\varphi$ and $\psi$,

**D1**: If $\mathsf{PA} \vdash \varphi$, then $\mathsf{PA} \vdash \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$.
**D2**: $\mathsf{PA} \vdash \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \to \psi \urcorner) \to (\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \psi \urcorner))$.
**D3**: $\mathsf{PA} \vdash \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \urcorner)$.
**D4**: If $\mathsf{PA} \vdash \neg\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$, then $\mathsf{PA} \vdash \neg\varphi$.

Notice that every Rosser provability predicate satisfies **D1**, and that **D4** is the converse implication of the statement of Proposition 2.1. If $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ satisfies both **D2** and **D3**, then the usual proof of the second incompleteness theorem can be applied for $\mathsf{Pr}^R_{\mathsf{PA}}(x)$, and we have $\mathsf{PA} \nvdash \neg\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \overline{0} = \overline{1} \urcorner)$. This is not the case as we have already mentioned above. Therefore each $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ does not satisfy at least one of **D2** and **D3**. Kreisel and Takeuti [6] asked whether $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ satisfies **D2**. Guaspari and Solovay [2] proved that there is a Rosser provability predicate which does not satisfy **D2** and **D3**. They established a method of constructing a new proof predicate with required properties from a given proof predicate by reordering nonstandard proofs. Arai [1] proved the existence of a Rosser provability predicate satisfying **D2** and a Rosser provability predicate satisfying **D3**. Also Guaspari and Solovay [2] mentioned that there is a Rosser provability predicate such that the condition **D4** does not hold. Kurahashi [7] proved the existence of a Rosser provability predicate with the property **D4**.

For each Rosser provability predicate with the property **D2**, the following proposition, corresponding to a principle known in modal logic, holds:

Proposition 3.2. *If* $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ *satisfies* **D2**, *then for any sentences* $\varphi_0, \dots, \varphi_{n-1}$, $\mathsf{PA} \vdash \bigwedge_{i<n} \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi_i \urcorner) \leftrightarrow \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \bigwedge_{i<n} \varphi_i \urcorner)$.

The condition **D3** says that $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ is closed under the operation $\varphi \mapsto \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$. Moreover, we investigate properties of the conditions **D2** and **D4**.

Proposition 3.3. *The following are equivalent*:

1. $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ *satisfies* **D2**.
2. *The theory* $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ *is consistent for any model* $M$ *of* PA.

PROOF. (1 ⇒ 2): Suppose that $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ satisfies **D2**. Let $M$ be any model of PA. Towards a contradiction, we assume that the theory $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ is inconsistent, that is, there are sentences $\varphi_0, \ldots, \varphi_{n-1} \in \mathsf{Thm}^R_{\mathsf{PA}}(M)$ such that $\vdash \neg(\varphi_0 \wedge \cdots \wedge \varphi_{n-1})$. Then $M \models \bigwedge_{i<n} \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi_i \urcorner)$ and $\mathsf{PA} \vdash \neg(\varphi_0 \wedge \cdots \wedge \varphi_{n-1})$. We have $\mathsf{PA} \vdash \neg \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi_0 \wedge \cdots \wedge \varphi_{n-1} \urcorner)$ by Proposition 2.1. Since $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ satisfies **D2**, $M \models \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi_0 \wedge \cdots \wedge \varphi_{n-1} \urcorner)$ by Proposition 3.2. This contradicts the assumption that $M$ is a model of PA.

(2 ⇒ 1): We prove the contraposition. Suppose that $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ does not satisfy **D2**, that is, there are sentences $\varphi$ and $\psi$ such that $\mathsf{PA} \nvdash \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \to \psi \urcorner) \to (\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \psi \urcorner))$. Then there is a model $M$ of PA such that $M \models \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \to \psi \urcorner) \wedge \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \wedge \neg \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \psi \urcorner)$. Since $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \to \psi \urcorner) \wedge \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$, we have $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$ by **D2** for $\mathsf{Pr}_{\mathsf{PA}}(x)$. Hence $M \models \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \neg\psi \urcorner)$ since $\psi$ and $\neg\neg\psi$ are identical. Thus the theory $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ contains $\varphi \to \psi$, $\varphi$ and $\neg\psi$. Therefore the theory $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ is inconsistent. ⊣

Hence, if $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ satisfies **D2**, then $\mathsf{Thm}^R_{\mathsf{PA}}(M)$ is a consistent complete extension of PA for any insane model $M$ of PA by Proposition 2.2.

PROPOSITION 3.4. *If* $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ *satisfies* **D2**, *then the following are equivalent*:

1. $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ *satisfies* **D4**.
2. *For any consistent extension* $T$ *of* PA, *the theory* $\mathsf{PA} + \{\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) : \varphi \in T\}$ *is consistent*.

PROOF. Assume that $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ satisfies **D2**.

(1 ⇒ 2): Suppose that $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ satisfies **D4**. Let $T$ be any consistent extension of PA. Towards a contradiction, suppose that the theory $U_T := \mathsf{PA} + \{\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) : \varphi \in T\}$ is inconsistent. Then there are sentences $\varphi_0, \ldots, \varphi_{n-1} \in T$ such that $\mathsf{PA} \vdash \neg \bigwedge_{i<n} \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi_i \urcorner)$. By **D2** for $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ and Proposition 3.2, we have $\mathsf{PA} \vdash \neg \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi_0 \wedge \cdots \wedge \varphi_{n-1} \urcorner)$. Also by **D4**, we obtain $\mathsf{PA} \vdash \neg(\varphi_0 \wedge \cdots \wedge \varphi_{n-1})$. This contradicts the consistency of $T$. Therefore the theory $U_T$ is consistent.

(2 ⇒ 1): We prove the contraposition. Suppose that $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ does not satisfy **D4**, that is, there is a sentence $\psi$ such that $\mathsf{PA} \vdash \neg \mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \psi \urcorner)$ and $\mathsf{PA} \nvdash \neg\psi$. Then the theory $T := \mathsf{PA} + \psi$ is consistent but the theory $\mathsf{PA} + \{\mathsf{Pr}^R_{\mathsf{PA}}(\ulcorner \varphi \urcorner) : \varphi \in T\}$ is inconsistent. ⊣

## §4. Universal Rosser predicates.

In this section, we prove the existence of a Rosser provability predicate which defines any complete consistent extension of PA in some insane model. We say such a Rosser provability predicate universal.

DEFINITION 4.1. A Rosser provability predicate $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ is *universal* if for any consistent complete extension $T$ of PA, there exists a model $M$ of PA such that $\mathsf{Thm}^R_{\mathsf{PA}}(M) = \{\varphi : T \vdash \varphi\}$.

For every universal Rosser predicate $\mathsf{Pr}^R_{\mathsf{PA}}(x)$, there exists a model $M$ such that $\mathsf{Thm}^R_{\mathsf{PA}}(M) = \mathsf{TA}$. Note that such a model $M$ must be insane because for any sane model $N$, $\mathsf{Thm}_{\mathsf{PA}}(N) = \mathsf{Thm}^R_{\mathsf{PA}}(N)$ and this set is not complete.

Since each Rosser provability predicate $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ is defined by using a proof predicate $\mathsf{Prf}_{\mathsf{PA}}(x, y)$, whether $\mathsf{Pr}^R_{\mathsf{PA}}(x)$ is universal or not depends on the choice of $\mathsf{Prf}_{\mathsf{PA}}(x, y)$. In the following, we show that any given proof predicate $\mathsf{Prf}_{\mathsf{PA}}(x, y)$ can be redefined into a new proof predicate $\mathsf{Prf'}_{\mathsf{PA}}(x, y)$ whose Rosser provability predicate is universal.

https://doi.org/10.1017/jsl.2016.19 Published online by Cambridge University Press

Actually, we prove the existence of a proof predicate whose Rosser provability predicate satisfies both **D2** and **D4** by modifying Arai's proof. As in Arai's proof, we also use the notion of *valuations* $V$ which are functions from the set of all formulas to $\{0, 1\}$ satisfying the usual conditions for propositional connectives such as $V(\neg\varphi) = 1 - V(\varphi)$. Since every valuation satisfies the condition corresponding to **D2**, valuations are used to obtain a new Rosser provability predicate satisfying **D2**. We say that a set $T$ of formulas is *satisfiable* if there exists a valuation $V$ such that $V(\varphi) = 1$ for all $\varphi \in T$. These notions can be formalized in PA. We assume that PA can prove "for every finite set $T$ of formulas and a formula $\varphi$, if $T \cup \{\varphi\}$ is not satisfiable, then $\bigwedge T \to \neg\varphi$ is provable in PA" (see Arai [1] for details).

In the following, $\mathsf{Pr}'_{\mathsf{PA}}(x)$ and $\mathsf{Pr}'^{R}_{\mathsf{PA}}(x)$ always denote the provability predicate and the Rosser provability predicate defined by using a proof predicate $\mathsf{Prf}'_{\mathsf{PA}}(x, y)$ of PA, respectively.

THEOREM 4.2. *There exists a proof predicate* $\mathsf{Prf}'_{\mathsf{PA}}(x, y)$ *of* PA *satisfying the following conditions*:

1. $\mathsf{PA} \vdash \forall x(\mathsf{Pr}_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'_{\mathsf{PA}}(x))$;
2. $\mathsf{Pr}'^{R}_{\mathsf{PA}}(x)$ *satisfies* **D2** *and* **D4**.

PROOF. We define a PA-provably recursive function $f$ in stages, which enumerates all theorems of PA. A new proof predicate $\mathsf{Prf}'_{\mathsf{PA}}(x, y)$ of PA can be taken as a $\Delta_1$ formula naturally representing the relation $x = f(y)$. In the construction of $f$, the value of $f(m)$ is defined by referring to the values of $f(0), \ldots, f(m-1)$. As in Guaspari and Solovay [2], we prepare a bell which plays a role of a flag in the construction of $f$. The construction of $f$ consists of Procedure 1 and Procedure 2, and the bell may ring during the execution of Procedure 1. Until the bell rings, the value of $f(m)$ is defined in each stage $m$. When the bell rings, the construction of $f$ goes to Procedure 2, and $f$ eventually enumerates all formulas. In the definition of $f$, we identify each formula with its Gödel number. Also in the definition of $f$, we can use the formula $\mathsf{Prf}'_{\mathsf{PA}}(x, y)$ itself by using Fixed-point Theorem or Diagonal Lemma[1] (see Lindström [8] or Hájek and Pudlák [4]). Here we start defining $f$.

Procedure 1: The bell has not rung yet.

Stage $m$: If $m$ is not a proof of any formula, that is, there is no $x \leq m$ with $\mathsf{Prf}_{\mathsf{PA}}(x, \overline{m})$, then let $f(m)$ be $\overline{0} = \overline{0}$ and go to the stage $m + 1$.

If $m$ is a proof of a formula $\varphi$, then we distinguish the following two cases:

CASE (a): If $\{f(0), \ldots, f(m-1), \varphi\}$ is satisfiable, then let $f(m)$ be $\varphi$. In this case, we also distinguish the following two cases:

CASE (a.1): If $\varphi$ is a formula of the form $\neg\mathsf{Pr}'^{R}_{\mathsf{PA}}(\ulcorner\psi\urcorner)$ for some $\psi$ and $\{f(0), \ldots, f(m), \psi\}$ is satisfiable, then let $f(m+1)$ be $\psi$. Ring the bell and go to Procedure 2.

CASE (a.2): Otherwise, go to the stage $m + 1$.

CASE (b): If $\{f(0), \ldots, f(m-1), \varphi\}$ is not satisfiable, then ring the bell and go to Procedure 2.

[1] In the definition of $f$, we actually define a formula $\xi(x, y, z)$ with three parameters $x, y$, and $z$, and the required proof predicate $\mathsf{Prf}'_{\mathsf{PA}}(x, y)$ is taken as a formula satisfying the equivalence $\mathsf{PA} \vdash \forall x \forall y(\mathsf{Prf}'_{\mathsf{PA}}(x, y) \leftrightarrow \xi(x, y, \ulcorner\mathsf{Prf}'_{\mathsf{PA}}(x, y)\urcorner))$.

https://doi.org/10.1017/jsl.2016.19 Published online by Cambridge University Press

Procedure 2: The bell has rung. We take a valuation $V$ and a number $k$ according to the following two cases:

1. The bell rang at the stage $m$ in Case (a.1): In this case, $m$ is a proof of a formula of the form $\neg \mathsf{Pr}'^R_{\mathsf{PA}}(\ulcorner \psi \urcorner)$. Then let $V$ be a valuation satisfying $\{f(0), \ldots, f(m), \psi\}$ and let $k$ be $m+2$.
2. The bell rang at the stage $m$ in Case (b): Then let $V$ be a valuation satisfying $\{f(0), \ldots, f(m-1)\}$ and let $k$ be $m$.

Then $f$ enumerates all the formulas in the following way: let $\{\varphi_i\}_{i \in \omega}$ be an effective enumeration of all formulas.

$$f(k+2i) = \begin{cases} \varphi_i & \text{if } V(\varphi_i) = 1, \\ \neg\varphi_i & \text{otherwise.} \end{cases}$$

$$f(k+2i+1) = \begin{cases} \neg\varphi_i & \text{if } V(\varphi_i) = 1, \\ \varphi_i & \text{otherwise.} \end{cases}$$

The definition of $f$ has just been finished.

Claim 4.3. $\mathsf{PA} \vdash$ "the bell rings" $\leftrightarrow \neg \mathsf{Con}_{\mathsf{PA}}$.

Proof. ($\rightarrow$): Reason in PA. Suppose that the bell rang at some stage $m$. Then there are the following two cases:

1. The bell rang at the stage $m$ in Case (a.1): Then $m$ is a proof of a formula of the form $\neg \mathsf{Pr}'^R_{\mathsf{PA}}(\ulcorner \psi \urcorner)$ and $\{f(0), \ldots, f(m), \psi\}$ is satisfiable. In this case, none of $f(0), \ldots, f(m)$ is $\neg\psi$ because of the satisfiability, and $f(m+1)$ is $\psi$ by the definition. Thus $\mathsf{Pr}'^R_{\mathsf{PA}}(\ulcorner \psi \urcorner)$ holds. Since this sentence is $\Sigma_1$, it is provable in PA by **$\Sigma_1$-compl**. Thus PA is inconsistent.
2. The bell rang at the stage $m$ in Case (b): Then $m$ is a proof of some $\varphi$, and $\{f(0), \ldots, f(m-1), \varphi\}$ is not satisfiable. Then $\neg\varphi$ is a tautological consequence of $\{f(0), \ldots, f(m-1)\}$. By the assumption, the sentence $f(0) \land \cdots \land f(m-1) \rightarrow \neg\varphi$ is provable, and hence $\neg\varphi$ is also provable because $f(0), \ldots, f(m-1)$ are all provable. It follows that PA is inconsistent.

($\leftarrow$): Argue in PA. Suppose that PA is inconsistent. Then even if the bell does not ring in Case (a.1), there is a proof $m$ of some $\varphi$ such that $\{f(0), \ldots, f(m-1), \varphi\}$ is not satisfiable. Therefore the bell must ring at some stage. ⊣

Claim 4.4. $\mathsf{PA} \vdash \forall x(\mathsf{Pr}_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'_{\mathsf{PA}}(x))$.

Proof. It is obvious that $\mathsf{PA} \vdash \neg$"the bell rings" $\rightarrow \forall x(\mathsf{Pr}_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'_{\mathsf{PA}}(x))$ by the definition of $f$.

Also $\mathsf{PA} \vdash \neg \mathsf{Con}_{\mathsf{PA}} \rightarrow \forall x(\mathsf{Fml}(x) \leftrightarrow \mathsf{Pr}_{\mathsf{PA}}(x))$ and $\mathsf{PA} \vdash$ "the bell rings" $\rightarrow \forall x(\mathsf{Fml}(x) \leftrightarrow \mathsf{Pr}'_{\mathsf{PA}}(x))$, where $\mathsf{Fml}(x)$ is a $\Delta_1$ formula naturally representing the set of all Gödel numbers of formulas. Thus $\mathsf{PA} \vdash$ "the bell rings" $\rightarrow \forall x(\mathsf{Pr}_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'_{\mathsf{PA}}(x))$ by Claim 4.3. ⊣

Therefore, $\mathsf{Prf}'_{\mathsf{PA}}(x)$ satisfies the first condition in the statement of the theorem.

Claim 4.5. $\mathsf{Pr}'^R_{\mathsf{PA}}(x)$ *satisfies* **D2**.

Proof. Since $\mathsf{PA} \vdash \mathsf{Con}_{\mathsf{PA}} \rightarrow \forall x(\mathsf{Pr}'_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'^R_{\mathsf{PA}}(x))$, we obtain $\mathsf{PA} \vdash \mathsf{Con}_{\mathsf{PA}} \rightarrow \forall x(\mathsf{Pr}_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'^R_{\mathsf{PA}}(x))$ by Claim 4.4. Thus $\mathsf{Pr}'^R_{\mathsf{PA}}(x)$ satisfies **D2** under the assumption of $\mathsf{Con}_{\mathsf{PA}}$ because $\mathsf{Pr}_{\mathsf{PA}}(x)$ satisfies **D2**.

https://doi.org/10.1017/jsl.2016.19 Published online by Cambridge University Press

We argue in PA: assume that the bell rings, then for each formula $\varphi$, $\mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ holds if and only if $V(\varphi) = 1$, where $V$ is a valuation taken in Procedure 2 in the construction of $f$. If $\mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\varphi \to \psi\urcorner)$ and $\mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ hold, then $V(\varphi \to \psi) = V(\varphi) = 1$ holds, and it implies $V(\psi) = 1$. Thus $\mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\psi\urcorner)$ holds.

Hence $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ satisfies **D2** under the assumption of "the bell rings". Therefore by Claim 4.3, $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ satisfies **D2** without any assumption. ⊣

Claim 4.6. $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ *satisfies* **D4**.

Proof. Let $\varphi$ be any sentence such that $\mathsf{PA} \vdash \neg\mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$. We argue in $\mathbb{N}$. Because PA is consistent, the bell does not ring at any stage by Claim 4.3. Let $p$ be the least proof of $\neg\mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ in PA, then $f(p)$ is $\neg\mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$. If $\{f(0), \dots, f(p), \varphi\}$ were satisfiable, then the bell rings at Stage $p$. Therefore, $\{f(0), \dots, f(p), \varphi\}$ is not satisfiable. Then $f(0) \land \cdots \land f(p) \to \neg\varphi$ is provable by our assumption. Since $f(0), \dots, f(p)$ are provable in PA, we conclude that $\neg\varphi$ is also provable in PA. ⊣

We have finished proving the theorem by these claims. ⊣

Let $\mathsf{Thm'}^R_{\mathsf{PA}}(M)$ be the set $\{\varphi : M \models \mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\varphi\urcorner)\}$.

Theorem 4.7. *There exists a proof predicate* $\mathsf{Prf'_{PA}}(x, y)$ *of* PA *satisfying the following conditions*:

1. $\mathsf{PA} \vdash \forall x(\mathsf{Pr_{PA}}(x) \leftrightarrow \mathsf{Pr'_{PA}}(x))$;
2. $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ *is universal.*

Proof. There is a proof predicate $\mathsf{Prf'_{PA}}(x, y)$ of PA satisfying the following conditions by Theorem 4.2:

1. $\mathsf{PA} \vdash \forall x(\mathsf{Pr_{PA}}(x) \leftrightarrow \mathsf{Pr'_{PA}}(x))$;
2. $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ satisfies **D2** and **D4**.

Let $T$ be any consistent complete extension of PA, then the theory $U_T = \mathsf{PA} + \{\mathsf{Pr'}^R_{\mathsf{PA}}(\ulcorner\varphi\urcorner) : \varphi \in T\}$ is consistent by Proposition 3.4 since $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ satisfies **D2** and **D4**. Thus, there exists a countable model $M$ of $U_T$. Then $\mathsf{Thm'}^R_{\mathsf{PA}}(M)$ is an extension of $T$. Since $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ satisfies **D2**, the theory $\mathsf{Thm'}^R_{\mathsf{PA}}(M)$ is consistent by Proposition 3.3. Therefore $\mathsf{Thm'}^R_{\mathsf{PA}}(M) = \{\varphi : T \vdash \varphi\}$ since $T$ is complete and consistent. This means that $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ is universal. ⊣

We say that an insane model $M$ is *going insane suddenly* if and only if for any initial segment $N$ of $M$ which is a sane model of PA, $\mathsf{Thm_{PA}}(N) = \mathsf{Thm_{PA}}(\mathbb{N})$. In Kikuchi and Kurahashi [5], the existence of an insane model which is going insane suddenly is proved. Here we give an alternative proof of this result.

Theorem 4.8. *There is a countable insane model which is going insane suddenly.*

Proof. By Theorem 4.7, there exists a proof predicate $\mathsf{Prf'_{PA}}(x, y)$ of PA satisfying the following two conditions:

1. $\mathsf{PA} \vdash \forall x(\mathsf{Pr_{PA}}(x) \leftrightarrow \mathsf{Pr'_{PA}}(x))$;
2. $\mathsf{Pr'}^R_{\mathsf{PA}}(x)$ is universal.

Let $M$ be any insane model of PA such that $\mathsf{Thm'}^R_{\mathsf{PA}}(M) = \mathsf{TA}$. Let $N$ be any initial segment of $M$ which is a sane model of PA. Towards a contradiction, suppose that $\mathsf{Thm_{PA}}(\mathbb{N}) \subsetneq \mathsf{Thm_{PA}}(N)$. Then there is a sentence $\varphi$ such that $\mathsf{PA} \nvdash \varphi$ and $N \models \mathsf{Pr_{PA}}(\ulcorner\varphi\urcorner)$. Let $\psi$ be the sentence $\mathsf{Pr_{PA}}(\ulcorner\varphi\urcorner)$, then $N \models \mathsf{Pr_{PA}}(\ulcorner\psi\urcorner)$ by **D3**.

https://doi.org/10.1017/jsl.2016.19 Published online by Cambridge University Press

Also we have $N \models \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg\psi \urcorner)$ because $N \models \mathsf{Con}_{\mathsf{PA}}$. Since $\mathsf{PA} \vdash \forall x(\mathsf{Pr}_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'_{\mathsf{PA}}(x))$, we have $N \models \mathsf{Pr}'_{\mathsf{PA}}(\ulcorner \psi \urcorner) \wedge \neg \mathsf{Pr}'_{\mathsf{PA}}(\ulcorner \neg\psi \urcorner)$. Therefore $N \models \mathsf{Pr}'^{R}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$. By the upward $\Sigma_1$-persistency, $M \models \mathsf{Pr}'^{R}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$, and thus $\psi \in \mathsf{Thm}'^{R}_{\mathsf{PA}}(M)$.

On the other hand, $\psi$ is false because $\varphi$ is not provable in PA. This contradicts the condition $\mathsf{Thm}'^{R}_{\mathsf{PA}}(M) = \mathsf{TA}$. Therefore, we conclude $\mathsf{Thm}_{\mathsf{PA}}(N) = \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$. ⊣

**§5. Antiuniversal Rosser predicates.** Finally, we prove the existence of a Rosser provability predicate $\mathsf{Pr}^{R}_{\mathsf{PA}}(x)$ such that $\mathsf{Thm}^{R}_{\mathsf{PA}}(M)$ is inconsistent for any insane model $M$ of PA. We say such a Rosser provability predicate antiuniversal.

DEFINITION 5.1. A Rosser provability predicate $\mathsf{Pr}^{R}_{\mathsf{PA}}(x)$ is *antiuniversal* if and only if $\mathsf{Thm}^{R}_{\mathsf{PA}}(M)$ is inconsistent for any insane model $M$ of PA.

THEOREM 5.2. *There exists a proof predicate* $\mathsf{Prf}'_{\mathsf{PA}}(x, y)$ *of* PA *satisfying the following conditions*:

1. $\mathsf{PA} \vdash \forall x(\mathsf{Pr}_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'_{\mathsf{PA}}(x))$;
2. $\mathsf{Pr}'^{R}_{\mathsf{PA}}(x)$ *is antiuniversal.*

PROOF. Let $\pi$ be any Rosser sentence of $\mathsf{Pr}^{R}_{\mathsf{PA}}(x)$, that is, $\pi$ be a $\Pi_1$ sentence satisfying $\mathsf{PA} \vdash \pi \leftrightarrow \neg \mathsf{Pr}^{R}_{\mathsf{PA}}(\ulcorner \pi \urcorner)$. Let $\psi$ be $\pi \wedge \overline{0} = \overline{0}$. Let $g$ be a recursive function enumerating all theorems of PA defined as follows:

$$g(m) = \begin{cases} \varphi & \text{if } m \text{ is a proof of } \varphi, \\ \overline{0} = \overline{0} & \text{if } m \text{ is not a proof of any formula.} \end{cases}$$

We define a function $f$ recursively as follows:

1. Let $f(m) = g(m)$ if none of $\pi$, $\neg\pi$, $\psi$, and $\neg\psi$ is among $g(0), \dots, g(m)$.
2. If $p$ is the least number such that $g(p)$ is one of $\pi$, $\neg\pi$, $\psi$, and $\neg\psi$, then let $f(p)$ be $\pi$ and let $f(p+1)$ be $\neg\psi$. After that, let $f(p+2+m) = g(p+m)$ for every $m$.

Let $\mathsf{Prf}'_{\mathsf{PA}}(x, y)$ be the $\Delta_1$ formula naturally representing $x = f(y)$.

CLAIM 5.3. $\mathsf{PA} \vdash \forall x(\mathsf{Pr}_{\mathsf{PA}}(x) \leftrightarrow \mathsf{Pr}'_{\mathsf{PA}}(x))$.

PROOF. Argue in PA. The required equivalence is obvious if there is no $m$ such that $g(m)$ is one of $\pi$, $\neg\pi$, $\psi$, and $\neg\psi$.

If there is $m$ such that $g(m)$ is one of $\pi$, $\neg\pi$, $\psi$, and $\neg\psi$, then $g(m)$ is provable in PA. Thus PA is inconsistent since the PA-provability of each of the above four sentences implies the inconsistency of PA by the formalization of Rosser's first incompleteness theorem. Thus both $f$ and $g$ eventually output all formulas. ⊣

Let $M$ be any insane model of PA. There is the least element $a$ of $M$ such that $M \models \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \rho \urcorner, a)$, where $\rho$ is one of $\pi$, $\neg\pi$, $\psi$, and $\neg\psi$. Then $M \models \mathsf{Prf}'_{\mathsf{PA}}(\ulcorner \pi \urcorner, a) \wedge \forall y \leq a \neg \mathsf{Prf}'_{\mathsf{PA}}(\ulcorner \neg\pi \urcorner, y)$ by the definition of $f$, and thus $M \models \mathsf{Pr}'^{R}_{\mathsf{PA}}(\ulcorner \pi \urcorner)$. Also we have $M \models \mathsf{Pr}'^{R}_{\mathsf{PA}}(\ulcorner \neg\psi \urcorner)$ since $\neg\neg\psi$ and $\psi$ are identical. Thus $\pi$ and $\neg\psi$ are in $\mathsf{Thm}'^{R}_{\mathsf{PA}}(M)$. Since $\mathsf{PA} \vdash \pi \wedge \neg\psi \rightarrow \overline{0} = \overline{1}$, $\mathsf{Thm}'^{R}_{\mathsf{PA}}(M)$ is inconsistent. Therefore, $\mathsf{Pr}'^{R}_{\mathsf{PA}}(x)$ is antiuniversal. ⊣

By Proposition 3.3 and Theorem 5.2, we obtain an alternative proof of the following Guaspari and Solovay's result.

Corollary 5.4. *There exists a Rosser provability predicate which does not satisfy* **D2**.

§**6. Acknowledgments.** This work was supported by JSPS KAKENHI Grant Numbers 24540125 and 26887045.

REFERENCES

[1] T. Arai, *Derivability conditions on Rosser's provability predicates*. ***Notre Dame Journal of Formal Logic***, vol. 31 (1990), no. 4, pp. 487–497.

[2] D. Guaspari and R. M. Solovay, *Rosser sentences*. ***Annals of Mathematical Logic***, vol. 16 (1979), no. 1, pp. 81–99.

[3] P. Hájek and P. Pudlák, *Two orderings of the class of all countable models of Peano arithmetic*, ***Model Theory of Algebra and Arithmetic***, Lecture Notes in Mathematics, vol. 834, Springer, Berlin Heidelberg, 1980, pp. 174–185.

[4] ———, ***Metamathematics of First-Order Arithmetic***, Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1993.

[5] M. Kikuchi and T. Kurahashi, *Illusory models of Peano arithmetic*, this Journal, vol. 81 (2016), pp. 1163–1175.

[6] G. Kreisel and G. Takeuti, *Formally self-referential propositions for cut free classical analysis and related systems*. ***Dissertationes Mathematicae (Rozprawy Matematyczne)***, vol. 118 (1974).

[7] T. Kurahashi, *Henkin sentences and local reflection principles for Rosser provability*. ***Annals of Pure and Applied Logic***, vol. 167 (2016), no. 2, pp. 73–94.

[8] P. Lindström, ***Aspects of Incompleteness***, second ed., Lecture Notes in Logic, vol. 10, A K Peters, Natick, MA, 2003.

[9] A. Mostowski, *A generalization of the incompleteness theorem*. ***Fundamenta Mathematicae***, vol. 49 (1961), pp. 205–232.

[10] J. B. Rosser, *Extensions of some theorems of Gödel and Church*, this Journal, vol. 1 (1936), no. 3, pp. 87–91.

GRADUATE SCHOOL OF SYSTEM INFORMATICS
KOBE UNIVERSITY, 1-1 ROKKODAI
NADA, KOBE 657-8501, JAPAN
*E-mail*: mkikuchi@kobe-u.ac.jp

DEPARTMENT OF NATURAL SCIENCES
NATIONAL INSTITUTE OF TECHNOLOGY
KISARAZU COLLEGE, 2-11-1 KIYOMIDAI-HIGASHI
KISARAZU, CHIBA 292-0041, JAPAN
*E-mail*: kurahashi@nebula.n.kisarazu.ac.jp