The Journal of Symbolic Logic
Volume 81, Number 3, September 2016

ILLUSORY MODELS OF PEANO ARITHMETIC

MAKOTO KIKUCHI AND TAISHI KURAHASHI

Abstract. By using a provability predicate of PA, we define $\mathsf{Thm}_{\mathsf{PA}}(M)$ as the set of theorems of PA in a model M of PA. We say a model M of PA is (1) illusory if $\mathsf{Thm}_{\mathsf{PA}}(M) \not\subseteq \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$, (2) heterodox if $\mathsf{Thm}_{\mathsf{PA}}(M) \not\subseteq \mathsf{TA}$, (3) sane if $M \models \mathsf{Con}_{\mathsf{PA}}$, and insane if it is not sane, (4) maximally sane if it is sane and $\mathsf{Thm}_{\mathsf{PA}}(M) \subseteq \mathsf{Thm}_{\mathsf{PA}}(N)$ implies $\mathsf{Thm}_{\mathsf{PA}}(M) = \mathsf{Thm}_{\mathsf{PA}}(N)$ for every sane model N of PA. We firstly show that M is heterodox if and only if it is illusory, and that $\mathsf{Thm}_{\mathsf{PA}}(M) \cap \mathsf{TA} \neq \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ for any illusory model M. Then we show that there exists a maximally sane model, every maximally sane model satisfies $\neg\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}$, and there exists a sane model of $\neg\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}$ which is not maximally sane. We define that an insane model is (5) illusory by nature if its every initial segment being a nonstandard model of PA is illusory, and (6) going insane suddenly if its every initial segment being a sane model of PA is not illusory. We show that there exists a model of PA which is illusory by nature, and we prove the existence of a model of PA which is going insane suddenly.

§1. Introduction. The provability in Peano arithmetic PA and the consistency of PA are formalized in the proofs of Gödel's incompleteness theorems by using a provability predicate $\mathsf{Pr}_{\mathsf{PA}}(x)$ of PA. By using this predicate, we define the set $\mathsf{Thm}_{\mathsf{PA}}(M)$ of theorems of PA in a model M of PA. That is, $\mathsf{Thm}_{\mathsf{PA}}(M) = \{\varphi : M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)\}$. It is clear that $\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ is equal to the set $\{\varphi : \mathsf{PA} \vdash \varphi\}$ of real theorems of PA, and $\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N}) \subseteq \mathsf{Thm}_{\mathsf{PA}}(M)$ for any nonstandard model M of PA since every nonstandard model of PA is an end-extension of $\mathbb{N}$ and $\mathsf{Pr}_{\mathsf{PA}}(x)$ is a Σ_1 formula. There may be new theorems of PA in nonstandard models of PA because it is possible that such models contain codes for nonstandard proofs of sentences from PA.

There exists a model M of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$ by Gödel's second incompleteness theorem, but $\mathsf{Thm}_{\mathsf{PA}}(M)$ is not a set of any interest when M is a model of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$ since $\mathsf{Thm}_{\mathsf{PA}}(M)$ is the set of all sentences if M satisfies $\neg\mathsf{Con}_{\mathsf{PA}}$. However, it does not mean that the structure of models of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$ are trivial nor that we know lucidly the mechanism how $0 = 1$ is derived in models of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$. Models of $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$ and proofs of $0 = 1$ in PA on such models are rather enigmatic and attractive objects which should be investigated and elucidated mathematically.

We introduce some terminology about models of PA in order to make exploration of nonstandard models of PA with new theorems of PA. We say that a model M of PA is *illusory* if $\mathsf{Thm}_{\mathsf{PA}}(M) \not\subseteq \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ and *heterodox* if $\mathsf{Thm}_{\mathsf{PA}}(M) \not\subseteq \mathsf{TA}$, where TA is the true arithmetic $\{\varphi : \mathbb{N} \models \varphi\}$. It is trivial that a heterodox model is illusory. We say that M is *sane* if $M \models \mathsf{Con}_{\mathsf{PA}}$ and *insane* if $M \models \neg\mathsf{Con}_{\mathsf{PA}}$.

Received March 9, 2015.
Key words and phrases. nonstandard models of arithmetic, provability predicates, Gödel's incompleteness theorems.

© 2016, Association for Symbolic Logic
0022-4812/16/8103-0022
DOI:10.1017/jsl.2015.52

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

Every insane model is illusory and heterodox since PA is consistent and every sentence is PA-provable in insane models of PA. We say that a sane model M is *maximally sane* if $\mathsf{Thm}_{\mathsf{PA}}(M) \subseteq \mathsf{Thm}_{\mathsf{PA}}(N)$ implies $\mathsf{Thm}_{\mathsf{PA}}(M) = \mathsf{Thm}_{\mathsf{PA}}(N)$ for every sane model N of PA.

In the following, we firstly show that every illusory model M of PA is heterodox. Hence M is heterodox if and only if it is illusory. We also prove that $\mathsf{Thm}_{\mathsf{PA}}(M) \cap \mathsf{TA} \neq \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ for any illusory model M. It means that there is a new true theorem of PA in any illusory model of PA. Then we show that there exists a maximally sane model, every maximally sane model satisfies $\neg\mathsf{Con}^2_{\mathsf{PA}}$ where $\mathsf{Con}^2_{\mathsf{PA}}$ is defined as the sentence $\neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\mathsf{Con}_{\mathsf{PA}}\urcorner)$ which means the consistency of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$, and there exists a sane model of $\neg\mathsf{Con}^2_{\mathsf{PA}}$ which is not maximally sane.

We can investigate the process of forfeiture of the consistency of PA and observe the history of madness in insane models of PA by considering initial segments of such insane models. We define that an insane model M of PA is *illusory by nature* if its every initial segment being a nonstandard model of PA is illusory, and we say that such an M is *going insane suddenly* if its every initial segment being a sane model of PA is not illusory. It is easy to show that there exists an insane model of PA which is not illusory by nature nor going insane suddenly. We prove that there exists a model of PA which is illusory by nature, and show that there exists a model of PA which is going insane suddenly by using results by Hájek, Krajíček and Pulák, and Solovay.

After the authors had finished preparation of publication of this paper, they have noticed that, under appropriate translations, some of the results in this paper had been proved as theorems, or are direct consequences of theorems about the E-tree which was introduced by Jensen and Ehrenfeucht [4]. In particular, Proposition 5.4 and Corollary 5.5 in this paper correspond to Proposition 2.3 and Theorem 3.1 in Misercque [9] respectively, and Theorem 5.6 in this paper corresponds to Theorem 7(iii) in Jensen and Ehrenfeucht [4]. However, the proofs of these propositions are essentially different from the known proofs, and the motivations and settings of this paper and the arguments on the E-tree do not coincide. See also Simmons [11] for more information on the E-tree. The authors would like to thank Volodya Shavrukov for valuable information and comments on the previous version of this paper.

§2. Preliminaries. In this paper, we consider only theories and structures in the language of first-order arithmetic $\{+, \times, 0, 1, <\}$. For each natural number n, $\bar{n}$ denotes the numeral for n. Without any notice, we use the fact that Peano arithmetic PA is sound, that is, every theorem of PA is true in the standard model $\mathbb{N}$ of arithmetic. Let TA be the theory $\{\varphi : \mathbb{N} \models \varphi\}$ which is called the *true arithmetic*. Thus TA is the only complete sound extension of PA. It is well-known that there are nonstandard models of TA, that is, models which are not isomorphic to $\mathbb{N}$.

Let M and N be models of PA. We say M is an end-extension of N, denoted by $N \subseteq_e M$, if and only if N is a substructure of M and for any $a, b \in M$, $b \in N$ whenever $M \models b < a$ and $a \in N$. In this case, we also say N is an initial segment of M. Every model M of PA always has an initial segment N which is isomorphic to $\mathbb{N}$, and thus we identify such an N and $\mathbb{N}$. We say a formula is Σ_1 (resp. Π_1) if and only if it is PA-provably equivalent to a formula which is in the strict

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

Σ_1 (resp. Π_1) form. Every Σ_1 formula is upward Σ_1-persistent with respect to the relation $\subseteq_e$, that is, for any models M and N of PA such that $N \subseteq_e M$, any elements $a_1, \ldots, a_n$ of N and any Σ_1 formula $\varphi(x_1, \ldots, x_n)$, if $N \models \varphi(a_1, \ldots, a_n)$, then $M \models \varphi(a_1, \ldots, a_n)$.

We say a formula is Δ_1 if and only if it is both a Σ_1 formula and a Π_1 formula. Let $\mathsf{Prf}_{\mathsf{PA}}(x, y)$ be a Δ_1 formula representing the relation that 'y is the Gödel number of a PA-proof of a sentence whose Gödel number is x'. Such a formula is called a *proof predicate* of PA. Note that a similar Δ_1 formula $\mathsf{Prf}_T(x, y)$ can be taken for any recursively enumerable extension T of PA by using Craig's trick [1].

Let $\mathsf{Pr}_{\mathsf{PA}}(x)$ be the Σ_1 formula $\exists y \mathsf{Prf}_{\mathsf{PA}}(x, y)$ which is called a *provability predicate* of PA. Then $\mathsf{Pr}_{\mathsf{PA}}(x)$ weakly represents the set of all Gödel numbers of theorems of PA, that is, for any natural number n, $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\bar{n})$ if and only if n is the Gödel number of a theorem of PA. Let $\mathsf{Con}_{\mathsf{PA}}$ be the Π_1 sentence $\neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner 0 = 1 \urcorner)$ asserting the consistency of PA, where $\ulcorner \varphi \urcorner$ is the numeral for the Gödel number of φ for each formula φ. In this paper, we assume that the formula $\mathsf{Pr}_{\mathsf{PA}}(x)$ satisfies the following conditions **D2** and **Σ_1-comp**:

D2: $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \to \psi \urcorner) \to (\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to (\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner))$.

Σ_1-comp: If φ is a Σ_1 sentence, then $\mathsf{PA} \vdash \varphi \to \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$.

Then Gödel's second incompleteness theorem holds for $\mathsf{Pr}_{\mathsf{PA}}(x)$, that is, $\mathsf{PA} \nvdash \mathsf{Con}_{\mathsf{PA}}$. Hence there is a nonstandard model M of $\mathsf{PA} + \neg \mathsf{Con}_{\mathsf{PA}}$ by the completeness theorem. Then the sentence $\exists y \mathsf{Prf}_{\mathsf{PA}}(\ulcorner 0 = 1 \urcorner, y)$ is true in such a model M, and thus M has a nonstandard proof of $0 = 1$.

Moreover, it follows from the following fact that every model of PA has an end-extension which is a model of $\mathsf{PA} + \neg \mathsf{Con}_{\mathsf{PA}}$.

FACT 2.1. *Let M be any model of* PA *and φ be any sentence. If $M \models \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg \varphi \urcorner)$, then there is a model N of $\mathsf{PA} + \varphi$ such that $M \subseteq_e N$.*

Fact 2.1 is a consequence of the arithmetized completeness theorem (see [7] and [12]).

Let M be a model of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$, then M is also a model of $\mathsf{PA} + \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \mathsf{Con}_{\mathsf{PA}} \urcorner)$ by the formalized version of Gödel's second incompleteness theorem. Then there is a model N of $\mathsf{PA} + \neg \mathsf{Con}_{\mathsf{PA}}$ such that $M \subseteq_e N$ by Fact 2.1.

Let $t(x)$ be a set of formulas whose only free variable is x. We say a theory T *locally realizes* $t(x)$ if and only if there is a formula $\xi(x)$ satisfying the following two conditions:

1. $T \nvdash \forall x \neg \xi(x)$,
2. for all formulas $\varphi(x)$ in $t(x)$, $T \vdash \xi(x) \to \varphi(x)$.

We say T *locally omits* $t(x)$ if and only if T does not locally realize $t(x)$. A model M of PA *omits* $t(x)$ if and only if for any $a \in M$, there exists $\varphi(x) \in t(x)$ such that $M \models \neg \varphi(a)$. The omitting types theorem states that if a theory T is consistent and locally omits a set $t(x)$ of formulas, then there is a countable model of T omitting $t(x)$.

For sentences φ and ψ, let α and β be the Π_1 sentences $\forall y(\mathsf{Prf}_{\mathsf{PA}}(\ulcorner \varphi \urcorner, y) \to \exists z \leq y \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \psi \urcorner, z))$ and $\forall z(\mathsf{Prf}_{\mathsf{PA}}(\ulcorner \psi \urcorner, z) \to \exists y < z \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \varphi \urcorner, y))$ respectively. Such sentences α and β are said to be witness comparison sentences, and satisfy the following proposition (see [2, 7]).

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

PROPOSITION 2.2.

1. $\mathsf{PA} \vdash \alpha \vee \beta$.
2. $\mathsf{PA} \vdash \alpha \wedge \beta \rightarrow \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \wedge \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$.

In this paper, we investigate the sets of theorems of PA proved in nonstandard models of PA. Here we define the sets $\mathsf{Thm}_{\mathsf{PA}}(M)$ which are main objects of this paper.

DEFINITION 2.3. Let M be any model of PA. Define $\mathsf{Thm}_{\mathsf{PA}}(M)$ to be the set $\{\varphi : \varphi \text{ is a (standard) sentence and } M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)\}$.

We can say that $\mathsf{Thm}_{\mathsf{PA}}(M)$ is the set of all (standard) theorems of PA proved in the model M. We obtain the following basic proposition about the sets $\mathsf{Thm}_{\mathsf{PA}}(M)$.

PROPOSITION 2.4. *Let M and N be any models of* PA.

1. $\mathsf{Thm}_{\mathsf{PA}}(M)$ *is closed under modus ponens.*
2. *If* $M \subseteq_e N$, *then* $\mathsf{Thm}_{\mathsf{PA}}(M) \subseteq \mathsf{Thm}_{\mathsf{PA}}(N)$. *In particular,* $\mathsf{Thm}_{\mathsf{PA}}(M)$ *always contains the set* $\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ *of all actual theorems of* PA.
3. $M \models \mathsf{Con}_{\mathsf{PA}}$ *if and only if there is a sentence* φ *not contained in* $\mathsf{Thm}_{\mathsf{PA}}(M)$.

PROOF.

1. Suppose that $\varphi \rightarrow \psi$ and φ are in $\mathsf{Thm}_{\mathsf{PA}}(M)$, then $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \rightarrow \psi \urcorner) \wedge \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$. By **D2**, we obtain $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$. Therefore $\psi \in \mathsf{Thm}_{\mathsf{PA}}(M)$.
2. Let φ be in $\mathsf{Thm}_{\mathsf{PA}}(M)$, then $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$. Since $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$ is a Σ_1 sentence and N is an end-extension of M, $N \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$ by the upward Σ_1-persistency. Then we obtain $\varphi \in \mathsf{Thm}_{\mathsf{PA}}(N)$. Therefore $\mathsf{Thm}_{\mathsf{PA}}(M) \subseteq \mathsf{Thm}_{\mathsf{PA}}(N)$.

   The latter statement is obvious since every model of PA is an end-extension of $\mathbb{N}$.
3. This follows from the observation that PA proves $\neg \mathsf{Con}_{\mathsf{PA}} \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$ for all sentences φ. ⊣

We summarize the terminology about models of PA in the previous section which will be used throughout this paper in the following definition.

DEFINITION 2.5. Let M be any model of PA.

1. M is *illusory* if and only if $\mathsf{Thm}_{\mathsf{PA}}(M) \not\subseteq \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$.
2. M is *heterodox* if and only if $\mathsf{Thm}_{\mathsf{PA}}(M) \not\subseteq \mathsf{TA}$.
3. M is *sane* if and only if $\mathsf{Thm}_{\mathsf{PA}}(M)$ does not contain some sentence, and M is *insane* if and only if M is not sane.

It is easy to see that every insane model is heterodox, and every heterodox model is illusory. By the previous proposition, M is sane if and only if $M \models \mathsf{Con}_{\mathsf{PA}}$. Every model of TA is obviously not illusory.

We say a model M of PA is *realistic* if and only if M is not illusory, and M is *orthodox* if and only if M is not heterodox, although we do not use these terminology in this paper.

§3. Illusory sane models. In this section, we consider the family $\mathcal{T} = \{\mathsf{Thm}_{\mathsf{PA}}(M) : M \text{ is a sane model of PA}\}$, and prove several basic properties concerning this family. First, we show that there is a sane model which is illusory.

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

Then we prove that such a model M has both a true Π_1 and a false Σ_1 theorems which are not actually provable. Moreover, we show the size of the family $\mathcal{T}$ is $2^{\aleph_0}$. Next, we investigate incompatible models of PA, that is, models whose theorems are mutually not included.

Recall that the formalized version of the second incompleteness theorem $\mathsf{PA} \vdash \mathsf{Con}_{\mathsf{PA}} \rightarrow \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \mathsf{Con}_{\mathsf{PA}} \urcorner)$ holds (see Lindström [7]). This means that for every sane model M of PA, $\mathsf{Con}_{\mathsf{PA}}$ is not contained in $\mathsf{Thm}_{\mathsf{PA}}(M)$. On the other hand, there is a sane model having $\neg \mathsf{Con}_{\mathsf{PA}}$ as a theorem.

PROPOSITION 3.1. *There is a sane model M of* PA *which is illusory.*

PROOF. Since PA is sound, we obtain $\mathsf{PA} \nvdash \mathsf{Con}_{\mathsf{PA}} \rightarrow \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \urcorner)$ by using Löb's theorem (cf. Lindström [7]). Hence there is a model M of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$ such that $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \urcorner)$. Then $\neg \mathsf{Con}_{\mathsf{PA}} \in \mathsf{Thm}_{\mathsf{PA}}(M)$. Since $\neg \mathsf{Con}_{\mathsf{PA}}$ is not a theorem of PA, M is illusory. ⊣

Because $\mathsf{Con}_{\mathsf{PA}}$ is true, such a sane model M contains a false Σ_1 theorem which is actually unprovable in PA. The following proposition says that if M is illusory, then $\mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ always contains both a true sentence and a false sentence.

PROPOSITION 3.2. *Suppose that M is illusory. Then*

1. $\mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ *contains a true* Π_1 *sentence*; *and*
2. $\mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ *contains a false* Σ_1 *sentence.*

PROOF. Let φ be a sentence in $\mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$.

1. Let π be a Π_1 sentence satisfying the following equivalence which is obtained by using the fixed-point lemma (see [7] for details):

$$\mathsf{PA} \vdash \pi \leftrightarrow \forall y(\mathsf{Prf}_{\mathsf{PA}}(\ulcorner \pi \urcorner, y) \rightarrow \exists z \leq y \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \varphi \urcorner, z)).$$

   We prove that π is true and contained in $\mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$.

   Define ξ to be the sentence $\exists z(\mathsf{Prf}_{\mathsf{PA}}(\ulcorner \varphi \urcorner, z) \wedge \forall y < z \neg \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \pi \urcorner, y))$. Then $\mathsf{PA} \vdash \xi \rightarrow \pi$ by Proposition 2.2.

   Suppose $\mathbb{N} \models \neg\pi$, then $\mathsf{PA} \vdash \neg\pi$ since $\neg\pi$ is a Σ_1 sentence. Since $\mathsf{PA} \vdash \neg\pi \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \pi \urcorner)$ by the choice of π, we obtain $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \pi \urcorner)$. By the soundness of PA, $\mathsf{PA} \vdash \pi$, this contradicts the consistency of PA. Therefore π is true.

   Suppose $\mathsf{PA} \vdash \pi$, then $\mathbb{N} \models \neg\pi$ by the choice of π because $\mathsf{PA} \nvdash \varphi$. This contradicts the above observation. Hence $\pi \notin \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$.

   By the choice of ξ, $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \wedge \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \pi \urcorner) \rightarrow \xi$. Because ξ is Σ_1, $\mathsf{PA} \vdash \xi \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \xi \urcorner)$. Also $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \xi \urcorner) \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \pi \urcorner)$ since $\mathsf{PA} \vdash \xi \rightarrow \pi$. By combining these observations, we obtain $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \wedge \neg \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \pi \urcorner) \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \pi \urcorner)$. Therefore $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \pi \urcorner)$. Since $\varphi \in \mathsf{Thm}_{\mathsf{PA}}(M)$, π is also in $\mathsf{Thm}_{\mathsf{PA}}(M)$.
2. Let σ be the Σ_1 sentence $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$. Since $\varphi \notin \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$, σ is false. Also σ is not in $\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$. Since $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \sigma \urcorner)$ by Σ_1**-comp**, $\sigma \in \mathsf{Thm}_{\mathsf{PA}}(M)$. ⊣

COROLLARY 3.3. *M is illusory if and only if M is heterodox.*

PROOF. We have already mentioned that each heterodox model is illusory. Assume that M is illusory. Then $\mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$ contains a false Σ_1 sentence by Proposition 3.2.2. Thus $\mathsf{Thm}_{\mathsf{PA}}(M) \nsubseteq \mathsf{TA}$, and hence T is heterodox. ⊣

The following FGH-principle tells us more things about the family $\mathcal{T}$.

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

FACT 3.4 (FGH-principle (see [7, 13])). *For any* Σ_1 *sentence* φ, *there is a sentence* ψ *such that* $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} \vdash \varphi \leftrightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$.

By the FGH-principle, it is not hard to check that for any sane models M and N of PA, $\mathsf{Thm}_{\mathsf{PA}}(M) = \mathsf{Thm}_{\mathsf{PA}}(N)$ if and only if M and N satisfy the same Σ_1 sentences. Then we show the size of the family $\mathcal{T}$ is $2^{\aleph_0}$ by using the following fact due to Mostowski.

DEFINITION 3.5. Let T be a consistent theory.

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

FACT 3.6 (Mostowski [10]). *Let* T *be a recursively enumerable consistent extension of* PA. *Then there is a* Σ_1 *formula* $\xi(x)$ *which is independent over* T.

PROPOSITION 3.7. *The size of the family* $\mathcal{T}$ *is* $2^{\aleph_0}$.

PROOF. Notice that the possible largest size of $\mathcal{T}$ is $2^{\aleph_0}$.

Let $T = \mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$, and $\xi(x)$ be a Σ_1 formula independent over T. Let f and g be any different functions from ω to $\{0, 1\}$. Then there are models M and N of $T + \{\xi(\bar{n})^{f(n)} : n \in \omega\}$ and $T + \{\xi(\bar{n})^{g(n)} : n \in \omega\}$ respectively. Since f and g are different, M and N do not satisfy the same Σ_1 sentences. Hence $\mathsf{Thm}_{\mathsf{PA}}(M)$ and $\mathsf{Thm}_{\mathsf{PA}}(N)$ are different elements of $\mathcal{T}$ by the FGH-principle.

Since there are $2^{\aleph_0}$ many functions from ω to $\{0, 1\}$, the family $\mathcal{T}$ contains $2^{\aleph_0}$ elements. ⊣

From the proof of Proposition 3.7, we obtain models M and N of PA such that $\mathsf{Thm}_{\mathsf{PA}}(M) \not\subseteq \mathsf{Thm}_{\mathsf{PA}}(N)$ and $\mathsf{Thm}_{\mathsf{PA}}(N) \not\subseteq \mathsf{Thm}_{\mathsf{PA}}(M)$. Such models M and N are said to be *incompatible*. Also we say that models M and N are *conflicting* if and only if there is a sentence φ such that $\varphi \in \mathsf{Thm}_{\mathsf{PA}}(M)$ and $\neg\varphi \in \mathsf{Thm}_{\mathsf{PA}}(N)$.

THEOREM 3.8. *Let* M *and* N *be models of* PA. *The following are equivalent*:

1. M *and* N *are conflicting and sane.*
2. M *and* N *are incompatible.*

PROOF. $(1 \Rightarrow 2)$: Suppose that M and N are conflicting sane models. Then there is a sentence φ such that $\varphi \in \mathsf{Thm}_{\mathsf{PA}}(M)$ and $\neg\varphi \in \mathsf{Thm}_{\mathsf{PA}}(N)$. Since M and N are sane, $\neg\varphi \notin \mathsf{Thm}_{\mathsf{PA}}(M)$ and $\varphi \notin \mathsf{Thm}_{\mathsf{PA}}(N)$. Therefore they are incompatible.

$(2 \Rightarrow 1)$: Suppose that M and N are incompatible. It is obvious that they are sane, and thus it suffices to show that they are conflicting.

Let $\varphi \in \mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(N)$ and $\psi \in \mathsf{Thm}_{\mathsf{PA}}(N) \setminus \mathsf{Thm}_{\mathsf{PA}}(M)$. Define ξ and η to be the Π_1 sentences $\forall y(\mathsf{Prf}_{\mathsf{PA}}(\ulcorner \varphi \urcorner, y) \to \exists z \leq y \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \psi \urcorner, z))$ and $\forall z(\mathsf{Prf}_{\mathsf{PA}}(\ulcorner \psi \urcorner, z) \to \exists y < z \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \varphi \urcorner, y))$ respectively.

Since $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \wedge \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$, we have $M \models \neg\xi$. Since $\neg\xi$ is Σ_1, $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg\xi \urcorner)$ by Σ_1**-comp**. Since $\mathsf{PA} \vdash \neg\xi \to \eta$ by Proposition 2.2, we obtain $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg\xi \urcorner) \to \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \eta \urcorner)$. Therefore $\eta \in \mathsf{Thm}_{\mathsf{PA}}(M)$. Also $N \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg\eta \urcorner)$ in a similar way, and thus $\neg\eta \in \mathsf{Thm}_{\mathsf{PA}}(N)$. Hence M and N are conflicting. ⊣

COROLLARY 3.9. *Let* M *and* N *be any incompatible models of* PA. *For any model* K *of* PA, *if* $\mathsf{Thm}_{\mathsf{PA}}(M) \cup \mathsf{Thm}_{\mathsf{PA}}(N) \subseteq \mathsf{Thm}_{\mathsf{PA}}(K)$, *then* K *is insane.*

COROLLARY 3.10. *For any incompatible models* M *and* N *of* PA, *there is no* $K \models \mathsf{PA}$ *such that* $\mathsf{Thm}_{\mathsf{PA}}(K) = \mathsf{Thm}_{\mathsf{PA}}(M) \cup \mathsf{Thm}_{\mathsf{PA}}(N)$.

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

PROOF. Let K be any model of PA satisfying $\mathsf{Thm}_{\mathsf{PA}}(M) \cup \mathsf{Thm}_{\mathsf{PA}}(N) \subseteq \mathsf{Thm}_{\mathsf{PA}}(K)$. By the previous Corollary, K satisfies $\neg\mathsf{Con}_{\mathsf{PA}}$.

Let $\varphi \in \mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(N)$ and $\psi \in \mathsf{Thm}_{\mathsf{PA}}(N) \setminus \mathsf{Thm}_{\mathsf{PA}}(M)$. Since $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \wedge \psi \urcorner) \leftrightarrow (\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \wedge \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner))$, the sentence $\varphi \wedge \psi$ is not contained in $\mathsf{Thm}_{\mathsf{PA}}(M) \cup \mathsf{Thm}_{\mathsf{PA}}(N)$. Therefore $\mathsf{Thm}_{\mathsf{PA}}(M) \cup \mathsf{Thm}_{\mathsf{PA}}(N) \subsetneq \mathsf{Thm}_{\mathsf{PA}}(K)$. ⊣

By using the FGH-principle, we also prove a similar result for the intersection of $\mathsf{Thm}_{\mathsf{PA}}(M)$ and $\mathsf{Thm}_{\mathsf{PA}}(N)$.

PROPOSITION 3.11. *For any incompatible models M and N of* PA, *there is no $K \models \mathsf{PA}$ such that* $\mathsf{Thm}_{\mathsf{PA}}(K) = \mathsf{Thm}_{\mathsf{PA}}(M) \cap \mathsf{Thm}_{\mathsf{PA}}(N)$.

PROOF. Let $\varphi \in \mathsf{Thm}_{\mathsf{PA}}(M) \setminus \mathsf{Thm}_{\mathsf{PA}}(N)$ and $\psi \in \mathsf{Thm}_{\mathsf{PA}}(N) \setminus \mathsf{Thm}_{\mathsf{PA}}(M)$. Since $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \vee \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$ is a Σ_1 sentence, there is a sentence δ such that $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \delta \urcorner) \leftrightarrow (\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \vee \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner))$ by the FGH-principle. Since M and N are models of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$, both M and N satisfy $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \delta \urcorner)$. Hence $\delta \in \mathsf{Thm}_{\mathsf{PA}}(M) \cap \mathsf{Thm}_{\mathsf{PA}}(N)$.

Suppose that a model K of PA satisfies $\mathsf{Thm}_{\mathsf{PA}}(K) = \mathsf{Thm}_{\mathsf{PA}}(M) \cap \mathsf{Thm}_{\mathsf{PA}}(N)$, then $\mathsf{Thm}_{\mathsf{PA}}(K)$ must contain δ and K satisfies $\mathsf{Con}_{\mathsf{PA}}$. Thus $K \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \delta \urcorner)$, and then $K \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \vee \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \psi \urcorner)$. Hence $\mathsf{Thm}_{\mathsf{PA}}(K)$ contains at least one of φ and ψ. However neither φ nor ψ is contained in $\mathsf{Thm}_{\mathsf{PA}}(M) \cap \mathsf{Thm}_{\mathsf{PA}}(N)$. This is a contradiction. ⊣

Next we show that the intersection of $\mathsf{Thm}_{\mathsf{PA}}(M)$ and $\mathsf{Thm}_{\mathsf{PA}}(N)$ may contain a nontrivial set $\mathsf{Thm}_{\mathsf{PA}}(K)$. Let $\mathsf{Con}^2_{\mathsf{PA}}$ be the sentence $\neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg\mathsf{Con}_{\mathsf{PA}} \urcorner)$. Notice that $\mathsf{PA} \vdash \mathsf{Con}^2_{\mathsf{PA}} \rightarrow \mathsf{Con}_{\mathsf{PA}}$ by using **Σ_1-comp**.

PROPOSITION 3.12. *For any model K of* $\mathsf{PA} + \mathsf{Con}^2_{\mathsf{PA}}$, *there are incompatible models M and N of* PA *such that $K \subseteq_e M$ and $K \subseteq_e N$. Consequently,* $\mathsf{Thm}_{\mathsf{PA}}(K) \subsetneq \mathsf{Thm}_{\mathsf{PA}}(M) \cap \mathsf{Thm}_{\mathsf{PA}}(N)$.

PROOF. Let K be any model of $\mathsf{PA} + \mathsf{Con}^2_{\mathsf{PA}}$. Let ξ be a Rosser sentence of the theory $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$, that is, ξ is a Π_1 sentence satisfying

$$\mathsf{PA} \vdash \xi \leftrightarrow \forall y(\mathsf{Prf}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}(\ulcorner \xi \urcorner, y) \rightarrow \exists z \leq y \mathsf{Prf}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}(\ulcorner \neg\xi \urcorner, z)).$$

Let η be the Π_1 sentence $\forall z(\mathsf{Prf}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}(\ulcorner \neg\xi \urcorner, z) \rightarrow \exists y < z \mathsf{Prf}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}(\ulcorner \xi \urcorner, y))$. Then $\mathsf{PA} \vdash \xi \vee \eta$. Also $\mathsf{PA} \vdash \mathsf{Con}^2_{\mathsf{PA}} \rightarrow \neg\mathsf{Pr}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}(\ulcorner \xi \urcorner) \wedge \neg\mathsf{Pr}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}(\ulcorner \eta \urcorner)$. Therefore $K \models \neg\mathsf{Pr}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}(\ulcorner \xi \urcorner) \wedge \neg\mathsf{Pr}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}(\ulcorner \eta \urcorner)$.

By Fact 2.1, there are end-extensions M and N of K such that $M \models \mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \neg\xi$ and $N \models \mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \neg\eta$. Then by **$\Sigma_1$-comp**, $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg\xi \urcorner)$ and $N \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \xi \urcorner)$ since $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg\eta \urcorner) \rightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \xi \urcorner)$. Thus M and N are conflicting. Since both M and N are sane, M and N are incompatible by Theorem 3.8. Also since M and N are end-extensions of K, $\mathsf{Thm}_{\mathsf{PA}}(K) \subseteq \mathsf{Thm}_{\mathsf{PA}}(M) \cap \mathsf{Thm}_{\mathsf{PA}}(N)$. By the previous proposition, this set inclusion is proper. ⊣

§4. Maximally sane models.

Corollary 3.3 says $\mathsf{Thm}_{\mathsf{PA}}(M)$ cannot be the true arithmetic TA. Then it is natural to ask whether there is a model M such that $\mathsf{Thm}_{\mathsf{PA}}(M)$ is consistent and complete. The negative answer to this question is easily obtained by considering the formalization of Rosser's first incompleteness theorem.

PROPOSITION 4.1. *Let M be any sane model of* PA. *Then* $\mathsf{Thm}_{\mathsf{PA}}(M)$ *is not complete.*

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

PROOF. Let π be a Rosser sentence of PA. By the formalized Rosser's incompleteness theorem, we have $\mathsf{PA} \vdash \mathsf{Con}_{\mathsf{PA}} \to \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\pi\urcorner) \wedge \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\pi\urcorner)$ (cf. Lindström [7]). Thus $M \models \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\pi\urcorner) \wedge \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\pi\urcorner)$, and hence neither π nor $\neg\pi$ is contained in $\mathsf{Thm}_{\mathsf{PA}}(M)$. ⊣

Thus no element $\mathsf{Thm}_{\mathsf{PA}}(M)$ of $\mathcal{T}$ is complete. However, this does not exclude a possibility of the existence of a maximal element of $\mathcal{T}$ with respect to the set inclusion.

In this section, we investigate maximally sane models of PA, that is, models M such that $\mathsf{Thm}_{\mathsf{PA}}(M)$ is maximal in $\mathcal{T}$. We prove that (1) there exists a maximally sane model of PA, moreover, $\mathcal{T}$ contains $2^{\aleph_0}$ maximal elements, (2) every maximally sane model M must be a model of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \neg\mathsf{Con}^2_{\mathsf{PA}}$, and (3) there is a model M of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \neg\mathsf{Con}^2_{\mathsf{PA}}$ which is not maximally sane.

DEFINITION 4.2. We say a model M of PA is *maximally sane* if and only if M is sane and there is no sane model N of PA such that $\mathsf{Thm}_{\mathsf{PA}}(N) \supsetneq \mathsf{Thm}_{\mathsf{PA}}(M)$.

For each model M of PA, let $\mathsf{Pr}_{\mathsf{PA}}(M)$ be the set $\{\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) : \varphi \in \mathsf{Thm}_{\mathsf{PA}}(M)\}$. Then it is easy to see that a model M of PA is maximally sane if and only if $M \models \mathsf{Con}_{\mathsf{PA}}$ and for any sentence φ, $\mathsf{PA} + \mathsf{Pr}_{\mathsf{PA}}(M) + \{\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner), \mathsf{Con}_{\mathsf{PA}}\}$ is inconsistent whenever $M \models \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$.

First, we prove the existence of a maximally sane model.

THEOREM 4.3. *For any sane model K of* PA, *there is a maximally sane model M of* PA *such that* $\mathsf{Thm}_{\mathsf{PA}}(K) \subseteq \mathsf{Thm}_{\mathsf{PA}}(M)$.

PROOF. Let $\{\varphi_i\}_{i\in\omega}$ be an enumeration of all the sentences. We define a sequence $\{T_i\}_{i\in\omega}$ of extensions of PA inductively as follows:

1. Let T_0 be $\mathsf{PA} + \mathsf{Pr}_{\mathsf{PA}}(K) + \mathsf{Con}_{\mathsf{PA}}$;
2. Suppose that we have already defined the theory T_i, and then

$$T_{i+1} := \begin{cases} T_i + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_i\urcorner) & \text{if this theory is consistent;} \\ T_i & \text{otherwise.} \end{cases}$$

Let T be $\bigcup_{i\in\omega} T_i$. Then it is easy to check that the theory T is consistent. Let M be any model of T, then M is trivially a model of $\mathsf{PA} + \mathsf{Pr}_{\mathsf{PA}}(K) + \mathsf{Con}_{\mathsf{PA}}$. Hence $\mathsf{Thm}_{\mathsf{PA}}(K) \subseteq \mathsf{Thm}_{\mathsf{PA}}(M)$.

We show that M is maximally sane. Let φ be any sentence such that $M \models \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$. Then $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ is not in T. This means that for $i \in \omega$ such that φ_i is φ, the theory $T_i + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_i\urcorner)$ is inconsistent by the definition of T. Since this theory is a subtheory of $U := \mathsf{PA} + \mathsf{Pr}_{\mathsf{PA}}(M) + \{\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner), \mathsf{Con}_{\mathsf{PA}}\}$, U is also inconsistent. Therefore M is maximally sane. ⊣

Then we prove there are $2^{\aleph_0}$ maximal elements in $\mathcal{T}$.

LEMMA 4.4. *For any recursively enumerable extension T of* PA, *if $T + \mathsf{Con}_{\mathsf{PA}}$ is consistent, then there is a Π_1 sentence φ such that both $T + \mathsf{Con}_{\mathsf{PA}} + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ and $T + \mathsf{Con}_{\mathsf{PA}} + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\varphi\urcorner)$ are consistent.*

PROOF. Suppose that $T + \mathsf{Con}_{\mathsf{PA}}$ is consistent. Let φ be a Π_1 sentence satisfying the equivalence

$$\mathsf{PA} \vdash \varphi \leftrightarrow \forall y(\mathsf{Prf}_T(\ulcorner\varphi_1\urcorner, y) \to \exists z \le y \mathsf{Prf}_T(\ulcorner\varphi_0\urcorner, z)),$$

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

where φ_0 and φ_1 are sentences $\mathsf{Con}_{\mathsf{PA}} \to \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ and $\mathsf{Con}_{\mathsf{PA}} \to \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\varphi\urcorner)$ respectively. Let ψ be the Π_1 sentence $\forall z(\mathsf{Prf}_T(\ulcorner\varphi_0\urcorner, z) \to \exists y < z\mathsf{Prf}_T(\ulcorner\varphi_1\urcorner, y))$. Then $\mathsf{PA} \vdash \neg\psi \to \varphi$ by Proposition 2.2.

Towards a contradiction, we suppose that at least one of φ_0 and φ_1 is provable in T. Let p be the least Gödel number of any proof of φ_0 or φ_1 in T. If p is the Gödel number of a proof of φ_0, then $\neg\psi$ is true in $\mathbb{N}$, and thus $\mathsf{PA} \vdash \neg\psi$ by the Σ_1-completeness of PA. Therefore $\mathsf{PA} \vdash \varphi$, and hence $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$. Since $T \vdash \mathsf{Con}_{\mathsf{PA}} \to \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$, $T + \mathsf{Con}_{\mathsf{PA}}$ is inconsistent. This is a contradiction. Also if p is the Gödel number of a proof of φ_1, we obtain a contradiction in a similar way.

Therefore neither φ_0 nor φ_1 is provable in T. This means both $T + \mathsf{Con}_{\mathsf{PA}} + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ and $T + \mathsf{Con}_{\mathsf{PA}} + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\varphi\urcorner)$ are consistent. ⊣

THEOREM 4.5. *The family $\mathcal{T}$ contains $2^{\aleph_0}$ maximal elements.*

PROOF. Let S be the set of all finite binary sequences, including the empty sequence ϵ. We define a family $\{T_\sigma\}_{\sigma\in S}$ of finite extensions of PA such that $T_\sigma + \mathsf{Con}_{\mathsf{PA}}$ is consistent for each $\sigma \in S$ recursively as follows:

1. Let T_ϵ be PA. Then $T_\epsilon + \mathsf{Con}_{\mathsf{PA}}$ is consistent.
2. Suppose that we have already defined the theory T_σ such that $T_\sigma + \mathsf{Con}_{\mathsf{PA}}$ is consistent. By Lemma 4.4, there is a Π_1 sentence φ_σ such that both $T_\sigma + \mathsf{Con}_{\mathsf{PA}} + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_\sigma\urcorner)$ and $T_\sigma + \mathsf{Con}_{\mathsf{PA}} + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\varphi_\sigma\urcorner)$ are consistent. Let $T_{\sigma 0}$ and $T_{\sigma 1}$ be theories $T_\sigma + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_\sigma\urcorner)$ and $T_\sigma + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\varphi_\sigma\urcorner)$ respectively. Then both $T_{\sigma 0} + \mathsf{Con}_{\mathsf{PA}}$ and $T_{\sigma 1} + \mathsf{Con}_{\mathsf{PA}}$ are consistent.

For each function $f : \omega \to \{0, 1\}$ and each natural number n, let $f \restriction n$ denote the finite binary sequence $f(0), f(1), \ldots, f(n)$. Define $T_f = \bigcup_{n\in\omega} T_{f\restriction n}$. By the definition of the family $\{T_\sigma\}_{\sigma\in S}$, $T_f + \mathsf{Con}_{\mathsf{PA}}$ is consistent. Thus there is a model M_f of $T_f + \mathsf{Con}_{\mathsf{PA}}$. By Theorem 4.3, there is a maximally sane model N_f of PA such that $\mathsf{Thm}_{\mathsf{PA}}(M_f) \subseteq \mathsf{Thm}_{\mathsf{PA}}(N_f)$.

It suffices to prove that for every distinct functions f and g from ω to $\{0, 1\}$, $\mathsf{Thm}_{\mathsf{PA}}(N_f) \neq \mathsf{Thm}_{\mathsf{PA}}(N_g)$. Let f and g be any distinct functions $\omega \to \{0, 1\}$. Let n be the least number such that $f(n) \neq g(n)$. Without loss of generality, we may assume that $f(n) = 0$ and $g(n) = 1$. Let σ be the sequence $f \restriction n - 1$, then $T_\sigma = T_{f\restriction n-1} = T_{g\restriction n-1}$. By the definition, $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_\sigma\urcorner) \in T_{\sigma 0} = T_{f\restriction n}$ and $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\varphi_\sigma\urcorner) \in T_{\sigma 1} = T_{g\restriction n}$. Thus $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_\sigma\urcorner) \in T_f$ and $\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\neg\varphi_\sigma\urcorner) \in T_g$, and hence $\varphi_\sigma \in \mathsf{Thm}_{\mathsf{PA}}(N_f)$ and $\neg\varphi_\sigma \in \mathsf{Thm}_{\mathsf{PA}}(N_g)$. Since N_f and N_g are models of $\mathsf{Con}_{\mathsf{PA}}$, we conclude $\mathsf{Thm}_{\mathsf{PA}}(N_f) \neq \mathsf{Thm}_{\mathsf{PA}}(N_g)$. ⊣

Next, we investigate models which are not maximally sane.

PROPOSITION 4.6. *No model of $\mathsf{PA} + \mathsf{Con}^2_{\mathsf{PA}}$ is maximally sane.*

PROOF. Let M be any model of $\mathsf{PA} + \mathsf{Con}^2_{\mathsf{PA}}$, then there exists a model N of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$ such that $M \subseteq_e N$ and $\mathsf{Thm}_{\mathsf{PA}}(M) \subsetneq \mathsf{Thm}_{\mathsf{PA}}(N)$ by Proposition 3.12. Thus M is not maximally sane. ⊣

Proposition 4.6 says that if M is maximally sane, then M is a model of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \neg\mathsf{Con}^2_{\mathsf{PA}}$. On the other hand, the following theorem says that the converse does not hold in general.

THEOREM 4.7. *There is a model of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \neg\mathsf{Con}^2_{\mathsf{PA}}$ which is not maximally sane.*

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

PROOF. Since PA is Σ_1-sound, the theory $T := \mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \neg\mathsf{Con}^2_{\mathsf{PA}}$ is consistent. Then there is a Σ_1 sentence ξ which is not provably equivalent to any Π_1 sentence in T (cf. Lindström [7] p.18). Define U to be the theory

$$\mathsf{PA} + \{\neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) : \mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \xi \vdash \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)\} + \{\neg\xi, \neg\mathsf{Con}^2_{\mathsf{PA}}\}.$$

We shall show that the theory U is consistent. Towards a contradiction, we suppose that U is inconsistent. Then there are sentences $\varphi_0, \ldots, \varphi_{k-1}$ such that $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \xi \vdash \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_0\urcorner) \wedge \cdots \wedge \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_{k-1}\urcorner)$ and $\mathsf{PA} + \neg\mathsf{Con}^2_{\mathsf{PA}} \vdash \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_0\urcorner) \wedge \cdots \wedge \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_{k-1}\urcorner) \to \xi$. Let π be the Π_1 sentence $\neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_0\urcorner) \wedge \cdots \wedge \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_{k-1}\urcorner)$, then $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \xi \vdash \pi$ and $\mathsf{PA} + \neg\mathsf{Con}^2_{\mathsf{PA}} \vdash \pi \to \xi$. Thus $T \vdash \xi \leftrightarrow \pi$, and ξ is provably equivalent to the Π_1 sentence π in T. This is a contradiction.

Therefore the theory U is consistent. By the completeness theorem, there is a model M of U. Then it is obvious that M is a model of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \neg\mathsf{Con}^2_{\mathsf{PA}}$ by the definition of U. Thus it suffices to show that M is not maximally sane.

Since ξ is a Σ_1 sentence, there is a sentence δ such that $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} \vdash \xi \leftrightarrow \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\delta\urcorner)$ by the FGH-principle. Then we have $M \models \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\delta\urcorner)$ because $M \models \neg\xi$. It also suffices to show that the theory S defined as

$$\mathsf{PA} + \mathsf{Pr}_{\mathsf{PA}}(M) + \{\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\delta\urcorner), \mathsf{Con}_{\mathsf{PA}}\}$$

is consistent.

Towards a contradiction, we suppose that S is inconsistent. Then there are formulas $\varphi_0, \ldots, \varphi_{k-1}$ such that $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_0\urcorner) \wedge \cdots \wedge \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_{k-1}\urcorner)$ and $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\delta\urcorner) \vdash \neg(\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_0\urcorner) \wedge \cdots \wedge \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi_{k-1}\urcorner))$. Let φ be $\varphi_0 \wedge \cdots \wedge \varphi_{k-1}$, then $M \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ and $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\delta\urcorner) \vdash \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$ by **D2**. Then $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} + \xi \vdash \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$, and thus $\neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \in U$ by the definition of U. Since M is a model of U, we have $M \models \neg\mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner)$. This is a contradiction.

Therefore the theory S is consistent. ⊣

§5. History of madness in insane models. Let M be any insane model of PA. Define $\mathcal{C}(M) := \{\mathsf{Thm}_{\mathsf{PA}}(N) : N \subseteq_e M \text{ and } N \models \mathsf{PA}\}$. Then $\mathcal{C}(M)$ forms a chain

$$\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N}) \subsetneq \cdots \subsetneq \mathsf{Thm}_{\mathsf{PA}}(N) \subsetneq \cdots \subsetneq \mathsf{Thm}_{\mathsf{PA}}(M),$$

and this chain represents the history of madness in the model M, that is, the situation that illusory models gradually increase their theorems and at last a proof of $0 = 1$ is obtained.

DEFINITION 5.1. Let M be any insane model of PA.

1. M is *going insane suddenly* if and only if $\mathcal{C}(M) = \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N}), \mathsf{Thm}_{\mathsf{PA}}(M)\}$, that is, for any initial segment K of M, K is not illusory whenever K is a sane model of PA.
2. M is *illusory by nature* if and only if every nonstandard initial segment of M is illusory.

In this section, we prove the existence of two such insane models: one is going insane suddenly and the other is illusory by nature. The existence of an insane model which does not satisfy both such properties is proved easily.

PROPOSITION 5.2. *There is an insane model M of* PA *which is neither going insane suddenly nor illusory by nature.*

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

PROOF. Let N be any nonstandard model of TA, then N is also a model of $\mathsf{PA} + \mathsf{Con}^2_{\mathsf{PA}}$. As in the proof of Proposition 4.6, there is a sane model K of PA such that $N \subsetneq_e K$ and $\mathsf{Thm}_{\mathsf{PA}}(N) \subsetneq \mathsf{Thm}_{\mathsf{PA}}(K)$. By Fact 2.1, there is an insane model M of PA such that $K \subsetneq_e M$. Then M is not going insane suddenly because $\mathcal{C}(M)$ contains at least three elements, and M is not illusory by nature since N is a nonstandard initial segment of M and N is not illusory. ⊣

First, we prove the existence of an insane model which is going insane suddenly. If M has a maximally sane initial segment N, then $\mathsf{Thm}_{\mathsf{PA}}(N)$ is the maximum element of $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$. On the other hand, even if M has no maximally sane initial segment, $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$ can have the maximum element. From the following fact due to Krajíček and Pudlák [6] (Hájek [3] and Solovay [14] also proved similar results), we prove that for any countable sane model N, $\mathsf{Thm}_{\mathsf{PA}}(N)$ can be the maximum element of $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$ for some insane model M.

FACT 5.3. *For any countable nonstandard model N of* PA *and any nonstandard elements a, b of N with $b \leq a$, there is a nonstandard model M of* PA *such that $N \restriction 2^a \subseteq M$, $N \restriction a \simeq M \restriction a$ and $M \models \exists y < 2^{a^b} \mathsf{Prf}_{\mathsf{PA}}(\ulcorner 0 = 1 \urcorner, y)$.*

Here $N \restriction a$ is the structure whose universe is $\{c \in N : N \models c \leq a\}$ and whose each operation is the restriction of the corresponding operation of N.

PROPOSITION 5.4. *For any countable sane nonstandard model N of* PA*, there exists an insane model M of* PA *such that $N \subsetneq_e M$ and $\mathsf{Thm}_{\mathsf{PA}}(N)$ is the maximum element of $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$.*

PROOF. Let N be any countable sane nonstandard model of PA. There exists a proper elementary end-extension N' of N, and then $\mathsf{Thm}_{\mathsf{PA}}(N) = \mathsf{Thm}_{\mathsf{PA}}(N')$. Let a be any element of $N' \setminus N$. By Fact 5.3, there is a nonstandard model M of PA such that $N' \restriction 2^{a^a} \subseteq M$, $N' \restriction a \simeq M \restriction a$ and $M \models \exists y < 2^{a^a} \mathsf{Prf}_{\mathsf{PA}}(\ulcorner 0 = 1 \urcorner, y)$. It is obvious that $N \subsetneq_e M$. We show that $\mathsf{Thm}_{\mathsf{PA}}(N)$ is maximum in $\mathcal{C}(M)$.

Let K be any initial segment of M such that $K \models \mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$. If $a \in K$, then $2^{a^a} \in K$ because K is a model of PA and K is closed under the exponentiation. Then $\exists y < 2^{a^a} \mathsf{Prf}_{\mathsf{PA}}(\ulcorner 0 = 1 \urcorner, y)$ is true in K, and thus K satisfies $\neg\mathsf{Con}_{\mathsf{PA}}$. This is a contradiction.

Therefore $a \notin K$. Since $N' \restriction a \simeq M \restriction a$, K is also an initial segment of N'. Hence $\mathsf{Thm}_{\mathsf{PA}}(K) \subseteq \mathsf{Thm}_{\mathsf{PA}}(N') = \mathsf{Thm}_{\mathsf{PA}}(N)$. Therefore $\mathsf{Thm}_{\mathsf{PA}}(N)$ is maximum in $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$. ⊣

COROLLARY 5.5. *There is an insane model of* PA *which is going insane suddenly.*

PROOF. Let N be any countable nonstandard model of TA, then there is an insane model M of PA such that $N \subsetneq_e M$ and $\mathsf{Thm}_{\mathsf{PA}}(N)$ is the maximum element of $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$ by Proposition 5.4. Since $\mathsf{Thm}_{\mathsf{PA}}(N) = \mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})$, we have $\mathcal{C}(M) = \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N}), \mathsf{Thm}_{\mathsf{PA}}(M)\}$. Therefore M is going insane suddenly. ⊣

Next we prove the existence of an insane model M which is illusory by nature. We also prove that $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})\}$ has no minimum element for such a model M.

THEOREM 5.6. *There is a countable insane model M of* PA *which is illusory by nature.*

PROOF. Let T be the theory $\mathsf{PA} + \neg\mathsf{Con}_{\mathsf{PA}}$ and let $t(x)$ be the set of formulas defined as

$$t(x) := \{\forall y \leq x \neg\mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg\mathsf{Con}_{\mathsf{PA}} \to \varphi \urcorner, y) : T \nvdash \varphi\} \cup \{x \geq \bar{n} : n \in \omega\}.$$

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

We shall prove that T locally omits $t(x)$. Let $\xi(x)$ be a formula such that $T \vdash \xi(x) \to \varphi(x)$ for all $\varphi(x) \in t(x)$. By the fixed-point lemma, there is a formula $\psi(x)$ such that

$$\mathsf{PA} \vdash \psi(x) \leftrightarrow \xi(x) \wedge \exists y \leq x \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \forall u \neg \psi(u) \urcorner, y). \tag{1}$$

If $T \nvdash \forall u \neg \psi(u)$, then $\forall y \leq x \neg \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \forall u \neg \psi(u) \urcorner, y)$ is in $t(x)$. Thus $T \vdash \xi(x) \to \forall y \leq x \neg \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \forall u \neg \psi(u) \urcorner, y)$ by the choice of $\xi(x)$. Since $T \vdash \neg \psi(x)$ by (1), we obtain $T \vdash \forall u \neg \psi(u)$. This is a contradiction.

Therefore $T \vdash \forall u \neg \psi(u)$. By (1) again, we have

$$T \vdash \xi(x) \to \forall y \leq x \neg \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \forall u \neg \psi(u) \urcorner, y).$$

Since $\mathsf{PA} \vdash \neg \mathsf{Con}_{\mathsf{PA}} \to \forall u \neg \psi(u)$, let p be a proof of the sentence $\neg \mathsf{Con}_{\mathsf{PA}} \to \forall u \neg \psi(u)$ in PA. Then $T \vdash \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \forall u \neg \psi(u) \urcorner, \bar{p})$. Hence $T \vdash \bar{p} \leq x \to \neg \xi(x)$. Also $T \vdash \xi(x) \to x \geq \bar{p}$ because $x \geq \bar{p}$ is in $t(x)$, and thus $T \vdash x < \bar{p} \to \neg \xi(x)$. Consequently, we obtain $T \vdash \forall x \neg \xi(x)$. Therefore T locally omits $t(x)$.

By the omitting types theorem, there is a countable model M of $T = \mathsf{PA} + \neg \mathsf{Con}_{\mathsf{PA}}$ omitting $t(x)$.

Let N be any initial segment of M which is a nonstandard model of PA, and let a be any nonstandard element of N. Then a is also an element of M. Since M omits $t(x)$, there is a formula $\varphi(x)$ in $t(x)$ such that $M \models \neg \varphi(a)$. Because a is nonstandard, we have $M \models a \geq \bar{n}$ for all $n \in \omega$. Hence there is a sentence ρ such that $T \nvdash \rho$ and $M \models \exists y \leq a \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \rho \urcorner, y)$. Since $\exists y \leq a \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \rho \urcorner, y)$ is Π_1, we obtain $N \models \exists y \leq a \mathsf{Prf}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \rho \urcorner, y)$ and thus $N \models \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \neg \mathsf{Con}_{\mathsf{PA}} \to \rho \urcorner)$. Therefore $\neg \mathsf{Con}_{\mathsf{PA}} \to \rho \in \mathsf{Thm}_{\mathsf{PA}}(N)$ and $\mathsf{PA} \nvdash \neg \mathsf{Con}_{\mathsf{PA}} \to \rho$. This means N is illusory, and thus M is illusory by nature. ⊣

COROLLARY 5.7. *There is a countable nonstandard insane model M of* PA *such that* $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})\}$ *has no minimum element*.

PROOF. Let M be a countable nonstandard insane model of PA as in Theorem 5.6. We show $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})\}$ has no minimum element.

Let N be any initial segment of M such that $\mathsf{Thm}_{\mathsf{PA}}(N) \in \mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})\}$, and φ be any sentence such that $\mathsf{PA} \nvdash \varphi$ and $\varphi \in \mathsf{Thm}_{\mathsf{PA}}(N)$. Then there is the least Gödel number p of any proof of φ in N. Such an element p of N must be nonstandard. Then let K be a nonstandard initial segment of N such that $K \models \mathsf{PA}$ and $p \notin K$, which is obtained by using indicators (cf. Kaye [5] p.204 and McAloon [8]). Since K does not have any proof of φ, $\varphi \notin \mathsf{Thm}_{\mathsf{PA}}(K)$, and thus $\mathsf{Thm}_{\mathsf{PA}}(K) \subsetneq \mathsf{Thm}_{\mathsf{PA}}(N)$.

By Theorem 5.6, there is a sentence ψ such that $\mathsf{PA} \nvdash \psi$ and $\psi \in \mathsf{Thm}_{\mathsf{PA}}(K)$. Therefore $\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N}) \subsetneq \mathsf{Thm}_{\mathsf{PA}}(K)$, and hence $\mathsf{Thm}_{\mathsf{PA}}(K) \in \mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})\}$. We conclude that $\mathsf{Thm}_{\mathsf{PA}}(N)$ is not the minimum element of $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})\}$. ⊣

By combining Proposition 5.4 and Corollary 5.7, we obtain the following corollary.

COROLLARY 5.8. *There is an insane model M of* PA *such that the set* $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$ *is infinite and contains the maximum element*.

PROOF. Let K be a countable insane model of PA as in Corollary 5.7. Also let N be any countable nonstandard initial segment of K which is a sane model of PA.

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press

Then $\mathcal{C}(N) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(\mathbb{N})\}$ has no minimum element by Corollary 5.7. By Proposition 5.4, there is an insane model M of PA such that $N \subseteq_e M$ and $\mathsf{Thm}_{\mathsf{PA}}(N)$ is maximum in $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$. Then $\mathcal{C}(M) \setminus \{\mathsf{Thm}_{\mathsf{PA}}(M)\}$ is also an infinite set and contains the maximum element. ⊣

§6. **Acknowledgments.** This work was supported by JSPS KAKENHI Grant Numbers 24540125, 12J00654, 26887045.

REFERENCES

[1] W. Craig, *On axiomatizability within a system*, this Journal, vol. 18 (1953), pp. 30–32.

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

[3] P. Hájek, *On a new notion of partial conservativity*, ***Computation and Proof Theory***, Lecture Notes in Mathematics, vol. 1104, Springer, Berlin, 1984, pp. 217–232.

[4] D. Jensen and A. Ehrenfeucht, *Some problem in elementary arithmetics*. ***Fundamenta Mathematicae***, vol. 92 (1976), pp. 223–245.

[5] R. Kaye, ***Models of Peano Arithmetic***, Oxford Logic Guides, vol. 15, Oxford Science Publications, New York, 1991.

[6] J. Krajíček and P. Pudlák, *On the structure of initial segments of models of arithmetic*. ***Archive for Mathematical Logic***, vol. 28 (1988), no. 2, pp. 91–98.

[7] P. Lindström, ***Aspects of Incompleteness***, Lecture Notes in Logic, vol. 10, Springer-Verlag, Berlin, 1997.

[8] K. McAloon, *On the complexity of models of arithmetic*, this Journal, vol. 47 (1982), no. 2, pp. 403–415.

[9] D. Misercque, *Branches of the E-tree which are not isomorphic*, ***Bulletin of the London Mathematical Society***, vol. 17 (1985), pp. 513–517.

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

[11] H. Simmons, *Large discrete parts of the E-tree*, this Journal, vol. 53 (1988), no. 3, pp. 980–984.

[12] C. Smoryński, *The incompleteness theorems*, ***Handbook of Mathematical Logic*** (J. Barwise, editor), North-Holland, Amsterdam, 1977, pp. 821–865.

[13] ———, *Fifty years of self-reference in arithmetic*. ***Notre Dame Journal of Formal Logic***, vol. 22 (1981), no. 4, pp. 357–378.

[14] R. M. Solovay, *Injecting inconsistencies into models of PA*. ***Annals of Pure and Applied Logic***, vol. 44 (1989), pp. 101–132.

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

https://doi.org/10.1017/jsl.2015.52 Published online by Cambridge University Press