THE REVIEW OF SYMBOLIC LOGIC
Volume 10, Number 4, December 2017

GENERALIZATIONS OF GÖDEL'S INCOMPLETENESS THEOREMS FOR Σ_n-DEFINABLE THEORIES OF ARITHMETIC

MAKOTO KIKUCHI
Graduate School of System Informatics, Kobe University
and
TAISHI KURAHASHI
Department of Natural Science, National Institute of Technology
Kisarazu College

Abstract. It is well known that Gödel's incompleteness theorems hold for Σ_1-definable theories containing Peano arithmetic. We generalize Gödel's incompleteness theorems for arithmetically definable theories. First, we prove that every Σ_{n+1}-definable Σ_n-sound theory is incomplete. Secondly, we generalize and improve Jeroslow and Hájek's results. That is, we prove that every consistent theory having Π_{n+1} set of theorems has a true but unprovable Π_n sentence. Lastly, we prove that no Σ_{n+1}-definable Σ_n-sound theory can prove its own Σ_n-soundness. These three results are generalizations of Rosser's improvement of the first incompleteness theorem, Gödel's first incompleteness theorem, and the second incompleteness theorem, respectively.

§1. Introduction. As it is inscribed in the title of the famous paper, Gödel's incompleteness theorems were proved for a particular system, Principia Mathematica PM. The proofs were based on the three facts—that PM is defined primitive recursively, PM is ω-consistent, and PM includes arithmetic. Hence, as Gödel had pointed out in the paper, Gödel's theorems are applicable to similar theories which satisfy these three conditions. Gödel's theorems have been generalized further and currently they are often stated as follows: for any extension T of Peano arithmetic PA, if T is Σ_1-definable and Σ_1-sound, then T is incomplete (the first incompleteness theorem), and if T is Σ_1-definable and consistent, then the consistency of T is not provable in T (the second incompleteness theorem).

The assumptions of Σ_1-definability and Σ_1-soundness in Gödel's theorems can be justified philosophically. PM and other similar theories have been constructed in order to formalize the whole of mathematics effectively. Σ_1-soundness should be satisfied by any theory which is intended to be a formalization of mathematics including arithmetic, and Σ_1-definability is an acceptable condition for the effectiveness, since Σ_1-definable theories are axiomatizable primitive recursively.

Mathematically, Σ_1-definability is the optimal condition for Gödel's theorems in the sense that we cannot generalize Gödel's theorems to Δ_2-definable theories. That is, there exists a Δ_2-definable consistent and complete extension of PA, and we can find a

Received: May 10, 2015.
2010 *Mathematics Subject Classification*: 03F30, 03F40.
Key words and phrases: Gödel's incompleteness theorems, Σ_n-definable theories.

doi:10.1017/S1755020317000235

Δ_2-definable theory whose consistency is provable in it (see Feferman (1960)). It is an interesting problem to investigate theories which are out of the range of Gödel's theorems, and, for example, nonrecursively-enumerable theories which prove their own consistency are investigated in Niebergall (2005) and Kaså (2012).

However, there have been various generalizations of Gödel's theorems. Concerning Σ_1-soundness, Rosser showed in Rosser (1936) that the Σ_1-soundness requirement on the theory T in Gödel's theorem can be weakened to the mere consistency of T, and this generalization is now called the Gödel-Rosser first incompleteness theorem.

Regarding Σ_1-definability, based on Carnap's analysis of nonconstructive rules, Rosser showed in Rosser (1937) that Gödel's theorems hold for certain extensions of PM which are not Σ_1-definable. By referring to Putnam's discussions of trial-and-error predicates, Jeroslow proved in Jeroslow (1975) that, for any consistent theory T including arithmetic and having a Δ_2-definable set of theorems, there is a true Π_1 sentence that is not provable in T. Also Jeroslow proved that for any Σ_1-sound extension T of arithmetic with a Σ_2-definable set of theorems and provability predicates $\mathsf{Pr}_T(x)$ of T satisfying certain additional conditions, the Σ_2-consistency of $\mathsf{Pr}_T(x)$ is not provable in T.[1] Hájek generalized Gödel's first incompleteness theorem in Hájek (1977) along the direction of Jeroslow's argument, proving that for any consistent extension T of PA whose set of theorems is PA-provably Δ_{n+2}, there exists a true Π_{n+1} sentence that is not provable in T. Hájek proved also that if an extension T of PA having a Π_{n+2} set of theorems is Σ_{n+2}-consistent, then there is a T-unprovable true Π_{n+1} sentence.

In this paper, we investigate further generalizations of Gödel's theorems to the case of arithmetically definable theories. We firstly discuss the first incompleteness theorem. We start with a generalization of the first incompleteness theorem to the statement that every Σ_{n+1}-definable consistent extension of PA has an unprovable true Π_{n+1} sentence. While this generalization itself is a consequence of Hájek's result, we shall give two stronger variations of this generalization.

The first such variation is an extension of Rosser's generalization. The Gödel-Rosser first incompleteness theorem cannot be generalized to Σ_n-definable theories directly, since, as we have mentioned before, there is a Δ_2-definable consistent and complete extension of PA. However, the Gödel-Rosser theorem can be restated as every Σ_1-definable and Σ_0-sound extension of PA is incomplete, and we prove that every Σ_{n+1}-definable and Σ_n-sound extension of PA is incomplete as well.[2]

Another variation is an extension of Jeroslow and Hájek's generalizations. Although Hájek's result is strongly related to Jeroslow's, the former is not a generalization of the latter because there is a Δ_2 set which is not PA-provably Δ_2. We prove that if T is a consistent extension of PA having a Π_{n+2} set of theorems, there exists a true Π_{n+1} sentence that is not provable in T. This is a generalization of Jeroslow's result and an improvement of Hájek's result, and it gives a negative answer to the following problem of Hájek given in Hájek (1977): Does there exist a consistent extension of PA having a Π_3 set of theorems

[1] Actually, Jeroslow stated these results in the terminology of experimental logics, and thus we describe adaptations of Jeroslow's results (see also Fact 5.7 below).

[2] The referee informed the authors that our article has overlap with the following preprint, which deals with a generalization of the Gödel-Rosser first incompleteness theorem: Salehi, S. & Seraji, P., Gödel-Rosser's incompleteness theorems for nonrecursively enumerable theories, http://arxiv.org/abs/1506.02790. See also Salehi, S. & Seraji, P., Gödel-Rosser's incompleteness theorem, generalized and optimized for definable theories, to appear in Journal of Logic and Computation.

that can prove all true Π_2 sentences? Using the result above, we show also that if T is a Σ_{n+1}-consistent theory having a Π_{n+2} set of theorems, then T is incomplete.

Next, we examine the second incompleteness theorem. We prove that no Σ_{n+1}-definable Σ_n-sound theory can prove its own Σ_n-soundness. In addition, we study the consistency statements for Σ_n-definable theories. We prove that for every Σ_{n+1}-definable and Σ_n-sound theory T, there is a consistency statement for some axiomatization of T which is independent of T. Thus appropriate consistency statements can be witnesses for the generalized version of the Gödel-Rosser first incompleteness theorem.

§2. Preliminaries. In this paper, we call a set of sentences a *theory*. Thus a theory is identified with its axiom set. We consider only theories in the language of first-order arithmetic $\{+, \times, 0, 1, <\}$. We assume that T and U always denote theories containing Peano arithmetic PA. Let ω be the set of all nonnegative integers. For each $n \in \omega$, $\bar{n}$ denotes the numeral for n. For each formula φ, $\mathsf{gn}(\varphi)$ is the Gödel number of φ, and $\ulcorner\varphi\urcorner$ denotes the numeral for $\mathsf{gn}(\varphi)$.

We recursively define the classes Σ_n and Π_n of formulas for every $n \in \omega$ as follows: $\Sigma_0 = \Pi_0$ is the class of all formulas all of whose quantifiers are bounded; Σ_{n+1} (resp. Π_{n+1}) is the class of all formulas of the form $\exists\vec{x}\varphi$ (resp. $\forall\vec{x}\varphi$) for some $\varphi \in \Pi_n$ (resp. $\varphi \in \Sigma_n$), and here quantifiers preceding φ are allowed to be absent. We say a formula is $\Sigma_n(\mathsf{PA})$ (resp. $\Pi_n(\mathsf{PA})$) if it is PA-provably equivalent to some formula in Σ_n (resp. Π_n). Throughout this paper, we sometimes omit '(PA)' if there is no danger of confusion. A formula is called $\Delta_n(\mathbb{N})$ (resp. $\Delta_n(T)$) if it is equivalent to both some Σ_n formula and some Π_n formula in $\mathbb{N}$ (resp. T). We suppose that the subscript n of Σ_n, Π_n and Δ_n ranges over ω unless otherwise stated.

We say a formula $\sigma(u)$ is a *definition* of a theory T if and only if $\{n \in \omega : \mathbb{N} \models \sigma(\bar{n})\} = \{\mathsf{gn}(\varphi) : \varphi \in T\}$. Let Γ be a class of formulas. A definition of T which is a Γ formula is called a Γ *definition* of T. A theory T having a Γ definition is said to be Γ-*definable*. Notice that distinct Γ definitions of a Γ-definable theory T need not be equivalent in T, and that every Σ_n-definable consistent theory always has two Σ_n definitions which are not T-equivalent (see Corollary 4.6 below).

We say a formula $\sigma(u)$ is a *binumeration* of a theory T in a theory U if and only if for any sentence φ, $U \vdash \sigma(\ulcorner\varphi\urcorner)$ whenever $\varphi \in T$, and $U \vdash \neg\sigma(\ulcorner\varphi\urcorner)$ whenever $\varphi \notin T$. When a binumeration $\sigma(u)$ is a Γ formula, we say $\sigma(u)$ a Γ *binumeration*. For each formula $\sigma(u)$, we can construct a formula $\mathsf{Prf}_\sigma(x, y)$ which states "a sentence with the code x has a proof with the code y from the set of all sentences satisfying $\sigma(u)$", and the formula $\mathsf{Prf}_\sigma(x, y)$ is called the *proof predicate* of $\sigma(u)$ (see Feferman (1960)). For $n > 0$, if $\sigma(u)$ is Σ_n (resp. Π_n), the resulting formula $\mathsf{Prf}_\sigma(x, y)$ is $\Sigma_n(\mathsf{PA})$ (resp. $\Pi_n(\mathsf{PA})$). Define $\mathsf{Pr}_\sigma(x)$ to be the formula $\exists y\mathsf{Prf}_\sigma(x, y)$ which is called the *provability predicate* of $\sigma(u)$. If $\sigma(u)$ is a definition of a theory T, then $\mathsf{Pr}_\sigma(x)$ is a definition of the theory $\{\varphi : T \vdash \varphi\}$.

For each definition $\sigma(u)$ of T, the consistency assertion Con_σ of $\sigma(u)$ is defined as $\neg\mathsf{Pr}_\sigma(\ulcorner\bar{0} = \bar{1}\urcorner)$, which expresses the consistency of T. If $\sigma(u)$ is Σ_n, then Con_σ is a $\Pi_n(\mathsf{PA})$ sentence. Let $(\sigma|x)(u)$ be the formula $\sigma(u) \wedge u < x$. Then for each $n \in \omega$, the formula $(\sigma|\bar{n})(u)$ is a definition of the finite subtheory $\{\varphi \in T : \mathsf{gn}(\varphi) < n\}$ of T.

The following facts hold (see Feferman (1960) and Lindström (1997)).

FACT 2.1. *Let T and U be theories, and $\sigma(x)$ be a binumeration of T in U.*

1. *If $p \in \omega$ is a code of a T-proof of φ, then $U \vdash \mathsf{Prf}_\sigma(\ulcorner\varphi\urcorner, \bar{p})$.*
2. *If $q \in \omega$ is not a code of any T-proof of φ, then $U \vdash \neg\mathsf{Prf}_\sigma(\ulcorner\varphi\urcorner, \bar{q})$.*

https://doi.org/10.1017/S1755020317000235 Published online by Cambridge University Press

FACT 2.2. *Let $\sigma(u)$ and $\tau(u)$ be any formulas.*

1. $\mathsf{PA} \vdash \forall u(\sigma(u) \to \tau(u)) \to \forall x(\mathsf{Pr}_\sigma(x) \to \mathsf{Pr}_\tau(x))$.
2. $\mathsf{PA} \vdash \forall u(\sigma(u) \to \tau(u)) \to (\mathsf{Con}_\tau \to \mathsf{Con}_\sigma)$.

FACT 2.3 (See Mostowski (1952)). *Let T be a subtheory of U. If $\sigma(u)$ is a binumeration of T in U, then $U \vdash \mathsf{Con}_{\sigma|\bar{n}}$ for any $n \in \omega$.*

Let Γ be either Σ_{n+1} or Π_{n+1}, then it is known that there is a Γ formula $\mathsf{True}_\Gamma(x)$ which is a truth-definition for sentences in Γ, that is, for any formula $\varphi(x) \in \Gamma$, $\mathsf{PA} \vdash \forall x(\varphi(x) \leftrightarrow \mathsf{True}_\Gamma(\ulcorner\varphi(\dot{x})\urcorner))$, where $\ulcorner\varphi(\dot{x})\urcorner$ is the standard dot notation, and notice that x is free in $\mathsf{True}_\Gamma(\ulcorner\varphi(\dot{x})\urcorner)$ (see Lindström (1997)). Then the formula $\mathsf{True}_\Gamma(x)$ is a Γ definition of the set $\mathsf{Th}_\Gamma(\mathbb{N}) := \{\varphi \in \Gamma : \mathbb{N} \models \varphi\}$ of all true sentences in Γ. On the other hand, Tarski's undefinability theorem says that there exists no formula defining the set $\mathsf{TA} := \{\varphi : \mathbb{N} \models \varphi\}$ of all true sentences. Also there is a $\Delta_1(\mathsf{PA})$ formula $\mathsf{True}_{\Sigma_0}(x)$ which is a truth-definition for sentences in Σ_0 (see Kaye (1991)).

Define $\mathsf{Pr}_{\sigma,n}(x)$ to be the formula $\exists v(\mathsf{True}_{\Sigma_{n+1}}(v) \wedge \mathsf{Pr}_\sigma(v \dot{\to} x))$. Then we have the following proposition (see Smoryński (1985)).

PROPOSITION 2.4. *Let $\sigma(x)$ be any Σ_{n+1} definition of a theory T.*

1. *If $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \forall x\varphi(x)$, then $\mathsf{PA} + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \forall x\mathsf{Pr}_{\sigma,n}(\ulcorner\varphi(\dot{x})\urcorner)$.*
2. $\mathsf{PA} \vdash \forall x(\mathsf{Pr}_{\sigma,n}(\ulcorner\varphi(\dot{x}) \to \psi(\dot{x})\urcorner) \to (\mathsf{Pr}_{\sigma,n}(\ulcorner\varphi(\dot{x})\urcorner) \to \mathsf{Pr}_{\sigma,n}(\ulcorner\psi(\dot{x})\urcorner)))$.
3. *If $\varphi(x)$ is Σ_{n+1}, then $\mathsf{PA} \vdash \forall x(\varphi(x) \to \mathsf{Pr}_{\sigma,n}(\ulcorner\varphi(\dot{x})\urcorner))$.*

§3. Notions related to consistency and completeness. In this section, we introduce some notions related to consistency and completeness of theories and show several properties of these notions.

DEFINITION 3.1. *Let T be a theory and Γ be a class of formulas.*

1. *T is Γ-sound if and only if for all Γ sentences φ, $\mathbb{N} \models \varphi$ whenever $T \vdash \varphi$.*
2. *T is sound if and only if T is Σ_n-sound for any $n \in \omega$.*
3. *T is Γ-consistent if and only if for all Γ formulas $\varphi(x)$, if $T \vdash \neg\varphi(\bar{k})$ for all $k \in \omega$, then $T \nvdash \exists x\varphi(x)$.*
4. *T is ω-consistent if and only if T is Σ_n-consistent for any $n \in \omega$.*
5. *T is Γ-complete if and only if for all Γ sentences φ, $T \vdash \varphi$ whenever $\mathbb{N} \models \varphi$.*
6. *T is Γ-decisive if and only if for all Γ sentences φ, either $T \vdash \varphi$ or $T \vdash \neg\varphi$ holds.*

It is well-known that every extension of PA is Σ_1-complete. It is easy to see that a theory T is complete if and only if T is Π_n-decisive for all $n \in \omega$, and that T is consistent if and only if T is Σ_0-sound. The notion of ω-consistency was introduced in Gödel (1931), and Π_{n-1}-consistency was originally introduced in Kreisel (1957) under the name 'n-consistency'.

We exhibit several properties of these notions.

PROPOSITION 3.2 (See Hájek (1977) and Smoryński (1977b)).

1. *T is Σ_n-sound if and only if T is Π_{n+1}-sound.*
2. *T is Π_n-consistent if and only if T is Σ_{n+1}-consistent.*
3. *T is Π_n-complete if and only if T is Σ_{n+1}-complete.*

https://doi.org/10.1017/S1755020317000235 Published online by Cambridge University Press

PROPOSITION 3.3 (See Smoryński (1977b)). *Let $n > 0$. If T is Σ_n-sound, then T is Σ_n-consistent.*

COROLLARY 3.4. *If T is sound, then T is ω-consistent.*

It is known that for $n = 1, 2$, the Σ_n-soundness of T is equivalent to the Σ_n-consistency of T. Also, an ω-consistent complete theory is deductively equivalent to TA (see Isaacson (2011) and Smoryński (1977b)). The following proposition is a stratified version of these results.

PROPOSITION 3.5.

1. *If $n \leq 2$ and T is Σ_n-consistent, then T is Σ_n-sound.*
2. *If $n \geq 3$, T is Σ_n-consistent and Π_{n-2}-decisive, then T is Σ_n-sound.*

Proof. We only prove clause 2. Actually, we prove the statement for $n \geq 2$ by induction on n. The statement for $n = 2$ is already obtained in clause 1. Suppose that the statement holds for n. Let T be any Σ_{n+1}-consistent and Π_{n-1}-decisive theory, and $\varphi(x, y)$ be any Σ_{n-1} formula. If $T \vdash \exists x \forall y \varphi(x, y)$, then $T \nvdash \neg \forall y \varphi(\bar{k}, y)$ for some $k \in \omega$ because T is Σ_{n+1}-consistent. For such k, $T \nvdash \neg \varphi(\bar{k}, \bar{l})$ for all $l \in \omega$. Since T is Π_{n-1}-decisive, $T \vdash \varphi(\bar{k}, \bar{l})$ for all $l \in \omega$. Since T is Σ_n-consistent and Π_{n-2}-decisive, T is Σ_n-sound by the induction hypothesis. Hence $\mathbb{N} \models \varphi(\bar{k}, \bar{l})$ for all $l \in \omega$. Therefore $\mathbb{N} \models \exists x \forall y \varphi(x, y)$. We have shown that T is Σ_{n+1}-sound. □

It is known that there exists a Σ_1-definable theory which is ω-consistent but not Σ_3-sound (cf. Lindström (1997) p. 36). Thus for $n \geq 3$, Σ_n-consistency does not imply Σ_n-soundness in general.

COROLLARY 3.6. *If T is ω-consistent and complete, then T is deductively equivalent to* TA.

We obtain the following relations between several properties a theory may have.

PROPOSITION 3.7. *For $n > 0$, the following are equivalent*:

1. *T is Π_n-complete and consistent*;
2. *T is Σ_n-sound and Π_n-decisive*;
3. *T is Σ_n-consistent and Π_n-decisive.*

Proof. ($1 \Rightarrow 2$): Suppose that T is a Π_n-complete consistent theory. Let φ be any Σ_n sentence.

First, we prove the Σ_n-soundness of T. If $T \vdash \varphi$, then $T \nvdash \neg\varphi$ by the consistency of T. Since $\neg\varphi$ is a Π_n sentence, we have $\mathbb{N} \not\models \neg\varphi$ by Π_n-completeness. Thus $\mathbb{N} \models \varphi$, and T is Σ_n-sound.

Secondly, we prove that T is Π_n-decisive. Suppose $T \nvdash \neg\varphi$, then $\mathbb{N} \models \varphi$ as we have seen above. Since T is also Σ_{n+1}-complete and φ is a Σ_n sentence, we obtain $T \vdash \varphi$, and thus T is Π_n-decisive.

($2 \Rightarrow 1$): Suppose that T is Σ_n-sound and Π_n-decisive. Obviously, T is consistent. Let φ be any Π_n sentence such that $\mathbb{N} \models \varphi$. By Σ_n-soundness, $T \nvdash \neg\varphi$. Then $T \vdash \varphi$ because T is Π_n-decisive. Therefore T is Π_n-complete.

($2 \Leftrightarrow 3$): This is immediate from Propositions 3.3 and 3.5. □

The following characterization of Σ_n-soundness is formally presented in Beklemishev (2005) Lemma 2.9.

https://doi.org/10.1017/S1755020317000235 Published online by Cambridge University Press

PROPOSITION 3.8. *A theory T is Σ_n-sound if and only if $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$ is consistent.*

Proof. ($\Rightarrow$): We show the contrapositive. Suppose that $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$ is inconsistent. Then there is a true Σ_{n+1} sentence φ such that $T \vdash \neg\varphi$. Since $\neg\varphi$ is a false Π_{n+1} sentence, T is not Π_{n+1}-sound. By Proposition 3.2.1, T is not Σ_n-sound.

($\Leftarrow$): We again show the contrapositive. Suppose that T is not Σ_n-sound, then T proves a false Σ_n sentence φ. Then $T + \neg\varphi$ is inconsistent, and $\neg\varphi \in \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$. Therefore $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$ is inconsistent. □

§4. The first incompleteness theorem. Gödel constructed in Gödel (1931) a true but T-unprovable Π_1 sentence, called the Gödel sentence of T, for each Σ_1-definable consistent theory T. Moreover, if T is ω-consistent, then such a sentence is not refutable in T, and therefore it is undecidable in T. This is Gödel's first incompleteness theorem. The ω-consistency assumption can be replaced by Σ_1-consistency in the proof of the first incompleteness theorem. We have seen in Propositions 3.3 and 3.5 that Σ_1-consistency is equivalent to Σ_1-soundness. Then we have

FACT 4.1 (Gödel's first incompleteness theorem).

1. *If T is Σ_1-definable and consistent, then T is not Π_1-complete.*
2. *If T is Σ_1-definable and Σ_1-sound, then T is not Π_1-decisive.*

There are two improvements of Gödel's first incompleteness theorem, which were obtained by Rosser and Jeroslow, respectively. Rosser improved in Rosser (1936) the second clause of Gödel's first incompleteness theorem by replacing the Σ_1-soundness assumption by the consistency of the theory.

FACT 4.2 (The Gödel-Rosser first incompleteness theorem). *If the theory T is Σ_1-definable and consistent, then T is not Π_1-decisive.*

Let $\mathsf{Th}(T)$ be the set of all theorems of T. Note that if T is Σ_1-definable, then $\mathsf{Th}(T)$ is also Σ_1-definable, and thus $\Delta_2(\mathbb{N})$-definable. Jeroslow improved the first clause of Gödel's first incompleteness theorem, which is Theorem 2 in Jeroslow (1975).

FACT 4.3 (See Jeroslow (1975)). *If $\mathsf{Th}(T)$ is $\Delta_2(\mathbb{N})$-definable and T is consistent, then T is not Π_1-complete.*

In the Gödel-Rosser first incompleteness theorem, the Σ_1-definability assumption of T cannot be replaced by the $\Delta_2(\mathbb{N})$-definability because of the following fact.

FACT 4.4 (See Jeroslow (1975); Smoryński (1977a)). *There exists a $\Delta_2(\mathbb{N})$-definable complete consistent theory T.*

Thus the Gödel-Rosser first incompleteness theorem cannot be extended to Σ_n-definable theories directly. On the other hand, Gödel's first incompleteness theorem is directly generalized to Σ_n-definable theories. We give a proof of such a generalization, however, later we improve it in two ways.

THEOREM 4.5.

1. *If T is Σ_{n+1}-definable and consistent, then T is not Π_{n+1}-complete.*
2. *If T is Σ_{n+1}-definable and Σ_{n+1}-sound, then T is not Π_{n+1}-decisive.*

Proof. Clause 2 is immediate from clause 1 by Proposition 3.7, thus it suffices to prove clause 1. Let T be a Σ_{n+1}-definable consistent theory. If T is not Π_n-complete, T is not

Π_{n+1}-complete. Thus we may assume that T is Π_n-complete. By Proposition 3.2, T is also Σ_{n+1}-complete.

Let $\sigma(u)$ be a Σ_{n+1} definition of T. The provability predicate $\mathsf{Pr}_\sigma(x)$ is a Σ_{n+1} formula. There is a Π_{n+1} sentence ψ satisfying $\mathsf{PA} \vdash \psi \leftrightarrow \neg \mathsf{Pr}_\sigma(\ulcorner \psi \urcorner)$ by Fixed-Point Lemma (see Lindström (1997) for details). If $T \vdash \psi$, then $\mathsf{Pr}_\sigma(\ulcorner \psi \urcorner)$ is a true Σ_{n+1} sentence. By our assumption, $T \vdash \mathsf{Pr}_\sigma(\ulcorner \psi \urcorner)$. Thus $T \vdash \neg\psi$. This contradicts the consistency of T.

Therefore $T \nvdash \psi$. Also $T \nvdash \neg \mathsf{Pr}_\sigma(\ulcorner \psi \urcorner)$. Then $\neg \mathsf{Pr}_\sigma(\ulcorner \psi \urcorner)$ is a Π_{n+1} sentence which is true but not T-provable. Therefore T is not Π_{n+1}-complete. □

From the first clause of Theorem 4.5, we obtain non$T$-equivalent Σ_{n+1} definitions of Σ_{n+1}-definable consistent theory T. For $n = 0$, this is well-known (see Feferman (1960)).

COROLLARY 4.6. *Let T be any Σ_{n+1}-definable consistent theory. Then there are Σ_{n+1} definitions $\sigma_0(u)$ and $\sigma_1(u)$ of T which are not equivalent in T.*

Proof. Let $\sigma_0(u)$ be any Σ_{n+1} definition of T. We may assume that $T \vdash \exists u \neg \sigma_0(u)$ (otherwise, replace $\sigma_0(u)$ by $\sigma_0(u) \wedge u \neq \ulcorner \bar{0} = \bar{1} \urcorner$). Since T is Σ_{n+1}-definable and consistent, there exists a true Π_{n+1} sentence φ which is not provable in T by Theorem 4.5.1. Define $\sigma_1(u)$ to be the Σ_{n+1} formula $\sigma_0(u) \vee \neg\varphi$. Then $\mathbb{N} \models \forall u(\sigma_0(u) \leftrightarrow \sigma_1(u))$ because $\mathbb{N} \models \varphi$, and hence $\sigma_1(u)$ is also a Σ_{n+1} definition of T. Suppose $T \vdash \forall u(\sigma_0(u) \leftrightarrow \sigma_1(u))$, then $T \vdash \neg\varphi \rightarrow \forall u \sigma_0(u)$. Since $T \vdash \exists u \neg \sigma_0(u)$, we have $T \vdash \varphi$. This is a contradiction. Therefore $T \nvdash \forall u(\sigma_0(u) \leftrightarrow \sigma_1(u))$. □

First, we improve the second clause of Theorem 4.5. Specifically, we prove that the assumption of Σ_{n+1}-soundness in the statement can be replaced by Σ_n-soundness. This is a generalized version of the Gödel-Rosser first incompleteness theorem. In our proof, we use a generalized version of Craig's trick.

FACT 4.7 (Craig's trick (see 2.2.C in Grzegorczyk, Mostowski, & Ryll-Nardzewski (1958))). *Every Σ_{n+1}-definable theory has a deductively equivalent Π_n-definable theory.*

THEOREM 4.8. *If T is Σ_{n+1}-definable and Σ_n-sound, then T is not Π_{n+1}-decisive.*

Proof. Suppose that T is Σ_{n+1}-definable and Σ_n-sound. It follows that $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$ is Σ_{n+1}-definable and consistent by Proposition 3.8. By Craig's trick, there is a Π_n-definable theory T' which is deductively equivalent to $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$. Let $\gamma(u)$ be a Π_n definition of T', then $\gamma(u)$ is a binumeration of T' in T' because T' knows all Σ_{n+1}-truth.

The proof predicate $\mathsf{Prf}_\gamma(x, y)$ of $\gamma(u)$ is a $\Delta_{n+1}(\mathsf{PA})$ formula. Let ψ be a Π_{n+1} sentence such that $\mathsf{PA} \vdash \psi \leftrightarrow \forall y(\mathsf{Prf}_\gamma(\ulcorner \psi \urcorner, y) \rightarrow \exists z < y \mathsf{Prf}_\gamma(\ulcorner \neg\psi \urcorner, z))$. Then neither ψ nor $\neg\psi$ is provable in T' by a usual argument of the proof of Rosser's incompleteness theorem. Since ψ is Π_{n+1}, T' is not Π_{n+1}-decisive. Hence also T is not Π_{n+1}-decisive. □

In the next section, we give an alternative proof of Theorem 4.8 (see Corollary 5.14). Theorem 4.8 can be slightly strengthened as follows.

THEOREM 4.9. *If T is Σ_{n+1}-definable and Σ_n-consistent, then T is not Π_{n+1}-decisive.*

Proof. The statement for $n = 0$ is exactly Gödel's first incompleteness theorem because Σ_0-consistency is equivalent to Σ_1-soundness by Propositions 3.2 and 3.5. We may assume $n > 0$. Let T be any Σ_{n+1}-definable Σ_n-consistent theory. If T were Π_{n+1}-decisive, then T is Σ_n-sound by Proposition 3.7. This contradicts Theorem 4.8. Therefore T is not Π_{n+1}-decisive. □

https://doi.org/10.1017/S1755020317000235 Published online by Cambridge University Press

Secondly, we improve the first clause of Theorem 4.5 along the direction of Jeroslow's improvement. One improvement like that has already been made by Hájek.

FACT 4.10 (See Hájek (1977)). *If* $\mathsf{Th}(T)$ *is* $\Delta_{n+2}(\mathsf{PA})$*-definable and* T *is consistent, then* T *is not* Π_{n+1}*-complete.*

Hájek also proved another generalization of the first incompleteness theorem.

FACT 4.11 (See Hájek (1977)). *If* $\mathsf{Th}(T)$ *is* Π_{n+2}*-definable and* T *is* Σ_{n+2}*-consistent, then* T *is not* Π_{n+1}*-complete.*

Facts 4.10 and 4.11 are Theorems 2.8 and 2.5 in Hájek (1977), respectively. Since $\Delta_2(\mathbb{N})$ sets are not always $\Delta_2(\mathsf{PA})$ in general, Fact 4.10 is not a generalization of Jeroslow's result.

We prove that the assumption of the Σ_{n+2}-consistency in Fact 4.11 can be replaced by consistency.

THEOREM 4.12. *If* $\mathsf{Th}(T)$ *is* Π_{n+1}*-definable and* T *is consistent, then* T *is not* Π_n*-complete.*

Proof. Let T be a consistent theory such that $\mathsf{Th}(T)$ is Π_{n+1}-definable, and let $\forall x\,\tau(u, x)$ be a Π_{n+1} definition of $\mathsf{Th}(T)$ where $\tau(u, x)$ is a Σ_n formula. Let φ be a Σ_{n+1} sentence satisfying the following equivalence:

$$\mathsf{PA} \vdash \varphi \leftrightarrow \exists x(\neg\tau(\ulcorner\varphi\urcorner, x) \wedge \forall y \leq x\ \tau(\ulcorner\neg\varphi\urcorner, y)).$$

Define ψ to be the Σ_{n+1} sentence $\exists x(\neg\tau(\ulcorner\neg\varphi\urcorner, x) \wedge \forall y < x\ \tau(\ulcorner\varphi\urcorner, y))$. Then it is easy to show $\mathsf{PA} \vdash \neg\varphi \vee \neg\psi$. Since T is consistent, at least one of $\varphi \notin \mathsf{Th}(T)$ or $\neg\varphi \notin \mathsf{Th}(T)$ holds. Thus $\mathbb{N} \models \exists x\neg\tau(\ulcorner\varphi\urcorner, x) \vee \exists x\neg\tau(\ulcorner\neg\varphi\urcorner, x)$. Hence we obtain $\mathbb{N} \models \varphi \vee \psi$.

Towards contradiction, we assume that T is Π_n-complete. Then T is also Σ_{n+1}-complete by Proposition 3.2. We distinguish two cases $\mathbb{N} \models \varphi$ and $\mathbb{N} \models \psi$.

If $\mathbb{N} \models \varphi$, then $T \vdash \varphi$ by our assumption. Thus $\varphi \in \mathsf{Th}(T)$. On the other hand, we have $\mathbb{N} \models \neg\forall x\,\tau(\ulcorner\varphi\urcorner, x)$ by the choice of φ. Then $\varphi \notin \mathsf{Th}(T)$ since $\forall x\,\tau(u, x)$ defines $\mathsf{Th}(T)$. This is a contradiction.

If $\mathbb{N} \models \psi$, then $T \vdash \psi$ by our assumption. Then $T \vdash \neg\varphi$, and hence $\neg\varphi \in \mathsf{Th}(T)$. On the other hand, we have $\mathbb{N} \models \neg\forall x\,\tau(\ulcorner\neg\varphi\urcorner, x)$ by the definition of ψ. Then $\neg\varphi \notin \mathsf{Th}(T)$. This is also a contradiction.

We conclude that T is not Π_n-complete. □

By Theorem 4.12, if $\mathsf{Th}(T)$ is Π_{n+2}-definable and T is consistent, then T is not Π_{n+1}-complete. This is a generalization of Jeroslow's result and an improvement of Hájek's results. Also the $n = 0$ case of Theorem 4.12 states that there is no consistent theory T such that $\mathsf{Th}(T)$ is Π_1-definable. This is an improvement of Remark 2.6(1) in Hájek (1977) which states that there is no Σ_1-consistent theory T such that $\mathsf{Th}(T)$ is Π_1-definable.

From Theorem 4.12 and Proposition 3.7, we immediately obtain the following corollary, which is also an improvement of the second clause of Theorem 4.5.

COROLLARY 4.13. *If* $\mathsf{Th}(T)$ *is* Π_{n+2}*-definable and* T *is* Σ_{n+1}*-consistent, then* T *is not* Π_{n+1}*-decisive.*

Hájek proposed the following problem, in Problem 2.9 of Hájek (1977): Does there exist a Π_2-complete consistent theory T such that $\mathsf{Th}(T)$ is Π_3-definable? Theorem 4.12 gives a negative answer to Hájek's problem.

From the following examples, we can see that Theorem 4.9, Theorem 4.12 and Corollary 4.13 are optimal.

EXAMPLE 4.14. *The theory* $\mathsf{PA} + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$ *is* Σ_{n+1}*-definable, sound,* Π_n*-complete and* Π_n*-decisive.*

EXAMPLE 4.15. *There are* $\Delta_{n+2}(\mathsf{PA})$*-definable,* Π_n*-complete,* Σ_n*-sound and complete theories.*

Every Lindenbaum completion of $\mathsf{PA} + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$ (see Lemma 3.2 in Hájek (1977)) witnesses this example.

§5. The second incompleteness theorem. Like Gödel's first incompleteness theorem, Gödel's second incompleteness theorem is also a theorem for Σ_1-definable theories. The second incompleteness theorem states that if T is Σ_1-definable and consistent, then T cannot prove its own consistency. However the statement described above is ambiguous because it is known that the unprovability of consistency statements depends on the underlying representation of T, and thus we must state Gödel's second incompleteness theorem more precisely.

FACT 5.1 (Gödel's second incompleteness theorem). *If* T *is* Σ_1*-definable and consistent, then for any* Σ_1 *definition* $\sigma(u)$ *of* T, $T \nvdash \mathsf{Con}_\sigma$.

Feferman showed that 'Σ_1 definition' in the statement of Gödel's second incompleteness theorem cannot be replaced by 'Π_1 definition'.

FACT 5.2 (Feferman, 1960). *If* T *is* Σ_1*-definable, then there is a* Π_1 *definition* $\tau(u)$ *of* T *such that* $T \vdash \mathsf{Con}_\tau$.

Therefore Gödel's second incompleteness theorem cannot be generalized to Σ_{n+1}-definable theories directly. On the other hand, the second incompleteness theorem can be seen as a theorem about soundness since the consistency of a theory is equivalent to its Σ_0-soundness.

For every definition $\sigma(u)$ of T, the *uniform* Σ_n *reflection principle* $\mathsf{RFN}_{\Sigma_n}(\sigma)$ of $\sigma(u)$ is the sentence $\forall x(\Sigma_n(x) \wedge \mathsf{Pr}_\sigma(x) \rightarrow \mathsf{True}_{\Sigma_n}(x))$ expressing the Σ_n-soundness of T, where $\Sigma_n(x)$ is the natural $\Delta_1(\mathsf{PA})$ binumeration of the set of all Σ_n sentences. The *uniform reflection principle* $\mathsf{RFN}(\sigma)$ of $\sigma(u)$ is the theory $\{\mathsf{RFN}_{\Sigma_n}(\sigma) : n \geq 1\}$ which expresses the soundness of T.

Let $\mathsf{Con}_{\sigma,n}$ be the sentence $\neg\mathsf{Pr}_{\sigma,n}(\ulcorner \bar{0} = \bar{1} \urcorner)$. If $\sigma(u)$ defines the theory T, then $\mathsf{Con}_{\sigma,n}$ can be seen as a formal consistency statement of $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$. By using the properties of partial truth-definitions $\mathsf{True}_{\Sigma_{n+1}}(x)$ and Proposition 2.4, Proposition 3.8 can be formalized in PA as follows.

PROPOSITION 5.3. *Let* $\sigma(u)$ *be a* Σ_{n+1} *definition of* T, *then* $\mathsf{PA} \vdash \mathsf{RFN}_{\Sigma_n}(\sigma) \leftrightarrow \mathsf{Con}_{\sigma,n}$.[3]

Since $\mathsf{RFN}_{\Sigma_0}(\sigma)$ and Con_σ are equivalent in PA for any Σ_1 definition $\sigma(u)$ of any Σ_1 definable theory, Gödel's second incompleteness theorem can be restated as follows.

THEOREM 5.4 (Gödel's second incompleteness theorem). *For any* Σ_1*-definable theory* T, *the following are equivalent*:

[3] The referee pointed out that the proof of Proposition 5.3 is not carried out if $\mathsf{RFN}_{\Sigma_n}(\sigma)$ is defined as the schema $\{\forall x(\mathsf{Pr}_\sigma(\ulcorner \varphi(\dot{x}) \urcorner) \rightarrow \varphi(x)) : \varphi(x) \in \Sigma_n\}$ because PA may not know that the theory defined by $\sigma(u)$ is sufficiently strong.

https://doi.org/10.1017/S1755020317000235 Published online by Cambridge University Press

1. *T is Σ_0-sound*;
2. *for all Σ_1 definitions $\sigma(u)$ of T, $T \nvdash \mathsf{RFN}_{\Sigma_0}(\sigma)$*;
3. *for some Σ_1 definition $\sigma(u)$ of T, $T \nvdash \mathsf{RFN}_{\Sigma_0}(\sigma)$*.

We generalize this version of the second incompleteness theorem. For our proof, we use the following improvement of Fact 2.3. Define $\mathsf{Pr}_{\emptyset}(x)$ to be the canonical Σ_1 provability predicate for pure predicate calculus.

FACT 5.5 (See Kreisel & Lévy (1968)). $\mathsf{PA} \vdash \mathsf{RFN}(\emptyset)$.

Here we prove a generalization of Gödel's second incompleteness theorem.

THEOREM 5.6. *For any Σ_{n+1}-definable theory T, the following are equivalent*:

1. *T is Σ_n-sound*;
2. *for all Σ_{n+1} definitions $\sigma(u)$ of T, $T \nvdash \mathsf{RFN}_{\Sigma_n}(\sigma)$*;
3. *for all Σ_{n+1} definitions $\sigma(u)$ of T, $T \nvdash \mathsf{RFN}(\sigma)$*.

Proof. $(1 \Rightarrow 2)$: Suppose that T is Σ_n-sound, then $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$ is consistent by Proposition 3.8. Let $\sigma(u)$ be any Σ_{n+1} definition of T. Then $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \nvdash \mathsf{Con}_{\sigma,n}$ by carrying out a usual proof of Gödel's second incompleteness theorem with Proposition 2.4. Since $\mathsf{PA} \vdash \mathsf{RFN}_{\Sigma_n}(\sigma) \rightarrow \mathsf{Con}_{\sigma,n}$ by Proposition 5.3, we have $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \nvdash \mathsf{RFN}_{\Sigma_n}(\sigma)$. Therefore $T \nvdash \mathsf{RFN}_{\Sigma_n}(\sigma)$.

$(2 \Rightarrow 3)$: Obvious.

$(3 \Rightarrow 1)$: We prove the contrapositive. Suppose that T is not Σ_n-sound. Then there is a Σ_n sentence φ such that $T \vdash \varphi$ and $\mathbb{N} \models \neg\varphi$. Let $\tau(u)$ be any Σ_{n+1} definition of T, and define $\sigma(u)$ to be the Σ_{n+1} formula $\tau(u) \wedge \neg\varphi$. Since $\mathbb{N} \models \neg\varphi$, $\mathbb{N} \models \forall u(\sigma(u) \leftrightarrow \tau(u))$, and thus $\sigma(u)$ is a Σ_{n+1} definition of T.

Since $\mathsf{PA} \vdash \varphi \rightarrow \forall u \neg\sigma(u)$, we obtain $\mathsf{PA} \vdash \varphi \rightarrow (\mathsf{Pr}_{\sigma}(x) \leftrightarrow \mathsf{Pr}_{\emptyset}(x))$ by Fact 2.2.1. Because $\mathsf{PA} \vdash \mathsf{RFN}(\emptyset)$ by Fact 5.5, for each $n \in \omega$,

$$\begin{aligned} \mathsf{PA} \vdash \varphi \wedge \Sigma_n(x) \wedge \mathsf{Pr}_{\sigma}(x) &\rightarrow \mathsf{Pr}_{\emptyset}(x) \\ &\rightarrow \mathsf{True}_{\Sigma_n}(x). \end{aligned}$$

Then we have $T \vdash \forall x(\Sigma_n(x) \wedge \mathsf{Pr}_{\sigma}(x) \rightarrow \mathsf{True}_{\Sigma_n}(x))$ since $T \vdash \varphi$. Therefore we conclude $T \vdash \mathsf{RFN}(\sigma)$. □

By Theorem 5.6, we can conclude that Gödel's second incompleteness theorem (Theorem 5.4 $(1 \Leftrightarrow 2)$) is the $n = 0$ case of the general property about the Σ_n-soundness of Σ_{n+1}-definable theories.

Under an appropriate interpretation, we can understand that Jeroslow proved a version of the second incompleteness theorem for a class of Σ_2-definable theories. That is, Jeroslow's proof of Theorem 6 in Jeroslow (1975) essentially showed the following fact.

FACT 5.7 (See Jeroslow (1975)). *Let T be a Σ_2-definable and Σ_1-sound theory. If there exists a Π_2 formula $\pi(x)$ such that $T \vdash \mathsf{Pr}_T(\ulcorner\varphi\urcorner) \leftrightarrow \pi(\ulcorner\varphi\urcorner)$ for all Σ_2 sentences φ, then T cannot prove its own Σ_2-soundness.*

The $n = 1$ case of Theorem 5.6 is an improvement of Fact 5.7.

If $n = 0$, the existence of a Σ_{n+1} definition $\sigma(u)$ of T with $T \nvdash \mathsf{RFN}_{\Sigma_n}(\sigma)$ implies the Σ_n-soundness of T. On the other hand, this is not the case for $n > 0$ in general.

PROPOSITION 5.8. *For $n > 0$, there exist a Σ_1-definable theory T and a Σ_1 definition $\sigma(u)$ of T such that T is consistent but not Σ_n-sound, and $T \nvdash \mathsf{RFN}_{\Sigma_{n-1}}(\sigma)$.*

https://doi.org/10.1017/S1755020317000235 Published online by Cambridge University Press

Proof. Let T be the theory $\mathsf{PA} + \neg\mathsf{Con}_{\tau,n-1}$ where $\tau(u)$ is a Σ_1 definition of PA. Then the formula $\sigma(u) \equiv \tau(u) \vee u = \ulcorner\neg\mathsf{Con}_{\tau,n-1}\urcorner$ is a Σ_1 definition of T. It follows from $\mathsf{PA} \vdash \mathsf{Con}_{\sigma,n-1} \rightarrow \mathsf{Con}_{\tau,n-1}$ and Proposition 5.3 that T and $\sigma(u)$ satisfy the required conditions. □

We obtain the following proposition.

PROPOSITION 5.9. *There is an ω-consistent Σ_1-definable theory T having a Σ_4 definition $\sigma(u)$ such that $T \vdash \mathsf{RFN}(\sigma)$.*

Proof. Let T be a Σ_1-definable theory which is ω-consistent but not Σ_3-sound (see our remark just after Proposition 3.5). Let $\sigma(u)$ be the Σ_4 definition of T from our proof of Theorem 5.6 $(3 \Rightarrow 1)$.[4] Then T proves $\mathsf{RFN}(\sigma)$. □

From this proposition, we obtain a Σ_3-consistent Σ_4-definable theory T having a Σ_4 definition $\sigma(u)$ such that T proves the Σ_3-consistency of $\sigma(u)$. Therefore the Σ_n-soundness assumption in the statement of Theorem 5.6 cannot be replaced by Σ_n-consistency throughout.

Finally, we investigate several properties of consistency statements. First, we give a characterization of the unprovability of the negation of consistency assertions.

THEOREM 5.10. *For any Σ_{n+1}-definable theory T, the following are equivalent*:

1. *T is Σ_{n+1}-sound*;
2. *for all Σ_{n+1} definitions $\sigma(u)$ of T, $T \nvdash \neg\mathsf{Con}_\sigma$.*

Proof. $(1 \Rightarrow 2)$: If T is Σ_{n+1}-sound, then the Π_{n+1} sentence Con_σ is true for any Σ_{n+1} definition $\sigma(u)$ of T. By the Σ_{n+1}-soundness of T, T does not prove $\neg\mathsf{Con}_\sigma$.

$(2 \Rightarrow 1)$: We show the contrapositive. Suppose that T is not Σ_{n+1}-sound, then there is a Σ_{n+1} sentence φ such that $T \vdash \varphi$ and $\mathbb{N} \models \neg\varphi$.

Let $\tau(u)$ be any Σ_{n+1} definition of T. Define $\sigma(u)$ to be the Σ_{n+1} formula $\tau(u) \vee \varphi$. Because $\mathbb{N} \models \neg\varphi$, we have $\mathbb{N} \models \forall u(\sigma(u) \leftrightarrow \tau(u))$. Thus $\sigma(u)$ is a Σ_{n+1} definition of T.

Since $\mathsf{PA} \vdash \varphi \rightarrow \forall u\sigma(u)$, $\mathsf{PA} \vdash \varphi \rightarrow \neg\mathsf{Con}_\sigma$. Therefore $T \vdash \neg\mathsf{Con}_\sigma$ since $T \vdash \varphi$. □

We obtain the following corollaries.

COROLLARY 5.11. *If T is Σ_{n+1}-definable and not Σ_n-sound, then there are Σ_{n+1} definitions $\sigma_1(u)$ and $\sigma_2(u)$ of T such that $T \vdash \mathsf{Con}_{\sigma_1}$ and $T \vdash \neg\mathsf{Con}_{\sigma_2}$.*

Proof. Suppose that T is Σ_{n+1}-definable and not Σ_n-sound. Then by Theorem 5.6, there exists a Σ_{n+1} definition $\sigma_1(u)$ of T such that $T \vdash \mathsf{RFN}(\sigma_1)$. Then $T \vdash \mathsf{Con}_{\sigma_1}$.

Since T is not Σ_{n+1}-sound, there exists a Σ_{n+1} definition $\sigma_2(u)$ of T such that $T \vdash \neg\mathsf{Con}_{\sigma_2}$ by Theorem 5.10. □

COROLLARY 5.12. *If T is Σ_{n+1}-definable and Π_{n+1}-decisive, then there are Σ_{n+1} definitions $\sigma_1(u)$ and $\sigma_2(u)$ of T such that $T \vdash \mathsf{Con}_{\sigma_1}$ and $T \vdash \neg\mathsf{Con}_{\sigma_2}$.*

Proof. This is immediate from Theorem 4.8 and Corollary 5.11. □

THEOREM 5.13. *If T is Π_n-definable and Σ_n-sound, then there exists a Σ_{n+1} definition $\sigma(u)$ of T such that $T \nvdash \mathsf{Con}_\sigma$ and $T \nvdash \neg\mathsf{Con}_\sigma$.*

[4] Moreover, in this case, the formula $\sigma(u)$ is in fact Π_3 by letting $\tau(u)$ in our proof of Theorem 5.6 be a Σ_1 definition of T.

https://doi.org/10.1017/S1755020317000235 Published online by Cambridge University Press

Proof. Suppose that T is Π_n-definable and Σ_n-sound. Then T is Π_{n+1}-sound by Proposition 3.2. Also by Proposition 3.8, $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$ is consistent.

Let $\tau(u)$ be any Π_n definition of T. Also let $\sigma(u)$ be a Σ_{n+1} formula satisfying the following equivalence:

$$\mathsf{PA} \vdash \sigma(u) \leftrightarrow [\tau(u) \vee \exists y(\mathsf{Prf}_\tau(\ulcorner \mathsf{Con}_\sigma \urcorner, y) \wedge \psi(y))] \wedge \psi(u),$$

where $\psi(x)$ is the formula $\forall z \leq x \neg \mathsf{Prf}_\tau(\ulcorner \neg \mathsf{Con}_\sigma \urcorner, z)$.

Towards contradiction, suppose $T \vdash \mathsf{Con}_\sigma$. Then $T \nvdash \neg \mathsf{Con}_\sigma$, and thus $\mathbb{N} \models \psi(\bar{n})$ holds for any $n \in \omega$. Hence $\mathbb{N} \models \sigma(\bar{n})$ holds for any $n \in \omega$ because $\mathbb{N} \models \exists y(\mathsf{Prf}_\tau(\ulcorner \mathsf{Con}_\sigma \urcorner, y) \wedge \psi(y))$. Then the formula $\sigma(u)$ is a definition of a trivially inconsistent theory, and thus we have $\mathbb{N} \models \neg \mathsf{Con}_\sigma$. This contradicts the Π_{n+1}-soundness of T because Con_σ is a Π_{n+1} sentence. Therefore $T \nvdash \mathsf{Con}_\sigma$.

Again towards contradiction, suppose $T \vdash \neg \mathsf{Con}_\sigma$ and let p be a natural number such that $\mathbb{N} \models \mathsf{Prf}_\tau(\ulcorner \neg \mathsf{Con}_\sigma \urcorner, \bar{p})$. Then $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \mathsf{Prf}_\tau(\ulcorner \neg \mathsf{Con}_\sigma \urcorner, \bar{p})$ because this sentence is true Σ_{n+1}. Hence

$$T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \psi(u) \to u < \bar{p}. \tag{1}$$

Since $T \nvdash \mathsf{Con}_\sigma$, the Σ_n sentence $\forall y < \bar{p} \neg \mathsf{Prf}_\tau(\ulcorner \mathsf{Con}_\sigma \urcorner, y)$ is true. Together with (1), this implies

$$T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \neg \exists y(\mathsf{Prf}_\tau(\ulcorner \mathsf{Con}_\sigma \urcorner, y) \wedge \psi(y)). \tag{2}$$

By (1) and (2), we have $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \sigma(u) \to \tau(u) \wedge u < \bar{p}$. By Fact 2.2.2, $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \mathsf{Con}_{\tau|\bar{p}} \to \mathsf{Con}_\sigma$. Since $\tau(u)$ is a binumeration of T in $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$, we have $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \mathsf{Con}_{\tau|\bar{p}}$ by Fact 2.3. Therefore we obtain $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N}) \vdash \mathsf{Con}_\sigma$. This contradicts the consistency of $T + \mathsf{Th}_{\Sigma_{n+1}}(\mathbb{N})$. We conclude $T \nvdash \neg \mathsf{Con}_\sigma$.

Then $\mathbb{N} \models \forall u(\sigma(u) \leftrightarrow \tau(u))$ because $T \nvdash \mathsf{Con}_\sigma$ and $T \nvdash \neg \mathsf{Con}_\sigma$. This means that $\sigma(u)$ is a Σ_{n+1} definition of T. □

Notice that the $n = 0$ case of Theorem 5.13 is a consequence of Theorem 7.4 in Feferman (1960).

By Craig's trick, we immediately obtain the following corollary.

COROLLARY 5.14. *If T is Σ_{n+1}-definable and Σ_n-sound, then there exists a Σ_{n+1} definition $\sigma(u)$ of some axiomatization of $\mathsf{Th}(T)$ such that $T \nvdash \mathsf{Con}_\sigma$ and $T \nvdash \neg \mathsf{Con}_\sigma$.*

Our generalization of the Gödel-Rosser first incompleteness Theorem (Theorem 4.8) follows from Corollary 5.14. Thus we obtain that the witnesses for Theorem 4.8 can be provided by appropriate consistency statements.

By combining Corollary 5.11 and Theorem 5.13, we obtain the following corollary.

COROLLARY 5.15. *If T is Π_n-definable and consistent, then there are Σ_{n+1} definitions $\sigma_1(u)$ and $\sigma_2(u)$ of T such that $T \nvdash \mathsf{Con}_{\sigma_1}$ and $T \nvdash \neg \mathsf{Con}_{\sigma_2}$.*

Proof. Suppose that T is Π_n-definable and consistent. If T is Σ_n-sound, then this is obvious by Theorem 5.13. If T is not Σ_n-sound, then there are Σ_{n+1} definitions $\sigma_1(u)$ and $\sigma_2(u)$ of T such that $T \vdash \neg \mathsf{Con}_{\sigma_1}$ and $T \vdash \mathsf{Con}_{\sigma_2}$ by Corollary 5.11. Since T is consistent, $T \nvdash \mathsf{Con}_{\sigma_1}$ and $T \nvdash \neg \mathsf{Con}_{\sigma_2}$. □

In contrast to Theorem 5.10, it follows from Corollary 5.15 that for Π_n-definable theory T, the existence of a Σ_{n+1} definition $\sigma(u)$ of T with $T \nvdash \neg \mathsf{Con}_\sigma$ does not imply any kind of soundness of T in general.

https://doi.org/10.1017/S1755020317000235 Published online by Cambridge University Press

§6. Acknowledgments. This work was partly supported by JSPS KAKENHI Grant Numbers 24540125, 17H02263, 26887045, and 16K17653. The authors would like to thank Hidenori Kurokawa for valuable discussions and helpful comments. The authors would also like to thank the referee for the valuable comments and suggestions on earlier versions of this paper.

BIBLIOGRAPHY

Beklemishev, L. D. (2005). Reflection principles and provability algebras in formal arithmetic. *Russian Mathematical Surveys*, **60**(2), 197–268.

Feferman, S. (1960). Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, **49**, 35–92.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. (in German). *Monatshefte für Mathematik und Physik*, **38**(1), 173–198. English translation in Kurt Gödel, *Collected Works*, Vol. 1. pp. 145–195.

Grzegorczyk, A., Mostowski, A., & Ryll-Nardzewski, C. (1958). The classical and the ω-complete arithmetic. *The Journal of Symbolic Logic*, **23**(2), 188–206.

Hájek, P. (1977). Experimental logics and Π^0_3 theories. *The Journal of Symbolic Logic*, **42**(4), 515–522.

Isaacson, D. (2011). Necessary and sufficient conditions for undecidability of the Gödel sentence and its truth. In DeVidi, D., Hallett, M., and Clark, P., editors. *Logic, Mathematics, Philosophy: Vintage Enthusiasms. Essays in Honour of John L. Bell*. The Western Ontario Series in Philosophy of Science, Vol. 75. Dordrecht: Springer, pp. 135–152.

Jeroslow, R. G. (1975). Experimental logics and Δ^0_2-theories. *Journal of Philosophical Logic*, **4**(3), 253–267.

Kaså, M. (2012). Experimental logics, mechanism and knowable consistency. *Theoria*, **78**(3), 213–224.

Kaye, R. (1991). *Models of Peano Arithmetic*. Oxford Logic Guides, Vol. 15. New York: Oxford Science Publications.

Kreisel, G. (1957). A refinement of ω-consistency (abstract). *The Journal of Symbolic Logic*, **22**, 108–109.

Kreisel, G. & Lévy, A. (1968). Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, **14**, 97–142.

Lindström, P. (1997). *Aspects of Incompleteness*. Lecture Notes in Logic, Vol. 10. Berlin: Springer-Verlag.

Mostowski, A. (1952). On models of axiomatic systems. *Fundamenta Mathematicae*, **39**, 133–158.

Niebergall, K.-G. (2005). "Natural" representations and extensions of Gödel's second theorem. In M. Baaz, S.-D. Friedman, and J. K., editors. *Logic Colloquium '01*. Urbana, IL, and Wellesley, MA: Association for Symbolic Logic/A K Peters, pp. 350–368.

Rosser, J. B. (1936). Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic*, **1**(3), 87–91.

Rosser, J. B. (1937). Gödel theorems for nonconstructive logics. *The Journal of Symbolic Logic*, **2**(3), 129–137.

Smoryński, C. (1977a). The incompleteness theorems. In Barwise, J., editor. *Handbook of Mathematical Logic*. Studies in Logic and the Foundations of Mathematics, Vol. 90. Amsterdam: North-Holland Publishing, pp. 821–865.

Smoryński, C. (1977b). ω-consistency and reflection. In *Colloque International de Logique: Clermont-Ferrand, 18–25 juillet 1975*. Paris: Editions du C.N.R.S., pp. 167–181.
Smoryński, C. (1985). *Self-reference and Modal Logic*. Universitext. New York: Springer-Verlag.

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