Gödelian sentences, Rosserian sentences and truth

ZIBA ASSADI, *Independent Scholar, Tabriz, Iran.*

SAEED SALEHI, *Department of Mathematical Sciences, University of Tabriz, P.O. Box 51666-16471, Tabriz, Iran.*

Abstract

There is a long-standing debate in the logico-philosophical community as to if/why the Gödelian sentences of a consistent and sufficiently strong theory are true (γ is a Gödelian sentence of T when γ is equivalent to the T-unprovability of γ inside T). The prevalent argument seems to be something like the following: *since every one of the Gödelian sentences of such a theory is equivalent to the consistency statement of the theory, even provably so inside the theory, the truth of those sentences follows from the consistency of the theory in question; so, Gödelian sentences of consistent and sufficiently strong theories are true*. In this paper, we critically examine this argument and present necessary and sufficient conditions for the truth of Gödelian sentences (and Rosserian sentences) of consistent and sufficiently strong arithmetical theories.

Keywords: Incompleteness theorem, Gödelian sentences, Rosser's trick, Rosserian sentences, soundness, consistency, Σ_n-soundness

1 Introduction

By the first incompleteness theorem of Gödel [9], for every consistent and sufficiently strong arithmetical theory there are sentences that are undecidable in the theory. Examples of such undecidable sentences are actually constructed in Gödel's original proof in a way that each of those sentences is equivalent to its own unprovability in the theory; see Definition 3.1 below. A natural question here is that while the theory in question cannot decide the truth of its Gödelian sentences, what about *us* (human beings)? Can we 'see' (or demonstrate) their truth? This question has attracted the attention of many philosophers, physicists, computer scientists, as well as mathematical logicians. As there are numerous papers and books on this subject, it is not possible to cite them all here; see Conclusions for a few.

If our theory (which is an RE set of sentences) is sufficiently strong and *sound*, then it proves the equivalence of each of its Gödelian sentences with the consistency statement of the theory (see Remark 2.3.III below). Since the consistency statement is true, and the theory is sound, then it follows that all the Gödelian sentences of the theory are true. Now, let us see an example of a false Gödelian sentence (of a consistent and sufficiently strong theory). Put T be a consistent and sufficiently strong RE theory; by Gödel's second incompleteness theoerm, T cannot prove its consistency statement. So, the theory T plus its *in*consistency statement is consistent; call it U. Now, U proves the inconsistency of T and so it proves the inconsistency of U, itself, as well. Therefore, U proves that the contradiction, $\bot$, is U-provable; so, $\bot$ is U-provably equivalent to the U-*un*provability of $\bot$. Thus, $\bot$ is a Gödelian sentence of U (note that U is an *un*sound theory; see also [17, 18]).

So, something must be wrong with the argument mentioned in the Abstract (that Gödelian sentences of consistent and sufficiently strong theories are true).

Vol. 00, No. 0,
https://doi.org/10.1093/logcom/exab075

The paper is organized as follows: in Section 2 we present some preliminaries necessary for following the upcoming arguments. In Section 3 we show that Gödelian (I-am-unprovable) sentences constitute all the unprovable sentences in a sense and provide necessary and sufficient conditions for their truth. In Section 4 we study Rosserian sentences; the sentences that express 'for every proof of me there is a smaller proof of my negation'. We also provide necessary and sufficient conditions for their truth. In Section 5 we conclude the paper with a diagram on the truth (and falsity) of Gödelian and Rosserian sentences by presenting some nice equivalent conditions on the underlying theory.

2 Preliminaries

We assume familiarity with the notions of Π_n and Σ_n formulas, Peano's arithmetic $\boldsymbol{PA}$ and its fragments, Robinson's Arithmetic $\boldsymbol{Q}$ and the fact that $\boldsymbol{Q}$ is a sound and Σ_1-complete theory (i.e. every $\boldsymbol{Q}$-provable sentence is true and every true Σ_1-sentence is $\boldsymbol{Q}$-provable). By the diagonal lemma of Gödel and Carnap, for every formula $\Psi(x)$ with the only free variable x, there exists a sentence θ such that $\theta \leftrightarrow \Psi(\#\theta)$ is true (in the standard model of natural numbers $\mathbb{N}$) and also provable in $\boldsymbol{Q}$; here $\#A$ denotes the numeral of the Gödel code of A, relative to a fixed Gödel numbering (arithmetization) of the syntax. Moreover, if $\Psi(x)$ is a Π_n-formula, for some $n \geqslant 1$, then θ can be taken to be a Π_n-sentence; and if $\Psi(x)$ is Σ_n, then θ can be taken to be Σ_n too. We provide more details in the following:

LEMMA 2.1 (The diagonal lemma).
Let $n \geqslant 1$. For every Π_n-formula $\Psi(x)$ there exists a Π_n-sentence θ such that $\boldsymbol{Q} \vdash \theta \leftrightarrow \Psi(\#\theta)$. And for every Σ_n-formula $\Psi(x)$ there exists a Σ_n-sentence θ with the same property.

PROOF. There is a primitive recursive partial function d that assigns to a given m, when m codes a formula with the only free variable x, the Gödel code of the sentence that results from substituting $\overline{m}$ for x (where $\overline{m}$ is the numeral of m, a term in the language of arithmetic representing m). There is a Σ_1-formula $\delta(x,y)$, in the language of arithmetic, that strongly represents d in $\boldsymbol{Q}$. This means that $\boldsymbol{Q} \vdash \forall y[\delta(\overline{m},y) \leftrightarrow y = \overline{\mathsf{d}(m)}]$ for every $m \in \mathbb{N}$.

If $\Psi(x)$ is a Π_n-formula, then put $\alpha(x) = \forall y[\delta(x,y) \rightarrow \Psi(y)]$ and let a be its Gödel code. Now, let $\theta = \alpha(\overline{\mathsf{a}})$; then θ is a Π_n-sentence and we have provably in $\boldsymbol{Q}$ that

$$\begin{array}{rcl} \theta & \leftrightarrow & \forall y[\delta(\overline{\mathsf{a}},y) \rightarrow \Psi(y)] \\ & \leftrightarrow & \forall y[y = \overline{\mathsf{d}(\mathsf{a})} \rightarrow \Psi(y)] \\ & \leftrightarrow & \forall y[y = \#\theta \rightarrow \Psi(y)] \\ & \leftrightarrow & \Psi(\#\theta). \end{array}$$

If $\Psi(x)$ is a Σ_n-formula, then put $\eta(x) = \exists y[\delta(x,y) \wedge \Psi(y)]$ and let e be its Gödel code. Now, let $\theta = \eta(\overline{\mathsf{e}})$; then θ is a Σ_n-sentence and we have provably in $\boldsymbol{Q}$ that

$$\begin{array}{rcl} \theta & \leftrightarrow & \exists y[\delta(\overline{\mathsf{e}},y) \wedge \Psi(y)] \\ & \leftrightarrow & \exists y[y = \overline{\mathsf{d}(\mathsf{e})} \wedge \Psi(y)] \\ & \leftrightarrow & \exists y[y = \#\theta \wedge \Psi(y)] \\ & \leftrightarrow & \Psi(\#\theta). \end{array}$$ □

The incompleteness theorem is usually stated for recursively enumerable (RE) theories that extend $\boldsymbol{Q}$; though it also holds for more general theories, see, e.g. [32] or [15]. For us a theory is a set of sentences. If T is an RE theory, then by [14, Corollary 3.4] the theory T is Σ_1-definable, in the sense that we have $T = \{\theta \mid \mathbb{N} \vDash \sigma(\#\theta)\}$ for a Σ_1-formula $\sigma(x)$, where θ ranges over the sentences.

By Craig's trick [4], every such RE theory can be axiomatized by a Δ_0-definable set of sentences (see [32, Lemma 2.4]): write $\sigma(x) \equiv \exists y\chi(y,x)$ where χ is a Δ_0-formula and consider the theory $T^* = \{\theta \wedge (\overline{n}=\overline{n}) \mid \mathbb{N} \vDash \chi(\overline{n}, \#\theta)\}$; now T^* is equivalent to T, and is definable by the Δ_0-formula $\tau(x) = \exists y,z \leqslant x[x = [y \& (\bar{z}=\bar{z})] \wedge \chi(z,y)]$.

Let $\tau(x)$ be an arbitrary Δ_0-formula. Put $\mathtt{Th}_\tau = \{\theta \mid \mathbb{N} \vDash \tau(\#\theta)\}$ to be the theory defined by τ, where θ ranges over the sentences. By the arithmetization of syntax with respect to a fixed Gödel coding, we can write a Σ_1-formula $\mathtt{prf}_\tau(y,x)$ stating that 'y is (the code of) a proof of the sentence (with code) x in the theory $\mathtt{Th}_\tau$'. Let us note that it suffices for y to be the code of a sequence of formulas, each of which is either a logical axiom or satisfies τ (is an axiom of $\mathtt{Th}_\tau$) or is derived from one or two earlier formulas by a logical rule (which could be Modus Ponens or Generalization). Since $\mathtt{prf}_\tau(y,x)$ is a decidable relation when τ is Δ_0 (which implies that $\mathtt{Th}_\tau$ is a decidable set of sentences), by [14, Corollary 3.5] it is equivalent to a Π_1-formula $\pi(y,x)$, and this equivalence is provable in the theory $\mathrm{I}\Sigma_1$, a fragment of ***PA*** in which the axiom scheme of induction is restricted to Σ_1-formulas (with parameters), by [11, Definition I.4.3], this is to say that we have $\mathrm{I}\Sigma_1 \vdash \forall x,y[\mathtt{prf}_\tau(y,x) \leftrightarrow \pi(y,x)]$. Let us note that the theory $\mathrm{I}\Sigma_1$ is finitely axiomatizable, and 'arithmetization of metamathematics' can be developed in it; see [11].

Throughout the paper, we consider Δ_0–definable theories that extend $\mathrm{I}\Sigma_1$.[1]

Thus, for a Δ_0-formula τ the *proof predicate* $\mathtt{prf}_\tau$ of τ is a $\mathrm{I}\Sigma_1$-provably Δ_1-formula; though it could be a Δ_0-formula by the techniques of [11, §V.3]. Let $\mathtt{Pr}_\tau(x) = \exists y\, \mathtt{prf}_\tau(y,x)$ be the *provability predicate* of τ and $\mathtt{Con}_\tau = \neg\mathtt{Pr}_\tau(\#[0 \neq 0])$ be its *consistency statement*. Note that $\mathtt{Pr}_\tau(x)$ is a Σ_1-formula, and $\mathtt{Con}_\tau$ is a Π_1-sentence. For our coding and arithmetization we expect the following to hold.

CONVENTION 2.2 For every Δ_0-formula $\tau(x)$ and every sentence φ the following hold:
(C1) $\mathtt{Th}_\tau \vdash \varphi \iff \boldsymbol{Q} \vdash \mathtt{prf}_\tau(\overline{m}, \#\varphi)$ for some $m \in \mathbb{N}$.
(C2) $\mathtt{Th}_\tau \nvdash \varphi \implies \mathrm{I}\Sigma_1 \vdash \neg\mathtt{prf}_\tau(\overline{n}, \#\varphi)$ for every $n \in \mathbb{N}$.
Also, the following *derivability conditions* hold for $\mathtt{Pr}_\tau(x)$, when τ is a Δ_0-formula such that $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$ and φ, ψ are sentences:
(D1) $\mathtt{Th}_\tau \vdash \varphi \iff \boldsymbol{Q} \vdash \mathtt{Pr}_\tau(\#\varphi)$.
(D2) $\mathrm{I}\Sigma_1 \vdash \mathtt{Pr}_\tau(\#[\varphi \to \psi]) \to [\mathtt{Pr}_\tau(\#\varphi) \to \mathtt{Pr}_\tau(\#\psi)]$.
(D3) $\mathrm{I}\Sigma_1 \vdash \mathtt{Pr}_\tau(\#\varphi) \to \mathtt{Pr}_\tau(\#[\mathtt{Pr}_\tau(\#\varphi)])$. ◇

REMARK 2.3
We will need the following consequences of the derivability conditions in Convention 2.2, where τ is a Δ_0-formula such that $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$ and φ is an arbitrary sentence:
(I) $\mathtt{Th}_\tau \vdash \neg\mathtt{Con}_\tau \to \mathtt{Pr}_\tau(\#\varphi)$.
(II) If $\mathtt{Th}_\tau \vdash \mathtt{Pr}_\tau(\#\varphi) \to \varphi$, then $\mathtt{Th}_\tau \vdash \varphi$.
(III) If $\mathtt{Th}_\tau \vdash \varphi \leftrightarrow \neg\mathtt{Pr}_\tau(\#\varphi)$, then $\mathtt{Th}_\tau \vdash \varphi \leftrightarrow \mathtt{Con}_\tau$. ◇

One can find proofs for Remark 2.3 in [38] and [2]; let us note that Remark 2.3.II is the so-called Löb's Rule.

[1] By a result of [40], instead of $\mathrm{I}\Sigma_1$ one can take the slightly weaker theory $\mathsf{EA} + \mathsf{B}\Sigma_1$, which is equivalent to $\mathrm{I}\Delta_1$ by a result of [35].

3 Gödelian sentences and their truth

Gödel's proof of his incompleteness theorem uses the diagonal lemma (2.1) for the negation of the provability predicate of the Δ_0-formula τ.

DEFINITION 3.1 (Gödelian Sentences).
A sentence γ is called a Gödelian sentence of the Δ_0-formula $\tau(x)$ when γ is equivalent to its unprovability in the theory defined by τ, i.e., we have $\mathtt{Th}_\tau \vdash \gamma \leftrightarrow \neg\mathtt{Pr}_\tau(\#\gamma)$. ◇

By Remark 2.3.III any two Gödelian sentences of τ, when $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$, are $\mathtt{Th}_\tau$-provably equivalent; so, many authors talk of *the* Gödel sentence of $\mathtt{Th}_\tau$. In some sources, the assumptions and definitions underlying (variants of) the following argument are not always made fully explicit. However, Gödel [9] and several authors after him argue that the Gödelian sentences of a consistent theory are true, since
(1) they are provably equivalent to their unprovability in the theory, and
(2) they are indeed unprovable in the theory; and so
(3) they must be true.
It is argued in [17] that this line of reasoning does not demonstrate the truth of Gödelian sentences, and indeed some Σ_1-unsound theories may have false Gödelian sentences. In fact, step (2) in the above argument is redundant:

LEMMA 3.2
Suppose that for the Δ_0-formula $\tau(x)$, the theory $\mathtt{Th}_\tau$ is consistent and contains $\mathrm{I}\Sigma_1$. For every sentence φ, if $\mathtt{Th}_\tau \vdash \varphi \to \neg\mathtt{Pr}_\tau(\#\varphi)$, then $\mathtt{Th}_\tau \nvdash \varphi$.

PROOF. Because $\mathtt{Th}_\tau \vdash \varphi$ would imply on the one hand $\mathtt{Th}_\tau \vdash \neg\mathtt{Pr}_\tau(\#\varphi)$ by the assumption, and on the other hand $\mathtt{Th}_\tau \vdash \mathtt{Pr}_\tau(\#\varphi)$ by Convention 2.2.D1. □

So, the question of the validity of the above reasoning for the truth of Gödelian sentences boils down to the following question:

Does $\mathtt{Th}_\tau \vdash \gamma \leftrightarrow \neg\mathtt{Pr}_\tau(\#\gamma)$ *imply* $\mathbb{N} \vDash \gamma$,
for a Δ_0*-formula* τ *with consistent* $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$?

Put in another way,
under which conditions are all the Gödelian sentences of τ *true?*

We answer this question in the present section, and in the next section we answer a similar question for the Rosserian sentences (of arithmetical theories). Let us start with an amusing result (cf. [18, Theorem 1]):

PROPOSITION 3.3 (Characterizing Gödelian sentences of super-theories).
Let $\tau(x)$ be a Δ_0-formula such that $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$. The following are equivalent for a sentence φ:

(1) φ is unprovable in $\mathtt{Th}_\tau$, i.e. $\mathtt{Th}_\tau \nvdash \varphi$;
(2) φ is a Gödelian sentence of some consistent extension of $\mathtt{Th}_\tau$;
(3) $\mathtt{Th}_\tau + [\varphi \leftrightarrow \neg\mathtt{Pr}_\tau(\#\varphi)]$ is consistent.

PROOF. ($1 \Rightarrow 2$): By Lemma 2.1, for a sentence ξ we have

$$\boldsymbol{Q} \vdash \xi \leftrightarrow [\varphi \leftrightarrow \neg\, \mathtt{Pr}_{\tau'}(\#\varphi)]$$

where $\tau'(x) = \tau(x) \vee (x = \#\xi)$. Then $\mathtt{Th}_{\tau'} \vdash \varphi \leftrightarrow \neg \mathtt{Pr}_{\tau'}(\#\varphi)$ and it remains to show that the theory $\mathtt{Th}_{\tau'}$ (which is $\mathtt{Th}_\tau + \xi$) is consistent. If not, then $\mathtt{Th}_\tau \vdash \neg\xi$. So, on the one hand we have (i) $\mathtt{Th}_\tau \vdash \neg[\varphi \leftrightarrow \neg \mathtt{Pr}_{\tau'}(\#\varphi)]$, and on the other hand $\mathtt{Th}_{\tau'} \vdash \varphi$, which implies (ii) $\boldsymbol{Q} \vdash \mathtt{Pr}_{\tau'}(\#\varphi)$ by Convention 2.2.D1. Now, (i) and (ii) imply that $\mathtt{Th}_\tau \vdash \varphi$, contradicting the assumption.

(2 $\Rightarrow$ 3): Suppose that the theory $\mathtt{Th}_\tau + [\varphi \leftrightarrow \neg \mathtt{Pr}_\tau(\#\varphi)]$ is not consistent; then we have $\mathtt{Th}_\tau \vdash \neg[\varphi \leftrightarrow \neg \mathtt{Pr}_\tau(\#\varphi)]$, and so $\mathtt{Th}_\tau \vdash \mathtt{Pr}_\tau(\#\varphi) \rightarrow \varphi$, which implies $\mathtt{Th}_\tau \vdash \varphi$ by Löb's Rule (Remark 2.3.II). So, for every extension $\mathtt{Th}_{\tau'}$ of $\mathtt{Th}_\tau$ we have $\mathtt{Th}_{\tau'} \vdash \varphi$, and so by Convention 2.2.D1, we have $\boldsymbol{Q} \vdash \mathtt{Pr}_{\tau'}(\#\varphi)$. Therefore, for every such $\mathtt{Th}_{\tau'}$ we have $\mathtt{Th}_{\tau'} \vdash \neg[\varphi \leftrightarrow \neg \mathtt{Pr}_{\tau'}(\#\varphi)]$, which contradicts the assumption.

(3 $\Rightarrow$ 1): If $\mathtt{Th}_\tau \vdash \varphi$, then, by Convention 2.2.D1, we have $\boldsymbol{Q} \vdash \mathtt{Pr}_\tau(\#\varphi)$, and so we should have also $\mathtt{Th}_\tau \vdash \neg[\varphi \leftrightarrow \neg \mathtt{Pr}_\tau(\#\varphi)]$. □

We now provide a necessary and sufficient condition for the truth of all the Gödelian Π_1-sentences (cf. [17, Theorem 3.4]):

THEOREM 3.4 (On the truth and independence of Gödelian Π_1-sentences).
We assume that for the Δ_0-formula $\tau(x)$ we have $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$.

If $\mathtt{Th}_\tau \vdash \neg \mathtt{Con}_\tau$, then every false Π_1-sentence is a Gödelian sentence of τ, and no Gödelian sentence of τ is independent from $\mathtt{Th}_\tau$.

If $\mathtt{Th}_\tau \nvdash \neg \mathtt{Con}_\tau$, then all the Gödelian Π_1-sentences of τ are true, and all the Gödelian sentences of τ are independent from $\mathtt{Th}_\tau$.

PROOF. If $\mathtt{Th}_\tau \vdash \neg \mathtt{Con}_\tau$, then by Remark 2.3.I we have $\mathtt{Th}_\tau \vdash \mathtt{Pr}_\tau(\#\varphi)$ for every sentence φ. So, for every Gödelian sentence γ of τ we have $\mathtt{Th}_\tau \vdash \neg\gamma$; thus no Gödelian sentence of τ can be independent from $\mathtt{Th}_\tau$. Now, let ϕ be an arbitrary false Π_1-sentence; then $\neg\phi$ is a true Σ_1-sentence, and so provable in $\boldsymbol{Q}$. Thus, $\mathtt{Th}_\tau \vdash \neg\phi$; and so from $\mathtt{Th}_\tau \vdash \mathtt{Pr}_\tau(\#\phi)$ we have $\mathtt{Th}_\tau \vdash \phi \leftrightarrow \neg \mathtt{Pr}_\tau(\#\phi)$, which means that ϕ is a (false) Gödelian Π_1-sentence of τ.

If $\mathtt{Th}_\tau \nvdash \neg \mathtt{Con}_\tau$, then Remark 2.3.III implies that for every Gödelian sentence γ of τ we have $\mathtt{Th}_\tau \nvdash \neg\gamma$; thus, γ is independent from $\mathtt{Th}_\tau$ (noting that $\mathtt{Th}_\tau$ is consistent and so we also have $\mathtt{Th}_\tau \nvdash \gamma$ by Lemma 3.2). If a Gödelian Π_1-sentence γ of τ is not true, then $\neg\gamma$ is a true Σ_1-sentence, and so should be $\boldsymbol{Q}$-provable; a contradiction with the $\mathtt{Th}_\tau$-independence of γ, proved above. □

If the theory $\mathtt{Th}_\tau$ is Σ_1-sound, then we have $\mathtt{Th}_\tau \nvdash \neg \mathtt{Con}_\tau$. If $\mathtt{Th}_\tau$ is inconsistent or we have $\tau(x) = \vartheta(x) \vee (x = \#[\neg \mathtt{Con}_\vartheta])$ for a Δ_0-formula ϑ such that $\mathtt{Th}_\vartheta$ is a consistent extension of $\mathrm{I}\Sigma_1$, then $\mathtt{Th}_\tau \vdash \neg \mathtt{Con}_\tau$ (noting that $\mathtt{Th}_\tau = \mathtt{Th}_\vartheta + \neg \mathtt{Con}_\vartheta$); in the latter case $\mathtt{Th}_\tau$ is consistent by Gödel's second incompleteness theorem. Thus, by Theorem 3.4, a necessary and sufficient condition for the truth of all the Gödelian Π_1-sentences of τ is the consistency of $\mathtt{Th}_\tau$ with $\mathtt{Con}_\tau$, a condition obviously implied by ω-consistency, although this condition is stronger than the mere consistency of $\mathtt{Th}_\tau$; see [13, Theorem 36].

For investigating on the truth of Gödelian Π_{n+1}-sentences (and Σ_{n+1}-sentences) we make a definition and an observation. Before that let us note that no Gödelian Σ_1-sentence of a consistent Δ_0-definable extension of $\mathrm{I}\Sigma_1$ can be true:

PROPOSITION 3.5 (On the truth of Gödelian Σ_1-sentences).
For every Δ_0-formula $\tau(x)$, no Gödelian Σ_1-sentence of τ can be true if $\mathtt{Th}_\tau$ is consistent and contains $\mathrm{I}\Sigma_1$.

PROOF. If a Gödelian Σ_1-sentence of τ were true, then it would have been provable in $\boldsymbol{Q}$, and this would have contradicted Lemma 3.2 for consistent $\mathtt{Th}_\tau$. □

DEFINITION 3.6 (Υ-Soundness).
Let Υ be a class of sentences. A theory S is called Υ-sound when every S-provable Υ-sentence is true. ◇

The following lemma has been proved for $\Upsilon = \Sigma_1, \Sigma_2$ in [13, Theorems 25, 27, 30 and 32]:

LEMMA 3.7 (On extensions of Υ-sound theories).
Let Υ be a class of sentences that is closed under disjunction. If T is a Υ-sound theory, then for every sentence φ, either $T+\varphi$ or $T+\neg\varphi$ is Υ-sound.

PROOF. If neither $T+\varphi$ nor $T+\neg\varphi$ is Υ-sound, then for some false Υ-sentences ξ and ξ' we have $T+\varphi \vdash \xi$ and $T+\neg\varphi \vdash \xi'$. Thus, $T \vdash \xi \vee \xi'$, and $\xi \vee \xi'$ is a false Υ-sentence; a contradiction with the Υ-soundness of T. □

One of our main results is the following necessary and sufficient condition for the truth of Gödelian (Π_{n+1}- and Σ_{n+1}-) sentences:

THEOREM 3.8 (On the truth of Gödelian Π_{n+1}- and Σ_{n+1}-sentences).
Let $n \geqslant 1$, and let τ be a Δ_0-formula such that $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$.
All the Gödelian Π_{n+1}-sentences of τ are true if and only if $\mathtt{Th}_\tau$ is Π_{n+1}-sound.
All the Gödelian Σ_{n+1}-sentences of τ are true if and only if $\mathtt{Th}_\tau$ is Σ_{n+1}-sound.

PROOF. Let Υ be either of the two classes of sentences (either Π_{n+1} or Σ_{n+1}).
First, suppose that $\mathtt{Th}_\tau$ is Υ-sound, and let γ be a Gödelian Υ-sentence of τ. By Lemma 3.2 and Convention 2.2.D1 we have $\mathbb{N} \vDash \neg\mathtt{Pr}_\tau(\#\gamma)$, and so $\mathtt{Pr}_\tau(\#\gamma)$ is a false Σ_1-sentence. Now, $\mathtt{Th}_\tau + \neg\gamma \vdash \mathtt{Pr}_\tau(\#\gamma)$, and so $\mathtt{Th}_\tau + \neg\gamma$ is not Σ_1-sound; hence, it is not Υ-sound either. Thus, by Lemma 3.7, the theory $\mathtt{Th}_\tau + \gamma$ should be Υ-sound. Therefore, γ must be true.[2]
Now, suppose that all the Gödelian Υ-sentences of τ are true. We show that the theory $\mathtt{Th}_\tau$ is Υ-sound. Assume that $\mathtt{Th}_\tau \vdash \xi$ for a Υ-sentence ξ. We prove that ξ is true. By Lemma 2.1 there exists a Υ-sentence γ such that $\boldsymbol{Q} \vdash \gamma \leftrightarrow [\xi \wedge \neg\mathtt{Pr}_\tau(\#\gamma)]$. Thus, from $\mathtt{Th}_\tau \vdash \xi$ we have $\mathtt{Th}_\tau \vdash \gamma \leftrightarrow \neg\mathtt{Pr}_\tau(\#\gamma)$, and so γ is a Gödelian Υ-sentence of τ. Hence, γ is true, and so, by the soundness of $\boldsymbol{Q}$, we have $\mathbb{N} \vDash \xi$. □

Hence, all the Gödelian sentences of a theory are true if and only if the theory is sound; cf. [36, Theorem 24.7].

REMARK 3.9 (On the hierarchy of Π_n- and Σ_n-soundness).
Let us note that an extension of $\boldsymbol{Q}$ is consistent if and only if it is Π_1-sound: indeed, no consistent extension of $\boldsymbol{Q}$ can prove a false Π_1-sentence, since the negation of such a sentence would be a true Σ_1-sentence and so would be provable in $\boldsymbol{Q}$.
One can also show that a theory is Σ_n-sound if and only if it is Π_{n+1}-sound: if the theory S is Σ_n-sound and $S \vdash \pi$, where π is a Π_{n+1}-sentence, then write $\pi = \forall x\, \sigma(x)$ for a Σ_n-formula σ; since

[2] Another proof (without appeal to Lemma 3.7): If γ is a Gödelian Υ-sentence of τ, then $\mathtt{Th}_\tau \vdash \gamma \vee \mathtt{Pr}_\tau(\#\gamma)$ and so $\mathbb{N} \vDash \gamma \vee \mathtt{Pr}_\tau(\#\gamma)$ since $\gamma \vee \mathtt{Pr}_\tau(\#\gamma)$ is a Υ-sentence and $\mathtt{Th}_\tau$ is Υ-sound; as $\mathbb{N} \nvDash \mathtt{Pr}_\tau(\#\gamma)$ by Lemma 3.2 and Convention 2.2.D1, we should have $\mathbb{N} \vDash \gamma$. *QED*

for every $k\in\mathbb{N}$ we have $S\vdash\sigma(\overline{k})$, and $\sigma(\overline{k})$ is a Σ_n-sentence, then $\mathbb{N}\vDash\sigma(\overline{k})$ for every $k\in\mathbb{N}$, so $\mathbb{N}\vDash\forall x\,\sigma(x)=\pi$.

The hierarchy of Σ_n-sound theories is strict, since there exist some Σ_n-sound theories, which are not Σ_{n+1}-sound; see, e.g. [32, Theorem 2.5] or [15, Theorem 4.8]. Therefore, the truth of (even all) the Gödelian Π_{n+1}-sentences (respectively, Σ_{n+1}-sentences) of a theory does not necessarily imply the truth of its Gödelian Π_{n+2}-sentences (respectively, Σ_{n+2}-sentences). ◇

4 Rosserian sentences and their truth

In Theorem 3.4 we saw that (all of the) Gödelian sentences of some theories could be refutable in those theories (though, they are always unprovable in consistent theories, see Lemma 3.2). Rosser's trick [31] constructs an independent sentence for a given theory when it is consistent (recall that our theories are RE extensions of $\boldsymbol{Q}$). Before going into Rosser's construction, let us note that no construction similar to Gödel's can result in an independent sentence.

DEFINITION 4.1 (Pseudo-Gödelian sentences).
Let τ be a Δ_0-formula. Let us call the sentence ψ a *pseudo-Gödelian sentence* of τ when there exist some propositional formulas $C_1(p),\cdots,C_n(p)$, over the propositional variable p, and there exists a propositional formula $B(p_1,\cdots,p_n)$, over the propositional variables $p_1,\cdots,p_n$, such that we have $\mathtt{Th}_\tau\vdash\psi\leftrightarrow B\big(\mathtt{Pr}_\tau[\#C_1(\psi)],\cdots,\mathtt{Pr}_\tau[\#C_n(\psi)]\big)$. ◇

For example, the sentences $\mathscr{P}$ and $\mathscr{R}$ for which we have

$$\mathtt{Th}_\tau\vdash\mathscr{P}\leftrightarrow\big[\neg\,\mathtt{Pr}_\tau(\#\mathscr{P})\wedge\neg\,\mathtt{Pr}_\tau(\#[\neg\mathscr{P}])\big]$$

and

$$\mathtt{Th}_\tau\vdash\mathscr{R}\leftrightarrow\big[\,\mathtt{Pr}_\tau(\#\mathscr{R})\rightarrow\mathtt{Pr}_\tau(\#[\neg\mathscr{R}])\big]$$

are both some peudo-Gödelian sentences of τ.

For a Δ_0-formula τ such that $\mathtt{Th}_\tau\supseteq\mathrm{I}\Sigma_1$ is consistent, let $\upsilon(x)=\tau(x)\vee(x=\#[\neg\mathtt{Con}_\tau])$. The theory $\mathtt{Th}_\upsilon$, which is $\mathtt{Th}_\tau+\neg\mathtt{Con}_\tau$, is consistent by Gödel's second incompleteness theorem. Now, from $\mathtt{Th}_\upsilon\vdash\neg\mathtt{Con}_\upsilon$, and Remark 2.3.I, we have $\mathtt{Th}_\upsilon\vdash\mathtt{Pr}_\upsilon(\#\theta)$ for every sentence θ. Hence, $\mathtt{Th}_\upsilon$ decides every pseudo-Gödelian sentence, and so we have the following Proposition (4.2); cf. [38, Exercise 1, p.149].

PROPOSITION 4.2 (On the decidability of pseudo-Gödelian sentences).
Let τ be a Δ_0-formula, and suppose that the theory $\mathtt{Th}_\tau$ is consistent and contains $\mathrm{I}\Sigma_1$. Let $\upsilon(x)=\tau(x)\vee(x=\#[\neg\mathtt{Con}_\tau])$. Then, no pseudo-Gödelian sentence of υ can be independent from the theory $\mathtt{Th}_\upsilon$. □

In the above examples, it can be seen that $\mathtt{Th}_\upsilon\vdash\neg\mathscr{P}$ and $\mathtt{Th}_\upsilon\vdash\mathscr{R}$ (noting that we have $\mathtt{Th}_\upsilon=\mathtt{Th}_\tau+\neg\mathtt{Con}_\tau$). Thus, for getting independent sentences (of consistent theories) one should go beyond the (pesudo-)Gödelian sentences.

DEFINITION 4.3 (Rosserian provability and Rosserian sentences).
The Rosserian provability predicate of a Δ_0-formula τ is
$\mathtt{R.Pr}_\tau(x)=\exists y\,[\mathtt{prf}_\tau(y,x)\wedge\forall z<y\,\neg\mathtt{prf}_\tau(z,\neg x)]$.
If $\mathtt{Th}_\tau\vdash\rho\leftrightarrow\neg\mathtt{R.Pr}_\tau(\#\rho)$, then ρ is called a Rosserian sentence of τ. ◇

Let us note that $\mathtt{R.Pr}_\tau(x)$ is an $\mathrm{I}\Sigma_1$-provably Σ_1-formula, when $\tau(x)$ is Δ_0; so, ρ is an $\mathrm{I}\Sigma_1$-provably Π_1-sentence, cf. [11, Remark III.4.19]. The independence of the Rosserian sentences (from the theory in question) follows from the following basic properties of the Rosserian provability:

LEMMA 4.4
If $\mathtt{Th}_\tau$ is consistent and contains $\mathrm{I}\Sigma_1$ for a Δ_0-formula τ, then for every sentence φ we have
(1) $\mathtt{Th}_\tau \vdash \varphi \iff \mathrm{I}\Sigma_1 \vdash \mathtt{R.Pr}_\tau(\#\varphi)$.
(2) $\mathtt{Th}_\tau \vdash \neg\varphi \implies \mathrm{I}\Sigma_1 \vdash \neg\mathtt{R.Pr}_\tau(\#\varphi)$.

PROOF. For (1) it suffices to note that for consistent $\mathtt{Th}_\tau$ we have $\mathtt{Th}_\tau \vdash \varphi$ if and only if the $\mathrm{I}\Sigma_1$-provably Σ_1-sentence $\mathtt{R.Pr}_\tau(\#\varphi)$ is true. For (2) suppose that $\mathtt{Th}_\tau \vdash \neg\varphi$; then by Convention 2.2.C1 we have $\boldsymbol{Q} \vdash \mathtt{prf}_\tau(\overline{m}, \#[\neg\varphi])$ for some m. Now, reason inside $\mathrm{I}\Sigma_1$:

> For any y with $\mathtt{prf}_\tau(y, \#\varphi)$ we have $y > \overline{m}$, since no $i \leqslant \overline{m}$ (which are $i = 0, \cdots, \overline{m}$) could satisfy $\mathtt{prf}_\tau(\overline{i}, \#\varphi)$ by Convention 2.2.C2, and so for some $z < y$, which is $z = \overline{m}$, we have $\mathtt{prf}_\tau(z, \#[\neg\varphi])$. Thus, $\forall y[\mathtt{prf}_\tau(y, \#\varphi) \rightarrow \exists z < y\, \mathtt{prf}_\tau(z, \#[\neg\varphi])]$ holds, and so $\neg\mathtt{R.Pr}_\tau(\#\varphi)$. □

Now, we can characterize the Rosserian sentences of super-theories:

PROPOSITION 4.5 (Characterizing Rosserian sentences of super-theories).
Let τ be a Δ_0-formula such that $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$. The following are equivalent for a sentence φ:

(1) φ is independent from $\mathtt{Th}_\tau$, i.e. $\mathtt{Th}_\tau \nvdash \varphi$ and $\mathtt{Th}_\tau \nvdash \neg\varphi$;
(2) φ is a Rosserian sentence of some consistent extension of $\mathtt{Th}_\tau$;

and are implied by the following:

(3) $\mathtt{Th}_\tau + [\varphi \leftrightarrow \neg\mathtt{R.Pr}_\tau(\#\varphi)]$ is consistent.

PROOF. First we show the equivalence of (1) and (2).

(1 ⇒ 2): By Lemma 2.1 we have $\boldsymbol{Q} \vdash \xi \leftrightarrow [\varphi \leftrightarrow \neg\mathtt{R.Pr}_{\tau'}(\#\varphi)]$ for some sentence ξ where $\tau'(x) = \tau(x) \vee (x = \#\xi)$. Then $\mathtt{Th}_{\tau'} \vdash \varphi \leftrightarrow \neg\mathtt{R.Pr}_{\tau'}(\#\varphi)$, which shows that φ is a Rosserian sentence of τ'. We show that the theory $\mathtt{Th}_{\tau'}$ is consistent. If not, then $\mathtt{Th}_\tau \vdash \neg\xi$. Thus, we have $(*)$ $\mathtt{Th}_\tau \vdash \neg[\varphi \leftrightarrow \neg\mathtt{R.Pr}_{\tau'}(\#\varphi)]$. Also, $\mathtt{Th}_{\tau'} \vdash \varphi$ and $\mathtt{Th}_{\tau'} \vdash \neg\varphi$, and so by Convention 2.2.C1 there are $m, n \in \mathbb{N}$ such that $\boldsymbol{Q} \vdash \mathtt{prf}_{\tau'}(\overline{m}, \#\varphi)$ and $\boldsymbol{Q} \vdash \mathtt{prf}_{\tau'}(\overline{n}, \#[\neg\varphi])$; we can assume that m and n are the least such numbers.

(i) If $m \leqslant n$, then $\mathrm{I}\Sigma_1 \vdash \mathtt{prf}_{\tau'}(\overline{m}, \#\varphi) \wedge \forall z < \overline{m} \neg\mathtt{prf}_{\tau'}(z, \#[\neg\varphi])$ and so $\mathrm{I}\Sigma_1 \vdash \mathtt{R.Pr}_{\tau'}(\#\varphi)$, which implies by $(*)$ that $\mathtt{Th}_\tau \vdash \varphi$, contradicting (1).

(ii) If $n < m$, then we have $\mathrm{I}\Sigma_1$-provably that for every y:

$$\begin{array}{lll} & \mathtt{prf}_{\tau'}(y, \#\varphi) & \\ \rightarrow & y \geqslant \overline{m} & \text{since } m \text{ is the least with } \mathtt{prf}_{\tau'}(m, \#\varphi) \\ \rightarrow & \exists z < y\ \mathtt{prf}_{\tau'}(z, \#[\neg\varphi]) & \text{since one can take } z = \overline{n} (< \overline{m} \leqslant y). \end{array}$$

So, $\mathrm{I}\Sigma_1 \vdash \neg\mathtt{R.Pr}_{\tau'}(\#\varphi)$, which implies by $(*)$ that $\mathtt{Th}_\tau \vdash \neg\varphi$, and this contradicts (1).

Thus, $\mathtt{Th}_{\tau'}$ must be consistent.

(2 ⇒ 1): Suppose that $\mathtt{Th}_{\tau'}$ is a consistent extension of $\mathtt{Th}_\tau$ such that φ is a Rosserian sentence of it. It suffices to show that φ is independent from $\mathtt{Th}_{\tau'}$. If $\mathtt{Th}_{\tau'} \vdash \varphi$, then we should have on the one hand $\mathtt{Th}_{\tau'} \vdash \mathtt{R.Pr}_{\tau'}(\#\varphi)$ by Lemma 4.4.1 and on the other hand $\mathtt{Th}_{\tau'} \vdash \neg\mathtt{R.Pr}_{\tau'}(\#\varphi)$ by Definition 4.3; contradicting the consistency of $\mathtt{Th}_{\tau'}$. Also, $\mathtt{Th}_{\tau'} \vdash \neg\varphi$ would imply on the one hand $\mathtt{Th}_{\tau'} \vdash \neg\mathtt{R.Pr}_{\tau'}(\#\varphi)$ by Lemma 4.4.2 and on the other hand $\mathtt{Th}_{\tau'} \vdash \mathtt{R.Pr}_{\tau'}(\#\varphi)$ by Definition 4.3; contradicting $\mathtt{Th}_{\tau'}$'s consistency again.

Now, we show that (3) implies (1), and so (2) too.

$(3 \Rightarrow 1)$: Note that $\mathtt{Th}_\tau$ is consistent by the assumption. If $\mathtt{Th}_\tau \vdash \varphi$, then $\mathtt{Th}_\tau \vdash \mathtt{R.Pr}_\tau(\#\varphi)$ by Lemma 4.4.1, and so $\mathtt{Th}_\tau \vdash \neg[\varphi \leftrightarrow \neg\mathtt{R.Pr}_\tau(\#\varphi)]$. If $\mathtt{Th}_\tau \vdash \neg\varphi$, then Lemma 4.4.2 would imply that $\mathtt{Th}_\tau \vdash \neg\mathtt{R.Pr}_\tau(\#\varphi)$, and so $\mathtt{Th}_\tau \vdash \neg[\varphi \leftrightarrow \neg\mathtt{R.Pr}_\tau(\#\varphi)]$ would hold again. □

REMARK 4.6 (Löb's rule for Rosserian provability).
Let us note that the contraposition of the implication $(1 \Rightarrow 3)$ in Proposition 4.5 says that if $\mathtt{Th}_\tau \vdash \varphi \leftrightarrow \mathtt{R.Pr}_\tau(\#\varphi)$, i.e. if φ is a *Rosser-type Henkin sentence*,[3] so-called in [16], then φ is not independent from T. Actually, it is shown in [16] that there are *standard* proof predicates (i.e. those satisfying Convention 2.2), which have independent Rosser-type Henkin sentences, and there are standard proof predicates none of whose Rosser-type Henkin sentences are independent. The latter proof predicates satisfy $(1 \Rightarrow 3)$ in Proposition 4.5 and satisfy a Löb-like rule for Rosserian provability, while the former ones do not satisfy $(1 \Rightarrow 3)$ in Proposition 4.5 and do not satisfy any Löb-like rule for Rosserian provability. So, the implication $(1 \Rightarrow 3)$ in Proposition 4.5 depends on $\mathtt{prf}_\tau(y,x)$, and is not robust. ◇

Unlike Gödelian Π_1-sentences, all the Rosserian Π_1-sentences of consistent theories are true, and like Gödelian Σ_1-sentence, all of their Rosserian Σ_1-sentences are false:

THEOREM 4.7 (On the truth of Rosserian Π_1- and Σ_1-sentences).
For an arbitrary Δ_0-formula $\tau(x)$, every Rosserian Π_1-sentence of τ is true and every Rosserian Σ_1-sentence of τ is false, if $\mathtt{Th}_\tau$ is consistent and contains $\mathrm{I}\Sigma_1$.

PROOF. If a Rosserian Π_1-sentence of τ were false, then its negation would be a true Σ_1-sentence, and so would be provable in $\boldsymbol{Q}$; contradicting Rosser's theorem on the independence of Rosserian sentences (see Proposition 4.5). If a Rosserian Σ_1-sentence were true, then it would be provable in $\boldsymbol{Q}$; contradicting the unprovability of Rosserian sentences. □

However, for $n \geqslant 1$, the truth of all the Gödelian Π_{n+1}-sentences is equivalent to the truth of all the Rosserian Π_{n+1}-sentences and the truth of all the Gödelian Σ_{n+1}-sentences is equivalent to the truth of all the Rosserian Σ_{n+1}-sentences:

THEOREM 4.8 (On the truth of Rosserian Π_{n+1}- and Σ_{n+1}-sentences).
Let $n \geqslant 1$, and let τ be a Δ_0-formula such that $\mathtt{Th}_\tau \supseteq \mathrm{I}\Sigma_1$.

All the Rosserian Π_{n+1}-sentences of τ are true iff $\mathtt{Th}_\tau$ is Π_{n+1}-sound.

All the Rosserian Σ_{n+1}-sentences of τ are true iff $\mathtt{Th}_\tau$ is Σ_{n+1}-sound.

PROOF. Let Υ be either of the two classes of sentences (either Π_{n+1} or Σ_{n+1}). If $\mathtt{Th}_\tau$ is Υ-sound and ρ is a Rosserian Υ-sentence of τ, then $\mathtt{R.Pr}_\tau(\#\rho)$ is a false $\mathrm{I}\Sigma_1$-provably Σ_1-sentence by Proposition 4.5 and Lemma 4.4.1. Since $\mathtt{Th}_\tau \vdash \rho \vee \mathtt{R.Pr}_\tau(\#\rho)$ and $\rho \vee \mathtt{R.Pr}_\tau(\#\rho)$ is a Υ-sentence, then $\mathbb{N} \vDash \rho \vee \mathtt{R.Pr}_\tau(\#\rho)$, and so ρ is true. Now, suppose that all the Rosserian Υ-sentence of τ are true and $\mathtt{Th}_\tau \vdash \xi$, where ξ is a Υ-sentence. By Lemma 2.1 there is a Υ-sentence ρ such that $\mathrm{I}\Sigma_1 \vdash \rho \leftrightarrow [\xi \wedge \neg\mathtt{R.Pr}_\tau(\#\rho)]$. So, ρ is a Rosserian Υ-sentence of τ; thus, it is true by the assumption. Therefore, by the soundness of $\mathrm{I}\Sigma_1$ the sentence ξ is true too. Hence, $\mathtt{Th}_\tau$ is Υ-sound. □

Therefore, all the Rosserian sentences of τ are true if and only if $\mathtt{Th}_\tau$ is sound; cf. also [36, Theorem 24.7].

[3] The sentence ψ is a *Henkin sentence* (of τ) when it is equivalent to its own provability in the theory, i.e. when we have $\mathtt{Th}_\tau \vdash \psi \leftrightarrow \mathtt{Pr}_\tau(\#\psi)$. A *Rosser-type Henkin sentence* φ is equivalent to its own Rosserian provability in the theory, i.e. we have that $\mathtt{Th}_\tau \vdash \varphi \leftrightarrow \mathtt{R.Pr}_\tau(\#\varphi)$.

5 Conclusions

The first one who talked about the truth of Gödelian sentences was Gödel himself [9]. This turned into a serious debate with [10] in which (what we call now) the *Gödel Disjunction* was announced; see [8] and [12], and the references therein. The so called *Anti-Mechanism Thesis*, or the *Lucas-Penrose Argument*, started with [19] and was popularized by [25]; see also [24] and [29]. After that, there has been a large discussion on the truth of Gödelian sentences; see, e.g. [7], [37, 38], [1], [34], [20, 21], [39], [30], [26], [22], [33], [13], [3], [5, 6], [27, 28] and [23].

As argued above, the consistency of a theory does not imply the truth of (all of) its Gödelian Π_1-sentences, but does imply the truth of its all Rosserian Π_1-sentences. One may wonder why the proponents of the anti-mechanism thesis have not used the Rosserian Π_1-sentences in their arguments, given that the truth of those sentences is straightforward (and immediately follows from the consistency of the theory). Though, the opponents have argued that actually for 'seeing' the truth of Gödelian Π_1-sentences one should 'see' (at least) the consistency of the theory (and indeed, more than that).

The following diagram summarizes our old and new results. Note that the conditions get (strictly) stronger from bottom to top. As the diagram shows, if one faces the question as to whether a given Gödelian sentence γ of a consistent and sufficiently strong RE theory T is true or not, then one should consider the complexity of the sentence: if γ is Σ_1, then it is false; if γ is Π_1, then it is true when T is consistent with the consistency statement of T; if γ is Π_2, then it is true when T is Σ_1-sound; and, finally, if γ is Σ_{n+1} or Π_{n+2} for some $n \geqslant 1$, then it is true when T is Σ_{n+1}-sound. Let ρ be a Rosserian sentence of such a theory T; if ρ is Σ_1, then it is false; if ρ is Π_1, then it is true; if ρ is Π_2, then it is true when T is Σ_1-sound; and if ρ is Σ_{n+1} or Π_{n+2} for some $n \geqslant 1$, then it is true when T is Σ_{n+1}-sound.

Soundness	≡	Truth of all the Gödelian sentences
	≡	Truth of all the Rosserian sentences
....	.	....
$(n \geqslant 1)$ Σ_{n+1}-soundness	≡	Π_{n+2}-soundness
	≡	Truth of all the Gödelian Σ_{n+1}-sentences
	≡	Truth of all the Gödelian Π_{n+2}-sentences
	≡	Truth of all the Rosserian Σ_{n+1}-sentences
	≡	Truth of all the Rosserian Π_{n+2}-sentences
....	.	....
....	.	....
Σ_1-soundness	≡	Π_2-soundness
	≡	Truth of all the Gödelian Π_2-sentences
	≡	Truth of all the Rosserian Π_2-sentences
....	.	....
Consistency with $\mathtt{Con}_T$	≡	Truth of all the Gödelian Π_1-sentences
....	.	....
Consistency	≡	Π_1-soundness
	≡	Truth of all the Rosserian Π_1-sentences
	≡	*Falsity* of all the Gödelian Σ_1-sentences
	≡	*Falsity* of all the Rosserian Σ_1-sentences

Acknowledgements

This research is supported by the grant 98013437 of the Iran National Science Foundation (INSF). The authors warmly thank Kaave Lajevardi for the most helpful discussions and comments. This is a continuation of a project that he started a while ago, which have resulted in e.g. [17, 18].

References

[1] G. Boolos. On "seeing" the truth of the Gödel sentence. *Behavioral and Brain Sciences*, **13**, 655–656, 1990. Reprinted in: R. Jeffrey, ed., G. Boolos, *Logic, Logic, and Logic*, pp. 389–391. Harvard University Press, 1999.

[2] G. Boolos. *The Logic of Provability*. Cambridge University Press, 1994.

[3] J. Boyer and G. Sandu. Between proof and truth. *Synthese*, **187**, 821–832. Erratum: ibid, 973–974, 2012.

[4] W. Craig. On axiomatizability within a system. *The Journal of Symbolic Logic*, **18**, 30–32, 1953.

[5] V. Drăghici. How do we know that G is true? *Logos Architekton, Journal of Logic and Philosophy of Science*, **6**, 39–65, 2012.

[6] V. Drăghici. Is G true by Gödel's theorem? *Logos Architekton, Journal of Logic and Philosophy of Science*, **7**, 53–59, 2013.

[7] M. Dummett. The philosophical significance of Gödel's theorem. *Ratio*, **5**, 140–155, 1963. Reprinted in *Truth and Other Enigmas*, M. Dummett, pp. 186–201. Harvard University Press (1978, 6th print), 1996.

[8] S. Feferman. Are there absolutely unsolvable problems? Godel's dichotomy. *Philosophia Mathematica*, **14**, 134–152, 2006.

[9] K. Gödel. Über Formal Unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme I. *Monatshefte für Mathematik und Physik*, **38**, 173–198, 1931 (in German). English Translation: On Formally Undecidable Propositions of Principia Mathematica and Related Systems I. In Kurt Gödel Collected Works, Volume I: Publications 1929–1936, S. Feferman, et al., eds., pp. 135–152. Oxford University Press, 1986.

[10] K. Gödel. Some basic theorems on the foundations of mathematics and their implications. In *Kurt Gödel Collected Works, Volume III: Unpublished Essays and Lectures*, S. Feferman, et al., eds., pp. 304–323. Oxford University Press (1995), 1951.

[11] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic*. Springer, 1993.

[12] L. Horsten and P. Welch, eds., *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*. Oxford University Press, 2016.

[13] D. Isaacson. Necessary and sufficient conditions for undecidability of the Gödel sentence and its truth. In *Logic, Mathematics, Philosophy, Vintage Enthusiasms*, D. DeVidi, M. Hallett and P. Clarke, eds., pp. 135–152. Springer, 2011.

[14] R. Kaye. *Models of Peano Arithmetic*. Oxford University Press, 1991.

[15] M. Kikuchi and T. Kurahashi. Generalizations of Gödel's incompleteness theorems for Σ_n-definable theories of arithmetic. *The Review of Symbolic Logic*, **10**, 603–616, 2017.

[16] T. Kurahashi. Henkin sentences and local reflection principles for Rosser provability. *Annals of Pure and Applied Logic*, **167**, 73–94, 2016.

[17] K. Lajevardi and S. Salehi. On the arithmetical truth of self-referential sentences. *Theoria: A Swedish Journal of Philosophy*, **85**, 8–17, 2019.

[18] K. Lajevardi and S. Salehi. There may be many arithmetical Gödel Sentences. *Philosophia Mathematica*, **29**, 278–287, 2021.

[19] J. R. Lucas. Minds, machines and Gödel. *Philosophy*, **36**, 112–127, 1961. Reprinted in: *The Modeling of Mind*, K.M. Sayre and F.J. Crosson, eds., pp. 255–271. University of Notre Dame Press, 1963; and in *Minds and Machines*, A.R. Anderson, ed., pp. 43–59. Prentice Hall, 1964.
[20] S. McCall. Can a Turing machine know that the Gödel sentence is true? *The Journal of Philosophy*, **96**, 525–532, 1999. Reprinted in: *The Consistency of Arithmetic, and Other Essays*, S. McCall, pp. 28–35. Oxford University Press, 2014.
[21] S. McCall. On "seeing" the truth of the Gödel sentence. *Facta Philosophica*, **3**, 25–29, 2001. Reprinted in: *The Consistency of Arithmetic, and Other Essays*, S. McCall, pp. 36–40. Oxford University Press, 2014.
[22] P. Milne. On Gödel sentences and what they say. *Philosophia Mathematica*, **15**, 193–226, 2007.
[23] E. Moriconi. Some remarks on *true* undecidable sentences. In *Truth, Existence and Explanation*, M. Piaza and G. Pulcini, eds., pp. 3–15. Springer, 2018.
[24] E. Nagel and J. R. Newman. *Gödel's Proof*. New York University Press, 1958 (revised: 3rd edn. Routledge, 2005).
[25] R. Penrose. *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics*. Oxford University Press, 1989 (republished with a new preface, 1999).
[26] J. Peregrin. Kurt Gödel, completeness, incompleteness. *Journal of Physics: Conference Series*, **82**, pp. 1–4, 2007.
[27] M. Piazza and G. Pulcini. A deflationary account of the truth of the Gödel sentence $\mathcal{G}$. In *From Logic to Practice*, G. Lolli, M. Panza and G. Venturi, eds., pp. 71–90. Springer, 2015.
[28] M. Piazza and G. Pulcini. What's so special about the Gödel sentence $\mathcal{G}$? In *Objectivity, Realism, and Proof*, F. Boccuni and A. Sereni, eds., pp. 245–263. Springer, 2016.
[29] H. Putnam. Minds and machines. In *Dimensions of Mind: A Symposium*, S. Hood, ed., pp. 138–164. New York University Press, 1960. Reprinted in: Philosophical Papers, Volume 2: Mind, Language and Reality, H. Putnam, pp. 362–385. Harvard University Press, 1975.
[30] P. Raatikainen. On the philosophical relevance of Gödel's incompleteness theorems. *Revue Internationale de Philosophie*, **59**, 513–534, 2005.
[31] B. Rosser. Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic*, **1**, 87–91, 1936.
[32] S. Salehi and P. Seraji. Gödel–Rosser's incompleteness theorem, generalized and optimized for definable theories. *Journal of Logic and Computation*, **27**, 1391–1397, 2017.
[33] G. Serény. How do we know that the Gödel sentence of a consistent theory is true? *Philosophia Mathematica*, **19**, 47–73, 2011.
[34] S. Shapiro. Induction and indefinite extensibility: the Gödel sentence is true, but did someone change the subject? *Mind*, **107**, 597–624, 1998.
[35] T. A. Slaman. Σ_n-bounding and Δ_n-induction. *Proceedings of The American Mathematical Society*, **132**, 2449–2456, 2004.
[36] P. Smith. *An Introduction to Gödel's Theorems*, 2nd edn. Cambridge University Press, 2013.
[37] C. Smoryński. The incompleteness theorems. In *Handbook of Mathematical Logic*, J. Barwise, ed., pp. 821–865. North Holland, 1977.
[38] C. Smoryński. *Self-Reference and Modal Logic*. Springer, 1985.
[39] N. Tennant. On Turing machines knowing their own Gödel sentences. *Philosophia Mathematica*, **9**, 72–79, 2001.
[40] A. Visser. Another look at the second incompleteness theorem. *The Review of Symbolic Logic*, **13**, 269–295, 2020.

Received 1 July 2021