278–287
10.1093/philmat/nkaa041
Philosophia Mathematica
Advance Access Publication on February 14, 2021

There May Be Many Arithmetical Gödel Sentence*s*[†]

Kaave Lajevardi* and Saeed Salehi**

*La Société des Philosophes Chômeurs, P.O. Box 13197-73587, Téhéran, Iran.
**Research Institute for Fundamental Sciences, University of Tabriz, Tabriz, Iran.

ABSTRACT

We argue that, under the usual assumptions for sufficiently strong arithmetical theories that are subject to Gödel's First Incompleteness Theorem, one cannot, without impropriety, talk about *the* Gödel sentence of the theory. The reason is that, without violating the requirements of Gödel's theorem, there could be a true sentence and a false one each of which is provably equivalent to its own unprovability in the theory if the theory is unsound.

1. INTRODUCTION AND PRELIMINARIES

In the course of proving what is now known as the *First Incompleteness Theorem* for every arithmetical theory T satisfying certain conditions, Gödel constructs a sentence which "says about itself" that it is T-unprovable. He then shows that that sentence is unprovable if the theory is consistent, and is irrefutable if the theory is ω-consistent.[1]

Year after year, and decade after decade, whether authoring textbook presentations of Gödel's theorem or writing research papers, a considerable number of world-class logicians and philosophers seem to have called each and every such sentence (*i.e.*, each and every sentence which is provably equivalent to its

† We are most grateful to the two anonymous referees of this journal for their stimulating and fruitful comments and suggestions. Saeed Salehi is supported by a grant from the Office of the Vice Chancellor for Research and Technology, University of Tabriz, Iran.

* Orcid.org/0000-0003-3710-237X. E-mail: Kaave.Lajevardi@gmail.com.
** Orcid.org/0000-0001-6961-5749. E-mail: root@SaeedSalehi.ir

[1] Simple consistency is obviously sufficient for showing the unprovability of that sentence; Gödel evidently introduced the stronger assumption of ω-consistency in order to show its irrefutability. As is well known, Rosser later weakened the requirement of ω-consistency to that of simple consistency — though, of course, Rosser shows the independence of *another* sentence (different from the one constructed by Gödel).

own unprovability), or a variant thereof, *the Gödel sentence* of the theory. In this short note we aim at showing that this is a misnomer. More specifically, we argue that the use of the definite article here is unjustified — and this we hold despite the initial plausibility of the *the* talk.

One main reason for speaking of *the* Gödel sentence, with its implied uniqueness, seems to be the well-known fact that for every theory T presented in a certain way, if each of the sentences A and B is, in the eye of T, equivalent to *its* own T-unprovability, then A and B are, in the eye of T, equivalent — in fact, A and B are both equivalent to a quite distinguished sentence, one which is normally taken to state the consistency of T. Many authors who write about Gödel's theorems are of course aware of this logical fact — a fact thus stated as Remark 2.2.5 in one of the technically most elegant papers on the incompleteness theorems:

> By the proof of the Second [Incompleteness] Theorem, the self-referential sentence which asserts its own unprovability is equivalent to the sentence asserting consistency. Hence, this sentence is unique up to provable equivalence and one may correctly speak of *the* sentence that asserts its own unprovability. [Smoryński, 1977, p. 829, emphasis original]

While the equivalence mentioned by Smoryński is of course the case, we think it does not warrant the *the* talk — the fact that there is, up to T-equivalence, only one sentence which asserts its own unprovability does not guarantee that there is just one Gödelian sentence, not even up to truth value. Or so we will argue.

How could that be?

As is well known, while in his introductory section Gödel takes theories to be *sound* (*i.e.*, having all their theorems true in the standard model $\mathbb{N}$ of natural numbers), in the official statement of the First Incompleteness Theorem the theories in question need not be sound. In fact, Gödel makes it explicit that one purpose for giving formally precise proofs of what he informally argues for in his Section 1 is to replace the requirement of soundness with a "much weaker" one, namely ω-consistency [1931, p. 151].

Fact 1 *Theories subject to the First Incompleteness Theorem need not be sound. Hence some of their theorems may be false* (i.e., *false in* $\mathbb{N}$).

Fact 2 *There are unsound theories which are ω-consistent.*

Gödel does not elaborate on how or why ω-consistency is weaker than soundness; for some examples of ω-consistent theories which are unsound, see [Isaacson, 2011, Proposition 19] (credited to Kreisel) or [Lindström, 1997, Chapter 2, Exercise 7(d)]. The oldest example of an ω-consistent and unsound theory seems to go back to Rosser [1937, Theorem 1], who shows that if $\neg\omega\text{-}\mathtt{Con}_T$ is a sentence saying that the theory T is not ω-consistent, and if T is ω-consistent, then $T \cup \{\neg\omega\text{-}\mathtt{Con}_T\}$ is ω-consistent too. Of course, $T \cup \{\neg\omega\text{-}\mathtt{Con}_T\}$ is not sound if T is ω-consistent; see also [Boolos, 1993, Chap. 15].

Now suppose that U is an ω-consistent but unsound arithmetical theory which is sufficiently strong. Let A and B be two sentences each of which is

U-provably equivalent to its own U-unprovability. Then of course we have $U \vdash A \leftrightarrow B$. However, *the equivalence of A and B might be one of those theorems of our unsound U which are false*, in which case one of A and B is true and the other false, so that "in reality" they are not equivalent. Even setting aside the realist talk, when one of the Gödelian sentences of U is true in $\mathbb{N}$ while the other false, it seems bizarre to talk about *the* Gödel sentence of the theory U.

To see some of the many places where the expression 'the Gödel sentence' (with definite article) obtrudes, we ask our readers to look at the titles of otherwise excellent works of Boolos [1990]; Shapiro [1998]; Serény [2011]; Isaacson [2011], and also [Raatikainen, 2005, p. 520]. As far as we know, only Milne [2007] talks about Gödel sentences (in the plural). It was noted by Lajevardi and Salehi [2019, pp. 12–13] that some unsound theories may have both true Gödel sentences and false ones. While the argument there was based on a particular unsound theory, here we demonstrate that this applies to *every* unsound theory.

In this note we elaborate on these points. Readers who are more philosophically inclined may skip Subsection 1.1 and the entire Section 2, where we fix our notation and prove a number of technical results which we summarize here without using technical jargon. Theorem 1 shows that a sentence G is unprovable in a theory T if and only if there exists a consistent extension T' of T such that G is equivalent to the T'-unprovability of G inside T'. (Half of this theorem follows from Gödel's proof: all the sentences that are provably equivalent to their unprovability in a consistent sufficiently strong arithmetical theory are unprovable in that theory.) We then show in Theorem 2 that for a sufficiently strong arithmetical theory T, the following are equivalent: (i) the truth of *every* sentence which is T-equivalent to its own T-unprovability (recall that the sentence constructed by Gödel is *one of* these), and (ii) the soundness of the theory. Therefore, given an unsound theory, there is a false sentence which is provably equivalent to its own unprovability. To complete our argument, we go on to show (Corollary 3) that every unsound theory has a true such sentence as well.

At the end of our technical discussion we provide an alternative proof for Theorem 1 and derive a corollary which may be seen as a cute refinement of Theorem 2: not only for every unsound theory is there a false I-am-unprovable sentence, but every false sentence whatsoever is an I-am-unprovable one with respect to some consistent and sufficiently strong arithmetical theory.

A word of caution might be in order before getting into the details. Beware of what we are *not* saying. We do not claim that an appeal to provable equivalence under an arithmetical theory — even if done properly and even if the theory is sound — will warrant the *the* talk. For all we know, the process of "dividing out" (*i.e.*, basically thinking of *the Gödel sentence* of a theory T as the set of all sentences that are T-provably equivalent to their T-unprovability) may have its own problems or shortcomings: for instance, while it does make sense, and it may be of some interest, to say that the number of symbols in the original sentence constructed by Gödel is less than the one constructed by Rosser, many such differences will be gone (or at least marginalized) after dividing out. All we say is this: (1) it seems that one alleged way of justifying the use of the definite article in the phrase 'the Gödel sentence' is to use the mathematical fact that all the I-am-unprovable sentences are provably equivalent, and (2) this provable equivalence does not help if the theory is unsound. We do *not* claim that we have solved the question of how to define *the* Gödel sentence (neither are we sure that it is possible at all).

1.1. Terminology and Background.

Let us fix our language of arithmetic as in [Boolos, 1993, Chapter 2], where PA (Peano's Arithmetic) and a Gödel coding $A \mapsto \ulcorner A \urcorner$ are defined. We will use the terminology of [Boolos, 1993] with a minor update: we write '$\mathtt{Pr}$' instead of his '$\mathtt{Bew}$' for provability. Boolos's whole discussion is about PA, but can be easily generalized to any recursively enumerable (RE) super-theory of it. It is well known that a set is RE if and only if it is definable by a Σ_1-formula [Feferman, 1960]. For a formula $\sigma(x)$, let $\mathtt{Th}_\sigma = \{A \mid \mathbb{N} \vDash \sigma(\ulcorner A \urcorner)\}$ be the *theory* defined by σ, where A ranges over the sentences of the language.[2] *Throughout this short note, we work only with Σ_1-definable theories that extend* PA.[3]

For a given Σ_1-formula $\sigma(x)$, there exists a Σ_1-formula $\mathtt{Pr}_\sigma(x)$ that defines the provable sentences of the theory $\mathtt{Th}_\sigma$; see [Boolos, 1993, p. 44] or [Feferman, 1960, Theorem 4.5]. One particular such provability predicate can be constructed in a way that it satisfies the following:

Fact 3 (The Derivability Conditions, and more). *For every Σ_1-formula $\sigma(x)$ with $\mathtt{Th}_\sigma \supseteq \mathsf{PA}$, the following hold for every sentences A and B:*

(D_0) *If* $\mathbb{N} \vDash \mathtt{Pr}_\sigma(\ulcorner A \urcorner)$, *then* $\mathtt{Th}_\sigma \vdash A$;

(D_1) *If* $\mathtt{Th}_\sigma \vdash B$, *then* $\mathsf{PA} \vdash \mathtt{Pr}_\sigma(\ulcorner B \urcorner)$;

(D_2) $\mathsf{PA} \vdash \mathtt{Pr}_\sigma(\ulcorner A \rightarrow B \urcorner) \rightarrow [\mathtt{Pr}_\sigma(\ulcorner A \urcorner) \rightarrow \mathtt{Pr}_\sigma(\ulcorner B \urcorner)]$;

(D_3) $\mathsf{PA} \vdash \mathtt{Pr}_\sigma(\ulcorner A \urcorner) \rightarrow \mathtt{Pr}_\sigma(\ulcorner \mathtt{Pr}_\sigma(\ulcorner A \urcorner) \urcorner)$.

Let $\sigma \boldsymbol{\cdot} A(x)$ abbreviate the formula $\sigma(x) \vee (x = \ulcorner A \urcorner)$. Note that the formula $\sigma \boldsymbol{\cdot} A$ defines the theory $\mathtt{Th}_\sigma \cup \{A\}$. We have:

(i) $\mathsf{PA} \vdash \mathtt{Pr}_{\sigma \cdot A}(\ulcorner B \urcorner) \leftrightarrow \mathtt{Pr}_\sigma(\ulcorner A \rightarrow B \urcorner)$.

Put $\mathtt{Con}_\sigma = \neg \mathtt{Pr}_\sigma(\ulcorner \bot \urcorner)$, which is normally understood as saying that the theory $\mathtt{Th}_\sigma$ is consistent. Now, we have:

(ii) $\mathsf{PA} \vdash \neg \mathtt{Con}_\sigma \rightarrow \mathtt{Pr}_\sigma(\ulcorner A \urcorner)$.

For a proof of Fact 3(ii) it suffices to note that $\bot \rightarrow A$ is $\mathtt{Th}_\sigma$-provable, and so by Fact 3(D_1) the sentence $\mathtt{Pr}_\sigma(\ulcorner \bot \rightarrow A \urcorner)$ is PA-provable; so is $\mathtt{Pr}_\sigma(\ulcorner \bot \urcorner) \rightarrow \mathtt{Pr}_\sigma(\ulcorner A \urcorner)$ by Fact 3(D_2). Fact 3(i) appears in [Feferman, 1960,

---

[2] Note that our use of the term 'theory' might be considered non-standard as $\mathtt{Th}_\sigma$ need not be deductively closed.

[3] Indeed, one can work with consistent Σ_1-definable extensions of a sufficiently strong sub-theory of PA such as $\mathrm{I}\Sigma_1$ or $\mathsf{EA} + \mathrm{B}\Sigma_1 (\equiv \mathrm{I}\Delta_1)$. Theories without $\mathrm{B}\Sigma_1$ would not be suitable since, as a referee of this journal remarked, the usual constructions of $\mathtt{Pr}_\sigma$ (for a given Σ_1-formula σ) need not be equivalently Σ_1 if Σ_1-collection is not provable in the base theory; see Visser [2020]. For the readers' convenience, we confine ourselves to Σ_1-definable extensions of PA, as their properties are more accessible in the literature. Let us note that Σ_1 is denoted simply by 'Σ' in [Boolos, 1993].

Theorem 4.8]. Let us say, by way of definition, that A is a *Gödelian sentence* of σ when $\mathtt{Th}_\sigma \vdash A \leftrightarrow \neg\mathtt{Pr}_\sigma(\ulcorner A \urcorner)$. Famously, [Gödel, 1931, Section 2] constructs, by means of what is now called the Diagonal Lemma, a Gödelian sentence G. Then Gödel shows $\mathtt{Th}_\sigma \nvdash G$ if $\mathtt{Th}_\sigma$ is consistent, and also $\mathtt{Th}_\sigma \nvdash \neg G$ if $\mathtt{Th}_\sigma$ is (also) ω-consistent. Note that we call G a Gödelian sentence *of the formula* $\sigma(x)$ and not *of the theory* $\mathtt{Th}_\sigma$ since a theory may have different presentations (defining formulas) that are not equivalent even inside the theory itself; see [Feferman, 1960, Theorem 7.5].

2. MANY GÖDELIAN SENTENCES: SOME TRUE, SOME FALSE.

Let us start with a theorem which is interesting in its own right, as it characterizes the Gödelian sentences of super-theories.

Theorem 1 (Unprovable sentences are Gödelian). *Let $\sigma(x)$ be a Σ_1-formula such that $\mathtt{Th}_\sigma \supseteq \mathsf{PA}$. Any sentence G is unprovable in $\mathtt{Th}_\sigma$ if and only if G is a Gödelian sentence of a consistent super-theory of $\mathtt{Th}_\sigma$.*

Proof. First, suppose that for a Σ_1-formula $\tau(x)$ the theory $\mathtt{Th}_\tau$ is consistent and $\mathtt{Th}_\tau \supseteq \mathtt{Th}_\sigma$, and let G be a Gödelian sentence of $\mathtt{Th}_\tau$. Then, $(*)$ $\mathtt{Th}_\tau \vdash G \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner G \urcorner)$. We show that $\mathtt{Th}_\tau \nvdash G$. Assume that $\mathtt{Th}_\tau \vdash G$. Then $\mathtt{Th}_\tau \vdash \mathtt{Pr}_\tau(\ulcorner G \urcorner)$ by Fact 3 (D_1), and so $\mathtt{Th}_\tau \vdash \neg G$ by $(*)$, contradicting the consistency of $\mathtt{Th}_\tau$. As a result, $\mathtt{Th}_\tau \nvdash G$; hence $\mathtt{Th}_\sigma \nvdash G$, *i.e.*, G is unprovable in $\mathtt{Th}_\sigma$.

Secondly, suppose that G is unprovable in $\mathtt{Th}_\sigma$. By the Diagonal Lemma there exists some sentence A such that[4]

$$\mathsf{PA} \vdash A \leftrightarrow [G \leftrightarrow \neg\mathtt{Pr}_{\sigma \cdot A}(\ulcorner G \urcorner)].$$

Let $\tau = \sigma \cdot A$. Then $\mathtt{Th}_\tau = \mathtt{Th}_\sigma \cup \{A\}$, and so we have $\mathtt{Th}_\tau \vdash G \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner G \urcorner)$; thus G is a Gödelian sentence of τ. It remains to show that $\mathtt{Th}_\tau$ is consistent. If not, then $\mathtt{Th}_\sigma \vdash \neg A$, and so $(**)$ $\mathtt{Th}_\sigma \vdash \neg[G \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner G \urcorner)]$. On the other hand, $\mathtt{Th}_\sigma \vdash A \rightarrow G$ follows from $\mathtt{Th}_\sigma \vdash \neg A$, and so we have $\mathsf{PA} \vdash \mathtt{Pr}_\sigma(\ulcorner A \rightarrow G \urcorner)$ by Fact 3(D_1), therefore $\mathsf{PA} \vdash \mathtt{Pr}_\tau(\ulcorner G \urcorner)$ by Fact 3(i). Thus, $\mathtt{Th}_\sigma \vdash G$ follows from $(**)$, a contradiction. □

Theorem 2 (The Gödelian sentences of only sound theories are all true) *Let $\sigma(x)$ be a Σ_1-formula such that $\mathtt{Th}_\sigma \supseteq \mathsf{PA}$. Then all of the Gödelian sentences of σ are collectively true if and only if $\mathtt{Th}_\sigma$ is a sound theory.*[5]

Proof. If $\mathtt{Th}_\sigma$ is sound and G is a Gödelian sentence of σ, then $\mathbb{N} \vDash G \leftrightarrow \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)$. On the other hand, from Fact 3(D_0) and $\mathtt{Th}_\sigma \nvdash G$ (Theorem 1) we have $\mathbb{N} \vDash \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)$. Thus, $\mathbb{N} \vDash G$.

Now, suppose that all the Gödelian sentences of σ are true. The theory $\mathtt{Th}_\sigma$ is consistent, since otherwise every false sentence would be a Gödelian sentence

---

[4] *Cf.* [McGee, 1992, p. 238].

[5] *Cf.* [Smith, 2013, Theorem 24.7].

of σ. We show that $\mathtt{Th}_\sigma$ is sound. Suppose that $\mathtt{Th}_\sigma \vdash A$ for a sentence A; we aim at showing the truth of A, *i.e.*, that $\mathbb{N} \vDash A$. By the Diagonal Lemma there exists a sentence G such that

$$\mathsf{PA} \vdash G \leftrightarrow [A \leftrightarrow \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)].$$

Hence $\mathtt{Th}_\sigma \vdash G \leftrightarrow \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)$, and so G is a Gödelian sentence of σ. Thus $\mathbb{N} \vDash G$ by the assumption. Assuming that PA is sound, we have $\mathbb{N} \vDash A \leftrightarrow \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)$. On the other hand, by Theorem 1, $\mathtt{Th}_\sigma \nvdash G$ and so $\mathbb{N} \vDash \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)$ by Fact 3(D_0). Therefore, $\mathbb{N} \vDash A$. □

So, an unsound theory does have some false Gödelian sentences. Below we show that it has some true Gödelian sentences as well.

Corollary 3 (Gödelian sentences of unsound theories: Some true, some false) *Every Σ_1-formula $\sigma(x)$ whose $\mathtt{Th}_\sigma$ is an unsound extension of* PA *has at least one true and one false Gödelian sentence.*

Proof. If $\mathtt{Th}_\sigma$ is inconsistent, then every sentence is a Gödelian sentence of σ; so, suppose that the unsound theory $\mathtt{Th}_\sigma$ is consistent. Then σ has a false Gödelian sentence by Theorem 2. By the Diagonal Lemma we have $\mathsf{PA} \vdash G \leftrightarrow \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)$ for some sentence G. Now G is a Gödelian sentence of σ, and so by Theorem 2 and Fact 3(D_0) we have $\mathbb{N} \vDash \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)$. Thus $\mathbb{N} \vDash G$ (by the assumed soundness of PA), and so σ has a true Gödelian sentence G. □

Before concluding this section, let us see another proof for Theorem 1 and an interesting consequence of it:

Theorem 1. *Let $\sigma(x)$ be a Σ_1-formula such that $\mathtt{Th}_\sigma \supseteq \mathsf{PA}$. Any sentence G is unprovable in $\mathtt{Th}_\sigma$ if and only if G is a Gödelian sentence of a consistent super-theory of $\mathtt{Th}_\sigma$.*

(Alternative) Proof. The "if" part is as before (all Gödelian sentences of consistent theories are unprovable). For the "only if" part, suppose $\mathtt{Th}_\sigma \nvdash G$. Let $\tau = \sigma \cdot \neg G$. Then the theory $\mathtt{Th}_\tau = \mathtt{Th}_\sigma \cup \{\neg G\}$ is consistent, and so is the theory $\mathtt{Th}_\tau \cup \{\neg\mathtt{Con}_\tau\}$ by Gödel's Second Incompleteness Theorem. Put $\nu = \tau \cdot \neg\mathtt{Con}_\tau$. Then $\mathtt{Th}_\nu = \mathtt{Th}_\sigma \cup \{\neg G, \neg\mathtt{Con}_\tau\}$ is a consistent super-theory of $\mathtt{Th}_\sigma$, and it remains to show that G is a Gödelian sentence of ν. By $\mathtt{Th}_\nu \vdash \neg\mathtt{Con}_\tau$ and Fact 3(ii) we have $\mathtt{Th}_\nu \vdash \mathtt{Pr}_\tau(\ulcorner \neg\mathtt{Con}_\tau \rightarrow G \urcorner)$, and so Fact 3($i$) implies $\mathtt{Th}_\nu \vdash \mathtt{Pr}_\nu(\ulcorner G \urcorner)$. Now we have $\mathtt{Th}_\nu \vdash \neg G$ and $\mathtt{Th}_\nu \vdash \mathtt{Pr}_\nu(\ulcorner G \urcorner)$; therefore, $\mathtt{Th}_\nu \vdash G \leftrightarrow \neg\mathtt{Pr}_\nu(\ulcorner G \urcorner)$. This proves that G is a Gödelian sentence of ν. □

Corollary 4. (All the false sentences are Gödelian) *Every false sentence is a Gödelian sentence of a consistent theory.*

Proof. If a sentence is false, then (by the assumed soundness of PA) it is PA-unprovable, and so by Theorem 1 it is a Gödelian sentence of a consistent super-theory of PA. □

Indeed, every false sentence is a Gödelian sentence of a consistent super-theory of an arbitrary sound extension of PA.

Remark 5. (Gödelian sentences relative to a sound sub-theory). So far, we have dealt with the issue of *the* Gödel sentence in a mono-theoretic way, in the sense of having one and the same theory $\mathtt{Th}_\sigma$ functioning as both the theory with respect to which we define the provability predicate ($\mathtt{Pr}_\sigma$) and the theory within which the equivalence of a Gödelian sentence to its unprovability in $\mathtt{Th}_\sigma$ is proved. One may think of going *di-theoretically* (or "bitheoretically" as [Detlefsen, 2001] puts it), namely choosing a different theory S (preferably a sub-theory of $\mathtt{Th}_\sigma$) as *meta-theory* and defining a sentence G to be a *σ-Gödelian sentence relative to* the theory S when we have $S \vdash G \leftrightarrow \neg\mathtt{Pr}_\sigma(\ulcorner G \urcorner)$. Now, if S is taken to be a *sound* sub-theory of $\mathtt{Th}_\sigma$, then our objection to the use of definite article seems to be defused. One may think of taking S to be Peano's Arithmetic PA itself, which is reasonably supposed to be sound; so, talking of *the* Gödel sentences of such theories relative to PA does not seem to be unreasonable here.[6] Let us note that this is the method employed by many textbooks for defining (and proving the existence of) Gödelian sentences; see [Lajevardi and Salehi, 2019, Appendix].

3. CONCLUDING REMARKS, AND THE (IN)-SIGNIFICANCE OF OUR OBSERVATION

Corollary 3 is the reason for our discomfort at the use of definite article in the phrase 'the Gödel sentence'. Let us anticipate a rejoinder, to the effect that a logician who talks about *the* Gödel sentence of a theory may have some particular syntactic object in mind — *e.g.*, she may be thinking of the exact same sentence constructed by Gödel [1931, p. 173], called 'v Gen r'.

Logically speaking, we think that this is not a very exciting way to go. For one thing, there is an arbitrariness in such a choice, as it depends on a number of inessential things like the particular coding (so why not use Quine's more elegant coding instead?). Insofar as the First Incompleteness Theorem is concerned, it seems that what is crucial for the particular sentence constructed by Gödel is its possession of certain properties *à propos* of provability, not that it is constructed in this or that particular way.[7] To concentrate on a *particular* syntactic object is to overlook the main idea.

More importantly, the suggestion of the first paragraph of this section may result in the unwelcome situation that while arguing for, *e.g.*, the truth of *the Gödel sentence of* PA, different logicians or philosophers are, strictly speaking, talking past each other: one is dealing with the particular sentence constructed

[6] We acknowledge that some of the authors we referred to in Section 1 might have had some such ideas in mind — see, in particular, [Smoryński, 1977, §2] or [Isaacson, 2011, §2,§3].

[7] Even its being self-referential is inessential for Gödel's theorem — see below.

by Gödel [1931], another is considering a particular sentence constructed by Feferman [1960], and so on.

Let us also note that there is of course another reason that makes the use of the definite article problematic here: there are different ways of arithmetizing the syntax, and, *a fortiori*, different Gödelian sentences — and this seems to be what Serény [2011, p. 68] is worried about. Of course our concern is different: we have argued that *even with a fixed arithmetization*, there might be true Gödelian sentences and false ones.[8]

Let us recapitulate. We say that a sentence A is a *Gödelian sentence* of a theory T iff A is T-provably equivalent to its own T-unprovability (with respect to a fixed arithmetization). Insofar as Gödel's theorem is concerned, our theories may be taken to be unsound. If our theory is not just unsound but even inconsistent, then every sentence is a Gödelian sentence of it, hence the theory has, in an insipid way, true Gödelian sentences as well as false ones. If the theory is unsound but (ω-)consistent, then Corollary 3 provides a false Gödelian sentence and a true one. Of course all these sentences are equivalent in the eye of the theory, but in this situation it is rather odd to talk about *the* Gödel sentence of the theory.

Does this in any way affect the validity of Gödel's incompleteness theorems as presented in textbooks or papers which talk about *the* Gödel sentence? Not really. In fact, *every* Gödelian sentence of a consistent RE extension of Peano's Arithmetic is unprovable in that theory. Yet perhaps one should be more careful and avoid talking as if there is only one such sentence (even up to truth value). For the sheer purpose of proving the First Incompleteness Theorem, it does not matter if we get this right. Our point in this note is a modest (and perhaps pedantic) one: the correction of an impropriety of speech.[9]

Compare to a related issue, that of self-reference. In many of those same textbooks and papers, authors speak of Gödelian sentences each of which "says about itself" that it is unprovable. Now it is by no means an easy task to give a satisfactory analysis of the notion of a sentence saying something about itself; for the purpose of proving the First Incompleteness Theorem, however, we need not trouble ourselves with that notion — see [Smoryński, 1991, p. 122] and [Kripke, 2014, p. 239].

Having said that, let us conclude by drawing attention to the fact that our observation here may well affect some philosophical discussions concerning the

[8] For a recent investigation of a similar worry concerning Gödel's second theorem see [Grabmayr, forthcoming].

[9] We wholeheartedly agree with a referee of this journal who wrote, 'Provable equivalence in a theory is not a criterion of synonymy (else all theorems of the theory mean the same!) so why people make an exception in the case of Gödel sentences is unclear'. As another referee commented, although identifying provably equivalent sentences turns out to be mathematically fruitful with respect to some background theories (as evidenced by the notion of Lindenbaum–Tarski Algebra, or Magari Algebra in the case of provability logic), provable equivalence does not guarantee truth preservation of Gödelian sentences, as illustrated in our note; thus we seem to have, philosophically speaking, further qualms about taking provable equivalence as a criterion for synonymy.

truth of Gödelian sentences, as discussed in many logico-philosophical texts. But that is the topic of another, much longer, paper than this short note.

REFERENCES

Boolos, George [1990]: 'On "seeing" the truth of the Gödel sentence', *Behavioral and Brain Sciences* **13**, 655–656. doi.org/10.1017/S0140525X00080687. Reprinted in: G. Boolos, *Logic, Logic, and Logic*. R. Jeffrey, ed., pp. 389–391. Harvard University Press, 1999.

______ [1993]: *The Logic of Provability*. Cambridge University Press.

Detlefsen, Michael [2001]: 'What does Gödel's second theorem say?', *Philosophia Mathematica* (3) **9**, 37–71. doi.org/10.1093/philmat/9.1.37.

Feferman, Solomon [1960]: 'Arithmetization of metamathematics in a general setting', *Fundamenta Mathematicæ* **49**, 35–92. doi.org/10.2307/2269834.

Gödel, Kurt [1931]: 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* **38**, 173–198. English translation 'On formally undecidable propositions of *Principia Mathematica* and related systems I', in S. Feferman *et al.*, eds, *Kurt Gödel: Collected Works*, Vol. I, pp. 135–152. Oxford University Press, 1986.

Grabmayr, Balthasar [forthcoming]: 'On the invariance of Gödel's second theorem with regard to numberings', *The Review of Symbolic Logic*. doi.org/10.1017/S1755020320000192.

Isaacson, Daniel [2011]: 'Necessary and sufficient conditions for undecidability of the Gödel sentence and its truth', in D. DeVidi, M. Hallett, P. Clarke, eds, *Logic, Mathematics, Philosophy: Vintage Enthusiasms: Essays in Honour of John L. Bell*, pp. 135–152. Springer. doi.org/10.1007/978-94-007-0214-1_7.

Kripke, Saul A. [2014]: 'The road to Gödel', in J. Berg, ed., *Naming, Necessity, and More: Explorations in the Philosophical Work of Saul Kripke*, pp. 223–241. Basingstoke: Palgrave Macmillan. doi.org/10.1057/9781137400932_11.

Lajevardi, Kaave, and Saeed Salehi [2019]: 'On the arithmetical truth of self-referential sentences', *Theoria: A Swedish Journal of Philosophy* **85**, 8–17. doi.org/10.1111/theo.12169.

Lindström, Per [1997]: *Aspects of Incompleteness*. Springer. Repr. AK Peters/CRC Press, 2003; Cambridge University Press, 2017.

McGee, Vann [1992]: 'Maximal consistent sets of instances of Tarski's schema (T)', *Journal of Philosophical Logic* **21**, 235–241. doi.org/10.1007/BF00260929.

Milne, Peter [2007]: 'On Gödel sentences and what they say', *Philosophia Mathematica* (3) **15**, 193–226. doi.org/10.1093/philmat/nkm015.

Raatikainen, Panu [2005]: 'On the philosophical relevance of Gödel's incompleteness theorems', *Revue Internationale de Philosophie* **59**, 513–534.

Rosser, Barkley [1937]: 'Gödel theorems for non-constructive logics', *Journal of Symbolic Logic* **2**, 129–137. doi.org/10.2307/2266293.

Serény, György [2011]: 'How do we know that the Gödel sentence of a consistent theory is true?', *Philosophia Mathematica* (3) **19**, 47–73. doi.org/10.1093/philmat/nkq028.

Shapiro, Stewart [1998]: 'Induction and indefinite extensibility: The Gödel sentence is true, but did someone change the subject?', *Mind* **107**, 597–624. 10.1093/mind/107.427.597.

Smith, Peter [2013]: *An Introduction to Gödel's Theorems*. 2nd ed. Cambridge University Press.

Smoryński, Craig [1977]: 'The incompleteness theorems', in J. Barwise, ed., *Handbook of Mathematical Logic*, pp. 821–865. Amsterdam: North-Holland. doi.org/10.1016/S0049-237X(08)71123-6.

——— [1991]: 'The development of self-reference: Löb's theorem', in T. Drucker, ed., *Perspectives on the History of Mathematical Logic*, pp. 110–133. Basel: Birkhäuser. Reprinted 2008. doi.org/10.1007/978-0-8176-4769-8_9.

Visser, Albert [2020]: 'Another look at the second incompleteness theorem', *Review of Symbolic Logic* **13**, 269–295. doi.org/10.1017/S1755020319000248.