Tarski's Undefinability Theorem and the Diagonal Lemma

SAEED SALEHI*, *Research Institute for Fundamental Sciences, University of Tabriz, 51666-16471, Tabriz, Iran.*

Abstract

We prove the equivalence of the semantic version of Tarski's theorem on the undefinability of truth with the semantic version of the diagonal lemma and also show the equivalence of a syntactic version of Tarski's undefinability theorem with a weak syntactic diagonal lemma. We outline two seemingly diagonal-free proofs for these theorems from the literature and show that the syntactic version of Tarski's theorem can deliver Gödel–Rosser's incompleteness theorem.

Keywords: The diagonal lemma, diagonal-free proofs, Gödel's incomplteness theorem, Rosser's theorem, self-reference, Tarski's undefinability theorem.

1 Introduction

One of the cornerstones of modern logic (and theory of incompleteness after Gödel) is the diagonal lemma (aka self-reference or fixed-point lemma) due to Gödel and Carnap (see [14], and the references therein). The lemma states that (when $\alpha \mapsto \ulcorner\alpha\urcorner$ is a suitable Gödel coding that assigns the closed term $\ulcorner\alpha\urcorner$ to a syntactic expression or object α) for a given formula $\Psi(x)$ with the only free variable x, there exists some sentence θ such that the equivalence $\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$ holds; 'holding' could mean either being true in the standard model of natural numbers $\mathbb{N}$ or being provable in a suitable theory T (which is usually taken to be a consistent extension of Robinson's arithmetic). When the equivalence $\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$ holds in $\mathbb{N}$ we call it *the semantic diagonal lemma* (studied in Section 2); when T proves the equivalence, we call it *the syntactic diagonal lemma*. A version of the diagonal lemma, which is called the *weak diagonal lemma* here, states the consistency of the sentence $\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$ with T, for some sentence θ that depends on the given arbitrary formula $\Psi(x)$ and the theory T (studied in Section 3).

The diagonal lemma has been used in proving many fundamental theorems of mathematical logic, such as Gödel's first and (also) second incompleteness theorems, Rosser's (strengthening of Gödel's incompleteness) theorem and Tarski's theorem (on the undefinability of truth). One problem with the diagonal lemma is its standard proof, which is a kind of magic (or 'pulling a rabbit out of the hat'; see, e.g. [20]); indeed it is not easy to remember its typical proof, even after several years of teaching it. Here, we quote some texts on the proof of this lemma from the literature:

(1998) S. Buss writes in [1] that the proof of the diagonal lemma is 'quite simple but rather tricky and difficult to conceptualize.'

(2002) V. McGee states in [10] that by the diagonal (aka self-referential) lemma there exists a sentence ϕ for a given formula $\Psi(x)$ such that $\Psi(\ulcorner\phi\urcorner) \leftrightarrow \phi$ is provable in Robinson's arithmetic. 'You would hope that such a deep theorem would have an insightful proof. No such luck. I am going to write down a sentence ϕ and verify that it works. What I won't

*E-mail: saesal@gmail.com

Vol. 00, No. 0,
doi:10.1093/jigpal/jzab016

do is give you a satisfactory explanation for why I write down the particular formula I do. I write down the formula because Gödel wrote down the formula, and Gödel wrote down the formula because, when he played the logic game he was able to see seven or eight moves ahead, whereas you and I are only able to see one or two moves ahead. I don't know anyone who thinks he has a fully satisfying understanding of why the Self-referential Lemma works. It has a rabbit-out-of-a-hat quality for everyone.'

(2004) H. Kotlarski [9] said that the diagonal lemma 'being very intuitive in the natural language, is highly unintuitive in formal theories like Peano arithmetic. In fact, the usual proof of the diagonal lemma ... is short, but tricky and difficult to conceptualize. The problem was to eliminate this lemma from proofs of Gödel's result.'

(2006) G. Serény [16] attempts to make 'the proof of the lemma completely transparent by showing that it is simply a straightforward translation of the Grelling paradox into first-order arithmetic.'

(2006) H. Gaifman mentions in [4] that the proof of the diagonal lemma is 'extremely short'. However, the 'brevity of the proof does not make for transparency; it has the aura of a magician's trick.'

In this paper, we attempt at giving some explanations and motivations for this basic lemma, in a way that we will have a satisfactory understanding for at least some weaker versions of it. For that purpose, we will first see the equivalence of the semantic form of the diagonal lemma with Tarski's theorem on the undefinability of arithmetical truth (in Section 2). In other words, the diagonal lemma holds in a model just in case the set of (the Gödel codes of) the true sentences of that model is not definable in that model (see Theorem 2.5 below). As a matter of fact, different proofs for Tarski's undefinability theorem can lead to different proofs for the semantic version of this lemma. We will review two such proofs (presented in [2, 7–9, 15]) that are supposedly diagonal-free. Having different proofs will, hopefully, shed some new light on the nature of this lemma and will increase our understanding about it. Then, secondly, we will see that a syntactic version of Tarski's theorem is equivalent to a weak (syntactic) version of the diagonal lemma (in Section 3). This weak form of the diagonal lemma is still sufficiently strong to prove Gödel–Rosser's incompleteness theorem. So, different proofs of the syntactic version of Tarski's theorem will provide some seemingly diagonal-free proofs for Rosser's theorem (cf. [19], in which Gödel's second incompleteness theorem is derived from Tarski's undefinability theorem by some circular-free arguments).

2 The Diagonal Lemma, Semantically

Let us fix the language of arithmetic as $\mathcal{L}_{ar} = \{0, 1, +, \times\}$ and let $\ulcorner \cdot \urcorner$ be a fixed Gödel coding that maps $\mathcal{L}_{ar}$-sentences to closed $\mathcal{L}_{ar}$-terms in a computable and injective way. Let $\#\eta$ denote the value (the standard interpretation) of the closed term $\ulcorner \eta \urcorner$.

DEFINITION 2.1 (Semantic Diagonal Lemma).
The following statement is called the *semantic diagonal lemma*:

For every $\mathcal{L}_{ar}$-formula $\Psi(x)$ there exists an $\mathcal{L}_{ar}$-sentence θ such that $\mathbb{N} \vDash \Psi(\ulcorner \theta \urcorner) \leftrightarrow \theta$.

This (weaker) form of the diagonal lemma serves to prove the semantic version of Gödel's incompleteness theorem (see [17, Theorem 6.3]):

THEOREM 2.2 (Gödel's Incompleteness Theorem for Sound and Definable Theories).
For every definable and sound theory T in the language of $\mathcal{L}_{ar}$ there exists a true $\mathcal{L}_{ar}$-sentence, which is independent from T.

PROOF. If T is definable then there exists an $\mathcal{L}_{ar}$-formula $\mathtt{Pr}_T(x)$ such that for every $\mathcal{L}_{ar}$-sentence η we have $T \vdash \eta$ if and only if $\mathbb{N} \vDash \mathtt{Pr}_T(\ulcorner \eta \urcorner)$. Now, by the semantic diagonal lemma (Definition 2.4) we have $\mathbb{N} \vDash \gamma \leftrightarrow \neg \mathtt{Pr}_T(\ulcorner \gamma \urcorner)$ for some $\mathcal{L}_{ar}$-sentence γ. It can be seen that $T \nvdash \gamma$, since $T \vdash \gamma$ implies on the one hand that $\mathbb{N} \vDash \mathtt{Pr}_T(\ulcorner \gamma \urcorner)$ and on the other hand (by the soundness of T) that $\mathbb{N} \vDash \gamma$ and so $\mathbb{N} \vDash \neg \mathtt{Pr}_T(\ulcorner \gamma \urcorner)$, a contradiction. So, $T \nvdash \gamma$, therefore $\mathbb{N} \vDash \neg \mathtt{Pr}_T(\ulcorner \gamma \urcorner)$, hence $\mathbb{N} \vDash \gamma$, which also implies (by the soundness of T) that $T \nvdash \neg \gamma$. □

Also, Tarski's theorem on the undefinability of (arithmetical) truth follows from the semantic diagonal lemma (see [6, Exercise 3.7] and cf. [6, Chapter 9]):

THEOREM 2.3 (Tarski's Theorem on the Undefinability of Arithmetical Truth).
The Gödel codes of the set of true sentences, i.e. $\{\#\eta \in \mathbb{N} \mid \mathbb{N} \vDash \eta\}$, is not definable in $\mathbb{N}$.

PROOF. If $\{\#\eta \mid \mathbb{N} \vDash \eta\}$ is definable by some $\mathcal{L}_{ar}$-formula $\Upsilon(x)$, then (i) $\mathbb{N} \vDash \Upsilon(\ulcorner \eta \urcorner) \leftrightarrow \eta$ holds for every $\mathcal{L}_{ar}$-sentence η. Now, by the semantic diagonal lemma we have (ii) $\mathbb{N} \vDash \neg \Upsilon(\ulcorner \lambda \urcorner) \leftrightarrow \lambda$ for some $\mathcal{L}_{ar}$-sentence λ. So, from (i) and (ii) we have $\mathbb{N} \vDash \lambda \leftrightarrow \neg\lambda$, a contradiction. □

In fact, as we show below, the semantic diagonal lemma is equivalent to Tarski's theorem on the undefinability of truth and to the semantic incompleteness theorem of Gödel (cf. [13, Remark 2.6]). To set the stage for a genuine equivalence we use the following definition.[1]

DEFINITION 2.4 (Gödel Coding and Definability).
For a first-order language $\mathcal{L}$, a Gödel coding is a mapping $\ulcorner \cdot \urcorner : \mathtt{Sent}_{\mathcal{L}} \to \mathtt{ClTerms}_{\mathcal{L}}$ from $\mathcal{L}$-sentences to closed $\mathcal{L}$-terms, which is computable and injective (see [5, §2]). An $\mathcal{L}$-theory T (i.e. a deductively closed set of $\mathcal{L}$-sentences) is definable in an $\mathcal{L}$-structure $\mathcal{M}$ when for an $\mathcal{L}$-formula $\Phi(x)$ we have $T = \{\eta \in \mathsf{Sent}_{\mathcal{L}} \mid \mathcal{M} \vDash \Phi(\ulcorner \eta \urcorner)\}$.

THEOREM 2.5 (Semantic Diagonal Lemma ≡ Semantic Gödel's Theorem ≡ Tarski's Theorem).
For every first-order language $\mathcal{L}$ with a Gödel coding $\ulcorner \cdot \urcorner$ and every $\mathcal{L}$-structure $\mathcal{M}$ the following are equivalent:

- $\mathbb{D}_{\mathcal{M}}$ (Diagonal Lemma for $\mathcal{M}$): For every $\mathcal{L}$-formula $\Psi(x)$ there exists an $\mathcal{L}$-sentence θ such that $\mathcal{M} \vDash \Psi(\ulcorner \theta \urcorner) \leftrightarrow \theta$.
- $\mathbb{G}_{\mathcal{M}}$ (Gödel's Theorem for $\mathcal{M}$): If an $\mathcal{L}$-theory $T \subseteq \mathrm{Th}(\mathcal{M})$ is definable in $\mathcal{M}$, then T is incomplete.
- $\mathbb{T}_{\mathcal{M}}$ (Tarski's Theorem for $\mathcal{M}$): The complete theory $\mathrm{Th}(\mathcal{M})$ of $\mathcal{M}$ is not definable in $\mathcal{M}$.

PROOF. ($\mathbb{D}_{\mathcal{M}} \Rightarrow \mathbb{G}_{\mathcal{M}}$): This can be proved exactly in the same lines of the proof of Theorem 2.2 (and the implication $\mathbb{D}_{\mathcal{M}} \Rightarrow \mathbb{T}_{\mathcal{M}}$ is essentially proved in Theorem 2.3).
($\mathbb{G}_{\mathcal{M}} \Rightarrow \mathbb{T}_{\mathcal{M}}$): If $\mathrm{Th}(\mathcal{M})$ were definable in $\mathcal{M}$, then by $\mathbb{G}_{\mathcal{M}}$ it should have been incomplete; a contradiction.
($\mathbb{T}_{\mathcal{M}} \Rightarrow \mathbb{D}_{\mathcal{M}}$): Suppose that (by $\mathbb{T}_{\mathcal{M}}$) the set $\{\ulcorner \eta \urcorner \mid \eta \in \mathsf{Sent}_{\mathcal{L}}, \mathcal{M} \vDash \eta\}$ is not definable by any $\mathcal{L}$-formula. Then for a given $\mathcal{L}$-formula $\Psi(x)$, with x free, the $\mathcal{L}$-formula $\neg\Psi(x)$ cannot define this

[1] Let us compare this to another theorem of Tarski [18], which states that the proposition $\forall^{\mathrm{inf}}\kappa(\kappa^2 = \kappa)$, saying that for every infinite cardinal number κ we have $\kappa \cdot \kappa = \kappa$, implies AC, the axiom of choice (the converse implication was known before). As is reported in [12], 'Tarski... tried to publish his theorem... but Fréchet and Lebesgue refused to present it. Fréchet wrote that an implication between two well known propositions is not a new result. Lebesgue wrote that an implication between two false propositions is of no interest.' With the further development of Zermelo–Frankel set theory ZF and thanks to the results of Gödel (1938) and Cohen (1963) we now know that the theories $\mathsf{ZF} + \mathtt{AC}$ and $\mathsf{ZF} + \neg\mathtt{AC}$ are both consistent, and the result of Tarski (1924) makes much sense when formulated as $\mathsf{ZF} \vdash \mathtt{AC} \leftrightarrow \forall^{\mathrm{inf}}\kappa(\kappa^2 = \kappa)$.

set, and so we cannot have $[[\mathcal{M} \vDash \neg\Psi(\ulcorner\eta\urcorner) \iff \mathcal{M} \vDash \eta$, for all $\mathcal{L}$-sentences $\eta]]$; hence there should exist some $\mathcal{L}$-sentence θ such that $\mathcal{M} \nvDash \neg\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$. Now, by the classical propositional tautology $\neg(p \leftrightarrow q) \equiv (\neg p \leftrightarrow q)$ we have $\mathcal{M} \vDash \Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$. So, for every $\Psi(x)$ there exists some θ for which the equivalence $\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$ holds in $\mathcal{M}$. □

Let us note that the result of Theorem 2.5 is quite general, as the model $\mathcal{M}$ can even be finite, thus it may verify that all the closed terms denote the same object; the definition of Gödel coding requires only an injective mapping of sentences into the set of closed terms of the language. It is instructive to have a look at the direct proof of $(\mathbb{T}_{\mathcal{M}} \Rightarrow \mathbb{G}_{\mathcal{M}})$: if $\mathbb{G}_{\mathcal{M}}$ does not hold, then there exists a complete theory $T \subseteq \mathrm{Th}(\mathcal{M})$, which is definable in $\mathcal{M}$. Since complete theories are maximally consistent, then we should have $T = \mathrm{Th}(\mathcal{M})$, and so the full theory $\mathrm{Th}(\mathcal{M})$ of $\mathcal{M}$ is definable in $\mathcal{M}$, and this contradicts $\mathbb{T}_{\mathcal{M}}$.

REMARK 2.6 (Making Sense of the Equivalences in Theorem 2.5).
For some $\langle\mathcal{L}, \ulcorner\cdot\urcorner, \mathcal{M}\rangle$'s all of the statements $\mathbb{D}_{\mathcal{M}}$, $\mathbb{G}_{\mathcal{M}}$, and $\mathbb{T}_{\mathcal{M}}$ hold: for e.g. $\mathcal{L}_{ar}$ with a standard Gödel coding $\ulcorner\cdot\urcorner$ and $\mathcal{M} = \langle\mathbb{N}; 0, 1, +, \times\rangle$.

For other $\langle\mathcal{L}, \ulcorner\cdot\urcorner, \mathcal{M}\rangle$'s none of those three statements hold: let $\mathcal{L}_+ = \{1, +\}$ and $\mathfrak{M} = \langle\mathbb{N}; 1, +\rangle$, and for a standard Gödel coding $\ulcorner\cdot\urcorner$ let for every $\mathcal{L}_+$-sentence η,

$$\llcorner\eta\lrcorner = \begin{cases} (\overline{\#\eta}) + (\overline{\#\eta}) & \text{if } \mathfrak{M} \vDash \eta, \\ 1 + [(\overline{\#\eta}) + (\overline{\#\eta})] & \text{if } \mathfrak{M} \nvDash \eta, \end{cases}$$

where (as we recall) $\#\eta$ is the value (natural number) of the closed term $\ulcorner\eta\urcorner$, and $\overline{m}$ (for $m \in \mathbb{N}$) is the standard term representing m, i.e. $1 + \cdots + 1$ where 1 appears for m times. Since $\mathfrak{M}$ is a decidable structure by Presburger's theorem (see e.g. [3,Theorem 32E]), the mapping $\llcorner\cdot\lrcorner$ (from $\mathcal{L}_+$-sentences to closed $\mathcal{L}_+$-terms) is computable (and also injective, since $\ulcorner\cdot\urcorner$ is so). Now, for $\Phi(x) = \exists y(x = y + y)$ and $\Psi(x) = \neg\Phi(x)$ the proposition $\mathbb{D}_{\mathfrak{M}}$ does not hold since for every $\mathcal{L}_+$-sentence η we have $\mathfrak{M} \vDash \eta$ iff $\mathfrak{M} \vDash \Phi(\llcorner\eta\lrcorner)$, and so $\mathfrak{M} \vDash \eta \leftrightarrow \Phi(\llcorner\eta\lrcorner)$; thus, for no $\mathcal{L}_+$-sentence θ could $\mathfrak{M} \vDash \Psi(\llcorner\theta\lrcorner) \leftrightarrow \theta$ hold. Hence, neither $\mathbb{G}_{\mathfrak{M}}$ nor $\mathbb{T}_{\mathfrak{M}}$ holds, as the complete theory $\mathrm{Th}(\mathfrak{M})$ is definable in $\mathfrak{M}$ by the $\mathcal{L}_+$-formula $\Phi(x)$.

So, after all, Tarski's undefinability theorem (in its semantic form) is not very much different from the diagonal lemma (in the semantic form). Therefore, it may seem at the first glance that the only way to prove Tarski's theorem is to use the diagonal lemma (as is done in almost all the textbooks). But as a matter of fact, there are some, supposedly, diagonal-free proofs for Tarski's theorem in the literature (see, e.g. [7]) which by Theorem 2.5 can give us some diagonal-free proofs for the diagonal lemma itself! We will outline two of them below.

Assume that the $\mathcal{L}_{ar}$-formula $\Upsilon(x)$ defines truth in $\langle\mathbb{N}; \mathcal{L}_{ar}\rangle$,

$$\text{i.e. } \mathbb{N} \vDash \Upsilon(\ulcorner\xi\urcorner) \leftrightarrow \xi \text{ for all } \mathcal{L}_{ar}\text{-formulas } \xi. \qquad (\mathfrak{T})$$

2.1 The first proof

Berry's paradox is rephrased as 'the least integer admitting no name involving less than two hundred words in English' by the late Kotlarski [9, §4]. The first proof uses the idea of this paradox, for which we need to make a convention.

CONVENTION 2.7
Let us make the convention that all the individual variables of our syntax are $x, x', x'', x''', \cdots$ whose lengths are $1, 2, 3, 4, \cdots$, respectively. By this convention, there will be at most finitely many $\mathcal{L}_{ar}$-formulas with length n for a given natural number $n \in \mathbb{N}$ (otherwise the formulas $x=x,\ y=y,\ z=z, \cdots$ all would have length three).

DEFINITION 2.8 ($\mathtt{len}(x), \overline{n}, \mathtt{definability}, \mathtt{D}(x), \mathtt{Def}_{\Upsilon}^{<z}(y), \mathtt{Berry}_{\Upsilon}^{<v}(u), \ell_{\Upsilon}, \mathtt{B}_{\Upsilon}(x)$).

- Let $\mathtt{len}(x)$ denote the *length* of the $\mathcal{L}_{ar}$-formula with Gödel code x; let us note that $\mathtt{len}$ is an $\mathcal{L}_{ar}$-definable function.
- For $n \in \mathbb{N}$, let $\overline{n}$ be the standard $\mathcal{L}_{ar}$-term that *represents* the number n, i.e. $\overline{0} = 0$, $\overline{1} = 1$ and for every $m \geqslant 1$ we have $\overline{m+1} = 1 + (\overline{m})$.
- We say that a number $n \in \mathbb{N}$ is *definable* by the $\mathcal{L}_{ar}$-formula $\varphi(x)$, in which x is the only free variable, when $\forall x[\varphi(x) \leftrightarrow x=\overline{n}]$ is true (in $\mathbb{N}$).
- Let $\mathtt{D}(x, x')$ be the Gödel code of the $\mathcal{L}_{ar}$-formula, which states that the formula with Gödel code x defines the number x'; let us note that $\mathtt{D}$ is an $\mathcal{L}_{ar}$-definable function and for every $\mathcal{L}_{ar}$-formula $\varphi(x)$ and every $m \in \mathbb{N}$ we have $\mathtt{D}(\ulcorner\varphi\urcorner, \overline{m}) = \ulcorner\forall x[\varphi(x) \leftrightarrow x=\overline{m}]\urcorner$.
- Let $\mathtt{Def}_{\Upsilon}^{<x'}(x)$ be the $\mathcal{L}_{ar}$-formula $\exists x''\ \big(\mathtt{Formula}(x'')\ \wedge\ \mathtt{len}(x'') < x'\ \wedge\ \Upsilon[\mathtt{D}(x'', x)]\big)$ which states that *the number x is definable by a formula with length less than x'*, if Υ is a truth predicate; needless to say, $\mathtt{Formula}(x'')$ states that x'' is the Gödel code of a formula.
- Let $\mathtt{Berry}_{\Upsilon}^{<x'}(x)$ be the $\mathcal{L}_{ar}$-formula $\neg\mathtt{Def}_{\Upsilon}^{<x'}(x)\ \wedge\ \forall x'' < x\, \mathtt{Def}_{\Upsilon}^{<x'}(x'')$, which states that x *is the least number not defined by a formula with length less than x'*.
- Let ℓ_{Υ} be the length of the $\mathcal{L}_{ar}$-formula $\mathtt{Berry}_{\Upsilon}^{<x'}(x)$. Let $\mathfrak{q}_{\Upsilon}$ be the $\mathcal{L}_{ar}$-term $(\overline{6})\times(\overline{\ell_{\Upsilon}})$; note that the value of $\mathfrak{q}_{\Upsilon}$ is the number $6\ell_{\Upsilon}$.
- Let $\mathtt{B}_{\Upsilon}(x)$ be the $\mathcal{L}_{ar}$-formula $\exists x'\big[x'=\mathfrak{q}_{\Upsilon}\ \wedge\ \mathtt{Berry}_{\Upsilon}^{<x'}(x)\big]$.

Here is an alternative proof (from [2, 9, 15]) for contradicting ($\mathbb{T}$):

PROOF. The length of $\mathtt{B}_{\Upsilon}(x)$ is less than $6\ell_{\Upsilon}$; since it is $14+\mathtt{len}(\overline{6})+\mathtt{len}(\overline{\ell_{\Upsilon}})+\ell_{\Upsilon} = 32+5\ell_{\Upsilon}$, as we have $\mathtt{len}(\overline{m}) = 4m-3$ for every $m \geqslant 1$. So, the $\mathcal{L}_{ar}$-formula $\mathtt{B}_{\Upsilon}(x)$ with length less than $6\ell_{\Upsilon}$ states that x is the least number that is not definable by any $\mathcal{L}_{ar}$-formula with length less than $6\ell_{\Upsilon}$. Hence, if $\mathtt{B}_{\Upsilon}(\alpha)$ holds, then α should not be definable by $\mathtt{B}_{\Upsilon}(x)$ itself. But this is a contradiction, since if $\mathtt{B}_{\Upsilon}(\alpha)$ holds, then α is definable by $\mathtt{B}_{\Upsilon}(x)$. That is because $\mathtt{B}_{\Upsilon}(\alpha)$ implies $\forall x[\mathtt{B}_{\Upsilon}(x) \leftrightarrow x=\alpha]$ by the sentence $\forall x, x'[\mathtt{B}_{\Upsilon}(x) \wedge \mathtt{B}_{\Upsilon}(x') \rightarrow x=x']$, which follows in turn from the sentence $\forall x, x', x''[\mathtt{Berry}_{\Upsilon}^{<x''}(x) \wedge \mathtt{Berry}_{\Upsilon}^{<x''}(x') \rightarrow x=x']$ that can be proved from the basic laws of the order relation. So, for no α can $\mathtt{B}_{\Upsilon}(\alpha)$ hold. Now, in reality, there exists a least number $\mathfrak{b} \in \mathbb{N}$ that is not definable by any $\mathcal{L}_{ar}$-formula of length less than $6\ell_{\Upsilon}$ (since by our convention there are only finitely many $\mathcal{L}_{ar}$-formulas with length less than $6\ell_{\Upsilon}$). So, $\neg\mathtt{Def}_{\Upsilon}^{<\mathfrak{q}_{\Upsilon}}(\overline{\mathfrak{b}})$ is true, and since it is the least such number then $\forall x'' < \overline{\mathfrak{b}}\, \mathtt{Def}_{\Upsilon}^{<\mathfrak{q}_{\Upsilon}}(x'')$ is true too. Thus, $\mathtt{Berry}_{\Upsilon}^{<\mathfrak{q}_{\Upsilon}}(\overline{\mathfrak{b}})$ is true and so is $\mathtt{B}_{\Upsilon}(\overline{\mathfrak{b}})$, which is a contradiction. □

This proof of Tarski's theorem (2.3) does not use the diagonal lemma (and so it can be called diagonal-free in a way), though it can be debated whether the proof is genuinely circular-free or not. By incorporating the proof of Theorem 2.5 into this proof (e.g. by taking $\Upsilon \equiv \neg\Psi$), one can get a proof for the semantic diagonal lemma, which is different from the standard (textbook) proofs (see [14]).

2.2 The second proof

DEFINITION 2.9 (Definable and Dominating Functions).
A function $f\colon \mathbb{N} \to \mathbb{N}$ is called *definable* whenever there exists an $\mathcal{L}_{ar}$-formula $\varphi(x,x')$ such that for every $m,n \in \mathbb{N}$ we have $f(m)=n \iff \mathbb{N} \vDash \varphi(\overline{m},\overline{n})$.
A function $F\colon \mathbb{N} \to \mathbb{N}$ is said to *dominate* a function $f\colon \mathbb{N} \to \mathbb{N}$, whenever there exists some $n \in \mathbb{N}$ such that $F(x) > f(x)$ holds for all $x \geqslant n$.

Indeed, for a given countably indexed family of functions $\{f_i\colon \mathbb{N} \to \mathbb{N}\}_{i\in\mathbb{N}}$ one can find a function that dominates all the functions of this family: put $F(x)=1+\max_{i\leqslant x} f_i(x)$; then for every $k \in \mathbb{N}$ and every $x \geqslant k$ we have $f_k(x) \leqslant \max_{i\leqslant x} f_i(x) < [1+\max_{i\leqslant x} f_i(x)] = F(x)$. This idea is used in the following proof of Tarski's theorem (2.3); cf. [8, 9]:

PROOF. Define the function $F\colon \mathbb{N} \to \mathbb{N}$ as

$$F(x) = \min\{x' \mid \forall \xi \leqslant x[\exists x'' \xi(x,x'') \to \exists x'' < x' \xi(x,x'')]\},$$

where ξ ranges over (the codes of) $\mathcal{L}_{ar}$-formulas with two free variables (whose codes are non-greater than x). We show that the function F dominates every $\mathcal{L}_{ar}$-definable function, but is itself $\mathcal{L}_{ar}$-definable if ($\mathfrak{T}$) holds; and this is a contradiction (since no function can dominate itself). To see that F dominates the family of all $\mathcal{L}_{ar}$-definable functions, assume that a function $f\colon \mathbb{N} \to \mathbb{N}$ is definable by an $\mathcal{L}_{ar}$-formula $\varphi(x,x')$. Now, for every $m \geqslant \#\varphi$ we show that $F(m) > f(m)$ holds: if $F(m) \leqslant f(m)$, then from $\varphi(\overline{m},\overline{f(m)})$ we have $\exists x''\varphi(\overline{m},x'')$ and so $\exists x'' < \overline{F(m)}\colon \varphi(\overline{m},x'')$ by the definition of F, which implies $\exists x'' < \overline{f(m)}\colon \varphi(\overline{m},x'')$ by the assumption $F(m) \leqslant f(m)$; but for every $k \neq f(m)$ we have $\neg\varphi(\overline{m},\overline{k})$, and so $\forall x'' < \overline{f(m)}\colon \neg\varphi(\overline{m},x'')$, a contradiction. Now, if ($\mathfrak{T}$) holds for Υ, then F is definable by the $\mathcal{L}_{ar}$-formula $\psi(x,x') \wedge \forall x'' < x' \neg\psi(x,x'')$ where ψ is the formula $\forall \xi \leqslant x[\exists x'' \Upsilon(\ulcorner \xi(x,x'') \urcorner) \to \exists x'' < x' \Upsilon(\ulcorner \xi(x,x'') \urcorner)]$. □

As a matter of fact, the function F used by Kotlarski [8, 9] is defined as

$$F(x) = \min\{x' \mid \forall \xi,\alpha \leqslant x[\exists x'' \, \xi(\alpha,x'') \to \exists x'' < x' \, \xi(\alpha,x'')]\},$$

which corresponds to $F(x)=1+\max_{i,j\leqslant x} f_i(j)$ that dominates $\{f_i\colon \mathbb{N}\to\mathbb{N}\}_{i\in\mathbb{N}}$.

3 The Diagonal Lemma, Syntactically

The diagonal lemma is usually stated as the provability of $\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$ in a theory like Robinson's arithmetic, for some sentence θ which depends on the given formula $\Psi(x)$. Let us call this the *syntactic diagonal lemma*. A syntactic version of Tarski's theorem on the undefinability of truth is as follows (where $\mathcal{L}$ is a first-order language with a computable injective Gödel coding $\ulcorner\cdot\urcorner : \mathtt{Sent}_{\mathcal{L}} \to \mathtt{ClTerms}_{\mathcal{L}}$).

DEFINITION 3.1 (Syntactic Version of Tarski's Theorem).
For an $\mathcal{L}$-formula $\Phi(x)$, let TB^{Φ} be the set of all truth biconditionals $\Phi(\ulcorner\eta\urcorner) \leftrightarrow \eta$, where η ranges over all the $\mathcal{L}$-sentences. That is $\mathsf{TB}^{\Phi} = \{\Phi(\ulcorner\eta\urcorner) \leftrightarrow \eta \mid \eta \in \mathsf{Sent}_{\mathcal{L}}\}$.
The following statement is called the *syntactic version of Tarski's theorem* on a consistent theory T in the language of $\mathcal{L}$:
For every $\mathcal{L}$-formula $\Phi(x)$ we have $T \not\supseteq \mathsf{TB}^{\Phi}$.

DEFINITION 3.2 (Weak Diagonal Lemma).
The following statement is called the *weak diagonal lemma* for a consistent $\mathcal{L}$-theory T:

For every $\mathcal{L}$-formula $\Psi(x)$ there exists an $\mathcal{L}$-sentence θ such that T is consistent with the sentence $[\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta]$.

We show that the syntactic version of Tarski's theorem is equivalent to the weak (syntactic) diagonal lemma.

THEOREM 3.3 (Weak Diagonal Lemma $\equiv$ Syntactic Version of Tarski's Theorem).
The weak diagonal lemma is equivalent to the syntactic version of Tarski's theorem.

PROOF. First, suppose that the weak diagonal lemma holds for a consistent theory T. Take any formula $\Phi(x)$; we show that $T \not\supseteq \mathsf{TB}^{\Phi}$. By the assumption, there exists a sentence θ such that the theory T is consistent with $[\neg\Phi(\ulcorner\theta\urcorner) \leftrightarrow \theta]$. Thus, $T \nvdash \Phi(\ulcorner\theta\urcorner) \leftrightarrow \theta$ (since $[\neg\Phi(\ulcorner\theta\urcorner) \leftrightarrow \theta] \equiv \neg[\Phi(\ulcorner\theta\urcorner) \leftrightarrow \theta]$) and so $T \not\supseteq \mathsf{TB}^{\Phi}$. Second, suppose that $T \not\supseteq \mathsf{TB}^{\Phi}$ for all formulas $\Phi(x)$. Take any formula $\Psi(x)$; we show the existence of some θ such that T is consistent with $\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$. Since $T \not\supseteq \mathsf{TB}^{\neg\Psi}$, there should exist some sentence θ such that $T \nvdash \neg\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$. Therefore, T is consistent with the sentence $\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta$ (which is equivalent to $\neg[\neg\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta]$). □

REMARK 3.4 (Making Sense of the Equivalences in Theorem 3.3).
For some $\langle\mathcal{L}, \ulcorner\cdot\urcorner, T\rangle$s both the weak diagonal lemma and the syntactic version of Tarski's theorem hold: when, e.g. our language is $\mathcal{L}_{ar}$ and $\ulcorner\cdot\urcorner$ is a classic Gödel coding and T is Robinson's arithmetic. For some $\langle\mathcal{L}, \ulcorner\cdot\urcorner, T\rangle$s neither of them holds: let $\mathcal{L}_{+} = \{1, +\}$ and $T = \mathrm{Th}(\mathfrak{M})$, where $\mathfrak{M}$ and $\llcorner\cdot\lrcorner$ are as defined in Remark 2.6. For the $\mathcal{L}_{+}$-formulas $\Phi(x)$ and $\Psi(x)$ defined in Remark 2.6, we have $T \supseteq \mathsf{TB}^{\Phi}$, and for no $\mathcal{L}_{+}$-sentence θ can the sentence $[\Psi(\ulcorner\theta\urcorner) \leftrightarrow \theta]$ be consistent with the theory T.

As a matter of fact, the weak diagonal lemma *cannot* show the independence of Gödelian sentences (even when the theory is sound):

REMARK 3.5 (Weak Diagonal Lemma vs. Gödel's Proof).
For a consistent and recursively enumerable theory T extending Robinson's arithmetic, the consistency of $\neg\mathtt{Pr}_T(\ulcorner\theta\urcorner) \leftrightarrow \theta$ with T implies that θ is unprovable in T, but does not imply that θ is independent from T (even if T is ω-consistent):

(1) If $T \vdash \theta$, then $T \vdash \mathtt{Pr}_T(\ulcorner\theta\urcorner)$ and so $T + [\neg\mathtt{Pr}_T(\ulcorner\theta\urcorner) \leftrightarrow \theta] \vdash \neg\theta$; therefore the theory $T + [\neg\mathtt{Pr}_T(\ulcorner\theta\urcorner) \leftrightarrow \theta]$ cannot be consistent.
(2) For a contradictory sentence like $\delta = (0 \neq 0)$, the sentence $\neg\mathtt{Pr}_T(\ulcorner\delta\urcorner) \leftrightarrow \delta$ is consistent with T (by Gödel's Second Incompleteness Theorem), but δ is not independent from T (as T proves its negation).

It is stated in [11, p. 202] that 'if T is also sound, then $T \nvdash \neg\sigma$' where σ is a sentence with the property $T \vdash \sigma \iff T \vdash \neg\mathtt{Pr}_T(\ulcorner\sigma\urcorner)$. Unfortunately, this is not true since e.g. for $\sigma = (0 \neq 0)$ we have $T \vdash \sigma \iff T \vdash \neg\mathtt{Pr}_T(\ulcorner\sigma\urcorner)$ by Gödel's second incompleteness theorem, but trivially $T \vdash \neg\sigma$. If we replace $\mathtt{Pr}_T(x)$ with Rosser's provability predicate $R\mathtt{Pr}_T(x)$, then it is true that for every ϱ that satisfies $T \vdash \varrho \iff T \vdash \neg R\mathtt{Pr}_T(\ulcorner\varrho\urcorner)$ we have $T \nvdash \varrho, \neg\varrho$ if T is (only) consistent; see the next theorem.

Indeed, the weak diagonal lemma is sufficiently strong to prove Rosser's theorem:

THEOREM 3.6 (Weak Diagonal Lemma $\Longrightarrow$ Rosser's Theorem).
If the weak diagonal lemma holds for a consistent and recursively enumerable theory that extends Robinson's arithmetic, then there exists a sentences, which is independent from that theory.

PROOF. For such a theory T, suppose that $\mathtt{prf}_T(x,y)$ is its proof predicate (stating that x is the Gödel code of a proof of the sentence with Gödel code y in T). By the weak diagonal lemma there exists a sentence ρ such that the following theory is consistent:

$$U = T + \Big(\forall x\big[\mathtt{prf}_T(x,\ulcorner\rho\urcorner) \to \exists y{<}x\,\mathtt{prf}_T(y,\ulcorner\neg\rho\urcorner)\big] \longleftrightarrow \rho\Big).$$

The standard proof of Rosser's theorem can show that ρ is independent from T:

- If $T \vdash \rho$, then $T \vdash \mathtt{prf}_T(\overline{k},\ulcorner\rho\urcorner)$ for some $k\in\mathbb{N}$ and so $U \vdash \exists y{<}\overline{k}\,\mathtt{prf}_T(y,\ulcorner\neg\rho\urcorner)$, by the definition of U, which contradicts $\bigwedge\!\!\bigwedge_{m\in\mathbb{N}} U \vdash \neg\mathtt{prf}_T(\overline{m},\ulcorner\neg\rho\urcorner)$ (that holds by $T \nvdash \neg\rho$).
- If $T \vdash \neg\rho$, then $T \vdash \mathtt{prf}_T(\overline{k},\ulcorner\neg\rho\urcorner)$ for some $k\in\mathbb{N}$. Reason inside U:
  for some x we have $\mathtt{prf}_T(x,\ulcorner\rho\urcorner) \wedge \forall y < x\,\neg\mathtt{prf}_T(y,\ulcorner\neg\rho\urcorner)$; now $\overline{k} < x$ is impossible, and so $x \leqslant \overline{k}$, hence $\bigvee_{i\leqslant k}(x=\bar{i})$, therefore $\bigvee_{i\leqslant k}\mathtt{prf}_T(\bar{i},\ulcorner\rho\urcorner)$.
  Hence, $U \vdash \bigvee_{i\leqslant k}\mathtt{prf}_T(\bar{i},\ulcorner\rho\urcorner)$, but this contradicts $\bigwedge\!\!\bigwedge_{m\in\mathbb{N}} U \vdash \neg\mathtt{prf}_T(\overline{m},\ulcorner\rho\urcorner)$ (that holds by $T \nvdash \rho$).

Therefore, $T \nvdash \rho, \neg\rho$. □

So, the weak diagonal lemma is worthy of studying further. Unfortunately, the second proof (Subsection 2.2) for Tarski's theorem cannot be carried over to the syntactic version of Tarski's theorem (cf. [19]). However, the first proof (Section 2.1) can be adapted for it:

THEOREM 3.7 (Syntactic Tarski's Theorem).
If T is a consistent extension of Robinson's arithmetic, then for no formula $\Psi(x)$ can we have $T \supseteq \mathsf{TB}^{\Psi}$.

PROOF. Assume that the consistent theory T contains Robinson's arithmetic, and it also contains the set TB^{Υ} for a formula $\Upsilon(x)$. We work with the Convention 2.7 (and also Definition 2.8). Fix a number $n\in\mathbb{N}$; and reason inside the theory T:

> Assume $\mathtt{B}_\Upsilon(\overline{n})$; so, $\mathtt{Berry}_\Upsilon^{<\mathfrak{q}_\Upsilon}(\overline{n})$ thus (1) $\neg\mathtt{Def}_\Upsilon^{<\mathfrak{q}_\Upsilon}(\overline{n})$ and (2) $\forall x''<\overline{n}\,\mathtt{Def}_\Upsilon^{<\mathfrak{q}_\Upsilon}(x'')$ hold. Fix x; if $\mathtt{B}_\Upsilon(x)$ holds, then (i) $\neg\mathtt{Def}_\Upsilon^{<\mathfrak{q}_\Upsilon}(x)$ and (ii) $\forall x''<x\,\mathtt{Def}_\Upsilon^{<\mathfrak{q}_\Upsilon}(x'')$. We also have either $x<\overline{n}$ or $x=\overline{n}$ or $x>\overline{n}$ (which holds in Robinson's arithmetic). Now, $x<\overline{n}$ contradicts (i) and (2), and $x>\overline{n}$ contradicts (1) and (ii). Hence, $x=\overline{n}$; which shows that $\forall x[\mathtt{B}_\Upsilon(x) \leftrightarrow x=\overline{n}]$ holds. Thus, $\Upsilon(\mathtt{D}(\ulcorner\mathtt{B}_\Upsilon\urcorner,\overline{n}))$ holds by TB^{Υ}, and since we have $\mathtt{len}(\mathtt{B}_\Upsilon(x)) < \mathfrak{q}_\Upsilon$, then $\mathtt{Def}_\Upsilon^{<\mathfrak{q}_\Upsilon}(\overline{n})$; which contradicts (1). Therefore, the assumption $\mathtt{B}_\Upsilon(\overline{n})$ leads to a contradiction. Hence, $\neg\mathtt{B}_\Upsilon(\overline{n})$ holds, so $\neg\mathtt{Berry}_\Upsilon^{<\mathfrak{q}_\Upsilon}(\overline{n})$ holds too, thus we have (∗) $\bigwedge\!\!\bigwedge_{i<n}\mathtt{Def}_\Upsilon^{<\mathfrak{q}_\Upsilon}(\bar{i}) \to \mathtt{Def}_\Upsilon^{<\mathfrak{q}_\Upsilon}(\overline{n})$.

Therefore, $T \vdash \mathtt{Def}_\Upsilon^{<\mathfrak{q}_\Upsilon}(\overline{n})$ can be shown by induction on $n\in\mathbb{N}$ from (∗). Let $\mathfrak{m}\in\mathbb{N}$ be greater than all the Gödel codes of formulas with length less than $6\ell_\Upsilon$. Therefore, for all $n\in\mathbb{N}$ we have $T \vdash \exists\xi<\overline{\mathfrak{m}}\,\Upsilon(\ulcorner\mathtt{D}(\xi,\overline{n})\urcorner)$. So, inside T for any $n\in\mathbb{N}$ there exists some formula $\xi_n(x)$ that defines n, i.e. $\forall x\,(\xi_n(x)\leftrightarrow x=\overline{n})$ holds by TB^{Υ}, and the Gödel codes of all ξ_n's are less than $\mathfrak{m}$. This contradicts the Pigeonhole's Principle a version of which is provable in Robinson's arithmetic: since both of the sentences $\forall x(x \not< 0)$ and $\forall x(x<\overline{k+1} \to x\leqslant\overline{k})$ are provable in this arithmetic, then for every $\{\xi_k<\mathfrak{m}\}_{k\leqslant\mathfrak{m}}$ there should exist some $i<j\leqslant\mathfrak{m}$ such that $\xi_i=\xi_j$. Reason inside T again:

There are some $\bar{i} < \bar{j} \leqslant \overline{\mathfrak{m}}$ and some formula $\varphi(x)$ for which we have $\Upsilon(\ulcorner \mathtt{D}(\ulcorner \varphi \urcorner, \bar{i}) \urcorner)$ and $\Upsilon(\ulcorner \mathtt{D}(\ulcorner \varphi \urcorner, \bar{j}) \urcorner)$. So, by TB^{Υ}, both $\forall x\,(\varphi(x) \leftrightarrow x = \bar{i})$ and $\forall x\,(\varphi(x) \leftrightarrow x = \bar{j})$ hold, combining which implies that $\bar{i} = \bar{i} \rightarrow \varphi(\bar{i}) \rightarrow \bar{i} = \bar{j}$, a contradiction.

So, T is inconsistent, which contradicts the assumption. □

4 Conclusion

The semantic form of Tarski's undefinability theorem, that the set $\{\#\eta \mid \mathbb{N} \vDash \eta\}$ is not definable in arithmetic, is equivalent to the semantic form of the diagonal lemma, that for a given $\Psi(x)$ there exists a sentence θ such that $\mathbb{N} \vDash \Psi(\#\theta) \leftrightarrow \theta$. We outlined two seemingly diagonal-free proofs for these equivalent theorems. The syntactic version of Tarski's theorem, that no consistent extension of Robinson's arithmetic contains the set of truth biconditionals $\mathsf{TB}^{\Phi} = \{\Phi(\ulcorner \eta \urcorner) \leftrightarrow \eta \mid \eta \text{ is a sentence}\}$ for any formula $\Phi(x)$, is equivalent to the weak (syntactic) diagonal lemma, that for every $\Psi(x)$ there exists a sentence θ such that $\Psi(\ulcorner \theta \urcorner) \leftrightarrow \theta$ is consistent with such a theory. Even though Gödel's proof does not work with the weak diagonal lemma, the weak lemma is sufficiently strong to prove Rosser's theorem. So, the syntactic version of Tarski's theorem can derive Gödel and Rosser's incompleteness theorem.

Acknowledgements

The most helpful comments and suggestions of the two referees for this journal are highly appreciated. This research is supported by the Office of the Vice Chancellor for Research and Technology, University of Tabriz, Iran.

References

[1] S. Buss. First-order proof theory of arithmetic. In *Handbook of Proof Theory (Chapter II)*, S. Buss., ed, pp. 79–147. Elsevier, 1998. Available on the net at https://bit.ly/321RiGM.

[2] X. Caicedo and L. P. de Berry Revisitada. O la indefinibilidad de la definibilidad y las limitaciones de los formalismos (in Spanish). In *Lecturas Matemáticas*, vol. **14**, pp. 37–48, 1993. Revised (2004), available on the net at https://bit.ly/3jf6GY1.

[3] H. B. Enderton. *A Mathematical Introduction to Logic*, 2nd edn. Academic Press, 2001.

[4] H. Gaifman. Naming and diagonalization, from Cantor to Gödel to Kleene. *Logic Journal of the IGPL*, **14**, 709–728, 2006. doi: 10.1093/jigpal/jzl006.

[5] B. Grabmayr and A. Visser. Self-reference upfront: a study of self-referential Gödel numberings. 1–40, 2020. arXiv:2006.12178v2 [math.LO]; available at https://arxiv.org/abs/2006.12178.

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

[7] R. Kossak. Undefinability of truth and nonstandard models. *Annals of Pure and Applied Logic*, **126**, 115–123, 2004. doi: 10.1016/j.apal.2003.10.011.

[8] H. Kotlarski. Other proofs of old results. *Mathematical Logic Quarterly*, **44**, 474–480, 1998. doi: 10.1002/malq.19980440406.

[9] H. Kotlarski. The incompleteness theorems after 70 years. *Annals of Pure and Applied Logic*, **126**, 125–138, 2004. doi: 10.1016/j.apal.2003.10.012.

[10] V. McGee. The first incompleteness theorem. In *Handouts of the Course "Logic II"*, 2002. Available on the net at https://bit.ly/301QLTA.

[11] Y. N. Moschovakis. Kleene's amazing second recursion theorem. *The Bulletin of Symbolic Logic*, **16**, 189–239, 2010. doi: 10.2178/bsl/1286889124.

[12] J. Mycielski. A system of axioms of set theory for the rationalists. *Notices of the American Mathematical Society*, **53**, 206–213, 2006. Available on the net at https://bit.ly/2KtlI0K.

[13] S. Salehi. Theorems of Tarski's Undefinability and Gödel's Second Incompleteness—Computationally. 1–12, 2019. arXiv:1509.00164v3 [math.LO]; available at https://arxiv.org/abs/1509.00164.

[14] S. Salehi. On the diagonal lemma of Gödel and Carnap. *The Bulletin of Symbolic Logic*, **26**, 80–88, 2020. doi: 10.1017/bsl.2019.58.

[15] G. Serény. Boolos-style proofs of limitative theorems. *Mathematical Logic Quarterly*, **50**, 211–216, 2004. doi: 10.1002/malq.200310091.

[16] G. Serény. The diagonal lemma as the formalized Grelling paradox. In *Collegium Logicum 9, Gödel Centenary 2006*, M. Baaz and N. Preining., eds, pp. 63–66. Kurt Gödel Society, Vienna, 2006. Available on the net at http://cds.cern.ch/record/965085.

[17] P. Smith. An Introduction to Gödel's Theorems, 2nd ed. Cambridge University Press, 2013.

[18] A. Tajtelbaum-Tarski. Sur Quelques Théorèmes Qui Équivalent à l'Axiome du Choix. *Fundamenta Mathematicæ*, **5**, 147–154, 1924. doi: 10.4064/fm-5-1-147-154.

[19] A. Visser. From Tarski to Gödel—or how to derive the second incompleteness theorem from the undefinability of truth without self-reference. *Journal of Logic and Computation*, **29**, 595–604, 2019. doi: 10.1093/logcom/exz004. The preprint version (2018): arXiv:1803.03937 [math.LO] (pp. 1–7); available at https://arxiv.org/abs/1803.03937.

[20] W. B. Wasserman. It is "Pulling a Rabbit Out of the Hat": typical diagonal lemma "proofs" beg the question. *Social Science Research Network*, 1–11, 2008. doi: 10.2139/ssrn.1129038.

Received 15 September 2020