THE BULLETIN OF SYMBOLIC LOGIC
Volume 26, Number 1, March 2020

ON THE DIAGONAL LEMMA OF GÖDEL AND CARNAP

SAEED SALEHI

Abstract. A cornerstone of modern mathematical logic is the diagonal lemma of Gödel and Carnap. It is used in, for example, the classical proofs of the theorems of Gödel, Rosser, and Tarski. From its first explication in 1934, just essentially one proof has appeared for the diagonal lemma in the literature; a proof that is so tricky and hard to relate that many authors have tried to avoid the lemma altogether. As a result, some so-called diagonal-free proofs have been given for the above-mentioned fundamental theorems of logic. In this paper, we provide new proofs for the semantic formulation of the diagonal lemma, and for a weak version of the syntactic formulation of it.

§1. Introduction. Gödel's original proof, in his seminal paper [13], for the first incompleteness theorem constructed a sentence, nowadays denoted by $\mathtt{G}$, such that $\mathtt{G}$ is equivalent to $\neg\mathsf{Pr}_T(\ulcorner\mathtt{G}\urcorner)$, that is, to the sentence that asserts the nonprovability of $\mathtt{G}$ in the theory T. Here $\ulcorner\mathtt{G}\urcorner$ denotes a term in the language of T that represents the Gödel code of $\mathtt{G}$. Later on, as was also confirmed by Gödel in [14, footnote 23, p. 363], Carnap [7, Section 35] realized that for any formula $\Psi(x)$, with the only free variable x, there exists a sentence θ such that the biconditional sentence $\theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)$ holds (is true in the standard model of natural numbers $\mathbb{N}$, and is even provable in certain weak arithmetical theories). This statement, now called the diagonal lemma (of Gödel and Carnap), has essentially only one proof in the literature. A proof that Buss [5, p. 119] describes as "quite simple but rather tricky and difficult to conceptualize," and McGee writes about it in [22] that we "would hope that such a deep theorem would have an insightful proof. No such luck. I am going to write down a sentence ... and verify that it works. What I won't 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." This tricky

Received July 16, 2019.
2020 *Mathematics Subject Classification*. Primary 03F40, Secondary 03F30.
Keywords and phrases. diagonal lemma, self-reference, Gödel's incompleteness theorem, Tarski's undefinability theorem, Berry's paradox, Carnap.

1079-8986/20/2601-0004
DOI:10.1017/bsl.2019.58

and magical proof has prompted some authors to try to either demystify it, such as in, for example [28], or even abandon the lemma altogether and look for some other (the so called diagonal-free) proofs; see, for example, [20, p. 126] and the references therein.

Here, we give a new proof for the semantical version of the diagonal lemma. This semantical form is sufficiently strong to prove a semantic version of Tarski's theorem on the undefinability of arithmetical truth, and the incompleteness of sound and definable arithmetical theories in the sense of Gödel and Smullyan. We also study a weak syntactic version of this lemma and provide an alternative proof for it. We will see that this weaker form is still sufficiently strong to prove certain formulations of the incompleteness theorems of Gödel and Rosser. Before going into the details, let us review the general strategy of the proofs.

A key observation is that if some formula $\Psi(x)$ does not have a fixed point in $\mathbb{N}$ (i.e., for no sentence θ, the biconditional $\theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)$ holds true in $\mathbb{N}$), then $\mathbb{N} \vDash \theta \leftrightarrow \neg\Psi(\ulcorner\theta\urcorner)$ holds for all sentences θ; in other words, the formula $\neg\Psi(x)$ is a truth definition, contradicting Tarski's theorem. We employ a paradox that is named after George Godfrey Berry (1867–1928), a junior librarian at Oxford's Bodleian Library, by Russell (see, e.g. [25, p. 223]). As the history goes, this is Russell's version of Berry's original paradox that is nowadays called *the Berry paradox* (see [9, pp. 8, 9]). The paradox is this:

"The least natural number that cannot be described by less than 15 words" describes a number, uniquely, by less than 15 words. Since there are finitely many sentences with less than 15 words, such a number exists and is unique. But the above description has less than 15 words and does describe that natural number; a contradiction.

In fact, Berry's paradox was first used by Chaitin [8] in 1970 for his proof of the first incompleteness theorem. Later, Boolos [2] gave another proof for the first incompleteness theorem of Gödel, in 1989, which was based on Berry's paradox too; see also [1], [3], and [4, Section 17.3]. Berry's paradox has been used for proving Tarski's undefinability theorem as well, see [6] and [27, Corollary 2]. The research on Berry-based proofs is a live topic, the two most recent publications on which are [17] and [26]. Before Chaitin and Boolos, Rosser [24] (in 1936) and Kleene [18, 19] (in 1936 and 1950) had given alternative proofs for the first incompleteness theorem. Their proofs did not use Berry's paradox (see [26] and the references therein), and, instead, used ideas of computability theory. The computability theory approach to proving the incompleteness has the conceptual advantage of linking the study of incompleteness with the study of computability in a smooth way. Another important connection with computability theory is that Kleene's second recursion theorem implies a slightly weaker form of the syntactic form of the diagonal lemma that is sufficient for proving the Gödel–Rosser theorem; see [23, Section 5] and also [12].

Gödel remarked in [13, footnote 14, p. 149] that any epistemological paradox "could be used for a similar proof of the existence of undecidable propositions"; as another example, the surprise examination paradox can be used to prove Gödel's (second) incompleteness theorem (see [21]). Gödel's proof is analogous to Richard's paradox, and "closely related to the" Liar paradox (see [13, p. 149]). Our proofs are similar to Boolos'.

§2. Diagonal lemma, semantically and syntactically. Let us fix the language of arithmetic as $\langle \mathfrak{s}, 0, +, \cdot, < \rangle$ (over first-order logic with the equality), where $\mathfrak{s}$ is a unary function symbol, interpreted as the successor function ($\mathfrak{s}(m) = m+1$ for each m), 0 is a constant symbol, $+$ and $\cdot$ are binary function symbols, and $<$ is a binary relation symbol, with their standard interpretations. For any $n \in \mathbb{N}$, let $\overline{n}$ denote the term $\mathfrak{s}(\ldots\mathfrak{s}(0))$, where $\mathfrak{s}$ appears n-times. Let us be given a fixed Gödel coding $\zeta \mapsto \ulcorner \zeta \urcorner$, where $\ulcorner \zeta \urcorner$ is the term $\overline{n}$ in the language of arithmetic when n is the Gödel number of ζ.

2.1. The semantic form of the diagonal lemma.

Convention. Let us make the convention that all the individual variables are $x, x', x'', x''', \ldots$ whose lengths are $1, 2, 3, 4, \ldots$, respectively. This way, there will be at most finitely many formulas with length less than n for all $n \in \mathbb{N}$; otherwise, the length of $x < x, x' < x', \ldots$ would be 3. ⊣

Definition 2.1 (Definability, and the formulas $\boldsymbol{\delta}(u, v)$, $\mathsf{F}_1(u)$, and $\mathsf{L}^{<v}(u)$). A formula $\varphi(x)$, with the only free variable x, defines the natural number n when the sentence $\forall x[\varphi(x) \leftrightarrow x = \overline{n}]$ is true (in $\mathbb{N}$). Let $\boldsymbol{\delta}(\ulcorner \varphi \urcorner, \overline{n})$ be an abbreviation for the sentence $\forall x[\varphi(x) \leftrightarrow x = \overline{n}]$.

Let $\mathsf{F}_1(u)$ be the formula indicating that u is the Gödel code of a formula whose only free variable is x. Let $\mathsf{L}^{<v}(u)$ be the formula indicating that u is the Gödel code of a formula with length less than v; here u and v are free variables (among $x, x', x'', \ldots$, above). ⊣

By Gödel's arithmetization and coding techniques, $\mathsf{F}_1(u)$ and $\mathsf{L}^{<v}(u)$ can be expressed by some formulas in the language of arithmetic; also the mapping $(y, z) \mapsto \ulcorner \boldsymbol{\delta}(y, z) \urcorner$, for the y's with $\mathsf{F}_1(y)$, can be represented in the language of arithmetic. We show in Theorem 2.3 that for any given formula $\Psi(x)$ with the only free variable x, there are some natural numbers $m, n \in \mathbb{N}$ such that $\mathbb{N} \vDash \boldsymbol{\delta}(\overline{m}, \overline{n}) \leftrightarrow \Psi(\ulcorner \boldsymbol{\delta}(\overline{m}, \overline{n}) \urcorner)$. So, let us fix $\Psi(x)$ as a given such formula.

Definition 2.2 ($\mathsf{D}^{<v}(u)$, $\boldsymbol{\beta}^{<v}(u)$, $\boldsymbol{\ell}$, and $\mathsf{B}(x)$ depending on Ψ).
Let $\mathsf{D}^{<v}(u)$ be the formula $\exists \phi[\mathsf{F}_1(\phi) \wedge \mathsf{L}^{<v}(\phi) \wedge \neg\Psi(\ulcorner \boldsymbol{\delta}(\phi, u) \urcorner)]$.
Let $\boldsymbol{\beta}^{<v}(u)$ be the formula $\neg \mathsf{D}^{<v}(u) \wedge \forall w < u\, \mathsf{D}^{<v}(w)$.
Let $\boldsymbol{\ell}$ be the length of the formula $\boldsymbol{\beta}^{<x'}(x)$.
Let $\mathsf{B}(x)$ be the formula $\exists x'[x' = \overline{5} \cdot \overline{\boldsymbol{\ell}} \wedge \boldsymbol{\beta}^{<x'}(x)]$. ⊣

The intuitive meaning of $\mathsf{D}^{<v}(u)$ is that the number u is definable by a formula with length less than v, if $\neg\Psi$ is a truth predicate. Then $\boldsymbol{\beta}^{<v}(u)$ says that u is the least number not definable by any formula with length less than v. It is rather easy to see that the length of the formula $\mathsf{B}(x)$ is less than $5\boldsymbol{\ell}$

(cf. [26]). So, the relation of the formulas $\boldsymbol{\beta}^{<\nu}(u)$ and $\mathsf{B}(u)$ with the Berry's paradox is apparent now.

THEOREM 2.3 (The semantic diagonal lemma). *For a given formula* $\Psi(x)$ *with the only free variable* x*, there are some natural numbers* $m,n\in\mathbb{N}$ *with* $\mathsf{F}_1(\overline{m})$ *such that* $\mathbb{N}\vDash\boldsymbol{\delta}(\overline{m},\overline{n})\leftrightarrow\Psi(\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner)$.

PROOF. Assume for the sake of a contradiction that for all numbers $m,n\in\mathbb{N}$ with $\mathsf{F}_1(\overline{m})$, we have $\mathbb{N}\nvDash\boldsymbol{\delta}(\overline{m},\overline{n})\leftrightarrow\Psi(\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner)$. Then $\neg\Psi$ is a truth predicate for the $\boldsymbol{\delta}(y,z)$ formulas: $\mathbb{N}\vDash\boldsymbol{\delta}(\overline{m},\overline{n})\leftrightarrow\neg\Psi(\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner)$ holds for every $m,n\in\mathbb{N}$ with $\mathsf{F}_1(\overline{m})$. Let $\mathfrak{b}\in\mathbb{N}$ be the least number that is not definable by any formula with length less than 5ℓ. Then $\mathbb{N}\vDash\mathsf{B}(\overline{\mathfrak{b}})$ holds by our assumption on Ψ. Since $\mathbb{N}\vDash\forall x,y[\mathsf{B}(x)\wedge\mathsf{B}(y)\rightarrow x=y]$ holds as well, we have $\mathbb{N}\vDash\forall x[\mathsf{B}(x)\leftrightarrow x=\overline{\mathfrak{b}}]$ too. So, $\mathfrak{b}$ is definable by $\mathsf{B}(x)$ whose length is less than 5ℓ; a contradiction. ⊣

This semantic form of the diagonal lemma is sufficiently strong for proving Gödel's first incompleteness theorem for sound and definable theories, and also for proving the semantic form of Tarski's theorem on the undefinability of the truth of the set of the (Gödel codes of the) arithmetical sentences:

COROLLARY 2.4 (Semantic theorems of Gödel and Tarski).
1. *If* T *is a sound theory whose set of axioms is arithmetically definable, then* T *is incomplete.*
2. *The set* $\{\ulcorner\xi\urcorner\mid\mathbb{N}\vDash\xi\}$*, where* ξ *ranges over the sentences in the language of arithmetic, is not arithmetically definable.*

2.2. The classical proof of the diagonal lemma. Let us compare the proof of Theorem 2.3 with the classical proof(s) of the diagonal lemma. The following argument appears in the lemma of [16], which, according to its author, "was discovered by the referee" of the *Journal of Symbolic Logic*. If our language contains a symbol for the primitive recursive function ϱ with the interpretation that $\varrho(\ulcorner\varphi(x)\urcorner)=\ulcorner\varphi(\ulcorner\varphi(x)\urcorner/x)\urcorner$, for any $\varphi(x)$ with $\mathsf{F}_1(\ulcorner\varphi(x)\urcorner)$, then for a formula $\Psi(x)$, let $\theta=\Psi(\varrho(\ulcorner\Psi(\varrho(x))\urcorner))$. Now,

$$\theta=\Psi(\ulcorner\Psi(\varrho(\ulcorner\Psi(\varrho(x))\urcorner))\urcorner)=\Psi(\ulcorner\theta\urcorner).$$

So, in Primitive Recursive Arithmetic, for any $\Psi(x)$, there exists a sentence θ such that θ is (not only equivalent with but also equal to) $\Psi(\ulcorner\theta\urcorner)$. If our language does not contain a symbol for the ϱ function, this primitive recursive function should be (strongly) representable by a formula such as $\boldsymbol{\sigma}(x,y)$ in the language of arithmetic. That is to say that the sentence $\forall y[\boldsymbol{\sigma}(\overline{n},y)\leftrightarrow y=\varrho(\overline{n})]$, for each natural number $n\in\mathbb{N}$, is true. Both of the following arguments appear in, for example, [29, Theorem 24.4]:
(A): The universal argument goes as follows:
Let $\alpha(x)=\forall y[\boldsymbol{\sigma}(x,y)\rightarrow\Psi(y)]$ and $\theta_\alpha=\alpha(\ulcorner\alpha(x)\urcorner/x)$. Then we have

$$\begin{aligned}\theta_\alpha &= \forall y[\boldsymbol{\sigma}(\ulcorner\alpha(x)\urcorner,y)\rightarrow\Psi(y)]\\ &\leftrightarrow\forall y[y=\varrho(\ulcorner\alpha(x)\urcorner)\rightarrow\Psi(y)]\end{aligned}$$

$$\leftrightarrow \forall y[y = \ulcorner\theta_\alpha\urcorner \rightarrow \Psi(y)]$$
$$\leftrightarrow \Psi(\ulcorner\theta_\alpha\urcorner).$$

(E): The existential argument goes as follows:
Let $\eta(x) = \exists y[\boldsymbol{\sigma}(x,y) \wedge \Psi(y)]$ and $\theta_\eta = \eta(\ulcorner\eta(x)\urcorner/x)$. Then we have

$$\begin{aligned}\theta_\eta &= \exists y[\boldsymbol{\sigma}(\ulcorner\eta(x)\urcorner, y) \wedge \Psi(y)] \\ &\leftrightarrow \exists y[y = \boldsymbol{\varrho}(\ulcorner\eta(x)\urcorner) \wedge \Psi(y)] \\ &\leftrightarrow \exists y[y = \ulcorner\theta_\eta\urcorner \wedge \Psi(y)] \\ &\leftrightarrow \Psi(\ulcorner\theta_\eta\urcorner).\end{aligned}$$

Let us note that the constructed sentences θ_α and θ_η, as fixed-points of $\Psi(y)$, are not necessarily equivalent with each other: if $\Psi(y)$ states that "the expression with the Gödel code y begins with a universal quantifier," then θ_α is true while θ_η is not; whence, $\theta_\alpha \leftrightarrow \theta_\eta$ does not hold.

Apart from the syntactic and qualitative differences, the classical proof and the proof of Theorem 2.3 differ from the constructivity point of view. The classical proof is constructive, that is, for a given $\Psi(x)$, it constructs a sentence θ such that $\theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)$ holds, but the proof of Theorem 2.3 is not; it only shows the mere existence of some $m,n \in \mathbb{N}$ such that the equivalence $\boldsymbol{\delta}(\overline{m},\overline{n}) \leftrightarrow \Psi(\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner)$ holds. The proof does not determine for which numbers $m,n \in \mathbb{N}$ this equivalence holds. As noted by a referee of this *Bulletin*, every sentence θ is equivalent with some $\boldsymbol{\delta}(\overline{m},\overline{n})$: take $\varphi(x) = [\theta \leftrightarrow x = \ulcorner\theta\urcorner]$, and let $m = \ulcorner\varphi(x)\urcorner$ and $n = \ulcorner\theta\urcorner$. Then $\theta \leftrightarrow \forall x[\theta \leftrightarrow (x = \ulcorner\theta\urcorner \leftrightarrow x = \ulcorner\theta\urcorner)]$, and so by the associativity of $\leftrightarrow$, we have $\theta \leftrightarrow \forall x[(\theta \leftrightarrow x = \ulcorner\theta\urcorner) \leftrightarrow x = \ulcorner\theta\urcorner]$, or equivalently, $\theta \leftrightarrow \forall x[\varphi(x) \leftrightarrow x = \ulcorner\theta\urcorner] = \boldsymbol{\delta}(\overline{m},\overline{n})$. It is worth noting that then the classical proof of the diagonal lemma proves Theorem 2.3 for extensional formulas $\Psi(x)$, that is, for formulas $\Psi(x)$ with the property that $\zeta \leftrightarrow \xi$ implies $\Psi(\ulcorner\zeta\urcorner) \leftrightarrow \Psi(\ulcorner\xi\urcorner)$, for all sentences ζ, ξ.

We should also point out that the Parametric Length Diagonalization Lemma [17, Lemma 3.2] implies Theorem 2.3 for arbitrary formulas $\Psi(x)$: there exists a formula $\mathsf{B}'(x)$ such that $\mathsf{B}'(x) \leftrightarrow \boldsymbol{\beta}^{<\mathfrak{s}(\|\mathsf{B}'(x)\|)}(x)$ is true, where $\|\zeta\|$ denotes the length of ζ. If we let $\mathfrak{b}' \in \mathbb{N}$ to be the least number that is not definable by any formula with length less than $\mathfrak{s}(\|\mathsf{B}'(x)\|)$, then the proof of Theorem 2.3 goes through with $\mathsf{B}'(x)$ and $\mathfrak{b}'$ in the place of $\mathsf{B}(x)$ and $\mathfrak{b}$, respectively (and $\mathfrak{s}(\|\mathsf{B}'(x)\|)$ in the place of $5\boldsymbol{\ell}$). Even this proof is nonconstructive; for extensional $\Psi(x)$, one can constructively find some $m,n \in \mathbb{N}$ such that $\boldsymbol{\delta}(\overline{m},\overline{n}) \leftrightarrow \Psi(\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner)$ holds, by first finding a fixed-point θ of $\Psi(x)$ by the classical proof, and then finding some $m,n \in \mathbb{N}$ such that $\theta \leftrightarrow \boldsymbol{\delta}(\overline{m},\overline{n})$ holds. For nonextensional formulas $\Psi(x)$, we do not know yet if there is a constructive way of finding some $m,n \in \mathbb{N}$ such that $\boldsymbol{\delta}(\overline{m},\overline{n}) \leftrightarrow \Psi(\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner)$ holds.

2.3. The syntactic form of the diagonal lemma. The syntactic version of the diagonal lemma asserts the existence of a sentence θ for a given formula $\Psi(x)$, with the only free variable x, such that $\theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)$ is (true and also) provable in a (sound) weak arithmetical theory, such as Robinson's Arithmetic $\mathcal{Q}$ (see [30]). Here, we assume familiarity with $\mathcal{Q}$, the notion of

Σ_1-formula, and the fact that $\mathcal{Q}$ is Σ_1-complete (i.e., can prove all the true Σ_1-sentences) and can, therefore, prove the Σ_0-sentence $\forall x<\overline{n}\,(\bigvee\!\!\!\bigvee_{i<n} x=\overline{i})$ for any $n\in\mathbb{N}$; see, for example, [29, Chapters 10 and 11]. As a result, a version of the Pigeonhole Principle $\forall\{x_i<\overline{n}\}_{i\leqslant n}\big(\bigvee\!\!\!\bigvee_{i<j\leqslant n} x_i=x_j\big)$ can be proved in $\mathcal{Q}$. Let us note that $\mathsf{F}_1(u)$ and $\mathsf{L}^{<v}(u)$, in Definition 2.1, can be written as Σ_1-formulas. Since $\mathsf{F}_1(\ulcorner\mathsf{B}\urcorner)$ and $\mathsf{L}^{<\overline{5}\cdot\overline{\ell}}(\ulcorner\mathsf{B}\urcorner)$ are true Σ_1-sentences, then they are both provable in $\mathcal{Q}$.

THEOREM 2.5 (The syntactic diagonal lemma). *For any given formula $\Psi(x)$ with the only free variable x, we have*

$$\mathcal{Q}\vdash \bigvee\!\!\!\bigvee_{i,j\leqslant\hbar}^{\mathsf{F}_1(i)} \big[\boldsymbol{\delta}(\overline{i},\overline{j}) \leftrightarrow \Psi\big(\ulcorner\boldsymbol{\delta}(\overline{i},\overline{j})\urcorner\big)\big],$$

where $\hbar=1+\max\{\ulcorner\varphi\urcorner \mid \mathsf{F}_1(\ulcorner\varphi\urcorner)\wedge\mathsf{L}^{<\overline{5}\cdot\overline{\ell}}(\ulcorner\varphi\urcorner)\}$.

PROOF. If $\mathcal{Q}\nvdash \bigvee\!\!\!\bigvee_{i,j\leqslant\hbar}^{\mathsf{F}_1(i)}\big[\boldsymbol{\delta}(\overline{i},\overline{j}) \leftrightarrow \Psi\big(\ulcorner\boldsymbol{\delta}(\overline{i},\overline{j})\urcorner\big)\big]$, then the following theory should be consistent:

$$\mathcal{Q}'=\mathcal{Q}+\bigwedge\!\!\!\bigwedge_{i,j\leqslant\hbar}^{\mathsf{F}_1(i)} \big[\boldsymbol{\delta}(\overline{i},\overline{j}) \leftrightarrow \neg\Psi\big(\ulcorner\boldsymbol{\delta}(\overline{i},\overline{j})\urcorner\big)\big].$$

We reach a contradiction by proving the inconsistency of the theory $\mathcal{Q}'$. We show that $\mathcal{Q}'\vdash\neg\mathsf{B}(\overline{n})$ holds for each $n\leqslant\hbar$. Reason inside $\mathcal{Q}'$:

> If $\mathsf{B}(\overline{n})$, then $\boldsymbol{\beta}^{<\overline{5}\cdot\overline{\ell}}(\overline{n})$ and so we have $\neg\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(\overline{n})$ and $\forall w<\overline{n}\,\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(w)$. For any x, if $\mathsf{B}(x)$ holds, then both $\neg\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(x)$ and $\forall w<x\,\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(w)$ should hold as well. Now, either $x<\overline{n}$ or $x=\overline{n}$ or $\overline{n}<x$. If $x<\overline{n}$, then we have a contradiction by $\forall w<\overline{n}\,\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(w)$ and $\neg\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(x)$; and if $\overline{n}<x$, then we have a contradiction by $\neg\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(\overline{n})$ and $\forall w<x\,\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(w)$. So, $x=\overline{n}$. Thus, $\forall x(\mathsf{B}(x)\leftrightarrow x=\overline{n})$ holds, or equivalently we have $\boldsymbol{\delta}(\ulcorner\mathsf{B}\urcorner,\overline{n})$, and so $\neg\Psi\big(\ulcorner\boldsymbol{\delta}(\ulcorner\mathsf{B}\urcorner,\overline{n})\urcorner\big)$. Therefore, by $\mathsf{F}_1(\ulcorner\mathsf{B}\urcorner)$ and $\mathsf{L}^{<\overline{5}\cdot\overline{\ell}}(\ulcorner\mathsf{B}\urcorner)$, we have $\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(\overline{n})$ and this is a contradiction with $\mathsf{B}(\overline{n})$.

As a consequence, for each $n\leqslant\hbar$, we have $\mathcal{Q}'\vdash\neg\boldsymbol{\beta}^{<\overline{5}\cdot\overline{\ell}}(\overline{n})$, and so we have that $\mathcal{Q}'\vdash\forall w<\overline{n}\,\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(w)\rightarrow\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(\overline{n})$. Whence, by induction on $n\leqslant\hbar$, one can show that $\mathcal{Q}'\vdash\mathsf{D}^{<\overline{5}\cdot\overline{\ell}}(\overline{n})$. Reason inside $\mathcal{Q}'$:

> For each $n\leqslant\overline{\hbar}$, we have $\mathsf{F}_1(\ulcorner\varphi_n\urcorner)\wedge\mathsf{L}^{<\overline{5}\cdot\overline{\ell}}(\ulcorner\varphi_n\urcorner)\wedge\neg\Psi\big(\ulcorner\boldsymbol{\delta}(\ulcorner\varphi_n\urcorner,\overline{n})\urcorner\big)$ for some φ_n. For each such φ_n, we have $\ulcorner\varphi_n\urcorner<\overline{\hbar}$, and so by the pigeonhole principle, there should exist some $i<j\leqslant\overline{\hbar}$ such that $\varphi_i=\varphi_j$. By $\neg\Psi\big(\ulcorner\boldsymbol{\delta}(\ulcorner\varphi_i\urcorner,\overline{i})\urcorner\big)$ and $\neg\Psi\big(\ulcorner\boldsymbol{\delta}(\ulcorner\varphi_j\urcorner,\overline{j})\urcorner\big)$, we should have $\boldsymbol{\delta}(\ulcorner\varphi_i\urcorner,\overline{i})$ and $\boldsymbol{\delta}(\ulcorner\varphi_j\urcorner,\overline{j})$, which, with $\varphi_i=\varphi_j$, imply that $\overline{i}=\overline{j}$.

Finally, since for any $i<j$, we have $\mathcal{Q}\vdash\overline{i}\neq\overline{j}$, then $\mathcal{Q}'$ is inconsistent. ⊣

This weaker form of the syntactic diagonal lemma can still be used to prove the Gödel–Rosser incompleteness theorem, by noting that complete theories have the disjunction property; that is, if T is complete and $T \vdash A \vee B$, then either $T \vdash A$ or $T \vdash B$. So, the standard incompleteness proofs of Gödel and Rosser should work as usual with Theorem 2.5 too.

COROLLARY 2.6 (Gödel–Rosser's Theorem). *Any recursively enumerable and consistent extension of $\mathcal{Q}$ is incomplete.*

Theorem 2.5 can also be used to prove a syntactic version of Tarski's theorem which states the undefinability of arithmetical truth:

COROLLARY 2.7 (Syntactic form of Tarski's Theorem). *If a theory T extends $\mathcal{Q}$, then for any formula $\Phi(x)$ with the only free variable x, the theory T is not consistent with the set $\{\xi \leftrightarrow \Phi(\ulcorner \xi \urcorner) \mid \xi$ is a sentence$\}$.*

§3. Conclusion. Alternative proofs for theorems strengthen our confidence in them and increase our understanding of them as well (see, e.g., [11] or its prequel [10]). In this short note, we presented a "different sort of reason" (in the words of Boolos [3]) for the semantical version of the diagonal lemma, and gave an alternative proof for a weak form of the syntactical diagonal lemma. Let us note that our alternative proof is nonconstructive, while the classical proof is constructive; and the new proof could seem even trickier than the classical one to some eyes. Of course, the problem of giving an entirely different proof for the following (stronger) syntactic diagonal lemma remains open:

"For a given $\Psi(x)$, there is a sentence θ such that $\mathcal{Q} \vdash \theta \leftrightarrow \Psi(\ulcorner \theta \urcorner)$."

Acknowledgments. This research is supported by the Office of the Vice Chancellor for Research and Technology, University of Tabriz, Iran.

I am grateful to two anonymous referees and an editor (Ali Enayat) of this *Bulletin* for their valuable comments and suggestions, which highly improved the presentation of the results. Some texts of this paper are the exact wordings of the editor and the two referees. Warm thanks go to my colleagues Kaave Lajevardi and Asadollah Fallahi for stimulating discussions and wholehearted encouragements.

REFERENCES

[1] J. BARWISE, *Editorial notes: This month's column*. ***Notices of the American Mathematical Society***, vol. 36 (1989), no. 4, p. 388.

[2] G. S. BOOLOS, *A new proof of the Gödel incompleteness theorem*. ***Notices of the American Mathematical Society***, vol. 36 (1989), no. 4, pp. 388–390. Reprinted in G. Boolos, ***Logic, Logic, and Logic***, Harvard University Press, Cambridge, MA, 1998, pp. 383–388.

[3] ———, *A letter from George Boolos*. ***Notices of the American Mathematical Society***, vol. 36 (1989), no. 6, p. 676.

[4] G. S. BOOLOS, J. P. BURGESS and R. C. JEFFREY, ***Computability and Logic***, fifth ed., Cambridge University Press, Cambridge, 2007.

[5] S. BUSS, *First-order proof theory of arithmetic*, ***Handbook of Proof Theory*** (S. Buss, editor), Elsevier Science, Amsterdam, 1998, pp. 79–147.

[6] X. Caicedo, *La paradoja de berry o la indefinibilidad de la definibilidad y las limitaciones de los formalismos*. ***Lecturas Matemáticas***, vol. 14 (1993), no. 1, pp. 37–48 (in Spanish). A 2004 revision (with some additional notes) is available at https://bit.ly/2XfecGR

[7] R. Carnap, ***Logische Syntax der Sprache***, Springer, New York, 1934. Translated into English by A. Smeaton (Countess von Zeppelin) as ***The Logical Syntax of Language,*** Kegan Paul, Trench, Trubner, 1937.

[8] G. J. Chaitin, *Computational complexity and Gödel's incompleteness theorem*. ***SIGACT News,*** vol. 9 (1971), pp. 11–12. ***Notices of the American Mathematical Society***, vol. 17 (1970), no. 6, p. 672, abstract.

[9] ———, ***The Unknowable***, Springer, New York, 1999.

[10] J. W. Dawson Jr., *Why do mathematicians re-prove theorems?*. ***Philosophia Mathematica***, vol. 14 (2006), no. 3, pp. 269–286.

[11] ———, ***Why Prove It Again? Alternative Proofs in Mathematical Practice***, Birkhäuser, Basel, 2015.

[12] H. Gaifman, *Naming and diagonalization, from cantor to Gödel to Kleene*. ***Logic Journal of the IGPL***, vol. 14 (2006), no. 5, pp. 709–728.

[13] K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I*. ***Monatshefte für Mathematik und Physik***, vol. 38 (1931), no. 1, pp. 173–198. English translation in [15, pp. 135–152].

[14] ———, *On undecidable propositions of formal mathematical systems*, ***Mimeographed Lecture Notes*** (taken by S. C. Kleene & J. B. Rosser), Institute for Advanced Study, Princeton University Press, 1934. Reprinted, with corrections and modifications, in [15, pp. 346–371].

[15] ———, ***Collected Works, vol. I, 1929–1936*** (S. Feferman, J. W. Dawson Jr., S. C. Kleene, G. H. Moore, R. M. Solovay and J. van Heijenoort, editors), Oxford University Press, Oxford, 1986.

[16] R. G. Jeroslow, *Redundancies in the Hilbert-Bernays derivability conditions for Gödel's second incompleteness theorem*. ***Journal of Symbolic Logic***, vol. 38 (1973), no. 3, pp. 359–367.

[17] M. Kikuchi, T. Kurahashi and H. Sakai, *On proofs of the incompleteness theorems based on Berry's Paradox by Vopěnka, Chaitin, and Boolos*. ***Mathematical Logic Quarterly***, vol. 58 (2012), no. 4–5, pp. 307–316.

[18] S. C. Kleene, *General recursive functions of natural numbers*. ***Mathematische Annalen***, vol. 112 (1936), no. 1, pp. 727–742.

[19] ———, *A symmetric form of Gödel's theorem*. ***Indagationes Mathematicae***, vol. 12 (1950), pp. 244–246.

[20] H. Kotlarski, *The incompleteness theorems after 70 years*. ***Annals of Pure and Applied Logic***, vol. 126 (2004), no. 1–3, pp. 125–138.

[21] S. Kritchman and R. Raz, *The surprise examination paradox and the second incompleteness theorem*. ***Notices of the American Mathematical Society***, vol. 57 (2010), no. 11, pp. 1454–1458.

[22] V. McGee, ***The first incompleteness theorem.*** Handouts of the course Logic II, 2002. Available at http://web.mit.edu/24.242/www/1stincompleteness.pdf

[23] Y. N. Moschovakis, *Kleene's amazing second recursion theorem*. ***Bulletin of Symbolic Logic***, vol. 16 (2010), no. 2, pp. 189–239.

[24] J. B. Rosser, *Extensions of some theorems of Gödel and Church*. ***Journal of Symbolic Logic***, vol. 1 (1936), no. 3, pp. 87–91.

[25] B. Russell, *Mathematical logic as based on the theory of types*. ***American Journal of Mathematics***, vol. 30 (1908), no. 3, pp. 222–262.

[26] S. Salehi and P. Seraji, *On constructivity and the Rosser property: A closer look at some Gödelean proofs*. ***Annals of Pure and Applied Logic***, vol. 169 (2018), no. 10, pp. 971–980.

[27] G. Serény, *Boolos-style proofs of limitative theorems*. ***Mathematical Logic Quarterly***, vol. 50 (2004), no. 2, pp. 211–216.

[28] ———, *The diagonal lemma as the formalized Grelling paradox*, ***Collegium Logicum 9, Gödel Centenary 2006*** (M. Baaz and N. Preining, editors), Kurt Gödel Society, Vienna, 2006, pp. 63–66.

[29] P. Smith, ***An Introduction to Gödel's Theorems***, second ed., Cambridge University Press, Cambridge, 2013.

[30] A. Tarski, (in collaboration with A. Mostowski and R. M. Robinson), ***Undecidable theories,*** North–Holland, 1953. Reprinted by Dover Publications (2010).

RESEARCH INSTITUTE FOR FUNDAMENTAL SCIENCES
UNIVERSITY OF TABRIZ, 29 BAHMAN BOULEVARD
P.O. BOX 51666-16471, TABRIZ, IRAN
E-mail: salehipour@tabrizu.ac.ir
URL: www.SaeedSalehi.ir