

A Reunion of Gödel, Tarski, Carnap, and Rosser

Saeed Salehi

Department of Mathematics, Statistics, and Computer Science, University of Tabriz,
Bahman 29th Boulevard, P.O.Box 51666–16471, Tabriz, IRAN. E-mail: salehipour@tabrizu.ac.ir

Abstract

We unify Gödel’s First Incompleteness Theorem (1931), Tarski’s Undefinability Theorem (1933), Gödel-Carnap’s Diagonal Lemma (1934), and Rosser’s (strengthening of Gödel’s first) Incompleteness Theorem (1936), whose proofs resemble much and use almost the same technique.

Keywords: Gödel’s First Incompleteness Theorem, Tarski’s Undefinability Theorem, Carnap’s Diagonal Lemma, Rosser’s Incompleteness Theorem, Chaitin’s Proof of the Incompleteness Theorem.

2020 AMS MSC: 03F40.

1 Introduction

Between 1930 and 1936, at the beginning of the Golden Age of Mathematical Logic, there appeared four fundamental theorems:

1. Gödel’s First Incompleteness Theorem in 1931; see [5].
2. (Gödel-)Tarski’s Truth-Undefinability Theorem in 1933; see [8] and [6, f. 25 on p. 363].
3. (Gödel-)Carnap’s Diagonal Lemma in 1934; see [11] and [6, f. 23 on p. 363].
4. (Gödel-)Rosser’s Incompleteness Theorem in 1936; see [9] and [6, p. 370].

A main part of the classic proofs of these theorems uses a common trick that constructs some suitable self-referential sentences [1]. So, it is natural to conjecture that these theorems are equivalent, in the sense that there is a sufficiently general framework in which either all these four theorems hold together, or none holds. This paper is a continuation of [11] where some semantic forms of 1–3 were proved to be equivalent. Here, we present some syntactic formulations of 1–4 and prove their equivalence (in §2); we also provide a framework in which none of 1–4 holds (it is too well known that they all hold for sufficiently strong theories). Having this equivalence has the advantage that one can translate a proof for any of 1–4 to get an alternative proof for another. This was done earlier in [10] where an alternative proof for the semantic Diagonal Lemma, and a weak syntactic formulation of it, was provided. Here, we will also answer a question left open there (in §4, the Appendix) and will see one more different proof for the weak syntactic diagonal lemma (in §3).

[1] One other basic result around that time (1938) which uses quite a similar technique was Kleene’s Recursion Theorem [7]; we do not consider it here, as we know of no equivalent formulation in the form of Theorem [2] below.
2 A Unification of the Four Theorems

Let us begin with a definition for the standard notion of Gödel coding, and fix our language.

**Definition 2.1** (Gödel coding, arithmetical languages, interpretation, representing numbers by terms):
For a first-order language \( \mathcal{L} \), a Gödel coding on \( \mathcal{L} \) is a computable injection from the syntactical expressions (finite strings) over \( \mathcal{L} \) into the set of natural number \( \mathbb{N} \).

Let \( \mathcal{L}^* = \{0, 1, \langle, \rangle, +, \nu\} \) be the first-order language that contains the constant symbols 0, 1, the binary relation symbols \( \langle \), the binary function symbol +, and the unary function symbol \( \nu \). Let the language of arithmetic be \( \{0, 1, \langle, \rangle, +, \times\} \), where \( \times \) is a binary function symbol.

The symbols 0, 1, \( \langle, \rangle, +, \times \) are interpreted as usual over \( \mathbb{N} \), and the interpretation of \( \nu \), which is a function \( \nu: \mathbb{N} \rightarrow \mathbb{N} \), depends on a fixed Gödel coding: \( \nu(n) \) is the code of \( \neg \sigma \) (the negation symbol \( \neg \) appended to the left side of \( \sigma \)) when \( n \) is the code of a sentence \( \sigma \) (and \( \nu(n) \) can be any arbitrary number when \( n \) is not the code of any sentence).

For a natural number \( n \in \mathbb{N} \), let \( \pi \) denote the closed term that represents \( n \), which is 0 if \( n = 0 \), is 1 if \( n = 1 \), and is \( 1 + \cdots + 1 \) (\( n \) times) if \( n > 1 \).

So, if \( \overline{\sigma} \) denotes the code of a sentence \( \sigma \), then \( \nu(\overline{\sigma}) = \overline{\neg \sigma} \).

**Definition 2.2** (arithmetical theories):
Let \( Q \) denote Robinson’s Arithmetic over the language of arithmetic.

Fix a Gödel coding \( \eta \mapsto \overline{\eta} \); for simplicity, let us denote the closed term \( \overline{\eta} \) by \( \overline{\eta} \). Let \( Q^- \) be the \( \mathcal{L}^* \)-theory axiomatized by

\[
\begin{align*}
(A_1): & \quad \forall x (x < \overline{\pi} \lor x = \overline{\pi} \lor \overline{\pi} < x), \text{ for every } n \in \mathbb{N}.
(A_2): & \quad \forall x (x < \overline{\pi} \leftrightarrow \bigwedge_{i < n} x = \overline{i}), \text{ for every } n \in \mathbb{N}.
(A_3): & \quad \nu(\overline{\sigma \pi}) = \overline{\neg \sigma}, \text{ for every } \mathcal{L}^*-\text{sentence } \sigma.
\end{align*}
\]

**Theorem 2.3** (Gödel\(_T\) \( \equiv \) Tarski\(_T\) \( \equiv \) Carnap\(_T\) \( \equiv \) Rosser\(_T\) ):
For every theory \( T \) that extends \( Q^- \), and a fixed coding, the following are equivalent:

1. **Gödel\(_T\)**: If \( U \) is a consistent extension of \( T \) such that for some formula \( \Psi(x) \), with the only variable \( x \), we have \( U \vdash \sigma \) iff \( U \vdash \Psi(\overline{\sigma \pi}) \) for every sentence \( \sigma \), then \( U \) is incomplete.
2. **Tarski\(_T\)**: For every formula \( \Upsilon(x) \), with the only free variable \( x \), the theory \( T \) is inconsistent with the set \( \{ \Upsilon(\overline{\sigma \pi}) \leftrightarrow \sigma \mid \sigma \text{ is a sentence}\} \).
3. **Carnap\(_T\)**: For every formula \( \Lambda(x) \), with the only free variable \( x \), there are finitely many sentences \( \{A_i\}_i \) such that \( T \vdash \bigwedge_i (\Lambda(\overline{A_i \pi}) \leftrightarrow A_i) \).
4. **Rosser\(_T\)**: If \( U \) is a consistent extension of \( T \) such that for some formula \( \Theta(x, y) \), with the shown free variables, we have for every sentence \( \sigma \) that (i) if \( U \vdash \sigma \) then \( U \vdash \Theta(\overline{\pi}, \overline{\sigma \pi}) \) for some number \( m \in \mathbb{N} \), and (ii) if \( U \nvdash \sigma \) then \( U \vdash \neg \Theta(\overline{\pi}, \overline{\sigma \pi}) \) for every number \( n \in \mathbb{N} \), then \( U \) is incomplete.

\( \diamond \)
Before proving the theorem, let us explain its content a bit.

(1) In Gödel$_T$, the formula $\Psi(x)$ is a kind of provability predicate, in the sense that if $U$ proves $\sigma$, then $U$ can verify that it proves $\sigma$; and conversely (here $U$ is assumed to possess some soundness) if $U$ proves that $\sigma$ is $U$-provable, then $\sigma$ is $U$-provable in reality.

(2) In Tarski$_T$, the formula $\Upsilon(x)$ is a kind of hypothetical truth predicate, and our version of the theorem is syntactic (Tarski’s theorem is usually formulated semantically, in the form that the truth predicate is not arithmetically definable; cf. [11]).

(3) In Carnap$_T$, the existence of finitely many (partial) fixed points for a given formula $\Lambda(x)$ has been claimed. This is a weak version of the syntactic Diagonal Lemma (see [10, Theorem 2.5]); the strong diagonal lemma states the existence of one fixed point sentence ($A$ such that $Q \vdash \Lambda(\#A) \leftrightarrow A$). Let us note that our weak version implies the Semantic Diagonal Lemma (see e.g. [10, Theorem 2.3] or [11, Definition 2.1]).

(4) In Rosser$_T$, the formula $\Theta(x, y)$ is a kind of proof predicate: $x$ codes a $U$-proof of $y$. The assumptions (i) and (ii) indicate that $U$-proofs are bi-representable in $U$: if $U$ proves $\sigma$ and $m$ is the code of its proof, then $U$ verifies this; and if $\sigma$ is not $U$-provable, then $U$ verifies that no number can code a $U$-proof of $\sigma$.

**Proof:**

(1 $\implies$ 2): If Tarski$_T$ does not hold, then let $\mathfrak{M}$ be a model of $T + \{ \Upsilon(\#\sigma) \leftrightarrow \sigma \mid \sigma$ is a sentence $\}$, and put $U = \text{Th}(\mathfrak{M})$. Now, for every sentence $\sigma$ we have $U \vdash \sigma$ iff $\mathfrak{M} \models U(\#\sigma)$ iff $U \vdash \Upsilon(\#\sigma)$. But $U$ is a complete extension of $T$; this contradicts Gödel$_T$.

(2 $\implies$ 3): For given $\Lambda(x)$, let $\Upsilon(x) = \neg \Lambda(x)$. By the inconsistency of the set $T + \{ \Upsilon(\#\sigma) \leftrightarrow \sigma \mid \sigma$ is a sentence $\}$, we have $T \vdash \neg \bigwedge_i \{ \Upsilon(\#A_i) \leftrightarrow A_i \}$ for some finitely many sentences $\{A_i\}_i$. By the propositional tautology $\neg \bigwedge_i (p_i \leftrightarrow q_i) \equiv \bigvee_i (p_i \leftrightarrow q_i)$, we have $T \vdash \bigvee_i (\Lambda(\#A_i) \leftrightarrow A_i)$.

(3 $\implies$ 4): Let $U$ and $\Theta$ satisfy the assumptions, and assume, for the sake of a contradiction, that $U$ is a complete theory. Let $\Lambda(x)$ be the formula $\forall y [\Theta(y, x) \rightarrow \exists z < y \Theta(z, \nu(x))]$. By Carnap$_T$, there are finitely many sentences $\{A_i\}_i$ such that $T \vdash \bigvee_i (\Lambda(\#A_i) \leftrightarrow A_i)$. Since complete theories have the disjunction property, then there exists one sentence $\rho$ such that (4) $U \vdash \Lambda(\#\rho) \leftrightarrow \rho$. Also, by the completeness of $U$ we have either (I) $U \vdash \rho$, or (II) $U \vdash \neg \rho$.

(I) If $U \vdash \rho$, then by (4) and (A3, Definition 2.2) we have $U \vdash \forall y [\Theta(y, \#\rho) \rightarrow \exists z < y \Theta(z, \#\neg \rho)]$. By (i) in the assumption, there is some $m \in \mathbb{N}$ such that $U \vdash \Theta(\#m, \#\rho)$. So, $U \vdash \exists z < m \Theta(z, \#\neg \rho)$. On the other hand, by $U \vdash \neg \rho$ and (ii) in the assumption, we have $U \vdash \neg \Theta(\#m, \#\neg \rho)$ for each $n \in \mathbb{N}$. Thus, by (A2, Definition 2.2) we have $U \vdash \forall z < m \neg \Theta(z, \#\neg \rho)$. Whence, $U$ is inconsistent; a contradiction.

(II) If $U \vdash \neg \rho$, then by (4) and (A3, Definition 2.2) we have $U \vdash \exists y [\Theta(y, \#\rho) \land \forall z < y \neg \Theta(z, \#\neg \rho)]$. By (i) in the assumption, there is some $m \in \mathbb{N}$ such that $U \vdash \Theta(\#m, \#\rho)$. So, (A1, Definition 2.2) implies that $U \vdash \exists y \leq m \Theta(y, \#\rho)$. On the other hand, by $U \not\vdash \rho$ and (ii) in the assumption, we have $U \vdash \neg \Theta(\#m, \#\rho)$ for each $n \in \mathbb{N}$. Thus, by (A2, Definition 2.2) we have $U \vdash \forall y \leq m \neg \Theta(y, \#\rho)$; a contradiction.

(4 $\implies$ 1): If Gödel$_T$ does not hold for the theory $U$ and formula $\Psi(x)$, let $\Theta(x, y)$ be $\Psi(y) \land (x = x)$. Now, by the assumption, for every sentence $\sigma$ we have (i) if $U \vdash \sigma$, then $U \vdash \Psi(\#\sigma)$, thus $U \vdash \Theta(\#m, \#\sigma)$ for every $m \in \mathbb{N}$. Also, since $U$ is (assumed to be) complete, then by the assumption we have (ii) if $U \not\vdash \sigma$, then $U \not\vdash \Psi(\#\sigma)$, so $U \vdash \neg \Theta(\#n, \#\sigma)$ for every $n \in \mathbb{N}$. This contradicts Rosser$_T$. \(\square\)

For the theorem to make sense, we should demonstrate a framework in which none of Gödel$_T$, Tarski$_T$, Carnap$_T$, or Rosser$_T$ holds. Let us fix a coding, which is due to Ackermann (1937), cf. [2 Example 7].
**Definition 2.4** (Ackermann (1937) coding):

The number 0 is not the code of anything; number 1 is the code of the empty string; and every symbol is coded by an odd number, in particular 3 is the code of \( \neg \). Code the finite string \( \langle a_1, \ldots, a_\ell \rangle \) by \( \sum_{k=1}^\ell 2^{\sum_{j=1}^k (a_j+1) = \sum_{k=1}^\ell 2^{a_1+1} + \ldots + 2^{a_\ell+1}} \).

\[ \Diamond \]

Let us note that this coding is computable and injective (cf. Definition 2.1), but not surjective (for example, 4 is not the code of anything).

**Lemma 2.5** (the code of negation):

If \( \eta \mapsto \llbracket \eta \rrbracket \) is Ackermann’s coding in Definition 2.4 then for every string \( \eta \) we have \( \llbracket \neg \eta \rrbracket = 16(1+\llbracket \eta \rrbracket) \).

**Proof:**

For \( \eta = \langle a_1, \cdots, a_\ell \rangle \) we have \( \llbracket \langle \neg, a_1, \cdots, a_\ell \rangle \rrbracket = 2^4 + \sum_{k=1}^\ell 2^{4+\sum_{j=1}^k (a_j+1)} = 16(1+\llbracket \eta \rrbracket) \).

**Definition 2.6** (a new framework):

Define the function \( \nu^* : \mathbb{N} \to \mathbb{N} \) by \( \nu^*(n) = 33 + 16n \) when \( n \) is even, and \( \nu^*(n) = 16 + 16n \) when \( n \) is odd. Let \( \mathcal{M}^* \) be the structure \( \langle \mathbb{N}; 0, 1, <, +, \nu \rangle \) where \( \nu \) is interpreted as \( \nu^* \) above.

Put \( T^* = Th(\mathcal{M}^*) \), and let \( \Psi^*(x) = \forall y (x = y + y) \), \( \Lambda^*(x) = \neg \Psi^*(x) \), and \( \Theta^*(x, y) = \Psi^*(y) \land (x = x) \).

Denote Ackermann’s coding in Definition 2.4 by \( \eta \mapsto \llbracket \eta \rrbracket \). Let \( \eta \mapsto \llbracket \eta \rrbracket \) be a new coding on \( \mathcal{L}^* \) (see Definition 2.1) defined as follows: \( \llbracket \eta \rrbracket \) is \( \mathcal{M}^* \)-true \( \mathcal{L}^* \)-sentence; and is \( 1+2\llbracket \eta \rrbracket \), otherwise. Let \( \lfloor E \rfloor \) denote the closed term \( \llbracket E \rrbracket \).

\[ \Diamond \]

Since \( \eta \mapsto \llbracket \eta \rrbracket \) (in Definition 2.4) is injective, the new coding \( \eta \mapsto \llbracket \eta \rrbracket \) (in Definition 2.6) is injective too. For its computability, we note that \( \mathcal{M}^* = \langle \mathbb{N}; 0, 1, <, +, \nu \rangle \) is decidable since \( \nu^* \) is \( \{1, +\} \)-definable and \( \langle \mathbb{N}; 0, 1, <, + \rangle \) is decidable by Presburger’s Theorem (cf. [2] Theorem 3.3). Thus, \( \eta \mapsto \llbracket \eta \rrbracket \) is computable as well (cf. Definition 2.1). We show that \( \nu^* \) calculates the negation of sentences in the new coding’s setting:

**Lemma 2.7** (\( \nu^* \) is the negation mapping of sentences in the setting of \( \llbracket \cdot \rrbracket \)):

For every \( \mathcal{L}^* \)-sentence \( \sigma \), we have \( \nu^*(\llbracket \sigma \rrbracket) = \llbracket \neg \sigma \rrbracket \).

**Proof:**

The sentence \( \sigma \) is either \( \mathcal{M}^* \)-true or \( \mathcal{M}^* \)-false. If \( \mathcal{M}^* \models \sigma \), then \( \llbracket \sigma \rrbracket = 2\llbracket \sigma \rrbracket \) and \( \neg \sigma \) is not \( \mathcal{M}^* \)-true. So, \( \llbracket \sigma \rrbracket \) is even, and by Lemma 2.5 we have \( \nu^*(\llbracket \sigma \rrbracket) = 33 + 16\llbracket \sigma \rrbracket = 33 + 32 \llbracket \sigma \rrbracket = 1+2 \cdot 16(1+\llbracket \sigma \rrbracket) = 1+2\llbracket \neg \sigma \rrbracket = \llbracket \neg \sigma \rrbracket \).

If \( \mathcal{M}^* \not\models \sigma \), then \( \llbracket \sigma \rrbracket = 1+2\llbracket \sigma \rrbracket \) and \( \neg \sigma \) is \( \mathcal{M}^* \)-true. So, \( \llbracket \sigma \rrbracket \) is odd, thus by Lemma 2.5 we have \( \nu^*(\llbracket \sigma \rrbracket) = 16 + 16\llbracket \sigma \rrbracket = 32 + 32 \llbracket \sigma \rrbracket = 2 \cdot 16(1+\llbracket \sigma \rrbracket) = 2\llbracket \neg \sigma \rrbracket = \llbracket \neg \sigma \rrbracket \).

\[ \text{QED} \]

Let us note that if the expression \( \eta \) is not a sentence, then Lemma 2.7 does not hold; indeed, in that case \( \nu^*(\llbracket \eta \rrbracket) + 1 = \llbracket \neg \eta \rrbracket \). Now, we have all the ingredients for constructing a case in which none of the equivalent statements Gödel\textsubscript{T}, Tarski\textsubscript{T}, Carnap\textsubscript{T}, Rosser\textsubscript{T} can hold (cf. [11] Remark 2.6).
**Theorem 2.8** (¬GoedelT*, ¬TarskiT*, ¬CarnapT*, ¬RosserT*):

The complete theory $T^*$ contains $Q^-$, and neither GoedelT*, nor TarskiT*, nor CarnapT*, nor RosserT*-holds for $U = T^*$ and, respectively, the formulas $\Psi^*(x), \Upsilon^*(x), \Lambda^*(x)$, and $\Theta^*(x, y)$, in Definition 2.6.

**Proof:**

The complete theory $T^*$ is decidable by Presburger’s theorem and the $\{1, +\}$-definability of $\nu^*$. Thus, the mapping $\eta \mapsto \bot_{\eta_j}$ is a (computable injective) coding. Trivially, the axioms $(A_1, A_2, \text{Definition } 2.2)$ hold in $\mathfrak{M}^*$, and Lemma 2.7 implies that $(A_3, \text{Definition } 2.2)$ holds too. Thus, $T^*$ contains $Q^-$. We now show that none of GoedelT*, TarskiT*, CarnapT*, or RosserT*- holds for $\Psi^*(x), \Upsilon^*(x), \Lambda^*(x)$, and $\Theta^*(x, y)$.

- GoedelT*: For every sentence $\sigma$, $T^* \vdash \sigma$ iff $\mathfrak{M}^* \models \sigma$ iff $\bot_{\sigma_j}$ is even iff $\mathfrak{M}^* \models \Psi^*(\sigma_\parallel)$ iff $T^* \vdash \Psi^*(\sigma_\parallel)$.

- TarskiT*: For every sentence $\sigma$ we have $T^* \vdash \Upsilon^*(\sigma_\parallel) \iff \sigma$ by ¬GoedelT*.

- CarnapT*: For any finitely many sentences $\{A_i\}_i$ we have $T^* \vdash \bigwedge_i (\neg \Lambda^*(\updownarrow_\parallel A_i) \leftrightarrow A_i)$ by ¬TarskiT*; so $T^* \not\vdash \bigwedge_i (\Lambda^*(\updownarrow_\parallel A_i) \leftrightarrow A_i)$, thus $T^* \not\vdash \bigwedge_i (\Lambda^*(\updownarrow_\parallel A_i) \leftrightarrow A_i)$ by the consistency of $T^*$.

- RosserT*: We saw that (i) if $T^* \vdash \sigma$, then $T^* \not\vdash \Psi^*(\sigma_\parallel)$, so $T^* \not\vdash \Theta^*(\overline{m}, \sigma_\parallel)$ for every $m \in \mathbb{N}$. Also note that (ii) if $T^* \not\vdash \sigma$, then $\bot_{\sigma_j}$ is odd, so $T^* \vdash \neg \Psi^*(\sigma_\parallel)$, thus $T^* \vdash \neg \Theta^*(\overline{m}, \sigma_\parallel)$ for every $n \in \mathbb{N}$. QED

3 Tarski’s Theorem and the Weak Syntactic Diagonal Lemma, à la Chaitin

Chaitin’s proof for the first incompleteness theorem appeared in [3]. There are several versions of it now; one was presented in [12, Theorem 3.3]. The proof was adapted for Rosser’s theorem in [12, Theorem 3.9].

Here, we prove Tarski’s undefinability theorem and the (weak syntactic) diagonal lemma of Carnap by the same method.

**Proposition 3.1** (Goedel-Tarski’s Truth-Undefinability Theorem):

For every formula $\Upsilon(x)$, the theory $Q$ is inconsistent with the set $\{ \Upsilon(\updownarrow_\parallel \sigma) \leftrightarrow \sigma \mid \sigma$ is a sentence $\}$.

**Proof:**

Assume not; fix a model $\mathfrak{M}$ of $Q + \{ \Upsilon(\updownarrow_\parallel \sigma) \leftrightarrow \sigma \mid \sigma$ is a sentence $\}$. Suppose that $\varphi_0^\Upsilon, \varphi_1^\Upsilon, \varphi_2^\Upsilon, \cdots$ is an effective enumeration of all the $\Upsilon$-computable (i.e., computable with oracle $\Upsilon$) unary functions. Define the $\Upsilon$-Kolmogorov-Chaitin complexity of $n \in \mathbb{N}$ to be $\mathcal{K}^\Upsilon(n) = \min \{i \mid \varphi^\Upsilon_i(0) = n \}$, the minimum index of the $\Upsilon$-computable function that outputs $n$ on input 0. By Kleene’s Recursion Theorem [7] there exists some $c \in \mathbb{N}$ such that $\varphi^\Upsilon_c(x) = \min z: \Upsilon(\langle \mathcal{K}^\Upsilon(x) > c \rangle)$, where $\langle \mathcal{K}^\Upsilon(x) > y \rangle$ is the arithmetical formula which says that “the $\Upsilon$-Kolmogorov-Chaitin complexity of $x$ is greater than $y$". By the Pigeonhole Principle (a version of which is provable in $Q$, see [12, Lemma 3.8]) there exists some element $u \leq c + 1$ in $\mathfrak{M}$ such that $\mathfrak{M} \models \langle \mathcal{K}^\Upsilon(\uparrow_\parallel u) > c \rangle$; note that $\{ \varphi_0^\Upsilon(0), \varphi_1^\Upsilon(0), \cdots, \varphi^\Upsilon_c(0) \}$ has at most $c + 1$ members, and $\{0, 1, \cdots, c + 1 \}$ has $c + 2$ members. For the least $u \in \mathfrak{M}$ with $\mathfrak{M} \models \langle \mathcal{K}^\Upsilon(\uparrow_\parallel u) > c \rangle$ we have $\mathfrak{M} \models \Upsilon(\langle \mathcal{K}^\Upsilon(\uparrow_\parallel u) > c \rangle)$ and $\mathfrak{M} \models \forall z < \uparrow_\parallel \Upsilon(\langle \mathcal{K}^\Upsilon(\uparrow_\parallel u) > c \rangle); \therefore \mathfrak{M} \models \langle \varphi^\Upsilon_c(\uparrow_\parallel u) \rangle = \uparrow_\parallel u$, thus $\mathfrak{M} \models \langle \mathcal{K}^\Upsilon(\uparrow_\parallel u) \leq c \rangle$, a contradiction. QED

(2) Notice that this very proof implies that the unary function $z \mapsto \mathcal{K}^\Upsilon(z)$ is not $\Upsilon$-computable; though, for every constant $c \in \mathbb{N}$, the function $x \mapsto \min z: \Upsilon(\langle \mathcal{K}^\Upsilon(x) > c \rangle)$ is clearly $\Upsilon$-computable.
The part \((2 \implies 3)\) in the proof of Theorem 2.3 immediately yields the weak syntactic diagonal lemma from the above proof. However, this Chaitin style argument can prove the weak lemma directly.

**Proposition 3.2 (weak syntactic Gödel-Carnap’s Diagonal Lemma):**

For every formula \(\Lambda(x)\) there are finitely many sentences \(\{A_i\}_i\) such that \(\forall i \models (\Lambda(iA) \leftrightarrow A_i)\).

**Proof:**

Put \(\Upsilon(x) = \neg \Lambda(x)\); with the notation of the proof of Proposition 3.1 let \(A_i = \langle \Upsilon(x) > i \rangle\) for \(i \leq c + 1\). If \(Q \not\vdash \forall_{i \leq c+1} (\Lambda(iA) \leftrightarrow A_i)\), then \(Q + \bigwedge_{i < c+1} (\Upsilon(iA) \leftrightarrow A_i)\) is consistent, and so has a model, say, \(\mathfrak{M}\). Now, continue the proof of Proposition 3.1 (after the footnote 2) for reaching to a contradiction. \(\Box\)

## 4 Appendix: Constructivity of an Alternative Proof of the Diagonal Lemma

For \(m, n \in \mathbb{N}\), let \(\delta(m, n)\) say that the formula with code \(m\) has exactly one free variable and defines the number \(n\); that is, if \(\varphi(x)\) is the formula with code \(m\) that has exactly one free variable \(x\), then the statement \(\forall x [\varphi(x) \leftrightarrow x = \overline{n}]\) holds. It was proved in [10] Theorem 2.3 that for every formula \(\Lambda(x)\) there are some \(m, n \in \mathbb{N}\) such that \(\mathbb{N} \models \Lambda(\overline{\delta(m, n)} \neg) \leftrightarrow \delta(m, n)\). The proof was not constructive, it only showed the mere existence of some \(m, n \in \mathbb{N}\) with the above property; it did not determine which \(m, n\). By a suggestion of a referee of [10], for a restricted class of \(\Lambda\) formulas, such \(m, n\) can be found constructively; but it was left open if there exists a constructive way of finding such \(m, n\) for every formula \(\Lambda(x)\). Here, we show that there is such a way, but with a very different method. Actually, the following proof is more similar to the classical one (rather than to the proof of Theorem 2.3 in [10]).

**Theorem 4.1 (strong syntactic Gödel-Carnap’s Diagonal Lemma):**

For every given formula \(\Lambda(x)\) one can effectively find \(m, n \in \mathbb{N}\) such that \(\forall \models \delta(m, n) \leftrightarrow \Lambda(\overline{\delta(m, n)}\overline{\theta})\).

**Proof:**

There is a formula \(\sigma(x, y)\) that strongly represents the diagonal function in \(Q\). That is to say that for every formula \(\alpha(y)\) we have \(\forall x [\sigma(\overline{x}, x) \leftrightarrow x = \overline{\alpha} \overline{\theta}]\), where \(\alpha = \overline{\gamma}\); the sentence \(\alpha(\overline{x})\) is called the diagonal of \(\alpha\). Let \(\zeta(x, y) = [\sigma(y, x) \rightarrow \Lambda(x)] \leftrightarrow (x = y)\) and \(\tau(y) = \forall x [\zeta(x, y) \leftrightarrow x = y]\); put \(n = \overline{\tau}\). Also, let \(\kappa(x) = \zeta(x, \overline{x})\) and put \(m = \overline{\kappa}\). We show that \(\delta(m, n) \leftrightarrow \Lambda(\overline{\delta(m, n)}\overline{\theta})\) is provable in \(Q\). Note that \(\delta(m, n) = \forall x [\kappa(x) \leftrightarrow x = \overline{n}] = \forall x [\zeta(x, \overline{x}) \leftrightarrow x = \overline{n}] = \tau(\overline{n}) = \text{the diagonal of } \tau\). Thus,

\[
Q \models \delta(m, n) \iff \\
\begin{align*}
\forall x [\zeta(x, \overline{x}) & \leftrightarrow x = \overline{n}] \\
& \iff \forall x [\left[ \left( \left[ \sigma(\overline{x}, x) \rightarrow \Lambda(x) \right] \leftrightarrow x = \overline{n} \right] \leftrightarrow x = \overline{n} \right] \\
& \iff \forall x [\left[ \sigma(\overline{x}, x) \rightarrow \Lambda(x) \right] \leftrightarrow [x = \overline{n} \leftrightarrow x = \overline{n}] ] \\
& \iff \forall x [\sigma(\overline{x}, x) \rightarrow \Lambda(x)] \\
& \iff \forall x [x = \overline{\tau(\overline{x})} \rightarrow \Lambda(x)] \\
& \iff \forall x [x = \overline{\theta\delta(\overline{m}, n)} \rightarrow \Lambda(x)] \\
& \iff \Lambda(\overline{\delta(\overline{m}, n)}\overline{\theta})
\end{align*}
\]

by what was shown above,

by the definition of \(\zeta\),

by the associativity of \(\leftrightarrow\),

by logic,

by \(n = \overline{\tau}\) and the property of \(\sigma\),

by \(\delta(m, n) = \tau(\overline{n})\) shown above,

by logic. \(\Box\)
References

[1] Rudolf Carnap (1934); Logische Syntax der Sprache, Springer (in German). Translated into English by Amethe Smeton (Countess von Zeppelin) as The Logical Syntax of Language, Kegan Paul, Trench, Trubner & Co Ltd (1937).

[2] Patrick Cégielski & Denis Richard (1999); On Arithmetical First-Order Theories Allowing Encoding and Decoding of Lists, Theoretical Computer Science 222:1-2, pp. 55–75.

[3] Gregory Chaitin (1971); Computational Complexity and Gödel’s Incompleteness Theorem, ACM SIGACT News 9, pp. 11–12. Abstract: Notices of the American Mathematical Society 17:4 (1970) p. 672.

[4] Monika Gruber (2016); Alfred Tarski and the “Concept of Truth in Formalized Languages”: A Running Commentary with Consideration of the Polish Original and the German Translation, Springer (Logic, Epistemology, and the Unity of Science 39).

[5] Kurt Gödel (1931); “On Formally Undecidable Propositions of Principia Mathematica and Related Systems, I”, in: S. Feferman, J.W. Dawson Jr., S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort (eds.), Kurt Gödel Collected Works, Volume I: Publications 1929–1936, Oxford University Press (1986) pp. 145–195.

[6] Kurt Gödel (1934); On Undecidable Propositions of Formal Mathematical Systems, Lecture Notes (with corrections, modifications, and additions) in: S. Feferman, J.W. Dawson Jr., S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort (eds.), Kurt Gödel Collected Works, Volume I: Publications 1929–1936, Oxford University Press (1986) pp. 346–371.

[7] Stephen Kleene (1938); On Notation for Ordinal Numbers, The Journal of Symbolic Logic 3:4, pp. 150–155.

[8] Roman Murawski (1998); Undefinability of Truth. The Problem of Priority: Tarski vs Gödel, History and Philosophy of Logic 19:3, pp. 153–160.

[9] Barkley Rosser (1936); Extensions of Some Theorems of Gödel and Church, The Journal of Symbolic Logic 1:3, pp. 87–91.

[10] Saeed Salehi (2020); On the Diagonal Lemma of Gödel and Carnap, The Bulletin of Symbolic Logic 26:1, pp. 80–88.

[11] Saeed Salehi (2022); Tarski’s Undefinability Theorem and the Diagonal Lemma, Logic Journal of the IGPL 30:3, pp. 489–498.

[12] Saeed Salehi & Payam Seraji (2018); On Constructivity and the Rosser Property: A Closer Look at Some Gödelian Proofs, Annals of Pure and Applied Logic 169:10, pp. 971–980.