A reunion of GÖDEL, TARSKI, CARNAP and ROSSER

SAEED SALEHI, *Department of Mathematics, Statistics, and Computer Science, University of Tabriz, Bahman 29th Boulevard, PO Box 51666–16471, Tabriz, Iran.*
E-mail: salehipour@tabrizu.ac.ir

Journal of Logic and Computation, Volume 34, Issue 6, 2024, Pages 1172–1178.

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

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 [4, 8] and [6, f. 25 on p. 363].
3. (GÖDEL-)CARNAP's diagonal lemma in 1934; see [1] 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 Section 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 Section 4, the Appendix) and will see one more different proof for the weak syntactic diagonal lemma (in Section 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 2.3 below.

Advance Access Publication on 25 February 2023 https://doi.org/10.1093/logcom/exad001

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

The symbols $\mathbf{0}, \mathbf{1}, <, +, \times$ are interpreted as usual over $\mathbb{N}$, and the interpretation of $\boldsymbol{\nu}$, which is a function $\nu\colon \mathbb{N} \to \mathbb{N}$, depends on a fixed GÖDEL coding: if $\ulcorner\sigma\urcorner$ denotes the code of a sentence σ, then $\nu(\ulcorner\sigma\urcorner) = \ulcorner\neg\sigma\urcorner$ (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 $\overline{n}$ denote the closed term that represents n, which is $\mathbf{0}$ if $n = 0$, is $\mathbf{1}$ if $n = 1$, and is $\mathbf{1} + \cdots + \mathbf{1}$ (n times) if $n > 1$.

DEFINITION 2.2 (Arithmetical theories).
Let Q denote Robinson's arithmetic over the language of arithmetic.

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

(A_1): $\forall x(x < \overline{n} \vee x = \overline{n} \vee \overline{n} < x)$, for every $n \in \mathbb{N}$.
(A_2): $\forall x(x < \overline{n} \leftrightarrow \bigvee\!\!\!\bigvee_{i<n} x = \overline{i})$, for every $n \in \mathbb{N}$.
(A_3): $\boldsymbol{\nu}(\ulcorner\!\ulcorner\sigma\urcorner\!\urcorner) = \ulcorner\!\ulcorner\neg\sigma\urcorner\!\urcorner$, for every $\mathcal{L}^*$-sentence σ.

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 $U \vdash \Psi(\ulcorner\!\ulcorner\sigma\urcorner\!\urcorner)$ for every sentence σ, 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 $\{\Upsilon(\ulcorner\!\ulcorner\sigma\urcorner\!\urcorner) \leftrightarrow \sigma \mid \sigma \text{ is a sentence}\}$.
3. CARNAP$_T$: For every formula $\Lambda(x)$, with the only free variable x, there are finitely many sentences $\{A_i\}_i$ such that $T \vdash \bigvee\!\!\!\bigvee_i \big(\Lambda(\ulcorner\!\ulcorner A_i\urcorner\!\urcorner) \leftrightarrow A_i\big)$.
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 σ that (i) if $U \vdash \sigma$ then $U \vdash \Theta(\overline{m}, \ulcorner\!\ulcorner\sigma\urcorner\!\urcorner)$ for some number $m \in \mathbb{N}$, and (ii) if $U \nvdash \sigma$ then $U \vdash \neg\Theta(\overline{n}, \ulcorner\!\ulcorner\sigma\urcorner\!\urcorner)$ for every number $n \in \mathbb{N}$, then U is incomplete. ◇

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 σ, then U can verify that it proves σ; and conversely, if U proves that σ is U-provable, then σ is U-provable in reality (so, here U is assumed to possess some soundness, in the sense that U does not prove the U-provability of U-unprovable sentences).

(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 $Q \vdash \Lambda(\ulcorner A \urcorner) \leftrightarrow A$). 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 σ and m is the code of its proof, then U verifies this; and if σ is not U-provable, then U verifies that no number can code a U-proof of σ.

PROOF. **(**1 $\Longrightarrow$ 2**)**: If TARSKI$_T$ does not hold, then let $\mathfrak{M}$ be a model of $T + \{\Upsilon(\ulcorner\sigma\urcorner) \leftrightarrow \sigma \mid \sigma$ is a sentence$\}$, and put $U = \mathrm{Th}(\mathfrak{M})$. Now, for every sentence σ, we have $U \vdash \sigma$ iff $\mathfrak{M} \vDash \sigma$ iff $\mathfrak{M} \vDash \Upsilon(\ulcorner\sigma\urcorner)$ iff $U \vdash \Upsilon(\ulcorner\sigma\urcorner)$. But U is a complete extension of T; this contradicts GÖDEL$_T$.

(2 $\Longrightarrow$ 3**)**: For given $\Lambda(x)$, let $\Upsilon(x) = \neg\Lambda(x)$. By the inconsistency of the theory T with the set $\{\Upsilon(\ulcorner\sigma\urcorner) \leftrightarrow \sigma \mid \sigma \text{ is a sentence}\}$, we have $T \vdash \neg \bigwedge\!\!\!\bigwedge_i \big(\Upsilon(\ulcorner A_i\urcorner) \leftrightarrow A_i\big)$ for some finitely many sentences $\{A_i\}_i$. By the propositional tautology $\neg \bigwedge\!\!\!\bigwedge_i(\neg p_i \leftrightarrow q_i) \equiv \bigvee\!\!\!\bigvee_i(p_i \leftrightarrow q_i)$, we have $T \vdash \bigvee\!\!\!\bigvee_i \big(\Lambda(\ulcorner A_i\urcorner) \leftrightarrow A_i\big)$.

(3 $\Longrightarrow$ 4**)**: Let U and Θ satisfy the assumptions, and assume, for the sake of a contradiction, that U is a complete theory. Let $\Lambda(x)$ be the formula $\forall y\big[\Theta(y,x) \rightarrow \exists z < y\, \Theta\big(z, \boldsymbol{\nu}(x)\big)\big]$. By CARNAP$_T$, there are finitely many sentences $\{A_i\}_i$ such that $T \vdash \bigvee\!\!\!\bigvee_i \big(\Lambda(\ulcorner A_i\urcorner) \leftrightarrow A_i\big)$. Since complete theories have the disjunction property, then there exists one sentence ρ such that $(*)$ $U \vdash \Lambda(\ulcorner\rho\urcorner) \leftrightarrow \rho$. 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 $(*)$ and (A_3, Definition 2.2) we have

$$U \vdash \forall y\big[\Theta(y,\ulcorner\rho\urcorner) \rightarrow \exists z < y\, \Theta(z,\ulcorner\neg\rho\urcorner)\big].$$

By the assumption (i), there is some $m \in \mathbb{N}$ such that $U \vdash \Theta(\overline{m},\ulcorner\rho\urcorner)$. So, $U \vdash \exists z < \overline{m}\, \Theta(z,\ulcorner\neg\rho\urcorner)$. On the other hand, by $U \nvdash \neg\rho$ and (ii) in the assumption, we have $U \vdash \neg\Theta(\overline{n},\ulcorner\neg\rho\urcorner)$ for each $n \in \mathbb{N}$. Thus, by (A_2, Definition 2.2), we have $U \vdash \forall z < \overline{m} \neg\Theta(z,\ulcorner\neg\rho\urcorner)$. Whence, U is inconsistent; a contradiction.

(II) If $U \vdash \neg\rho$, then by $(*)$ and (A_3, Definition 2.2) we have

$$U \vdash \exists y\big[\Theta(y,\ulcorner\rho\urcorner) \wedge \forall z < y \neg\Theta(z,\ulcorner\neg\rho\urcorner)\big].$$

By (i) in the assumption, there is some $m \in \mathbb{N}$ such that $U \vdash \Theta(\overline{m},\ulcorner\neg\rho\urcorner)$. So, ($\mathsf{A}_1$, Definition 2.2) implies that $U \vdash \exists y \leqslant \overline{m}\, \Theta(y,\ulcorner\rho\urcorner)$. On the other hand, by $U \nvdash \rho$ and (ii), in the assumption, we have $U \vdash \neg\Theta(\overline{n},\ulcorner\rho\urcorner)$ for each $n \in \mathbb{N}$. Thus, by (A_2, Definition 2.2) we have $U \vdash \forall y \leqslant \overline{m} \neg\Theta(y,\ulcorner\rho\urcorner)$; a contradiction.

(4 $\Longrightarrow$ 1**)**: If GÖDEL$_T$ does not hold for the theory U and formula $\Psi(x)$, let $\Theta(x,y)$ be the formula $\Psi(y) \wedge (x = x)$. Now, by the assumption, for every sentence σ we have (i) if $U \vdash \sigma$, then $U \vdash \Psi(\ulcorner\sigma\urcorner)$, thus $U \vdash \Theta(\overline{m},\ulcorner\sigma\urcorner)$ for every $m \in \mathbb{N}$. Also, since U is (assumed to be) complete, then by the assumption we have (ii) if $U \nvdash \sigma$, then $U \nvdash \Psi(\ulcorner\sigma\urcorner)$, so $U \vdash \neg\Psi(\ulcorner\sigma\urcorner)$, thus $U \vdash \neg\Theta(\overline{n},\ulcorner\sigma\urcorner)$ for every $n \in \mathbb{N}$. This contradicts ROSSER$_T$. ☐

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 $\langle a_1, \cdots, a_\ell \rangle$ by $\sum_{k=1}^{\ell} 2^{\sum_{j=1}^{k}(a_j+1)} = \sum_{k=1}^{\ell} 2^{(a_1+1)+\cdots+(a_k+1)} = 2^{(a_1+1)} + 2^{(a_1+1)+(a_2+1)} + \cdots + 2^{(a_1+1)+\cdots+(a_\ell+1)}$.

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 $\eta \mapsto \ulcorner\!\!\ulcorner \eta \urcorner\!\!\urcorner$ is ACKERMANN's coding in Definition 2.4, then for every string η we have $\ulcorner\!\!\ulcorner \neg\eta \urcorner\!\!\urcorner = 16(1+\ulcorner\!\!\ulcorner \neg\eta \urcorner\!\!\urcorner)$.

PROOF. For $\eta = \langle a_1, \cdots, a_\ell \rangle$ we have $\ulcorner\!\!\ulcorner \langle \neg, a_1, \cdots, a_\ell \rangle \urcorner\!\!\urcorner = 2^4 + \sum_{k=1}^{\ell} 2^{4+\sum_{j=1}^{k}(a_j+1)}$ $= 16(1+\ulcorner\!\!\ulcorner \eta \urcorner\!\!\urcorner)$. □

DEFINITION 2.6 (A new framework).
Define the function $\nu^* \colon \mathbb{N} \to \mathbb{N}$ by $\nu^*(n) = 33+16n$ when n is even, and $\nu^*(n) = 16+16n$ when n is odd.
Let $\mathfrak{M}^*$ be the structure $\langle \mathbb{N}; \mathbf{0}, \mathbf{1}, \boldsymbol{<}, \boldsymbol{+}, \boldsymbol{\nu} \rangle$ where $\boldsymbol{\nu}$ is interpreted as ν^* above and put $T^* = \mathrm{Th}(\mathfrak{M}^*)$.
Let $\Psi^*(x) = \Upsilon^*(x) = \exists y(x = y \boldsymbol{+} y)$, $\Lambda^*(x) = \neg\Psi^*(x)$ and $\Theta^*(x,y) = \Psi^*(y) \wedge (x = x)$.
Denote ACKERMANN's coding in Definition 2.4 by $\eta \mapsto \ulcorner\!\!\ulcorner \eta \urcorner\!\!\urcorner$. Let $\eta \mapsto \llcorner \eta \lrcorner$ be a new coding on $\mathcal{L}^*$ (see Definition 2.1) defined as follows: $\llcorner \eta \lrcorner$ is $2\ulcorner\!\!\ulcorner \eta \urcorner\!\!\urcorner$, when η is an $\mathfrak{M}^*$-true $\mathcal{L}^*$-sentence; and is $1+2\ulcorner\!\!\ulcorner \eta \urcorner\!\!\urcorner$, otherwise. Let $\llcorner\!\!\llcorner E \lrcorner\!\!\lrcorner$ denote the closed term $\overline{\llcorner E \lrcorner}$.

Since $\eta \mapsto \ulcorner\!\!\ulcorner \eta \urcorner\!\!\urcorner$ (in Definition 2.4) is injective, the new coding $\eta \mapsto \llcorner \eta \lrcorner$ (in Definition 2.6) is injective too. For its computability, we note that $\mathfrak{M}^* = \langle \mathbb{N}; \mathbf{0}, \mathbf{1}, \boldsymbol{<}, \boldsymbol{+}, \boldsymbol{\nu} \rangle$ is decidable since ν^* is $\{\mathbf{1}, \boldsymbol{+}\}$-definable and $\langle \mathbb{N}; \mathbf{0}, \mathbf{1}, \boldsymbol{<}, \boldsymbol{+} \rangle$ is decidable by PRESBURGER's Theorem (cf. [2, Theorem 3.3]). Thus, $\eta \mapsto \llcorner \eta \lrcorner$ is computable as well (cf. Definition 2.1). We show that ν^* calculates the negation of sentences in the new coding's setting:

LEMMA 2.7 (ν^* is the negation mapping of sentences in the setting of $\llcorner \cdot \lrcorner$).
For every $\mathcal{L}^*$-sentence σ, we have $\nu^*(\llcorner \sigma \lrcorner) = \llcorner \neg\sigma \lrcorner$.

PROOF. The sentence σ is either $\mathfrak{M}^*$-true or $\mathfrak{M}^*$-false. If $\mathfrak{M}^* \vDash \sigma$, then $\llcorner \sigma \lrcorner = 2\ulcorner\!\!\ulcorner \sigma \urcorner\!\!\urcorner$ and $\neg\sigma$ is not $\mathfrak{M}^*$-true. So, $\llcorner \sigma \lrcorner$ is even, and by Lemma 2.5 we have $\nu^*(\llcorner \sigma \lrcorner) = 33+16\llcorner \sigma \lrcorner = 33+32\ulcorner\!\!\ulcorner \sigma \urcorner\!\!\urcorner = 1+2\cdot 16(1+\ulcorner\!\!\ulcorner \sigma \urcorner\!\!\urcorner) = 1+2\ulcorner\!\!\ulcorner \neg\sigma \urcorner\!\!\urcorner = \llcorner \neg\sigma \lrcorner$. If $\mathfrak{M}^* \nvDash \sigma$, then $\llcorner \sigma \lrcorner = 1+2\ulcorner\!\!\ulcorner \sigma \urcorner\!\!\urcorner$ and $\neg\sigma$ is $\mathfrak{M}^*$-true. So, $\llcorner \sigma \lrcorner$ is odd, thus by Lemma 2.5 we have $\nu^*(\llcorner \sigma \lrcorner) = 16+16\llcorner \sigma \lrcorner = 32+32\ulcorner\!\!\ulcorner \sigma \urcorner\!\!\urcorner = 2\cdot 16(1+\ulcorner\!\!\ulcorner \sigma \urcorner\!\!\urcorner)$ $= 2\ulcorner\!\!\ulcorner \neg\sigma \urcorner\!\!\urcorner = \llcorner \neg\sigma \lrcorner$. □

Let us note that if the expression η is not a sentence, then Lemma 2.7 does not hold; in that case $\nu^*(\llcorner \eta \lrcorner)+1 = \llcorner \neg\eta \lrcorner$. Now, we have all the ingredients for constructing a case in which none of the equivalent statements GÖDEL$_T$, TARSKI$_T$, CARNAP$_T$ and ROSSER$_T$ can hold (cf. [11, Remark 2.6]).

THEOREM 2.8 ($\neg$GÖDEL$_{T^*}$, $\neg$TARSKI$_{T^*}$, $\neg$CARNAP$_{T^*}$, $\neg$ROSSER$_{T^*}$).
The complete theory T^* contains Q^-, and neither GÖDEL$_{T^*}$, nor TARSKI$_{T^*}$, nor CARNAP$_{T^*}$, nor ROSSER$_{T^*}$ 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 $\{\mathbf{1},+\}$-definability of ν^*. Thus, the mapping $\eta \mapsto \llcorner\eta\lrcorner$ is a (computable injective) coding. Trivially, the axioms ($\mathsf{A}_1, \mathsf{A}_2$, Definition 2.2) hold in $\mathfrak{M}^*$, and Lemma 2.7 implies that (A_3, Definition 2.2) holds too. Thus, T^* contains Q^-. We now show that none of GÖDEL_{T^*}, TARSKI_{T^*}, CARNAP_{T^*} or ROSSER_{T^*} holds for $\Psi^*(x)$, $\Upsilon^*(x)$, $\Lambda^*(x)$, and $\Theta^*(x,y)$.

$\neg\text{GÖDEL}_{T^*}$: For every sentence σ, $T^* \vdash \sigma$ iff $\mathfrak{M}^* \vDash \sigma$ iff $\llcorner\sigma\lrcorner$ is even iff $\mathfrak{M}^* \vDash \Psi^*(\llbracket\sigma\rrbracket)$ iff $T^* \vdash \Psi^*(\llbracket\sigma\rrbracket)$.

$\neg\text{TARSKI}_{T^*}$: For every sentence σ, we have $T^* \vdash \Upsilon^*(\llbracket\sigma\rrbracket) \leftrightarrow \sigma$ by $\neg\text{GÖDEL}_{T^*}$.

$\neg\text{CARNAP}_{T^*}$: For any finitely many sentences $\{A_i\}_i$, we have $T^* \vdash \bigwedge\!\!\!\bigwedge_i \big(\neg\Lambda^*(\llbracket A_i\rrbracket) \leftrightarrow A_i\big)$ by $\neg\text{TARSKI}_{T^*}$; so $T^* \vdash \neg \bigvee\!\!\!\bigvee_i \big(\Lambda^*(\llbracket A_i\rrbracket) \leftrightarrow A_i\big)$, thus $T^* \nvdash \bigvee\!\!\!\bigvee_i \big(\Lambda^*(\llbracket A_i\rrbracket) \leftrightarrow A_i\big)$ by the consistency of T^*.

$\neg\text{ROSSER}_{T^*}$: We saw that (i) if $T^* \vdash \sigma$, then $T^* \vdash \Psi^*(\llbracket\sigma\rrbracket)$, so $T^* \vdash \Theta^*(\overline{m},\llbracket\sigma\rrbracket)$ for every $m\in\mathbb{N}$. Also note that (ii) if $T^* \nvdash \sigma$, then $\llcorner\sigma\lrcorner$ is odd, so $T^* \vdash \neg\Psi^*(\llbracket\sigma\rrbracket)$, thus $T^* \vdash \neg\Theta^*(\overline{n},\llbracket\sigma\rrbracket)$ for every $n\in\mathbb{N}$. $\square$

3 TARSKI's theorem and the weak syntactic diagonal lemma, á 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 (GÖDEL-TARSKI's truth-undefinability theorem).
For every formula $\Upsilon(x)$, the theory Q is inconsistent with the set $\{\Upsilon(\ulcorner\sigma\urcorner)\leftrightarrow\sigma \mid \sigma \text{ is a sentence}\}$.

PROOF. Assume not; fix a model $\mathfrak{M}$ of $Q + \{\Upsilon(\ulcorner\sigma\urcorner) \leftrightarrow \sigma \mid \sigma \text{ is a sentence}\}$. Suppose that $\varphi_0^\Upsilon, \varphi_1^\Upsilon, \varphi_2^\Upsilon, \cdots$ is an effective enumeration of all the Υ-computable (i.e., computable with oracle Υ) unary functions. Define the Υ-KOLMOGOROV-CHAITIN complexity of $n\in\mathbb{N}$ to be $\mathscr{K}^\Upsilon(n) = \min\{i \mid \varphi_i^\Upsilon(0)=n\}$, the minimum index of the Υ-computable function that outputs n on input 0. By KLEENE's Recursion theorem [7] there exists some $\mathfrak{c}\in\mathbb{N}$ such that $\varphi_{\mathfrak{c}}^\Upsilon(x)=\min z\colon \Upsilon(\ulcorner\langle\mathscr{K}^\Upsilon(\overline{z})>\overline{\mathfrak{c}}\rangle\urcorner)$, where $\langle\mathscr{K}^\Upsilon(x)>y\rangle$ is the arithmetical formula which says that "the Υ-KOLMOGOROV-CHAITIN complexity of x is greater than y".[2] By the Pigeonhole principle (a version of which is provable in Q, see [12, Lemma 3.8]) there exists some element $u \leqslant \mathfrak{c}+1$ in $\mathfrak{M}$ such that $\mathfrak{M} \vDash \langle\mathscr{K}^\Upsilon(\overline{u})>\overline{\mathfrak{c}}\rangle$; note that $\{\varphi_0^\Upsilon(0), \varphi_1^\Upsilon(0), \cdots, \varphi_{\mathfrak{c}}^\Upsilon(0)\}$ has at most $\mathfrak{c}+1$ members, and $\{0,1,\cdots,\mathfrak{c}+1\}$ has $\mathfrak{c}+2$ members. For the least $\mathfrak{u}\in\mathfrak{M}$ with $\mathfrak{M} \vDash \langle\mathscr{K}^\Upsilon(\overline{\mathfrak{u}})>\overline{\mathfrak{c}}\rangle$ we have $\mathfrak{M} \vDash \Upsilon(\ulcorner\langle\mathscr{K}^\Upsilon(\overline{\mathfrak{u}})>\overline{\mathfrak{c}}\rangle\urcorner)$ and $\mathfrak{M} \vDash \forall z<\overline{\mathfrak{u}}\,\neg\Upsilon(\ulcorner\langle\mathscr{K}^\Upsilon(\overline{z})>\overline{\mathfrak{c}}\rangle\urcorner)$; so $\mathfrak{M} \vDash \langle\varphi_{\mathfrak{c}}^\Upsilon(\overline{0})=\overline{\mathfrak{u}}\rangle$, thus $\mathfrak{M} \vDash \langle\mathscr{K}^\Upsilon(\overline{\mathfrak{u}})\leqslant\overline{\mathfrak{c}}\rangle$, a contradiction. $\square$

The part ($2 \Longrightarrow 3$) 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 $Q \vdash \bigvee\!\!\!\bigvee_i \big(\Lambda(\ulcorner A_i\urcorner)\leftrightarrow A_i\big)$.

[2]Notice that this very proof implies that the unary function $z\mapsto\mathscr{K}^\Upsilon(z)$ is not Υ-computable; though, for every constant $c\in\mathbb{N}$, the function $x\mapsto \min z\colon \Upsilon(\ulcorner\langle\mathscr{K}^\Upsilon(\overline{z})>\overline{c}\rangle\urcorner)$ is clearly Υ-computable.

PROOF. Put $\Upsilon(x)=\neg\Lambda(x)$; with the notation of the proof of Proposition 3.1, let $A_i=\langle\mathscr{K}^{\Upsilon}(\bar{i})>\bar{\mathfrak{c}}\rangle$ for $i\leqslant\mathfrak{c}+1$. If $Q\nvdash\bigvee\!\!\!\bigvee_{i\leqslant\mathfrak{c}+1}\big(\Lambda(\ulcorner\!\ulcorner A_i\urcorner\!\urcorner)\leftrightarrow A_i\big)$, then $Q+\bigwedge\!\!\!\bigwedge_{i\leqslant\mathfrak{c}+1}\big(\Upsilon(\ulcorner\!\ulcorner A_i\urcorner\!\urcorner)\leftrightarrow A_i\big)$ 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. $\square$

4 Appendix: constructivity of an alternative proof of the diagonal lemma

For $m,n\in\mathbb{N}$, let $\boldsymbol{\delta}(m,n)$ 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 $\mathbb{N}\vDash\Lambda\big(\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner\big)\leftrightarrow\boldsymbol{\delta}(\overline{m},\overline{n})$. 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 Λ 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 two natural numbers $m,n\in\mathbb{N}$ such that $Q\vdash\boldsymbol{\delta}(\overline{m},\overline{n})\leftrightarrow\Lambda\big(\ulcorner\!\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner\!\urcorner\big)$.

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

$$
\begin{array}{llll}
Q\vdash\boldsymbol{\delta}(\overline{m},\overline{n}) & \longleftrightarrow & \forall x[\zeta(x,\overline{n})\leftrightarrow x=\overline{n}] & \text{by what was shown above,}\\
 & \longleftrightarrow & \forall x\Big[\big([\boldsymbol{\sigma}(\overline{n},x)\rightarrow\Lambda(x)]\leftrightarrow x=\overline{n}\big)\leftrightarrow x=\overline{n}\Big] & \text{by the definition of }\zeta,\\
 & \longleftrightarrow & \forall x\big([\boldsymbol{\sigma}(\overline{n},x)\rightarrow\Lambda(x)]\leftrightarrow[x=\overline{n}\leftrightarrow x=\overline{n}]\big) & \text{by the associativity of }\leftrightarrow,\\
 & \longleftrightarrow & \forall x[\boldsymbol{\sigma}(\overline{n},x)\rightarrow\Lambda(x)] & \text{by logic,}\\
 & \longleftrightarrow & \forall x[x=\ulcorner\!\ulcorner\tau(\overline{n})\urcorner\!\urcorner\rightarrow\Lambda(x)] & \text{by }n=\ulcorner\tau\urcorner\text{ and the property of }\boldsymbol{\sigma},\\
 & \longleftrightarrow & \forall x[x=\ulcorner\!\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner\!\urcorner\rightarrow\Lambda(x)] & \text{by }\boldsymbol{\delta}(\overline{m},\overline{n})=\tau(\overline{n})\text{ shown above,}\\
 & \longleftrightarrow & \Lambda\big(\ulcorner\!\ulcorner\boldsymbol{\delta}(\overline{m},\overline{n})\urcorner\!\urcorner\big) & \text{by logic.}
\end{array}
$$

$\square$

References

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

[2] P. Cégielski and D. Richard. On arithmetical first-order theories allowing encoding and decoding of lists. *Theoretical Computer Science*, **222**, 55–75, 1999.

[3] G. Chaitin. Computational complexity and Gödel's incompleteness theorem. *ACM SIGACT News*, **9**, 11–12, 1971. Abstract: *Notices of the American Mathematical Society* **17**, 672, 1970.

[4] M. Gruber. *Alfred Tarski and the "Concept of Truth in Formalized Languages": A Running Commentary with Consideration of the Polish Original and the German Translation*. Vol. 39 of *Logic, Epistemology, and the Unity of Science*. Springer, 2016.

[5] K. Gödel. On formally undecidable propositions of *principia mathematica* and related systems, I. In *Kurt Gödel Collected Works, Volume I: Publications 1929–1936*, in: S. Feferman, J. W. Dawson Jr., S. C. Kleene, G. H. Moore, R. M. Solovay and J. van Heijenoort, eds, pp. 145–195. Oxford University Press, 1931, 1986.

[6] K. Gödel. On undecidable propositions of formal mathematical systems. In *Kurt Gödel Collected Works, Volume I: Publications 1929–1936*, in: S. Feferman, J. W. Dawson Jr., S. C. Kleene, G. H. Moore, R. M. Solovay and J. van Heijenoort, eds, Lecture Notes (with corrections, modifications, and additions), pp. 346–371. Oxford University Press, 1934, 1986.

[7] S. Kleene. On notation for ordinal numbers. *The Journal of Symbolic Logic*, **3**, 150–155, 1938.

[8] R. Murawski. Undefinability of truth. The problem of priority: Tarski vs Gödel. *History and Philosophy of Logic*, **19**, 153–160, 1998.

[9] B. Rosser. Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic*, **1**, 87–91, 1936.

[10] S. Salehi. On the diagonal lemma of Gödel and Carnap. *The Bulletin of Symbolic Logic*, **26**, 80–88, 2020.

[11] S. Salehi. Tarski's undefinability theorem and the diagonal lemma. *Logic Journal of the IGPL*, **30**, 489–498, 2022.

[12] S. Salehi and P. Seraji. On constructivity and the Rosser property: a closer look at some Gödelean proofs. *Annals of Pure and Applied Logic*, **169**, 971–980, 2018.

Received 10 October 2022