The Diagonalization Lemma
Demystified Hopefully

Saeed Salehi

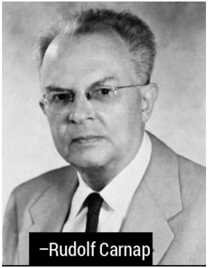
Celebrating 90 Years of Gödel's Incompleteness Theorems

Diagonalization

Nürtingen, Germany, 8 July 2021

The Diagonal Lemma.

Fix a sufficiently expressive language $\mathcal{L}$, with a computable injection (Gödel coding) $\alpha \mapsto \ulcorner \alpha \urcorner$ from $\mathcal{L}$-sentences to closed $\mathcal{L}$-terms.



Lemma (Gödel 1931 & Carnap 1934)

For every formula $\Psi(x)$ there exists a sentence θ such that

$$\boldsymbol{Q} \vdash \theta \leftrightarrow \Psi(\ulcorner \theta \urcorner).$$

The *Hocus-Pocus* Proof of the Diagonal Lemma.

Proof (Easy ?!):

If $\mathcal{L}$ contains the PR function $\boldsymbol{d}(\ulcorner\Phi(x)\urcorner) = \ulcorner\Phi(x/\ulcorner\Phi(x)\urcorner)\urcorner$, which maps the code of a formula with one free variable to the code of a sentence, then it suffices to put $\theta = \Psi(\boldsymbol{d}(\ulcorner\Psi(\boldsymbol{d}(x))\urcorner))$.

Note that then $\theta = \Psi(\ulcorner\theta\urcorner)$. ■

Proof (Difficult ‼):

If $\boldsymbol{d} \notin \mathcal{L}$, then some $\mathcal{L}$-formula $\delta(x, y)$ should (strongly) represent $\boldsymbol{d}$ in $\boldsymbol{Q}$, thus we have $\boldsymbol{Q} \vdash \forall y\,[\delta(\overline{n}, y) \leftrightarrow y = \boldsymbol{d}(\overline{n})]$ for every $n \in \mathbb{N}$, where $\overline{n} = 1 + \cdots + 1$ (n-times).

Now, one can show $\boldsymbol{Q} \vdash \theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)$ by taking either

(U) $\theta = \alpha(x/\ulcorner\alpha(x)\urcorner)$ where $\alpha(x) = \forall y\,[\delta(x, y) \rightarrow \Psi(y)]$, or

(E) $\theta = \eta(x/\ulcorner\eta(x)\urcorner)$ where $\eta(x) = \exists y\,[\delta(x, y) \wedge \Psi(y)]$. ■

What they say about the *proof* ...

(2002) MCGEE, VANN; The First Incompleteness Theorem, *Handouts of the Course "Logic II"*. https://bit.ly/301QLTA

> "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."

(2006) GAIFMAN, HAIM; Naming and Diagonalization, from Cantor to Gödel to Kleene, *Logic Journal of the IGPL* 14(5):709–728.

> "The brevity of the proof does not make for transparency; it has the aura of a magician's trick."

(2008) WASSERMAN, W. URBAN; It ***Is*** "Pulling a Rabbit Out of the Hat": Typical Diagonal Lemma "Proofs" Beg the Question, *Social Science Research Network*, 1–11. DOI: 10.2139/ssrn.1129038

Abracadabra & … The Diagonal-Free Proofs.

(2004) Kotlarski, Henryk; The Incompleteness Theorems After 70 Years, *Annals of Pure and Applied Logic* 126(1-3):125–138.

> "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 Göodel's result. This was achieved only in the 1990s".

- Kleene, S. (1936 & 50) for Gödel's (& Rosser's) Theorem
- Robinson, A. (1963) for Tarski's Theorem
- Chaitin, G. (1970) for Gödel's Theorem
- Boolos, G. (1989) for Gödel's Theorem
- Caicedo, X. (1993) for Tarski's Theorem
- Jech, Th. (1994) for Gödel's 2nd Theorem
- Kotlarski, H. (1994 & 96 & 98) for Gödel's & Tarski's Theorems

Demystifying the Diagonal Lemma (1).

Another Approach:

Not Removing the Lemma altogether (going diagonal-free), but

SEE(K)ING ALTERNATIVE PROOFS OF THE LEMMA — IF POSSIBLE.

- Proof Theory, Modal Logic and Reflection Principles Workshop (Wormshop'17), Steklov Mathematical Institute, Moscow, Russia, 17–20 October 2017.
  – Title: *Diagonal-Free Proofs of the Diagonal Lemma*.

- Tarski's Undefinability Theorem and the Diagonal Lemma, *Logic Journal of the IGPL* (forthcoming).
  DOI: 10.1093/jigpal/jzab016

Demystifying the Diagonal Lemma (2).

- (Semantic Tarski's Theorem) $\forall\Psi : \mathrm{Th}(\mathbb{N}) \neq \{\theta \mid \mathbb{N} \models \Psi(\ulcorner\theta\urcorner)\}$.
- (Semantic Diagonal Lemma) $\forall\Psi\exists\theta : \mathbb{N} \models \theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)$.

Theorem

Semantic Tarski's Theorem $\Longleftrightarrow$ *Semantic Diagonal Lemma.*

Proof.

¬Semantic Diagonal Lemma $\equiv \exists\Psi(x)\forall\theta\colon \mathbb{N} \not\models \theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)$

$\mathbb{N} \models \neg[\theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)]$

$\neg(p \leftrightarrow q) \equiv (p \leftrightarrow \neg q)$ $\qquad$ $\mathbb{N} \models \theta \leftrightarrow \neg\Psi(\ulcorner\theta\urcorner)$

$\Psi'(x) = \neg\Psi(x)$

¬Semantic Diagonal Lemma $\equiv \exists\Psi'(x)\forall\theta\colon \mathbb{N} \models \theta \leftrightarrow \Psi'(\ulcorner\theta\urcorner)$

$\equiv \exists\Psi'(x)\colon \mathrm{Th}(\mathbb{N}) = \{\theta \mid \mathbb{N} \models \Psi'(\ulcorner\theta\urcorner)\}$

$\equiv$ ¬Semantic Tarski's Theorem ■

[Almost] Everyone Loves Magic (1).

A new proof for the Semantic Diagonal Lemma, and a new proof for a weak syntactic version of the lemma was given in:

- On the Diagonal Lemma of Gödel and Carnap, *The Bulletin of Symbolic Logic* 26:1 (2020) 80–88.

The proof is based on Berry's paradox (rather than classical Liar's).

Lemma (Weak Syntactic Diagonal Lemma)

For every formula $\Psi(x)$ *there exist sentences* $\{\theta_j\}_{j<n}$ *such that*

$$\boldsymbol{Q} \vdash \bigvee\!\!\!\bigvee_{j<n} [\theta_j \leftrightarrow \Psi(\ulcorner \theta_j \urcorner)]. \qquad \blacksquare$$

Theorem

The Weak Lemma *implies Theorems of Gödel, Tarski, and Rosser.* ■

[Almost] Everyone Loves Magic (2).

Another (New) Proof for the (strong) Diagonal Lemma:

Proof.

Identify the formula $\Psi(x)$ with the set $\hat{\Psi} = \{n \in \mathbb{N} \mid \mathbb{N} \vDash \Psi(\overline{n})\}$.

Let $\varphi_0^\Psi, \varphi_1^\Psi, \cdots$ be an effective enumeration of all the unary computable functions with oracle $\hat{\Psi}$.

Let $\langle\!\langle \varphi_u^\Psi(v)\!\uparrow \rangle\!\rangle$ be a formula saying that "the Ψ-recursive function with code u does not halt at v".

Let $\Bbbk$ be a code of the Ψ-recursive $x \mapsto \boldsymbol{\mu} y.[\neg\Psi(\ulcorner\langle\!\langle \varphi_x^\Psi(x)\!\uparrow \rangle\!\rangle\urcorner)]$.

We have $\varphi_\Bbbk^\Psi(x)\!\uparrow \iff \Psi(\ulcorner\langle\!\langle \varphi_x^\Psi(x)\!\uparrow \rangle\!\rangle\urcorner)$, for every x.

Let $\theta = \langle\!\langle \varphi_\Bbbk^\Psi(\Bbbk)\!\uparrow \rangle\!\rangle$.

For $x = \Bbbk$ we get $\theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)$, and this is provable in $\boldsymbol{Q}$. ■

GÖDEL ≡ TARSKI ≡ DIAGONAL — Semantically.

Definition (Semantic Diagonal, GÖDEL, and TARSKI)

Recall $\ulcorner\cdot\urcorner\colon \mathcal{L}$-Formulas $\to \mathcal{L}$-ClosedTerms is a computable injection.

Let $\mathscr{M}$ be an $\mathcal{L}$-structure.

- GÖDEL$_{\mathscr{M}}$: $\forall\,\Psi(x)\forall T\subseteq \mathrm{Th}(\mathscr{M})$: $T=\{\theta \mid \mathscr{M}\vDash\Psi(\ulcorner\theta\urcorner)\}\Longrightarrow$ T is incomplete.
- TARSKI$_{\mathscr{M}}$: $\forall\,\Psi(x)$: $\mathrm{Th}(\mathscr{M})\neq\{\theta \mid \mathscr{M}\vDash\Psi(\ulcorner\theta\urcorner)\}$.
- DIAGONAL$_{\mathscr{M}}$: $\forall\,\Psi(x)\,\exists\theta$: $\mathscr{M}\vDash\theta\leftrightarrow\Psi(\ulcorner\theta\urcorner)$. ▲

Theorem

For every $\langle\mathcal{L},\ulcorner\cdot\urcorner,\mathscr{M}\rangle$ *we have* GÖDEL$_{\mathscr{M}}$ ≡ TARSKI$_{\mathscr{M}}$ ≡ DIAGONAL$_{\mathscr{M}}$.

For some $\langle\mathcal{L},\ulcorner\cdot\urcorner,\mathscr{M}\rangle$*'s all three hold* (*such as* $\langle\mathbb{N};1,+,\times\rangle$) *and*

for some $\langle\mathcal{L},\ulcorner\cdot\urcorner,\mathscr{M}\rangle$*'s none holds* (*such as* $\langle\mathbb{N};1,+\rangle$). ■

Gödel ≡ Rosser ≡ Tarski ≡ ^{w}Diagonal — Syntactically.

Definition

Fix an $\mathcal{L}$-theory T.

- Gödel$_T$: $\forall\Psi(x)\,\forall U\supseteq T$: if $U\nvdash\bot$ and $\forall\theta\in\mathcal{L}$-Sentences $U\vdash\theta\iff U\vdash\Psi(\ulcorner\theta\urcorner)$, then U is <u>in</u>complete.

- Rosser$_T$: $\forall\Phi(x,y)\,\forall U\supseteq T$: if $U\nvdash\bot$ and $\forall\theta\in\mathcal{L}$-Sentences
  $U\vdash\theta\Longrightarrow U\vdash\Phi(\overline{m},\ulcorner\theta\urcorner)$ for some $m\in\mathbb{N}$ and
  $U\nvdash\theta\Longrightarrow U\vdash\neg\Phi(\overline{n},\ulcorner\theta\urcorner)$ for each $n\in\mathbb{N}$, then U is <u>in</u>complete.

- Tarski$_T$: $\forall\Psi(x)$: $[T+\{\theta\leftrightarrow\Psi(\ulcorner\theta\urcorner)\mid\theta\in\mathcal{L}\text{-Sentences}\}]\vdash\bot$.

- ^{w}Diagonal$_T$: $\forall\Psi(x)\,\exists\{\theta_j\}_{j<n}$: $T\vdash\bigvee_{j<n}[\theta_j\leftrightarrow\Psi(\ulcorner\theta_j\urcorner)]$. ▲

The Magic Trick is Revealed.

Theorem

For every $\langle\mathcal{L},\ulcorner\cdot\urcorner,T\rangle$, $\text{GÖDEL}_T \equiv \text{ROSSER}_T \equiv \text{TARSKI}_T \equiv {}^{w}\text{DIAGONAL}_T$.

For some $\langle\mathcal{L},\ulcorner\cdot\urcorner,T\rangle$*'s all four hold (such as* Robinson Arithmetic*) and for some* $\langle\mathcal{L},\ulcorner\cdot\urcorner,T\rangle$*'s none holds (such as* Presburger Arithmetic*).* ■

The Proof on page 9 was a translation of Kleene's Proof for Gödel's First Incompleteness Theorem. That (somehow magically) worked (also) for the strong syntactic diagonal lemma $\llbracket\ Q \vdash \theta \leftrightarrow \Psi(\ulcorner\theta\urcorner)\ \rrbracket$.

Other proofs (of Chaitin, Boolos, and Kotlarski) are translated only to the weak syntactic diagonal lemma $\llbracket\ Q \vdash \bigvee_{j<n}[\theta_j \leftrightarrow \Psi(\ulcorner\theta_j\urcorner)]\ \rrbracket$.

Problem (Open)

Does the weak syntactic diagonal lemma imply Löb's Theorem/Rule? ▲

Thank You!

Diagonally Yours, SaeedSalehi.ir