Self-Reference and Diagonalization:

their difference and a short history

Saeed Salehi

The New York City Category Theory Seminar

CUNY Graduate Center 23 November 2022

Fixed-Points, Diagonalization, and Self-Reference

- ### Fixed Points

  There is a *mapping*, and **an object** is proved to exist that **is mapped to itself**, in the Theorem or in the Proof.

- ### Diagonalization

  The **diagonal of a matrix** is used (or referred to) in the Theorem or in the Proof.

- ### Self-Reference

  Something (an object, or a concept) **refers to** (the code, the name, or something of) **itself**, either in the Theorem or in the Proof.

Self-Referential

- **Something** (an object, or a concept) **refers to** (the code, the name, or something of) **itself**, either in the Theorem or in the Proof.

Theorem (BARBER's Paradox)

$F.O.Logic \vdash \neg\exists\,\text{barber}\,\forall x\,(\text{barber}\ \mathfrak{shaves}\ x \longleftrightarrow \neg[x\ \mathfrak{shaves}\ x]).$

Proof.

If $\exists\,\text{barber}\,\forall x\,(\text{barber}\ \mathfrak{shaves}\ x \longleftrightarrow \neg[x\ \mathfrak{shaves}\ x])$, then for $x = \text{barber}$ we get the contradiction (similar to the LIAR's paradox)

$\text{barber}\ \mathfrak{shaves}\ \text{barber} \longleftrightarrow \neg[\text{barber}\ \mathfrak{shaves}\ \text{barber}]$! ■

FIXED POINT? DIAGONAL?

THEOREM. $S.O.Logic \vdash \neg\exists X^{(2)}\exists\alpha\forall x\,[X(\alpha,x)\longleftrightarrow\neg X(x,x)]$.

QUESTION: What about YABLO's Paradox?

Fixed-Points

- There is a mapping, and **an object** is proved to exist that **is mapped to itself**, in the Theorem or in the Proof.

LAWVERE: In a cartesian closed category, if there is a point-surjective map $\mathfrak{h}\colon B \to A^B$ (for objects A, B), then every map $\mathfrak{f}\colon A \to A$ has a fixed point ($\mathfrak{s}\colon \mathbf{1} \to A$ such that $\mathfrak{s} = \mathfrak{f}\mathfrak{s}$).

KNASTER–TARSKI: Every monotonic function on a complete lattice has some fixed points (which constitute a complete lattice).

KLEENE: Every Scott-continuous function on a directed complete partial order with a least element, has a (least) fixed point.

SELF-REFERENTIAL? DIAGONAL?

Kleene's Recursion Theorem

For every computable $F(x,\vec{y})$ there is an e such that $\varphi_e(\vec{y}) \simeq F(e,\vec{y})$.
For every computable $\mathfrak{F}(x)$ there is an e such that $\varphi_e(\vec{y}) \simeq \varphi_{\mathfrak{F}(e)}(\vec{y})$.

Proof.

Let $\mathcal{S}(i,j)$ be a recursive index of $\vec{y} \mapsto \varphi_i(j,\vec{y})$. Consider the matrix $[F(\mathcal{S}(i,j),\vec{y})]_{i,j\in\mathbb{N}}$ and its diagonal $(x,\vec{y}) \mapsto F(\mathcal{S}(x,x),\vec{y})$, which is recursive and so has an index m; put $e = \mathcal{S}(m,m)$. Now, we have $\varphi_e(\vec{y}) \simeq \varphi_{\mathcal{S}(m,m)}(\vec{y}) \simeq \varphi_m(m,\vec{y}) \simeq F(\mathcal{S}(m,m),\vec{y}) \simeq F(e,\vec{y})$. ■

e may not be equal to $\mathfrak{F}(e)$, they just code the same function!

For $\mathbf{\Phi}(\hbar) = \varphi_{\mathfrak{F}(\#\hbar)}$ there is a fixed point $\mathfrak{y} = \mathbf{\Phi}(\mathfrak{y})$; and $e = \#\mathfrak{y}$.
But $\mathbf{\Phi}(\hbar)$ is *not* well-defined, unless $\varphi_i \simeq \varphi_j \Rightarrow \varphi_{\mathfrak{F}(i)} \simeq \varphi_{\mathfrak{F}(j)}$.

SELF-REFERENTIAL ✓ FIXED POINT X DIAGONAL ✓

Diagonalization

- The `diagonal of a matrix` is used (or referred to) in the Theorem or in the Proof.

WHO INVENTED/DISCOVERED THE DIAGONALIZATION?

- Georg CANTOR (1891)?
- Paul DU BOIS-REYMOND (1870,1872,1875)?
- René DESCARTES?[*]
- EUCLID OF ALEXANDRIA?
- PYTHAGORAS?

If *diagonalization* was not invented/discovered by CANTOR, it was surely matured by him! In a way that everyone after him, including RUSSELL, GÖDEL, TURING, and KLEENE, followed his footsteps.

---

[*]T. MEADOWS (2022), Did Descartes Make a Diagonal Argument?, *J.Phil.Log.* 51_2:219–47.

An Ancient Diagonalization (?)

Theorem (Infinitude of the Primes)

There are infinitely many prime numbers.

Proof.

For every finite number of primes $\mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_n$, there is a prime ((factor of $1 + \mathfrak{p}_1 \cdot \mathfrak{p}_2 \cdots \mathfrak{p}_n$, which is)) distinct from $\mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_n$. ■

A Diagonal Proof.

$n! = 1 \times 2 \times \cdots \times n$.

Let $a_{\langle n,m\rangle} = 1$ if all the primes factors of $m!+1$ are $\leqslant n$, and $a_{\langle n,m\rangle} = 0$ if some prime factor of $m!+1$ is $> n$. If all the primes are $\leqslant N$, then the Nth row is all 1. But the diagonal $\{a_{\langle n,n\rangle}\}_{n\in\mathbb{N}}$ is all 0, since no factor of $n!+1$ can be $\leqslant n$. A contradiction; so, there is no such N. ■

A Non-Diagonal Proof.

For every N, the N numbers $\{k \cdot N! + 1\}_{k=1}^{N}$ are pairwise coprime; so the number of primes cannot be $< N$ by the Pigeonhole Principle. ■

How was (CANTOR's) diagonalization discovered (?)

THEOREM. $\mathbb{R} \cap (0,1)$ *is uncountable.*

CANTOR's proofs:

Assume (for the sake of a contradiction) that $(0,1) = \{x_n\}_{n\in\mathbb{N}}$.

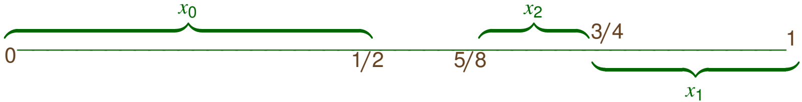
(1874): Let $b_0 = \min\{x_0, x_1\}$, $d_0 = \max\{x_0, x_1\}$, and inductively let $b_{m+1} < d_{m+1}$ be the first two elements of $\{x_n\}_{n\in\mathbb{N}}$ that lie inside (b_m, d_m). Then $\ell im\{b_m\}_{m\in\mathbb{N}} \in (0,1) \setminus \{x_n\}_{n\in\mathbb{N}}$, since $x_n \notin (b_n, d_n)$ for each $n \in \mathbb{N}$. (generalized in 1879)

(1884): Let $\mathtt{I}_0$ be a closed sub-interval of $(0,1)$ with length$<\frac{1}{2}$ that leaves out x_0. Inductively, let $\mathtt{I}_{m+1}$ be a closed sub-interval of $\mathtt{I}_m$ with length$<\frac{1}{2}$(length of $\mathtt{I}_m$) that leaves out x_{m+1}. Then $\bigcap_m \mathtt{I}_m$ is non-empty and disjoint from $\{x_n\}_{n\in\mathbb{N}}$.

(1891): **Diagonal Argument.** ⟦Nested Intervals⟧

A (Re-)Discivery of Diagonalization:

Ignore the (countable many) numbers $m/2^n$ and write the *infinite* binary expansion (0, 1's in the base 2) of x_n as $0.y_n^{\backslash\mathtt{0}\backslash}y_n^{\backslash\mathtt{1}\backslash}y_n^{\backslash\mathtt{2}\backslash}\cdots$.
Let $\mathtt{I}_0=[0,1]$; and inductively let $\mathtt{I}_{m+1}$ be the half of $\mathtt{I}_m$ that misses the point x_m (we have ignored the boundary x_n's). For example,



$$x_0=0.0y_0^{\backslash\mathtt{1}\backslash}y_0^{\backslash\mathtt{2}\backslash}y_0^{\backslash\mathtt{3}\backslash}\cdots,\ x_1=0.11y_1^{\backslash\mathtt{2}\backslash}y_1^{\backslash\mathtt{3}\backslash}\cdots,\ x_2=0.101y_2^{\backslash\mathtt{3}\backslash}\cdots.$$

So, if $\mathtt{I}_m=[b_m,d_m]$ let $c_m=(b_m+d_m)/2$; if $x_m\in[b_m,c_m]$ let $\mathtt{I}_{m+1}=[c_m,d_m]$, and if $x_m\in[c_m,d_m]$ let $\mathtt{I}_{m+1}=[b_m,c_m]$.
Note that in the first case $y_m^{\backslash\mathtt{m}\backslash}=0$, and in the second case $y_m^{\backslash\mathtt{m}\backslash}=1$.
If $\{x\}=\bigcap_{m\in\mathbb{N}}\mathtt{I}_m$, then $x\notin\{x_n\}_{n\in\mathbb{N}}$. Notice that $x=0.\widehat{y_0^{\backslash\mathtt{0}\backslash}}\,\widehat{y_1^{\backslash\mathtt{1}\backslash}}\,\widehat{y_2^{\backslash\mathtt{2}\backslash}}\cdots$.
In the example, $x=0.100yy'\cdots=$ the anti-diagonal of $[y_i^{\backslash\mathtt{j}\backslash}]_{i,\mathtt{j}\in\mathbb{N}}$.

Some History

CANTOR's 2nd Proof [of $\mathbb{R} \not\approx \mathbb{N}$] (almost missing):

- 1994 **A.M.M.**: "We begin by analyzing Cantor's original articles, his 1874 article that contains his first proof and his 1891 article that contains his diagonal proof." (... ?)
- 2010 **A.M.M.**: "In 1874, two years before the publication of his famous diagonalization argument, Georg Cantor's first proof of the uncountability of the real numbers appeared in print···." (X)
- 2010 *Mathematics Magazine* 83(4):283–9, Cantor's Other Proofs that $\mathbb{R}$ Is Uncountable, by J. FRANKS. (✓✓)

Fixed-Point⇒Diagonal⇒Self-Referential

Generalized (Relational) Fixed-Point ≡ Self-Referential:

There is a (binary) *relation*, and **an object** is proved to exist that **is *related* to itself**, in the Theorem or in the Proof.

- Fixed-Point⇒Diagonal:

  For $F\colon I \to I$, let $a_{\langle i,j\rangle} = \begin{cases} 1 & \text{if } F(i)=j \\ 0 & \text{if } F(i)\neq j \end{cases}$ and $M = [a_{\langle i,j\rangle}]_{i,j\in I}$

  The fixed-points of F are indexed on the diagonal with entry 1.

- Diagonal⇒Self-Referential:

  Given $[a_{\langle i,j\rangle}]_{i,j\in I}$ the diagonal entry $a_{\langle k,k\rangle}$ relates $k \in I$ to itself.

Self-Referential ¿⟹? Diagonal ¿⟹? Fixed-Point

- Self-Referential ¿⟹? Diagonal
  Liar's Paradox? Descarte's Cogito? Non-Trivial Diagonal?
  "I am lying" $\neg(\lambda \leftrightarrow \neg\lambda)$ Cogito, ergo sum ("I think, therefore I am")
- Diagonal ¿⟹? Fixed-Point
  For the matrix $M = [a_{\langle i,j\rangle}(\in\mathcal{A})]_{i,j\in I}$, if for $f\colon \mathcal{A}\to\mathcal{A}$ the function $g(x)=f(a_{\langle x,x\rangle})$ is a-definable [[i.e., $g(x)=a_{\langle \Bbbk,x\rangle}$, for some $\Bbbk\in I$, or $g(x)=a_{\langle x,\Bbbk\rangle}$]], then f has a fixed point [[which is $a_{\langle \Bbbk,\Bbbk\rangle}$]].

$$\begin{array}{ccc} I^2 & \xrightarrow{\ a\ } & \mathcal{A} \\ {\scriptstyle\Delta}\uparrow & & \downarrow{\scriptstyle f} \\ I & \xrightarrow[\ g\ ]{} & \mathcal{A} \end{array}$$

Lawvere (CT 1969) & Yanofsky (BSL 2003).

Self-Referential / Diagonal / Fixed-Point

- B. Buldt (2016); "On Fixed Points, Diagonalizatin, and Self-Reference", in: *Von Rang und Namen*, Brill, pp. 47–64.

"··· diagonalization need not result in fixed points and fixed points need not be self-referential." (p. 48)

$$\text{diagonalization} \Longrightarrow \text{fixed points} \Longleftarrow \text{(objectual) self-reference}$$
$$\Downarrow$$
$$\text{incompleteness}$$

(p. 63)

" Yanofsky (2003) shows how all the usual suspects (i.e., paradoxes and limitative theorems) can be couched in terms of this framework and then follow from the generalized Cantor theorem. "

Diagonal Lemma (of GÖDEL and CARNAP), popularly

- C. SMORYŃSKI (*forthcoming*); The Early History of Formal Diagonalization, *Logic Journal of IGPL*, online 15 July 2022.

> "Linguistic self-reference goes back at least as far as the Greeks$\cdots$[to] a variant of the Liar paradox.
> Self-reference in formal languages, however, originated in Gödel's paper of 1931.
> In it, as we know, he presented the construction for a formula $\neg Pr_{PM}(v_0)$ of a sentence φ such that $\mathcal{PM} \vdash \varphi \leftrightarrow \neg Pr_{PM}(\ulcorner\varphi\urcorner)$.
> He also noted that the construction held for any extension $\mathcal{T}$ of $\mathcal{PM}$ which was primitive recursively axiomatized."

- C. S. (NDJFL 1981); Fifty Years of Self-Reference in Arithmetic.
- C. S. (1991); The Development of Self-Reference: Löb's Theorem.

Diagonal Lemma of GÖDEL, originally

GÖDEL 1931 (Collected Works, Vol. 1):

Let's write **diag**(y) for $Sb(y^{19}_{Z(y)})$, which results from substituting (all) the free variable(s) of y with the Gödel code of y. Let $Q(x,y)$ say that 《x is not a proof-code for the diagonal of y》 (p. 175). Since Q is [primitive] recursive, there is a "relation sign" (formula) q such that

if m is not a proof-code for **diag**(n), then $PM \vdash q(\overline{m},\overline{n})$ (9)

if m is a proof-code for **diag**(n), then $PM \vdash \neg q(\overline{m},\overline{n})$ (10).

Let $p(y) = \forall x\, q(x,y)$ and $r(x) = q(x, \ulcorner p(y) \urcorner)$.

"Then we have" **diag**$(p) = \forall x\, q(x, \ulcorner p \urcorner) = \forall x\, r(x)$ $[= G]$;

"furthermore" $q(\overline{m}, \ulcorner p \urcorner) = r(\overline{m})$. Now, (9,10) for $n = \ulcorner p \urcorner$ become

if m is not a proof-code for $G[=\forall x\, r(x)]$, then $T \vdash r(\overline{m})$,

and if m is a proof-code for $G[=\forall x\, r(x)]$, then $T \vdash \neg r(\overline{m})$.

Now, if $T \vdash_m G$, then $T \vdash \neg r(\overline{m})$ and $T \vdash \forall x\, r(x)$; so T is inconsistent!

If $T \vdash \neg G$, then $T \vdash \neg \forall x\, r(x)$ and $\bigwedge\!\!\!\bigwedge_m T \vdash r(\overline{m})$; so T is ω-inconsistent!

What Happened to $Q \vdash G \leftrightarrow \neg\mathtt{Pr}(\ulcorner G \urcorner)$?

Did Gödel have a formula $\boldsymbol{\pi}(x, y)$ for proof predicate such that

if m is a proof-code for ψ, then $PM \vdash \boldsymbol{\pi}(\overline{m}, \ulcorner\psi\urcorner)$

and if m is not a proof-code for ψ, then $PM \vdash \neg\boldsymbol{\pi}(\overline{m}, \ulcorner\psi\urcorner)$?

Could he show then that $PM \vdash G \leftrightarrow \neg\exists x\, \boldsymbol{\pi}(x, \ulcorner G \urcorner)$???

If we start from $\boldsymbol{\pi}$, then $\mathtt{Pr}(y) = \exists x\, \boldsymbol{\pi}(x, y)$. But since **diag** is not a function symbol in our language, we need a formula $\boldsymbol{\delta}(x, y)$ such that

if** m **is the code of** $\varphi[\vec{v}/\ulcorner\varphi\urcorner]$, **then $PM \vdash \forall z(\boldsymbol{\delta}(z, \ulcorner\varphi\urcorner) \leftrightarrow z = \overline{m})$.

Thus, if m is not the code of $\varphi[\vec{v}/\ulcorner\varphi\urcorner]$, then $PM \vdash \neg\boldsymbol{\delta}(\overline{m}, \ulcorner\varphi\urcorner)$.

Now, let $q(x, y) = \forall z[\boldsymbol{\delta}(z, y) \rightarrow \neg\boldsymbol{\pi}(x, z)]$. Note that $q, r, G \in \Pi_1$.

Yes, $PM \vdash G \leftrightarrow \neg\mathtt{Pr}(\ulcorner G \urcorner)$**!** for $G = \forall x\, q(x, \ulcorner \forall x\, q(x, y) \urcorner)$.

Diagonal Lemma of CARNAP, originally

- R. CARNAP (1934); *Logische Syntax der Sprache*, Springer.
  English translation: A. SMEATON, *The Logical Syntax of Language*, Kegan Paul, Trench, Trubner & Co Ltd (1937). (page 130)

  " Let any syntactical property of expressions be chosen $\cdots$. Let $\mathfrak{S}_1$ be the sentence with the free variable '*x*' (for which we will take the term-number $\mathfrak{z}$) which expresses this property $\cdots$. Let $\mathfrak{S}_2$ be that sentence which results from $\mathfrak{S}_1$ if for '*x*' 'subst[*x*,$\mathfrak{z}$,str(*x*)]' is substituted. $\cdots$ Thus, if $\mathfrak{S}_2$ is given, the series-number of $\mathfrak{S}_2$ can be calculated; let it be designated by 'b' ('b' is a defined $\mathfrak{zz}$). Let the $^{\mathrm{SN}}$sentence subst[b,$\mathfrak{z}$,str(b)] be $\mathfrak{S}_3$; thus $\mathfrak{S}_3$ is the sentence which results from $\mathfrak{S}_2$ when the $\mathfrak{St}$ with the value b is substituted for '*x*'. It is easy to see that, syntactically interpreted, $\mathfrak{S}_3$ measn that $\mathfrak{S}_3$ itself has the chosen syntactical property."

Diagonal / Self-Referential Lemma

- GÖDEL: There exists a formula $r(x)$ such that for every $m \in \mathbb{N}$:
  if m is *not* a T-proof-code for $\forall x\, r(x)$, then $T \vdash r(\overline{m})$,
  and if m is a T-proof-code for $\forall x\, r(x)$, then $T \vdash \neg r(\overline{m})$.
- CARNAP: For every formula $\mathsf{F}(x)$ there is a sentence σ such that
  σ is true iff $\mathsf{F}(\ulcorner\sigma\urcorner)$ is true. (Semantic Diagonal Lemma)
- ROSSER(1936,37,39); KREISEL(1950,53); HENKIN(1952);
  TARSKI-MOSTOWSKI-ROBINSON(1953,68,71,2010[1938-9]);
  LÖB(1955); — MOSTOWSKI(1952).
- FEFERMAN(1960); MONTAGUE(1962);
  KREISEL-TAKEUTI(1974); SMORYŃSKI(1977) ...
  - For every formula $\mathsf{F}(x)$ there is a sentence σ such that
    $$Q \vdash \sigma \leftrightarrow \mathsf{F}(\ulcorner\sigma\urcorner).$$

More History

- B. ROSSER (1939); An Informal Exposition of Proofs of Gödel's Theorems and Church's Theorem, *J. Symbolic Logic* 4(2):53–60.

"**LEMMA 1.** *Let "x has the property Q" be expressible in L. Then for suitable L, there can be found a sentence F of L, with a number n, such that F expresses "n has the property Q." That is, F expresses "F has the property P."* [Formula has the property P iff its number has the property Q].
$\cdots$ "for suitable L" [means] that "$z=\phi(x,x)$" [is] expressible in $L \cdots$

DEFINITION. $\phi(x,y)$ is the number of the formula got by taking the formula with the number x and replacing all occurrences of v in it by the term of L which denotes the number of y.

[PROOF.] Let G be the formula of L which expresses "$\phi(x,x)$ has the property Q." G has a number, n. Now get F from G by replacing all v's of G by the term of L which denotes n. Then F denotes "$\phi(n,n)$ has the property Q" $\cdots$. However$\cdots$, $\phi(n,n)$ is the number of F, because F was got by taking the formula with the number n and replacing all occurrences of v in it by the term of L which denotes n. So F expresses "the number of F has the property Q," that is "F has the property P." "

Even More History

- G. KREISEL (1950); Note on Arithmetic Models for Consistent Formulae of the Predicate Calculus, *Fund. Math.* 37(1):265–85.

"$\cdots$ what Gödel [did was] to apply the diagonal definition to a system of predicates which are not *systematically decidable*, but quantified; now we must expect that the formal definition of the diagonal predicate is of the given sequence $\mathfrak{A}_n(m)$, say the p^{th}; then $\mathfrak{A}_p(p)$ is undecided in the system. This situation occurs in$\cdots$ Gödel's argument. $\cdots s(a, b)$ is a function whose value is the number of the expression got when the free variable in the expression with number b is replaced by the number a. Then Gödel orders all expressions of a formalism by his numbering, so that, say, $\mathfrak{A}_n(\alpha)$ with the free variable α has the number n. He considers the sequence of formulae $\exists y\, \mathbf{prf}[y, s(m, n)]$ which will be provable if $\mathfrak{A}_n(m)$ can be proved in the system. The [anti-]diagonal definition is $\forall y \neg \mathbf{prf}[y, s(n, n)]$ and$\cdots$; i.e. the [anti-]diagonal definition is one of the sequence, and here the diagonal argument establishes undecidability."

A Fixed-Point Lemma?

- For every formula $F(x)$ there is a sentence σ such that

$$Q \vdash \sigma \leftrightarrow F(\ulcorner\sigma\urcorner).$$

Looks Like a Fixed-Point!?

Consider $\psi \mapsto F(\ulcorner\psi\urcorner)$. Under monotone codings, $\ulcorner F(\ulcorner\psi\urcorner)\urcorner > \ulcorner\psi\urcorner$.

Let $\mathfrak{F}\colon \mathit{Sent}_{\mathfrak{T}} \to \mathit{Sent}_{\mathfrak{T}}$ be $\mathfrak{F}([\psi]_{\mathfrak{T}}) = [F(\ulcorner\psi\urcorner)]_{\mathfrak{T}}$.

A fixed-point is $[\sigma]_{\mathfrak{T}} = [F(\ulcorner\sigma\urcorner)]_{\mathfrak{T}}$, or $\mathfrak{T} \vdash \sigma \leftrightarrow F(\ulcorner\sigma\urcorner)$.

If $\mathfrak{F}$ is a well-defined function: $\mathfrak{T} \vdash \varphi \leftrightarrow \psi \;\Rightarrow\; \mathfrak{T} \vdash F(\ulcorner\varphi\urcorner) \leftrightarrow F(\ulcorner\psi\urcorner)$.

- GÖDEL's: $\mathfrak{T} \vdash \varphi \leftrightarrow \psi \;\Rightarrow\; \mathfrak{T} \vdash \neg\mathtt{Pr}(\ulcorner\varphi\urcorner) \leftrightarrow \neg\mathtt{Pr}(\ulcorner\psi\urcorner)$.
- CARNAP's: Let $H(x)$ say that "x starts with $\neg$", and let A be a $\neg$-free sentence. Then $A \equiv \neg\neg A$, but $H(\ulcorner A\urcorner)$ is false while $H(\ulcorner\neg\neg A\urcorner)$ is true. So, $[\psi]_{\mathfrak{T}} \mapsto [H(\ulcorner\psi\urcorner)]_{\mathfrak{T}}$ is not well-defined.

Strong Diagonal/Direct Self-Referential Lemma

LEMMA. *In a sufficiently expressive language* $\forall F(x)\exists\sigma\colon \sigma = F(\ulcorner\sigma\urcorner)$.

Proof. Recall $\mathtt{diag}(\ulcorner\varphi\urcorner) = \ulcorner\varphi[\vec{v}/\ulcorner\varphi\urcorner]\urcorner$. Let $n = \ulcorner F(\mathtt{diag}(x))\urcorner$ and $\sigma = F(\mathtt{diag}(\overline{n}))$. Then $\sigma = F(\ulcorner F(\mathtt{diag}(n))\urcorner) = F(\ulcorner\sigma\urcorner)$. ■

- R.G. JEROSLOW (1973); Redundancies in the Hilbert-Bernays Derivability Conditions for Gödel's 2nd Thm, *JSL* 38(3):359–67.

  "The··· lemma was discovered by the referee···."

LEMMA. *There are Gödel codings (computable injections* $\eta \mapsto \llcorner\eta\lrcorner$ *from strings to closed terms) such that* $\forall F(x)\exists\sigma\colon \sigma = F(\llcorner\sigma\lrcorner)$.

- S.A. KRIPKE (1975); Outline of a Theory of Truth, *The Journal of Philosophy* 72(19):690–716.
- A. VISSER (1989); "Semantics and the Liar Paradox", *Handbook of Philosophical Logic* **IV**, pp. 617–706 (2^{nd} ed. 2004, **11**, pp. 149–240).
- S.A. KRIPKE (*forthcoming*); Gödel's Theorem and Direct Self-Reference, *Review of Symbolic Logic*, online 02 December 2021.

Where is the Original GÖDEL-CARNAP Lemma?

- $\forall F(x)\exists\sigma\colon Q\vdash\sigma\leftrightarrow F(\ulcorner\sigma\urcorner)$.

Many sentences σ leak in.

- GÖDEL-CARNAP: write $F(x)=\forall y\,\theta(y,x)$ $[\theta=\neg\mathtt{prf}]$; let $q(y,z)=\theta(y,\mathtt{diag}(z))$,[†] $p(z)=\forall y\,q(y,z)$, $r(y)=q(y,\ulcorner p(z)\urcorner)$, and $\sigma=\forall y\,r(y)$. Then, we have $\mathtt{diag}(\ulcorner p(z)\urcorner)=\ulcorner\sigma\urcorner$,[‡] so $\sigma=\forall y\,\theta\big(y,\mathtt{diag}(\ulcorner p(z)\urcorner)\big)=\forall y\,\theta(y,\ulcorner\sigma\urcorner)=F(\ulcorner\sigma\urcorner)$.[§]

GÖDEL had **diag** at his disposal, but didn't use it!

- $\forall F(x)\exists\sigma\colon\sigma=F(\ulcorner\sigma\urcorner)$.

[†] $q(y,z)=\forall w\,[\boldsymbol{\delta}(w,z)\rightarrow\theta(y,w)]$ or $q(y,z)=\exists w\,[\boldsymbol{\delta}(w,z)\wedge\theta(y,w)]$,

[‡] $\boldsymbol{\delta}(\ulcorner\sigma\urcorner,\ulcorner p(z)\urcorner)$ $[\dashv Q]$.

[§] $\sigma\leftrightarrow F(\ulcorner\sigma\urcorner)$ $[\dashv Q]$.

THANK YOU!

Thanks to

The Participants ....For Listening $\cdots$

and

The Organizers, For Taking Care of Everything $\cdots$