TARSKI'S UNDEFINABILITY OF TRUTH THEOREM STRIKES AGAIN

Ali Enayat

May 7^{th} and 16^{th}, 2024

MOPA, CUNY

Outline

- **PART 1** revisits Tarski's Undefinability of Truth (**TUT**), as originally presented in the following.

  Andrzej Mostowski, Raphael Robinson, Alfred Tarski, *Decidability and Essential Undecidability in Arithmetic*, in **Undecidable Theories**, North Holland, Amsterdam, 1953.

  Alfred Tarski, *Two general theorems on undefinability and undecidability*, **Bulletin of American Mathematical Society** (1953), pp. 365-366.

- **PART 2** reports on the following paper in which TUT together with partial definable truth predicates join hands to deliver new incompleteness theorems.

  Ali Enayat and Albert Visser, *Incompleteness of boundedly axiomatizable theories*, arXiv:2311.14025 [math.LO]

Three relevant subtheories of PA

R1. $\vdash \underline{m} + \underline{n} = \underline{m+n}$
R2. $\vdash \underline{m} \cdot \underline{n} = \underline{m \cdot n}$
R3. $\vdash \underline{m} \neq \underline{n}$, for $m \neq n$
R4. $\vdash x \leq \underline{n} \to \bigvee_{i \leq n} x = \underline{i}$
R5. $\vdash x \leq \underline{n} \vee \underline{n} \leq x$

Q1. $\vdash \mathsf{S}x = \mathsf{S}y \to x = y$
Q2. $\vdash \mathsf{S}x \neq 0$
Q3. $\vdash x = 0 \vee \exists y \;\; x = \mathsf{S}y$
Q4. $\vdash x + 0 = x$
Q5. $\vdash x + \mathsf{S}y = \mathsf{S}(x + y)$
Q6. $\vdash x \cdot 0 = 0$
Q7. $\vdash x \cdot \mathsf{S}y = x \cdot y + x$

PA^-1. $\vdash x + 0 = x$
PA^-2. $\vdash x + y = y + x$
PA^-3. $\vdash (x + y) + z = x + (y + z)$
PA^-4. $\vdash x \cdot 1 = x$
PA^-5. $\vdash x \cdot y = y \cdot x$
PA^-6. $\vdash (x \cdot y) \cdot z = x \cdot (y \cdot z)$
PA^-7. $\vdash x \cdot (y + z) = x \cdot y + x \cdot z$
PA^-8. $\vdash x \leq y \vee y \leq x$
PA^-9. $\vdash (x \leq y \wedge y \leq z) \to x \leq z$
PA^-10. $\vdash x + 1 \not\leq x$
PA^-11. $\vdash x \leq y \to (x = y \vee x + 1 \leq y)$
PA^-12. $\vdash x \leq y \to x + z \leq y + z$
PA^-13. $\vdash x \leq y \to x \cdot z \leq y \cdot z$
PA^-14. $\vdash x \leq y \to \exists z \;\; x + z = y$

Semantic form of TUT

- Let $\mathcal{L}$ be a language (signature), and $\mathcal{M}$ be an $\mathcal{L}$-structure. Also let $\varphi \mapsto \#(\varphi) \in M$ is an arbitrary mapping of unary $\mathcal{L}$-formulae into M.
- **Theorem.** (Semantic Form of TUT) *There is no binary $\mathcal{L}$-formula $T(x, y)$ such that for all unary $\mathcal{L}$-formulae*: $\mathcal{M} \models \forall x \, (T\,(x, \#(\varphi)) \leftrightarrow \varphi(x))\,.$

  Proof. Suppose not and consider $R(x) = \neg T(x, x)$. Then:

  (1) $\mathcal{M} \models \forall x \, (T\,(x, \#(R)) \leftrightarrow R(x))\,.$ If $r := \#(R)$, by (1) and the definition of R we obtain:

  (2) $\mathcal{M} \models T(r, r) \leftrightarrow R(r) \leftrightarrow \neg T(r, r),$ contradiction. □
- The above proof is reminiscent of the proof of Russell's Paradox (1901), and of the proof of Cantor's theorem (1891) on nonexistence of a surjection of a set X onto $\mathcal{P}(X)$.

Kripke's formulation of Cantor's Theorem

Cantor's Diagonal Principle

A relation is called *arithmetical* if it is definable in L, the language of arithmetic. Since L contains RE, it follows that all r.e. relations are arithmetical. Also, since L contains negation, it follows that all complements of r.e. relations are arithmetical. That L contains negation also implies that the enumeration theorem fails for arithmetical sets, i.e. there is no arithmetical relation that enumerates all the arithmetical relations; similarly, there is no recursive relation that enumerates all the recursive relations.

The best way to see this is by proving a general theorem. As in the enumeration theorem for r.e. sets, if R is a two-place relation, we write R_x for {y: R(x, y)}. We give the following

Definition: Let X be a set, F be a family of subsets of X, and R a two place relation defined on X. R is said to *supernumerate* F iff for any $S \in F$, there is an $x \in X$ such that $S = R_x$. R is said to *enumerate* F iff R supernumerates F and for all $x \in X$, $R_x \in F$.

The content of the enumeration theorem is thus that there is an r.e. relation which enumerates the r.e. sets. Next we have

Cantor's Diagonal Principle: The following two conditions are incompatible:

(i) R supernumerates F

(ii) The complement of the*Diagonal Set* is in F (the Diagonal Set is $\{x \in X: R(x, x)\}$).

Proof: Suppose (i)-(ii) hold. Then by (ii) $X - \{x \in X: R(x, x)\} = \{x \in X: \sim R(x, x)\} \in F$. By (i), $\{x \in X: \sim R(x, x)\} = R_y$ for some y. But then R(y, x) iff ~R(x, x) for all $x \in X$, so in particular R(y, y) iff ~R(y, y), contradiction.

Source: p.66 of Kripke's Lecture Notes on Elementary Recursion Theorem, Princeton, 1996

Semantic form of TUT, cont'd

- Let $\mathrm{Form}^n_{\mathcal{L}}$ = the set of n-ary $\mathcal{L}$-formulae, and suppose $\mathcal{M}$ is an $\mathcal{L}$-structure with a pairing function $\langle \cdot, \cdot \rangle$. Also assume that the coding $\varphi \mapsto \#(\varphi) \in M$ is 1-1. In this context, the (codes of) sentences in the elementary diagram of $\mathcal{M}$ can be split into:

$$\mathrm{ED}^+(\mathcal{M}) = \{\langle \#(\varphi), \langle a_1, ..., a_n \rangle\rangle \in M : \mathcal{M} \models \varphi(a_1, ..., a_n), \varphi \in \mathrm{Form}^n_{\mathcal{L}}\},$$
$$\mathrm{ED}^-(\mathcal{M}) = \{\langle \#(\varphi), \langle a_1, ..., a_n \rangle\rangle \in M : \mathcal{M} \models \neg\varphi(a_1, ..., a_n), \varphi \in \mathrm{Form}^n_{\mathcal{L}}\}.$$

- **Corollary**. (Inseparability of positive and negative fragments ED) $\mathrm{ED}^+(\mathcal{M})$ *and* $\mathrm{ED}^-(\mathcal{M})$ *are definably inseparable in* $\mathcal{M}$*, i.e., there is no* $\mathcal{M}$*-definable* D *(parameters allowed) such that* $\mathrm{ED}^+(\mathcal{M}) \subseteq D$ *and* $\mathrm{ED}^-(\mathcal{M}) \cap D = \varnothing$.

Semantic form of TUT, concluded

- **Corollary**. (Incompleteness of PA)
  Let $\mathrm{TA} = \{\#(\varphi) : (\omega, +, \cdot) \models \varphi,\ \varphi \in \mathsf{Sent}_{\mathcal{L}_{\mathrm{PA}}} := \mathsf{Form}^0_{\mathcal{L}_{\mathrm{PA}}}\}$.
  TA *is not definable in* $(\omega, +, \cdot)$.
  In particular, TA *is not axiomatizable by a subtheory of itself the set of whose* $\#$-codes *is definable in* $(\omega, +, \cdot)$. *Hence* PA *is incomplete*.

Syntactic Formulation of TUT

- Let T be an $\mathcal{L}$-theory, and suppose $n \mapsto \mathbb{n}$ be an arbitrary mapping of ω (natural numbers) into the set of closed $\mathcal{L}$-terms (terms with no free variables).
- Fix an arbitrary 1-1 correspondence $\varphi \mapsto \#(\varphi)$ between $\mathrm{Form}_{\mathcal{L}}^{\leq 1}$ and ω, and let $n \longmapsto \varphi_n$ be its inverse.
- The *diagonal function* $\delta : \omega \to \omega$ is given by

$$\delta(n) = \#(\varphi_n(\mathbb{n})).$$

- A function $f : \omega \to \omega$ is said to be *T-definable* if there is an $\mathcal{L}$-formula $\theta(x, y)$ such that

$$\forall n \in \omega \;\; T \vdash \forall y \;\; [\theta(\mathbb{n}, y) \leftrightarrow y = \mathbb{f}(\mathbb{n})] \, .$$

- A subset P of ω is said to be *T-definable* if there is an $\mathcal{L}$-formula $\psi(x)$ such that:

$$\forall n \in P \;\; T \vdash \psi(\mathbb{n}) \quad \textbf{and} \quad \forall n \notin P \;\; T \vdash \neg\psi(\mathbb{n}).$$

Syntactic Formulation of TUT (cont'd)

- **Theorem 1.** (Syntactic formulation of TUT, ver 1.)
  Given a theory T, let $V_T = \{\#(\varphi) : T \vdash \varphi\}$. Assuming that T is consistent, then the diagonal function δ and the set V_T are not both T-definable.

- **Corollary.** (Syntactic formulation of TUT, ver. 2).
  If T is a consistent $\mathcal{L}$-theory such that δ is T-definable, then there is no $\mathcal{L}$-formula $\theta(x)$ such that for all $\mathcal{L}$-sentences φ we have:

  $$T \vdash \varphi \leftrightarrow \theta(\mathbb{n}), \text{ where } n = \#(\varphi).$$

- **Corollary.** *If T is a consistent theory such that all total recursive functions are T-representable, and $\varphi \mapsto \#(\varphi)$ is recursive, then V_T is not recursive. In particular, T is incomplete.*

- **Remark.** *If T interprets Robinson's* R *(let alone Robinson's* Q*) , then represents all total recursive functions are T-definable.*

Proof of version 1 of TUT

Suppose not, thus there are formulae θ and ψ such that the following hold:

(1) $\forall n \in \omega \quad T \vdash \forall y\,[\theta(\mathbb{n}, y) \leftrightarrow y = \mathbb{r}]$, where $\delta(n) = r$.
(2) $\forall n \in V_T \quad T \vdash \psi(\mathbb{n})$.

(3) $\forall n \notin V_T \quad T \vdash \neg\psi(\mathbb{n})$.

Choose $m \in \omega$ such that $\varphi_m(x) = \forall y\,(\theta(x, y) \to \neg\psi(y))$, hence:

(4) $\varphi_m(\mathbb{m}) = \forall y\ \ (\theta(\mathbb{m}, y) \to \neg\psi(y))$.

If $T \vdash \varphi_m(\mathbb{m})$, then by (1) and (4) we have $T \vdash \neg\psi(\mathbb{k})$, where $\delta(m) = k$.

If $T \nvdash \varphi_m(\mathbb{m})$, then $\#(\varphi_m(\mathbb{m})) \notin V_T$; and by the definition of δ,

(5) $\delta(m) = \#(\varphi_m(\mathbb{m}))$.

So by (3) in this case we can also conclude that $T \vdash \neg\psi(\mathbb{k})$. Thus we have shown:

(6) $T \vdash \neg\psi(\mathbb{k})$.

By (1) and (6), $T \vdash \forall y\,(\theta(\mathbb{m}, y) \to \neg\psi(y))$. So by (4) and (5) $\delta(m) \in V_T$ and therefore by (2),

(7) $T \vdash \psi(\mathbb{k})$.

This contradicts the assumption of consistency of T.

Tarski's 1953 abstract

418*t*. Alfred Tarski: *Two general theorems on undefinability and undecidability.*

This paper contains a generalization of ideas known from works of Gödel and other authors. See specifically Mostowzki, *Sentences undeciable* $\cdots$, Amsterdam, 1952; R. M. Robinson, Proceedings of the International Congress of Mathematicians, 1950, vol. 1; Tarski, Studia Philosophica vol. 1. Assumptions: $\mathfrak{T}$ is any formalized theory; S is a set of $\mathfrak{T}$-formulas including all axioms of predicate calculus with identity and closed under rules of inference; $\Delta_0, \Delta_1, \cdots, \Delta_n, \cdots$ are $\mathfrak{T}$-terms containing no variables; $\sim(\Delta_0 = \Delta_1)$ is in S; x, y are fixed $\mathfrak{T}$-variables. A function F on and to the integers is called S-definable (relative to Δ_n) if, for some formula Φ and every integer n, $(x = \Delta_n) \rightarrow [(y = \Delta_{F(n)} \leftrightarrow \Phi]$ is in S. Consider a quite arbitrary one-one correlation between $\mathfrak{T}$-expressions Ψ and integers n; $Nr\ (\Psi)$ is the integer correlated with Ψ, Ω_n is the expression correlated with n. Let $D(n) \equiv Nr[(x = \Delta_n) \rightarrow \Omega_n]$; let $P(n)$ be 0 if Ω_n is in S, and 1 otherwise. Theorem I. *If S is consistent, then functions D and P are not both S-definable.* New assumptions: $G(n) \equiv Nr(\Delta_n)$ and $H(n, p) \equiv Nr(\Omega_n \frown \Omega_p)$ (where $\frown$ is the concatenation symbol) are general recursive functions. Then Theorem I implies: Theorem II. *If all general recursive functions are S-definable, then S is inconsistent or essentially undecidable.* (Received January 16, 1953.)

Tarski's assessment

The idea of this reconstruction and the realization of its far-reaching implications is due to Gödel [7]. The present version of this reconstruction is distinguished by its generality and simplicity. It applies to arbitrary formalized theories, and not only to those in which a comprehensive fragment of the arithmetic of natural numbers can be developed; to a large extent it is independent of the way in which the notion of validity has been defined for a given theory, and in particular it does not involve the notion of a formal proof within this theory; it does not use the apparatus of recursive functions—although this apparatus will play a fundamental role in applications of Theorem 1 to the decision problem.

For discussion on the history of TUT, see:

- R. Murawski, *Undefinability of truth; the problem of priority: Tarski vs Gödel*, **History and Philosophy of Logic** (1988), Vol. 19, 153–160.
- J. Woleński, *Gödel, Tarski and Truth*, **Revue Internationale de Philosophie** (2005) Vol. 59, pp. 459–490.

(END OF PART 1)

A Question of Lempp and Rossegger

- PA^- is the finitely axiomatized fragment of PA whose axioms describe the non-negative substructure of discretely ordered rings (with no instance of the induction scheme, hence the minus superscript).
- **Question.** *Is there a consistent completion of* $T = \mathrm{PA}^-$ *that is axiomatized by a set of sentences of bounded quantifier complexity*?
- The above question was posed by Steffen Lempp and Dino Rossegger in the context of their recent joint work [AGLRZ] with Uri Andrews, David Gonzalez, and Hongyu Zhu, in which they establish:

  Theorem. *The following are equivalent for a complete first-order theory* T:

  (1) *The set of models of* T *is* Π^0_ω*-complete under Wadge reducibility (i.e., reducibility via continuous functions).*

  (2) T *does not admit a first-order axiomatization by formulae of bounded quantifier complexity.*

  [AGLRZ] *The Borel complexity of the class of models of first-order theories*, arXiv:2402.10029[math.LO]

Two relevant hierarchies of formulae

Bounded Quantifiers and Arithmetical Hierarchy. $(\exists x \leq y)\varphi$ is an abbreviation for $(\exists x)(x \leq y \,\&\, \varphi)$ and $(\forall x \leq y)$ is an abbreviation for $(\forall x)(x \leq y \rightarrow \varphi)$. By convention, x and y must be *distinct variables*. An L_0-formula is *bounded* if all quantifiers occuring in it are bounded, i.e. occur in a context as above. Furthermore, $(\forall x < y)\varphi$ is an abbreviation for $(\forall x \leq y)(x \neq y \rightarrow \varphi)$ and similarly for $(\forall x < y)$; $x \neq y$ is the same as $\neg(x = y)$.

We introduce a hierarchy of formulas called the *arithmetical hierarchy.* Σ_0-formulas = Π_0-formulas = bounded formulas; Σ_{n+1}-formulas have the form $(\exists x)\varphi$ where φ is Π_n, Π_{n+1}-formulas have the form $(\forall x)\varphi$ where φ is Σ_n. Thus a Σ_n-formula has a block of n alternating quantifiers, the first one being existential, and this block is followed by a bounded formula. Similarly for Π_n.

- $\Sigma_0^* := \Pi_0^* := \emptyset$.
- $\Sigma_{n+1}^* ::=$ AT $\mid \neg\,\Pi_{n+1}^* \mid (\Sigma_{n+1}^* \wedge \Sigma_{n+1}^*) \mid (\Sigma_{n+1}^* \vee \Sigma_{n+1}^*) \mid (\Pi_{n+1}^* \rightarrow \Sigma_{n+1}^*) \mid \exists v\, \Sigma_{n+1}^* \mid \forall v\, \Pi_n^*$.
- $\Pi_{n+1}^* ::=$ AT $\mid \neg\,\Sigma_{n+1}^* \mid (\Pi_{n+1}^* \wedge \Pi_{n+1}^*) \mid (\Pi_{n+1}^* \vee \Pi_{n+1}^*) \mid (\Sigma_{n+1}^* \rightarrow \Pi_{n+1}^*) \mid \forall v\, \Pi_{n+1}^* \mid \exists v\, \Sigma_n^*$.

Some background results (1)

- Let $\mathbb{N}$ be the standard model of PA. For $T = \mathrm{TA} = \mathrm{Th}(\mathbb{N})$ the answer to the Lempp-Rossegger question is in the negative. This follows from the Arithmetical Hierarchy Theorem of Kleene (1943) and Mostowski (1946) that states that $\Sigma_n^{\mathbb{N}} \subsetneq \Sigma_{n+1}^{\mathbb{N}}$ for each $n \in \omega$.
- For $T = \mathrm{PA}$, the negative answer follows from a theorem of Rabin (1961) that states that for each $n \in \omega$ no consistent extension of PA (in the same language) is axiomatized by a set of Σ_n-sentences.
- Rabin's result refines an earlier theorem of Ryll-Nardzewski (1952) that states that no consistent extension of PA is finitely axiomatizable. Ryll-Nardzewski and Rabin both employed model-theoretic arguments relying on **nonstandard** elements to prove the aforementioned results (Theorem 10.2 of Kaye's **Models of Peano Arithmetic** offers a modern treatment).

Some background results (2)

- Rabin's result can be also established with an argument that mixes proof-theoretic machinery, partial satisfaction classes, and Gödel's second incompleteness theorem (see Theorem 2.36 of Chapter III of Hájek and P. Pudlák's **Metamathematics of First Order Order Arithmetic**).
- As shown by Montague (1961) a similar result can be established for any *inductive* sequential theory T, i.e., a sequential theory that has the power to prove the full scheme of induction over its 'natural numbers' for all formulae in the language of T. In the setting of Montague's result the relevant hierarchy is based on quantifier-alternations-depth.
- Canonical examples of inductive sequential theories include all extensions of PA, Z (Zermelo set theory), Z_2 (second order arithmetic), and KM (Kelley-Morse theory of classes).

Sequential Theories

- At first approximation, a theory is sequential if it supports a modicum of coding machinery to handle finite sequences of all objects in the domain of discourse. Gödel (1931) used the Chinese Remainder Theorem to show that PA is sequential. Jeřábek (2012) showed that PA^{-} is sequential, and Visser (2008) showed that Q is not sequential.
- It is known that T is sequential iff T has a definitional extension to *Adjunctive Set Theory*. The original definition of sequentiality due to Pudlák is as follows: A theory T is sequential if there is a formula $N(x)$ (read as "x is a number") , together with appropriate formulae providing interpretations of equality, and the operations of successor, addition, and multiplication for elements satisfying $N(x)$ such that T proves the translations of the axioms of Q when relativized to $N(x)$; and additionally, there is a formula $\beta(x, i, w)$ (whose intended meaning is that x is the i-th element of a sequence w) such that T proves that every sequence can be extended by any given element of the domain of discourse.

Surprising power of sequential theories

- The definition of sequentiality is self-improving: the 'numbers' of a sequential theory can be required to satisfy $\mathsf{I}\Delta_0$, thanks to a result of Wilkie (1980s) that shows that Q can interpret $\mathsf{I}\Delta_0$ 'on a cut'.
- The following result was established by Visser (1993, 2019); this result refines the work of Pudlák (1984, 1998) in which *logical depth* (length of the longest branch in the formation tree of the formula) is used as a measure of complexity instead of the depth of quantifier alternations complexity.
- **Fact F.** *Suppose T is a sequential theory T formulated in a finite language $\mathcal{L}$, and fix $n \in \omega$. Fix some interpretation $\mathcal{N}$ of arithmetic in T satisfying* $\mathsf{I}\Delta_0$.

  (a) *There is a T-provable definable cut I_n* of $\mathcal{N}$ *and a formula* $\mathrm{Sat}_n(x, y)$ *such that, provably in T,* Sat_n *satisfies the Tarskian compositional clauses if* x is a Σ_n^**-formulae in I_n (and for all variable assignments y).* Therefore:

  (b) *There is a formula* $\mathrm{True}_n(x)$ *such that, provably in T,* $\mathrm{True}_n(x)$ *is extensional, i.e., it respects the equivalence relation representing equality in the interpretation $\mathcal{N}$; and for all models $\mathcal{M} \models T$, and for all Σ_n^*-sentences ψ, we have:* $\mathcal{M} \models (\psi \leftrightarrow \mathrm{True}_n(\ulcorner \psi \urcorner))$.

Commercial Break:Telegraphic History of Definable Partial Truth Predicates:

- Turing, Post, Kleene, Mostowski (1940s) – $0^{(n)}$ is Σ_n-complete.
- Mostowski (1952) – PA supports a definable truth predicate for Σ_n-formulae.
- Montague (1961) – Every inductive sequential theory supports a definable truth predicate for Σ_n^*-formulae.
- Levy (1965) – ZF supports a definable truth predicate for Σ_n^{Levy}-formulae.
- Gaifman and Dimitracopoulos (1980) – $\mathrm{I}\Delta_0 + \mathsf{Exp}$, supports a definable truth predicate for Σ_n-formulae.
- Pudlák (1984, 1998) – Every sequential theory supports a definable truth predicate for Depth_n-formulae.
- Visser (1994, 2019) – Every sequential theory supports a definable truth predicate for Σ_n^*-formulae.

Theorem A

Theorem A. *For any fixed* $n \in \omega$, *every consistent sequential theory formulated in a finite language that is axiomatized by a set of* Σ_n^**-sentences is incomplete*.

Proof of Theorem A. Suppose not, and let T be consistent completion of sequential theory formulated in a finite language $\mathcal{L}$. Then by the definition of sequentiality T is also sequential. Suppose to the contrary that for some $n \in \omega$, T is axiomatized by a set of Σ_n^* sentences, i.e., suppose (1) below:

(1) For $n \in \omega$, there is a set A of Σ_n^* sentences such that for all $\mathcal{L}$-sentences ψ, $\psi \in T$ iff $A \vdash \psi$.

Our proof by contradiction of Theorem A will be complete once we verify Claim ♡ below since it contradicts TUT.

CLAIM ♡. There is a unary $\mathcal{L}$-formula $\varphi(x)$ such that for all $\mathcal{L}$-sentences ψ, $T \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$.

Theorem A (cont'd)

Since T is sequential, we can find an $\mathcal{L}$-formula, denoted $\mathsf{Prf}_{\mathrm{True}_n}(\pi, x)$ such that for each *standard* $\mathcal{L}$-sentence ψ and *standard* π, and each model $\mathcal{M}$ of T, we have:

(2) $\mathcal{M} \models \mathsf{Prf}_{\mathrm{True}_n}(\pi, \ulcorner\psi\urcorner)$ iff π is (a code for) a proof of ψ from

$$\mathsf{True}_n^{\mathcal{M}} := \{\varphi : \mathcal{M} \models \mathsf{True}_n(\ulcorner\varphi\urcorner)\}.$$

Our proposed candidate of $\varphi(x)$ for establishing Claim ♡ is the following formula $\rho(x)$; our choice of the letter ρ indicates the fact that the formula expresses *Rosser-provability* (from the true Σ_n^* sentences).

$$\rho(x) := \exists y \left[\mathsf{Prf}_{\mathsf{True}_n}(y, x) \wedge \forall z < y \ \neg\mathsf{Prf}_{\mathsf{True}_n}(z, \neg x)\right].$$

Thus our goal is to show that for all $\mathcal{L}$-sentences ψ, $T \vdash \psi \leftrightarrow \rho(\ulcorner\psi\urcorner)$. It suffices to show that for each model $\mathcal{M}$ of T, $\mathcal{M} \models \psi \leftrightarrow \rho(\ulcorner\psi\urcorner)$. For the rest of the proof, let $\mathcal{M} \models T$. We will first show:

(3) For all $\mathcal{L}$-sentences ψ, $\mathcal{M} \models \psi \rightarrow \rho(\ulcorner\psi\urcorner)$. To show (3), assume ψ holds in $\mathcal{M}$. Let n and A be as in (1), and note that $A \subseteq \mathsf{True}_n^{\mathcal{M}}$.

Theorem A (concluded)

By the assumptions about T, there are finitely many sentences $\alpha_1, ..., \alpha_n$ in A such that $\{\alpha_1, ..., \alpha_n\} \vdash \psi$. Let $\pi_0 \in \omega$ be (the code of) a proof of ψ from $\{\alpha_1, ..., \alpha_n\}$. Thanks to (2) we have: $\mathcal{M} \models \mathsf{Prf}_{\mathsf{True}_n}(\pi_0, \ulcorner\psi\urcorner)$. The assumption of consistency of T coupled with (2) yields: $\mathcal{M} \models \forall z < \pi_0 \, \neg\mathsf{Prf}_{\mathsf{True}_n}(z, \ulcorner\neg\psi\urcorner)$. Hence (3) holds.

To complete the proof of CLAIM ♡, we need to show that $\mathcal{M} \models \neg\psi \rightarrow \neg\rho(\ulcorner\psi\urcorner)$ for all $\mathcal{L}$-sentences ψ. For this purpose assume $\mathcal{M} \models \neg\psi$. By putting (1) and the assumption that $\mathcal{M} \models \neg\psi$, we conclude that there is a *standard* proof π_0 of $\neg\psi$ from $\mathsf{True}_n^{\mathcal{M}}$, which by (2) implies:

(4) For some $\pi_0 \in \omega$, $\mathcal{M} \models \mathsf{Prf}_{\mathsf{True}_n}(\pi_0, \ulcorner\neg\psi\urcorner)$.

To see that $\mathcal{M} \models \neg\rho(\ulcorner\psi\urcorner)$ suppose to the contrary that $\mathcal{M} \models \rho(\ulcorner\psi\urcorner)$. By the choice of ρ, this means:

(5) For some $m_0 \in M$, $\mathcal{M} \models \mathsf{Prf}_{\mathsf{True}_n}(m, \ulcorner\psi\urcorner) \wedge \forall z < m_0 \, \neg\mathsf{Prf}_{\mathsf{True}_n}(z, \ulcorner\neg\psi\urcorner)$.

The key observation is that putting (2) with the assumption $\mathcal{M} \models \neg\psi$ allows us to conclude that the m_0 in (6) must be a *nonstandard element of* $\mathcal{M}$. Thus by standardness of π_0 of (4) and the ordering properties of 'natural numbers' in $\mathcal{M}$, $\mathcal{M} \models \pi_0 < m_0$, which contradicts the second conjunct of (5). □

Theorem B

Theorem B. *For each $n \in \omega$ every consistent extension of* $\mathrm{I}\Delta_0 + \mathrm{Exp}$ (*in the same language*) *that is axiomatized by a set of Σ_n-sentences is incomplete.*

Proof. As shown by Gaifman and Dimitracopoulos (1980) for each $n \in \omega$ there is a formula Sat_{Σ_n} such that, provably in $\mathrm{I}\Delta_0 + \mathrm{Exp}$, Sat_{Σ_n} satisfies compositional clauses for all Σ_n-formulae. In particular there is a formula $\mathrm{True}_{\Sigma_n}(x)$ such that for all models $\mathcal{M}$ of $\mathrm{I}\Delta_0 + \mathrm{Exp}$, and for all Σ_n-sentences ψ, $\psi \in \mathrm{True}^{\mathcal{M}}_{\Sigma_n}$ iff $\mathcal{M} \models \psi$. We can now repeat the proof strategy of Theorem A with the use of $\mathrm{True}^{\mathcal{M}}_{\Sigma_n}$ instead of $\mathrm{True}^{\mathcal{M}}_{n}$.

Alternatively, invoke the provability of the MRDP theorem on the Diophantine representability of computably enumerable sets in $\mathrm{I}\Delta_0 + \mathrm{Exp}$ (shown by Gaifman and Dimitracopoulos in the aforementioned paper). By the MRDP-theorem each Σ_n-formula is equivalent to a Σ^*_n-formula in $\mathrm{I}\Delta_0 + \mathrm{Exp}$, so Theorem A applies. □

MRDP=Matijasevic-Robinson-Davis-Putnam

A more general form of Theorem A

Theorem A$^+$. *Let T be a computably enumerable sequential theory formulated in a finite language $\mathcal{L}$ and suppose A is a collection of $\mathcal{L}$-sentences such that $A \subseteq \Sigma_n^*$ for some $n \in \omega$ and $T \cup A$ is consistent. Then $T \cup A$ is incomplete.*

Remark 1. Note that if $A = \varnothing$, then the proof strategy of Theorem A, when applied to the setting of Theorem A$^+$, goes through for all computably enumerable consistent extensions T of R, without the assumption of sequentiality of T.

Remark 2. We can obtain an analogous following strengthening of Theorem B. There is also an analogous theorem at work for set theories:

Theorem C. *For each $n \in \omega$ every consistent extension of* KP (*Kripke-Platek set theory*) *that is axiomatized by a set of Σ_n^{Levy}-sentences is incomplete.*

Emil Jeřábek's hit the same idea in 2016

> **Proposition:** Let T_0 be an r.e. theory interpreting Robinson's arithmetic, and Γ a set of sentences for which T_0 has a truth predicate $\mathrm{Tr}_\Gamma(x)$, that is,
>
> $$T_0 \vdash \phi \leftrightarrow \mathrm{Tr}_\Gamma(\ulcorner \phi \urcorner) \qquad (*)$$
>
> for all $\phi \in \Gamma$. Then no extension of T_0 by a set of Γ-sentences is a consistent complete theory.

SOURCE: A MATHOVERFLOW ANSWER BY EMIL (2016);

https://mathoverflow.net/questions/256785/a-completion-of-zfc

Why $\mathcal{L}$ cannot be infinite (1)

- Consider the theory $U = \mathsf{CT}^{-}_{\omega}[\mathrm{I}\Sigma_1]$ of ω-iterated compositional truth over $\mathrm{I}\Sigma_1$ (without any induction for formulae using nonarithmetical symbols, hence the minus superscript) formulated in an extension of the language $\mathcal{L}_{\mathsf{A}}$ of arithmetic with infinitely many predicates $\{\mathsf{T}_{n+1} : n \in \omega\}$, and Tarski-style compositional axioms that stipulate that T_{n+1} is compositional for all $\mathcal{L}_n$-formulae, with $\mathcal{L}_0 = \mathcal{L}_{\mathsf{A}}$ and $\mathcal{L}_{n+1} = \mathcal{L}_n \cup \{\mathsf{T}_{n+1}\}$.
- Since bi-conditionals of form $\varphi \longleftrightarrow \mathsf{T}_{n+1}(\ulcorner\varphi\urcorner)$ are provable in U for every $\mathcal{L}_n$-sentence (thanks to the available composition axioms) ANY complete extension V of U is axiomatized by U (which is of bounded complexity) together with atomic sentences of form $\mathsf{T}_{n+1}(\ulcorner\varphi\urcorner)$ where $\varphi \in V$ and φ is an $\mathcal{L}_n$-sentence, thus U axiomatizable by a set of axioms of bounded quantifier complexity.
- By adding one axiom (internal induction) to the above theory we can get a theory of bounded complexity whose deductive consequence includes PA, and every completion of which is boundely axiomatizable.

Why $\mathcal{L}$ cannot be infinite (2)

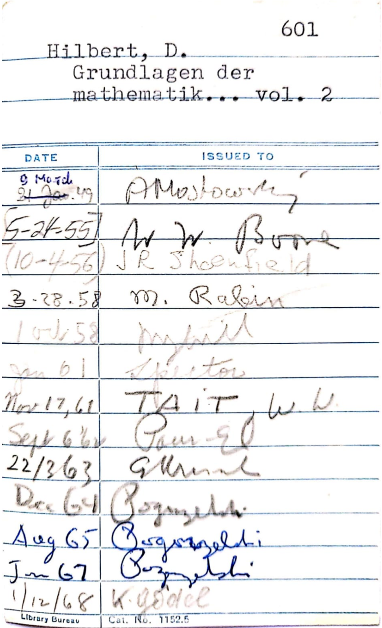
- Alternatively, starting with any theory T formulated in a language $\mathcal{L}$, we can apply a process known in model theory as *Morleyization/Atomization* to obtain an extension T^+ of T, formulated in an extension $\mathcal{L}^+$ of $\mathcal{L}$, such that T^+ is axiomatized by adding a collection of sentences of bounded quantifier depth to T, and T^+ has elimination of quantifiers in the sense that for each $\mathcal{L}^+$-formula $\varphi(x_1, ..., x_n)$, there is an n-ary predicate $P_\varphi \in \mathcal{L}^+$ such that the equivalence $\varphi(x_1, ..., x_n) \leftrightarrow P_\varphi(x_1, ..., x_n)$ is provable in T^+.

Off Shoots

- **Question.** *Is it possible for a consistent completion of* Q *to be axiomatized by a collection of sentences of bounded quantifier-depth*? Conjecture: Yes.
- In Theorem B, the theory $\mathrm{I}\Delta_0 + \mathsf{Exp}$ cannot be weakened to PA^-, i.e., for some $n \in \omega$ there is a consistent completion of PA^- (in the same language) that is axiomatized by single sentence together with a set of Σ_1-sentences. The proof of this and related results will appear in upcoming paper(s) with Albert Visser and Mateusz Łełyk.
- Albert Visser, in work in progress, studies the most basic forms of Rosser-provability and its sibling FGH-provability. (FGH = Friedman-Goldfarb-Harrington) under minimal demands on the meta-language and on the object-language. His work also expands on Saeed Salehi's [S] who has shown that incompleteness for predicates with salient properties are equivalent to their preconditions: roughly certain statements of incompleteness theorems are equivalent to a weak version of the Fixed Point Lemma.

[S] S. Salehi, *A reunion of Gödel, Tarski, Carnap and Rosser*, **Journal of Logic and Computation (2023)**.

Thanx!


601

Hilbert, D.
Grundlagen der
mathematik... vol. 2

DATE	ISSUED TO
9 March ~~21 Jan.~~ 49	A Mostowski
5-24-55	W. W. Boone
10-4-56	J R Shoenfield
3-28-58	M. Rabin
1 Oct 58	[illegible]
Jan 61	[illegible]
Nov 17, 61	TAIT, W. W.
Sept 6 '62	[illegible]
22/3 63	[illegible]
Dec 64	[illegible]
Aug 65	[illegible]
Jan 67	[illegible]
1/12/68	K. Gödel

Library Bureau   Cat. No. 1152.5