SOME RECENT NEWS ABOUT TRUTH THEORIES

Ali Enayat

University of Gothenburg

LOGIC COLLOQUIUM 2019, PRAGUE
SPECIAL SESSION ON REFLECTION PRINCIPLES AND MODAL LOGIC

August 15, 2019

Base theories and Truth theories

- A base theory B is a first order theory with "enough coding for handling finite sequences of objects", for example:
  (1) B = PA (Peano arithmetic).
  (2) B = ACA_0 (the predicative extension of PA).
  (3) B = ZF (Zermelo-Fraenkel set theory).
  (4) B = GB (the predicative extension of ZF).
- A truth theory over a base theory B is a theory of the form:

$$\mathrm{P[B]} = \mathrm{B} \cup \mathrm{P},$$

  where P (for prawda = truth in Polish) is a set of "truth axioms" formulated in the language $\mathcal{L}_{\mathrm{B}} \cup \{\mathrm{T}(x)\}$, where the intended interpretation of $\mathrm{T}(x)$ is "x is the Gödel number of a true $\mathcal{L}_{\mathrm{B}}$-sentence".
- We will refer to $\mathcal{L}_{\mathrm{B}} \cup \{\mathrm{T}\}$ as "the extended language".

Compositional Truth

- CT^- consists of the following axioms, where s and t range over closed terms of $\mathcal{L}_{\mathrm{B}}$; and ϕ and ψ range over formulae of $\mathcal{L}_{\mathrm{B}}$.

CT1 $T(s=t) \leftrightarrow \left(s^\circ = t^\circ\right)$.

CT2 $\left\{T\left(R(s_1,\ldots,s_n)\right) \leftrightarrow R({s_1}^\circ,\ldots,{s_n}^\circ) : R \in \mathcal{L}_{\mathrm{B}}\right\}$.

CT3 $T(\neg\phi) \leftrightarrow \neg T(\phi)$.

CT4 $T(\phi \vee \psi) \leftrightarrow T(\phi) \vee T(\psi)$.

CT5 $T(\exists v\ \phi) \leftrightarrow \exists x T(\phi(\underline{x}))$.

CT6 $\left(\bar{s}^\circ = \bar{t}^\circ \rightarrow T(\phi(\bar{s})) \leftrightarrow T(\phi(\bar{t}))\right)$.

- $\mathrm{CT}^- \vdash \mathrm{TB}^-$, where TB^- consist of Tarski bi-conditionals

$$\mathrm{T}(\phi) \leftrightarrow \phi$$

Friedman-Sheard untyped truth theory

FS^- consists of the following axioms, where s and t range over closed terms of $\mathcal{L}_{\mathrm{B}}$; and ϕ and ψ range over formulae of $\mathcal{L}_{\mathrm{B}} \cup \{\mathrm{T}(x)\}$.

FS1 $T(s = t) \leftrightarrow \big(s^\circ = t^\circ\big)$.

FS2 $\{T\big(R(s_1, \ldots, s_n)\big) \leftrightarrow R({s_1}^\circ, \ldots, {s_n}^\circ) : R \in \mathcal{L}_{\mathrm{B}}\}$.

FS3 $T(\neg\phi) \leftrightarrow \neg T(\phi)$.

FS4 $T(\phi \vee \psi) \leftrightarrow T(\phi) \vee T(\psi)$.

FS5 $T(\exists v\ \phi) \leftrightarrow \exists x T(\phi(\underline{x}))$.

FS6 $\Big(\bar{s}^\circ = \bar{t}^\circ \to T(\phi(\bar{s})) \leftrightarrow T(\phi(\bar{t}))\Big)$.

FS^- is equipped with the following additional derivation rules:

$$\frac{\phi}{T(\phi)}\ \text{(NEC)} \qquad\qquad \frac{T(\phi)}{\phi}\ \text{(CONEC)}$$

.

Kripke-Feferman untyped truth theory

- KF^- consists of the following axioms, where s and t range over closed terms of $\mathcal{L}_{\mathrm{B}}$; and ϕ and ψ range over formulae of $\mathcal{L}_{\mathrm{B}} \cup \{\mathrm{T}(x)\}$.

KF1 $T(s = t) \leftrightarrow (s^\circ = t^\circ)$.

KF2 $T(s \neq t) \leftrightarrow (s^\circ \neq t^\circ)$.

KF3 $\{T\big(R(s_1, \ldots, s_n)\big) \leftrightarrow R(s_1{}^\circ, \ldots, s_n{}^\circ) : R \in \mathcal{L}_{\mathrm{B}}\}$.

KF4 $\{T\big(\neg R(s_1, \ldots, s_n)\big) \leftrightarrow \neg R(s_1{}^\circ, \ldots, s_n{}^\circ) : R \in \mathcal{L}_{\mathrm{B}}\}$.

KF5 $T(\neg\neg\phi) \leftrightarrow T(\phi)$.

KF6 $T(\phi \vee \psi) \leftrightarrow T(\phi) \vee T(\psi)$.

KF7 $T(\neg(\phi \vee \psi)) \leftrightarrow T(\neg\phi) \wedge T(\neg\psi)$.

KF8 $T(\exists y\ \phi(y)) \leftrightarrow \exists x T(\phi(\underline{x}))$.

KF9 $T(\neg\exists y\ \phi(y)) \leftrightarrow \forall x T(\neg\phi(\underline{x}))$.

KF10 $\Big(\bar{s}^\circ = \bar{t}^\circ \rightarrow T(\phi(\bar{s})) \leftrightarrow T(\phi(\bar{t}))\Big)$.

KF11 $\Big(t^\circ = \phi \rightarrow T(T(t)) \leftrightarrow T(\phi)\Big)$.

KF12 $\Big(t^\circ = \phi \rightarrow T(\neg T(t)) \leftrightarrow T(\neg\phi)\Big)$.

The power of truth axioms in the presence of induction

- Given a truth theory P^-, let P be the result of strengthening P^- with all instances of induction in the extended language.
- (Tarski, 1935) CT[PA] proves Con(PA), but TB[PA] is conservative over PA.
- (Feferman, 1964) The arithmetical consequences of CT[PA] coincide with the arithmetical consequences of ACA.
- (Halbach, 2010) The arithmetical consequences of FS coincide with the arithmetical consequences of $\mathsf{RA}_{<\omega}$.
- (Feferman 1985, Cantini 1987) The arithmetical consequences of KF coincide with the arithmetical consequences of $\mathsf{RA}_{<\varepsilon_0}$.

A little induction goes very far

- Let CT_n be the fragment of CT[PA] in which the scheme of induction for formulae in the extended language is restricted to Σ_n formulae.
- $\mathrm{CT}_1 \vdash \mathrm{GR}_{\mathrm{PA}}$, where $\mathrm{GR}_{\mathrm{PA}}$ is the sentence in the extended language expressing: "all theorems of PA are true". Therefore $\mathrm{CT}_1 \vdash \mathrm{Con(PA)}$, since $\mathrm{CT}^- \vdash \neg \mathrm{T}(0 = 1)$.
- What about CT_0? In a 1986 paper, Kotlarski claimed that CT_0 proves $\mathrm{GR}_{\mathrm{PA}}$, but a serious gap was found in his proof outline around 2012 by Heck and Visser.
- Łełyk and Wcisło (2017) proved that CT_0 and $\mathrm{GR}_{\mathrm{PA}}$ have the same arithmetic consequences.
- Łełyk (PhD thesis, 2018) confirmed Kotlarski's hunch by verifying that $\mathrm{GR}_{\mathrm{PA}}$ is provable in CT_0.

Conservativity of typed truth

- Krajewski-Kotlarski-Lachlan (1981) showed that $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{IC}$ is conservative over PA, where IC is the axiom of Induction Correctness (also known as Internal Induction) asserting, as one sentence, that each instance of arithmetical induction is true.
- We now know that, more generally, $\mathsf{CT}^-[\mathsf{B}]$ is conservative over B for every base theory B, and that if τ is a "scheme template" such that B proves every instance of τ, then:
  $\mathsf{CT}^-[\mathsf{B}] + \text{"every instance of } \tau \text{ is true"}$ is conservative over B.
- These general results were established in the joint work of Visser and me (2014), using elementary model-theoretic ideas, and by Leigh (2015), using proof-theoretic machinery.

Conservativity of untyped truth

- Cantini (1989) showed that $KF^{-}[B]$ is conservative over B.

- Halbach (2011) noted that the conservativity of $FS^{-}[B]$ over B can be established by using:

  (1) the conservativity of $CT^{-}[B]$ over B for all B, and

  (2) the proof theoretic reduction of FS[PA] to $RT_{<\omega}$.

Between CT^- and CT_0

- Cieślinski (2010) and (2017) showed:
  (1) $CT^-PA] + VAL \subseteq T$ proves CT_0, where $VAL \subseteq T$ expresses:
  "T contains all arithmetical instances of theorems of first order logic".
  (2) $CT^-[PA] + Cl_{Prop(T)} \subseteq T$ proves CT_0, where $Cl_{Prop}(T) \subseteq T$ expresses:
  "T is closed under propositional logic".
- It is easy to verify that (1) $CT_0 \vdash VAL \subseteq T$,
  (2) $CT^- \vdash VAL \subseteq T \rightarrow Cl_{Prop}(T) \subseteq T$, and
  (3) $CT^-[PA] + Cl_{Prop} \subseteq T \vdash DC \wedge IC$, where
  DC (Disjunctive Correctness) is the axiom that expresses: a finite disjunction of arithmetical sentences is true iff one of the disjuncts is true.
- Open Question 1. Is $CT^- + VAL_{Prop} \subseteq T$ conservative over PA?
- I showed (2012) that $CT^-[PA] + DC + IC \vdash CT_0$.

The surprising power of Disjunctive Correctness (1)

- **Theorem** (Pakhomov, (2019)) $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{DC} \vdash \mathsf{IC}$.
- Coupled with the facts that (1) $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{DC} + \mathsf{IC}$ proves CT_0, and (2) CT_0 proves Con (PA), and (3) Gödel's second incompleteness theorem, Pakhomov's theorem shows that $\mathsf{CT}^-[\mathsf{PA}] + \mathsf{DC}$ is not conservative over PA.
- Pakhomov's proof is based on a generalization of Visser's theorem on the non-existence of infinite descending chains of truth definitions. Its proof uses (Löb's version) of Gödel's second incompleteness theorem, and therefore implicitly uses the Gödel-Carnap fixed point theorem.
- **Proof Outline.** Consider the theory ITB (iterated truth biconditionals) extending Robinson's Q + "$\prec$ is a transitive relation" plus the following biconditionals B_φ:

$$\mathsf{B}_\varphi := \forall\alpha(\mathsf{T}_\alpha(\ulcorner\phi\urcorner) \leftrightarrow \phi^{\prec\alpha}).$$

Lemma (1) *The following theory* DTB *is inconsistent:*

$$\mathsf{DTB} := \mathsf{ITB} + \forall\alpha\exists\beta(\beta \prec \alpha) + \exists\alpha(\alpha = \alpha).$$

The surprising power of Disjunctive Correctness (2)

Lemma (2) *Every finite subtheory of* DTB *is interpretable in* $\mathrm{CT}^-[\mathrm{I}\Delta_0 + \mathrm{Exp}] + \mathrm{DC} + \neg\, \mathrm{IC}$.

- Pakhomov later found a proof based on a result of Flumini and Sato concerning the relationship between well-foundedness and the existence of hierarchies in second order systems (2014) whose proof is fixed-point free.
- DC can be written as $\mathrm{DC}_{\mathrm{Elim}} + \mathrm{DC}_{\mathrm{Intro}}$.
- Recent joint work of Wcisło, Łełyk and E. shows that $\mathrm{DC}_{\mathrm{Intro}}$ can be conservatively added to: $\mathrm{CT}^-[\mathrm{PA}] + \mathrm{IC} + \{\forall x\, (\mathrm{True}_{\Sigma_n}(x) \rightarrow \mathrm{T}(x)) : n < \omega\}$.
- Open Problem 2. Is $\mathrm{CT}^-[\mathrm{PA}] + \mathrm{DC}_{\mathrm{Elim}}$ *conservative over* PA?
- Open Problem 3. Is $\mathrm{CT}^-[\mathrm{S}^1_2] + \mathrm{DC}$ *conservative over* PA?

How complex is the reduction of P[B] to B?

- Suppose P[B] is conservative over B. Among the commonly studied computational classes of functions $\mathcal{F}$, what is the *optimal complexity class* $\mathcal{F}$ that contains some f with the property that for all proofs π and all $\mathcal{L}_{\mathsf{B}}$-sentences ϕ, we have:

$$\mathsf{P[B]} \vdash_{\pi} \phi \Longrightarrow \mathsf{B} \vdash_{f(\pi)} \phi.$$

- Let Supexp-time be the class of functions that are computable by a Turing machine whose run time is bounded above by a function that is provably total in the fragment of PRA known as SEFA. An examination of Leigh's 2015 proof makes it clear that there is a Supexp-time computable function f such that for all proofs π and all $\mathcal{L}_{\mathsf{B}}$-sentences ϕ we have:

$$\mathsf{CT}^{-}[\mathsf{B}] \vdash_{\pi} \phi \Longrightarrow \mathsf{B} \vdash_{f(\pi)} \phi.$$

Examples where the Supexp upper bound is optimal

- By a theorem of Pudlák (1985), if T_1 is a sequential theory, and $T_2 \supseteq T_1$ and T_2 proves the consistency of T_1 on a cut, then T_2 has superexponential speed-up over T_1.
- Pudlák's above theorem readily implies that $\mathrm{CT}^-[\mathrm{B}]$ is not Exp-time reducible to B if B is finitely axiomatizable. Therefore, for finitely axiomatizable B, Leigh's Supexp upper bound is optimal.
- Pudlák's theorem also implies that Supexp is optimal for reducing $\mathrm{CT}^-[\mathrm{PA}] + \mathrm{IC}$ to PA.
- What about $\mathrm{CT}^-[\mathrm{PA}]$?

Feasible interpretations

- $I : \mathsf{B} \rhd \mathsf{P[B]}$ is a feasible interpretation if there is a P-time computable function $f(s)$ such that for all proofs π and all $\mathcal{L}_{\mathsf{P[B]}}$-formulae ϕ,

$$\mathsf{P[B]} \vdash_{\pi} \phi \Longrightarrow \mathsf{B} \vdash_{f(\pi)} \phi^{I}.$$

- Feasible interpretations were first systematically studied in the 1993 doctoral dissertation of Rineke Verbrugge. She showed, among many other things, that:

There is a sentence θ such that $\mathsf{PA} \rhd \mathsf{PA} + \theta$, but $\mathsf{PA} \not\rhd_{f} \mathsf{PA} + \theta$.

Feasibly neat interpretations (1)

- A family of interpretations $\{I_n\}_{n\in\mathbb{N}} : \mathrm{B} \rhd \mathrm{P[B]}$ is feasibly neat if there are P-time computable functions $f(s_0, s_1)$ and $g(s_0, s_1)$ such that the following two conditions hold:

  1. For every $k \in \mathbb{N}$, and every $\mathcal{L}_\mathrm{B}$-formula ϕ of length at most k,

  $$\mathrm{B} \vdash_{f(\mathrm{tal}(k),\phi)} \phi^{I_k} \rightarrow \phi,$$

  where $\mathrm{tal}(k)$ is the tally numeral $1+1+...+1$ (k times).

  2. For every $k \in \mathbb{N}$, and every proof π,

  $$\mathrm{P[B]} \vdash_\pi \phi \Longrightarrow \mathrm{B} \vdash_{g(\mathrm{tal}(k),\pi)} \phi^{I_k}.$$

Feasibly neat interpretations (2)

- **Theorem** (Joint with Mateusz Łełyk and Bartosz Wcisło) (2019). *Let* P *denote any of the truth theories* CT^{-}, FS^{-}, *and* KF^{-}. *There is a family* $\{I_n\}_{n\in\mathbb{N}} : \mathsf{PA} \rhd \mathsf{P}[\mathsf{PA}]$ *of feasibly neat interpretations.*

Corollaries of the Theorem

- Let P denote any of the truth theories CT^-, FS^- *and* KF^-.
- **Corollary 1.** P[PA] *is feasibly reducible to* PA *,i.e., there is a polynomial-time computable function* f *with the property that for all proofs* π *and all* $\mathcal{L}_{\mathrm{PA}}$*-sentences* ϕ*, we have*:

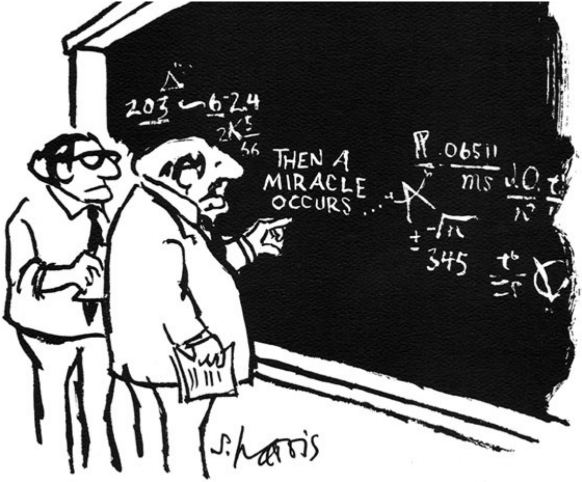
$$\mathrm{P[PA]} \vdash_{\pi} \phi \Longrightarrow \mathrm{PA} \vdash_{f(\pi)} \phi.$$

- **Corollary 2.** P[PA] *has at most polynomial speed-up over* PA, *i.e., there is a polynomial* $p(n)$ *such that for all* $n \in \mathbb{N}$

$$\mathrm{P[PA]} \vdash^{\leq n} \phi \Longrightarrow \mathrm{PA} \vdash^{\leq p(n)} \phi.$$

- Open Problem 4. Is the conservativity of $\mathrm{CT}^-[\mathrm{PA}]$ *over* PA *verifiable in* S^1_2?

Thank you for your attention



"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."