On the principle of disjunctive correctness

Cezary Cieśliński

joint work with Mateusz Łełyk and Bartosz Wcisło

Faculty of Philosophy
University of Warsaw
Poland

Dubrovnik 2021

Theories CT^- and CT

CT^- ('compositional truth') is an axiomatic truth theory obtained by turning into axioms Tarski's classical inductive clauses.

Definition

Apart from all the axioms of PA, CT^- contains the following truth axioms:

- $\forall s, t \in Tm^c \big(T(s = t) \equiv val(s) = val(t)\big)$
- $\forall \varphi \big(Sent_{L_{PA}}(\varphi) \rightarrow (T\neg\varphi \equiv \neg T\varphi)\big)$
- $\forall \varphi \forall \psi \big(Sent_{L_{PA}}(\varphi \rightarrow \psi) \rightarrow (T(\varphi \rightarrow \psi) \equiv (T\varphi \rightarrow T\psi))\big)$
- $\forall v \forall \varphi(x) \big(Sent_{L_{PA}}(\forall v \varphi(v)) \rightarrow (T(\forall v \varphi(v)) \equiv \forall x T(\varphi(\dot{x})))\big)$

Definition

CT is CT^- with induction for formulas of the extended language.

CT^- and CT - basic properties

Definition

A truth theory Th is conservative over its syntactic base theory B iff for every $\psi \in \mathrm{L_B}$, if $\mathrm{Th} \vdash \psi$, then $\mathrm{B} \vdash \psi$.

Observation

CT *is not conservative over* PA. *In particular, it proves that all theorems of* PA *are true (global reflection for* PA*).*

Theorem

CT^- *is conservative over* PA*.*

Source: H. Kotlarski, K. Krajewski, A. Lachlan "Construction of satisfaction classes for nonstandard models", *Canadian Mathematical Bulletin* 1981.

Tarski boundary programme

- General question: which truth axioms are responsible for the transition from conservative to non-conservative extensions of the base theory?
- The ‘line’ where this transition takes place was called the Tarski Boundary.
- Concrete open questions concern, e.g., systems with weak induction, non-inductive systems with collection principle, non-inductive systems with axioms guaranteeing some versions of ‘propositional soundness’ of truth.

Tarski boundary - some results

Theorem

The following axiomatisations are equivalent:

(a) $\mathsf{CT}^- + \forall\psi[\mathsf{Pr}_{\varnothing}(\psi) \rightarrow \mathsf{T}(\psi)]$,

(b) $\mathsf{CT}^- + \forall\psi[\mathsf{Pr}_{\mathsf{PA}}(\psi) \rightarrow \mathsf{T}(\psi)]$,

(c) $\mathsf{CT}^- + \forall\psi[\mathsf{Pr}_{\mathsf{T}}(\psi) \rightarrow \mathsf{T}(\psi)]$.

The expressions '$\mathsf{Pr}_{\varnothing}(x)$', '$\mathsf{Pr}_{\mathsf{PA}}(x)$' and '$\mathsf{Pr}_{\mathsf{T}}(x)$' mean (respectively) that x has a proof in pure logic, x is provable in PA and x has a proof from true premises.

The move from (c) to (a) is trivial. For the transition from (a) to (b), see C. Cieśliński (2010) "Truth, conservativeness, and provability". For the transition from (b) to (c), see C. Cieśliński (2017) *The Epistemic Lightness of Truth*.

Tarski boundary: additional (non-)conservativity results

Let $\mathrm{Pr}_{\mathrm{T}}^{\mathrm{Sent}}(x)$ be a formula of L_{T} stating that x has a purely propositional proof from true premises.

Definition

CT_0 is CT^- with Δ_0 induction for the extended language. $\mathrm{CT}^{\mathrm{PropCl}}$ is CT^- together with the axiom '$\forall\psi[\mathrm{Pr}_{\mathrm{T}}^{\mathrm{Sent}}(\psi) \rightarrow \mathrm{T}(\psi)]$'.

Theorem

(a) $\mathrm{CT}^{\mathrm{PropCl}} \subseteq \mathrm{CT}_0$,

(b) $\mathrm{CT}_0 \subseteq \mathrm{CT}^{\mathrm{PropCl}}$,

(c) CT_0 *is not conservative over* PA.

(a) is an easy exercise. For (b), see C. Cieśliński (2010) "Deflationary Truth and Pathologies". For (c) see M. Łełyk and B. Wcisło (2017) "Notes on bounded induction for the compositional truth predicate".

Disjunctive correctness

Let us introduce the following axiom of disjunctive correctness:

(DC) $\forall x[\forall y \in x\, y \in \mathrm{Sent}_{L_{PA}} \rightarrow \big(T(\bigvee x) \equiv \exists y \in x\, T(y)\big)]$.

We will discuss also the following weakenings of (DC):

(DC-in) $\forall x[\forall y \in x\, y \in \mathrm{Sent}_{L_{PA}} \rightarrow \big(\exists y \in x\, T(y) \rightarrow T(\bigvee x)\big)]$.

(DC-out) $\forall x[\forall y \in x\, y \in \mathrm{Sent}_{L_{PA}} \rightarrow \big(T(\bigvee x) \rightarrow \exists y \in x\, T(y)\big)]$.

Theorem

$\mathrm{CT}^- + (\mathrm{DC}) = \mathrm{CT}_0$, *hence* $\mathrm{CT}^- + (\mathrm{DC})$ *is not conservative over Peano arithmetic.*

Source: A. Enayat and F. Pakhomov “Truth, disjunction and induction”, *The Archive for Mathematical Logic* 2019.

Yablo's paradox

Universal version

Consider a sequence of sentences $\varphi_0, \varphi_1 \ldots$ such that each φ_n states 'for every number $k > n$, φ_k is false'. Then a simple reasoning produces a contradiction.

Existential version

Consider a sequence of sentences $\varphi_0, \varphi_1 \ldots$ such that each φ_n states 'there is a number $k > n$ such that φ_k is false'. Then a simple reasoning produces a contradiction.

Disjunctive correctness is Δ_0 induction

Theorem

$\mathrm{CT}^- + (\mathrm{DC}) \vdash \forall\psi \in \mathrm{Sent}_{\mathrm{Ar}}[\mathrm{Pr}_{\mathrm{T}}^{\mathrm{Sent}}(\psi) \to \mathrm{T}(\psi)].$

Working in $\mathrm{CT}^- + (\mathrm{DC})$, assume that $\mathrm{Pr}_{\mathrm{T}}^{\mathrm{Sent}}(\psi)$. Let $(\psi_0 \dots \psi_n)$ be a sentential proof of ψ from true premises. Define:

- $\varphi_0 := \ulcorner 0 = 0 \urcorner$,
- for $k > 0$, $\varphi_k := \neg\psi_k \to \bigvee_{i<k} \neg\varphi_i$.

Thus, for example, φ_3 is the sentence:

$$\neg\psi_3 \to \Big(\overbrace{0 \neq 0}^{\neg\varphi_0} \vee \overbrace{\neg(\neg\psi(1) \to 0 \neq 0)}^{\neg\varphi_1} \vee \overbrace{\neg(\neg\psi(2) \to (0 \neq 0 \vee \neg(\neg\psi(1) \to 0 \neq 0)))}^{\neg\varphi_2}\Big)$$

Disjunctive correctness is Δ_0 induction, cont.

We observe that:

(1) $\forall k \leq n \big[T(\neg\psi_k) \rightarrow \big(T(\varphi_k) \equiv \exists i < k T(\neg\varphi_i)\big)\big]$.

For the proof of (1), assume that $T(\neg\psi_k)$. Then $k > 0$, because $(\psi_0 \ldots \psi_n)$ is a proof from true premises.

For the implication from left to right, assume $T(\varphi_k)$. In other words, $T(\neg\psi_k \rightarrow \bigvee_{i<k} \neg\varphi_i)$, hence $T(\bigvee_{i<k} \neg\varphi_i)$ and by (DC-out) $\exists i < k T(\neg\varphi_i)$.

For the implication from right to left, assume $\exists i < k T(\neg\varphi_i)$. Then by (DC-in) $T(\bigvee_{i<k} \neg\varphi_i)$, therefore $T(\varphi_k)$. Thus the proof of (1) is finished.

Disjunctive correctness is Δ_0 induction, cont.

We now have:

(1) $\forall k \leq n\big[T(\neg\psi_k) \rightarrow \big(T(\varphi_k) \equiv \exists i < k T(\neg\varphi_i)\big)\big]$.

It follows that:

(2) $\forall k \leq n\big[T(\neg\psi_k) \rightarrow T(\varphi_k)\big]$.

For the proof of (2), assume that $T(\neg\psi_k)$. For a contradiction, assume also that $\neg T(\varphi_k)$. Therefore by (1), $\forall i < k T(\varphi(i))$. Fix $i < k$ such that $T(\neg\psi_i)$ (such an i must exist because ψ_k is false and it must have been obtained in our propositional proof by modus ponens). Hence $T(\varphi_i)$. By (1), take $j < i$ such that $\neg T(\varphi_j)$. But $j < k$, so $T(\varphi_j)$ - a contradiction.

Disjunctive correctness is Δ_0 induction, cont.

We have obtained:

(1) $\forall k \leq n\big[T(\neg\psi_k) \rightarrow \big(T(\varphi_k) \equiv \exists i < k T(\neg\varphi_i)\big)\big]$.

(2) $\forall k \leq n\big[T(\neg\psi_k) \rightarrow T(\varphi_k)\big]$.

We conclude that every sentence in the proof $(\psi_0 \ldots \psi_n)$ must be true. Otherwise let $k \leq n$ and $\neg T(\psi_k)$. Then $T(\varphi_k)$ by (2). From (1), take a number $i < k$ such that $\neg T(\varphi_i)$. Then $\neg T(\psi_i)$ (because φ_i has the form '$\neg\psi_i \rightarrow \ldots$'), hence $T(\varphi_i)$ by (2) and this contradiction ends the proof.

Stipulations

Convention

Let $(\varphi_0 \ldots \varphi_n)$ be a (coded) sequence of formulas. The expression $\bigvee_{i \leq n} \varphi_i$ denotes the disjunction $\varphi_n \vee (\varphi_{n-1} \vee (\varphi_{n-2} \vee (\ldots \varphi_0) \ldots))$.

In effect, $\bigvee_{i \leq n} \varphi_i$ is always $\varphi_n \vee \bigvee_{i \leq n-1} \varphi_i$.

Definition

- (DC) is the sentence: $\forall(\varphi_0 \ldots \varphi_n)[T(\bigvee_{i \leq n} \varphi_i) \equiv \exists k \leq n T(\varphi_k)]$.
- (DC-out) is the sentence: $\forall(\varphi_0 \ldots \varphi_n)[T(\bigvee_{i \leq n} \varphi_i) \rightarrow \exists k \leq n T(\varphi_k)]$.
- (DC-in) is the sentence: $\forall(\varphi_0 \ldots \varphi_n)[\exists k \leq n T(\varphi_k)] \rightarrow T(\bigvee_{i \leq n} \varphi_i)]$.

(DC-out) proves (DC-in)

Theorem

$\mathrm{CT}^- + (\text{DC- out}) \vdash (\text{DC- in})$.

Working in CT^-+ (DC-out), let $\vartheta_0 \ldots \vartheta_n$ be a sequence of sentences. Fix $m \leq n$ such that $T(\vartheta_m)$. We claim that $T(\bigvee_{i \leq n} \vartheta_i)$. We define:

- $\psi_k := \bigvee_{i \leq m+k} \vartheta_i$.

In effect, ψ_k is $\vartheta_{m+k} \vee \underbrace{\Big(\vartheta_{m+k-1} \vee \big(\vartheta_{m+k-2} \vee \ldots \vee \vartheta_0\big)\Big)}_{\psi_{k-1}}$.

Hence we have: if $\neg T(\psi_k)$, then $\neg T(\psi_{k-1})$.

Define:

- $\varphi_0 = \ulcorner 0 = 0 \urcorner$,
- for $k > 0$, $\varphi_k := \neg\psi_k \rightarrow \bigvee_{i<k} \neg\varphi_i$.

Then by (DC-out) we have:

(1) $\forall k \leq n\big[T(\neg\psi_k) \rightarrow \big(T(\varphi_k) \rightarrow \exists i < k \neg T(\varphi_i)\big)\big]$.

Now we claim that $\forall k \leq n T(\varphi_k)$.
Take any $k \leq n$ and suppose that $\neg T(\varphi_k)$. Then $T(\neg\psi_k)$ and $T(\neg \bigvee_{i<k} \neg\varphi_i)$. With our bracketing, this means that $T\big(\neg(\neg\varphi_{k-1} \vee \bigvee_{i<k-1} \neg\varphi_i)\big)$. Hence by compositional axioms of CT^-, $T(\varphi_{k-1})$ and $T(\neg \bigvee_{i<k-1} \neg\varphi_i)$. Since $T(\neg\psi_k)$, we have $T(\neg\psi_{k-1})$; in effect, $T(\varphi_{k-1})$ is the same as $T(\neg\psi_{k-1} \rightarrow \bigvee_{i<k-1} \neg\varphi_i)$. Therefore $T(\bigvee_{i<k-1} \neg\varphi_i)$, which is a contradiction.

Hence we have:

(1) $\forall k \leq n\big[T(\neg\psi_k) \rightarrow \big(T(\varphi_k) \rightarrow \exists i < k \neg T(\varphi_i)\big)\big]$,

(2) $\forall k \leq n T(\varphi_k)$.

As before, this leads to a contradiction. Let k be such that $T(\neg\psi_k)$. Then by (2) $T(\varphi_k)$. Hence by (1) $\exists i < k T(\neg\varphi_i)$, which contradicts (2) and thus the whole proof is finished.

Propositional tautologies: an open question

Problem

Let $\mathrm{CT}^{\mathrm{Prop}}$ be the theory: $\mathrm{CT}^{-} + \forall\psi[\mathrm{PropTaut}(\psi) \rightarrow \mathrm{T}(\psi)]$, with the formula 'PropTaut(x)' having a natural reading 'x is a propositional tautology'. Is $\mathrm{CT}^{\mathrm{Prop}}$ syntactically conservative over PA?

We know at the moment that the answer would be negative if (in $\mathrm{CT}^{\mathrm{Prop}}$) we could derive "$\mathrm{T}(\varphi_0 \wedge \ldots \wedge \varphi_a)$" from the information that $\forall i \leq a \mathrm{T}(\varphi_i)$ (indeed, this amounts to deriving (DC-out)).

THE END

Thanks for your attention!!!