Partial Reflection over Uniform Disquotational Truth

Mateusz Łełyk

Institute of Philosophy, University of Warsaw

Supported by Foundation for Polish Science (START programme)

MOPA

Online June 22, 2020

Outline

Introduction: Reflection & Truth

The main result (hopefully)

Reflection Principles

Let Th be any elementary theory (we want Th to be given by a $\Delta_0(\exp)$ formula). The uniform reflection principle for Th is a scheme

$$\forall x(\mathrm{Prov}_{\mathrm{Th}}(\phi(\dot{x})) \rightarrow \phi(x)).$$

where $\phi \in \mathcal{L}_{\mathrm{Th}}$. This scheme will be denoted with REF(Th). We shall consider also

- $\Gamma(\mathcal{L})$-REF(Th) – the restriction of REF(Th) to formulae from class Γ from language $\mathcal{L}$ (Think $\Gamma = \Sigma_n, \Pi_n$).

Reflection Principles

Let Th be any elementary theory (we want Th to be given by a $\Delta_0(\exp)$ formula). The uniform reflection principle for Th is a scheme

$$\forall x(\mathrm{Prov}_{\mathrm{Th}}(\phi(\dot{x})) \rightarrow \phi(x)).$$

where $\phi \in \mathcal{L}_{\mathrm{Th}}$. This scheme will be denoted with REF(Th). We shall consider also

- $\Gamma(\mathcal{L})$-REF(Th) – the restriction of REF(Th) to formulae from class Γ from language $\mathcal{L}$ (Think $\Gamma = \Sigma_n, \Pi_n$).
- $\Gamma(\mathcal{L})$-REF$^{\alpha}$(Th) – α-many iterations of $\Gamma(\mathcal{L})$-REF(Th).

Reflection Principles

Let Th be any elementary theory (we want Th to be given by a $\Delta_0(\exp)$ formula). The uniform reflection principle for Th is a scheme

$$\forall x(\mathrm{Prov}_{\mathrm{Th}}(\phi(\dot{x})) \to \phi(x)).$$

where $\phi \in \mathcal{L}_{\mathrm{Th}}$. This scheme will be denoted with REF(Th). We shall consider also

- $\Gamma(\mathcal{L})$-REF(Th) – the restriction of REF(Th) to formulae from class Γ from language $\mathcal{L}$ (Think $\Gamma = \Sigma_n, \Pi_n$).
- $\Gamma(\mathcal{L})$-REF$^{\alpha}$(Th) – α-many iterations of $\Gamma(\mathcal{L})$-REF(Th).

Reflection Principles

Let Th be any elementary theory (we want Th to be given by a $\Delta_0(\exp)$ formula). The uniform reflection principle for Th is a scheme

$$\forall x(\mathrm{Prov}_{\mathrm{Th}}(\phi(\dot{x})) \to \phi(x)).$$

where $\phi \in \mathcal{L}_{\mathrm{Th}}$. This scheme will be denoted with REF(Th). We shall consider also

- $\Gamma(\mathcal{L})$-REF(Th) – the restriction of REF(Th) to formulae from class Γ from language $\mathcal{L}$ (Think $\Gamma = \Sigma_n, \Pi_n$).
- $\Gamma(\mathcal{L})$-REF$^{\alpha}$(Th) – α-many iterations of $\Gamma(\mathcal{L})$-REF(Th).

For later purposes, let (for a formula $\phi(x)$)

$$\mathrm{Prov}^{\phi}_{\mathrm{Th}}, \mathrm{Prov}_{\mathrm{Th}+\phi}$$

denote the oracle-provability predicates, in which as axioms we can take arbitrary sentences from Th and arbitrary sentences satisfying $\phi(x)$.

Some metamathematics of reflection principles

Proposition

For every n, $\Pi_{n+1}(\mathcal{L})$-REF(Th) *and* $\Sigma_n(\mathcal{L})$-REF(Th) *coincide, provably over* EA.

Some metamathematics of reflection principles

Proposition

For every n, $\Pi_{n+1}(\mathcal{L})$-REF(Th) *and* $\Sigma_n(\mathcal{L})$-REF(Th) *coincide, provably over* EA.

Theorem (Kreisel-Lévy)

PA *coincide with* EA + REF(EA).

Some metamathematics of reflection principles

Proposition

For every n*,* $\Pi_{n+1}(\mathcal{L})$-REF(Th) *and* $\Sigma_n(\mathcal{L})$-REF(Th) *coincide, provably over* EA*.*

Theorem (Kreisel-Lévy)

PA *coincide with* $\mathrm{EA} + \mathrm{REF}(\mathrm{EA})$*.*

Theorem (Leivant)

For every $n > 0$*,* $I\Sigma_n$ *and* $\mathrm{EA} + \Sigma_{n+1}(\mathcal{L}_{\mathrm{PA}})$-REF(EA) *coincide.*

Basic truth theories

Definition

$\mathrm{UTB}^-(\mathrm{Th})$ is the $\mathcal{L}_T := \mathcal{L}_{\mathrm{Th}} \cup \{T\}$ theory extending Th with all the axioms of the form

$$\forall x_1, \ldots, x_n \quad T(\phi(\dot{x}_1, \ldots, \dot{x}_n)) \equiv \phi(x_1, \ldots, x_n),$$

for $\phi(\bar{x}) \in \mathcal{L}_{\mathrm{Th}}$.

Basic truth theories

Definition

$\mathrm{UTB}^-(\mathrm{Th})$ is the $\mathcal{L}_T := \mathcal{L}_{\mathrm{Th}} \cup \{T\}$ theory extending Th with all the axioms of the form

$$\forall x_1, \ldots, x_n \quad T(\phi(\dot{x}_1, \ldots, \dot{x}_n)) \equiv \phi(x_1, \ldots, x_n),$$

for $\phi(\bar{x}) \in \mathcal{L}_{\mathrm{Th}}$.

Definition

$\mathrm{CT}^-(\mathrm{Th})$ is the $\mathcal{L}_T$ theory extending Th with the following axioms

Basic truth theories

Definition

$\mathrm{UTB}^-(\mathrm{Th})$ is the $\mathcal{L}_T := \mathcal{L}_{\mathrm{Th}} \cup \{T\}$ theory extending Th with all the axioms of the form

$$\forall x_1, \ldots, x_n \quad T(\phi(\dot{x}_1, \ldots, \dot{x}_n)) \equiv \phi(x_1, \ldots, x_n),$$

for $\phi(\bar{x}) \in \mathcal{L}_{\mathrm{Th}}$.

Definition

$\mathrm{CT}^-(\mathrm{Th})$ is the $\mathcal{L}_T$ theory extending Th with the following axioms

1. $\forall s, t \quad T(\dot{R}(s, t)) \equiv R(s^\circ, t^\circ)$, $R \in \{=, <\}$.

Basic truth theories

Definition

$\mathrm{UTB}^{-}(\mathrm{Th})$ is the $\mathcal{L}_T := \mathcal{L}_{\mathrm{Th}} \cup \{T\}$ theory extending Th with all the axioms of the form

$$\forall x_1, \dots, x_n \quad T(\phi(\dot{x}_1, \dots, \dot{x}_n)) \equiv \phi(x_1, \dots, x_n),$$

for $\phi(\bar{x}) \in \mathcal{L}_{\mathrm{Th}}$.

Definition

$\mathrm{CT}^{-}(\mathrm{Th})$ is the $\mathcal{L}_T$ theory extending Th with the following axioms

1. $\forall s, t \quad T(\dot{R}(s, t)) \equiv R(s^{\circ}, t^{\circ}),\ R \in \{=, <\}$.
2. $\forall \phi \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\dot{\neg}\phi) \equiv \neg T(\phi)$.

Basic truth theories

Definition

$\mathrm{UTB}^{-}(\mathrm{Th})$ is the $\mathcal{L}_T := \mathcal{L}_{\mathrm{Th}} \cup \{T\}$ theory extending Th with all the axioms of the form

$$\forall x_1, \ldots, x_n \quad T(\phi(\dot{x}_1, \ldots, \dot{x}_n)) \equiv \phi(x_1, \ldots, x_n),$$

for $\phi(\bar{x}) \in \mathcal{L}_{\mathrm{Th}}$.

Definition

$\mathrm{CT}^{-}(\mathrm{Th})$ is the $\mathcal{L}_T$ theory extending Th with the following axioms

1. $\forall s, t \quad T(\dot{R}(s, t)) \equiv R(s^{\circ}, t^{\circ}),\ R \in \{=, <\}$.
2. $\forall \phi \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\dot{\neg}\phi) \equiv \neg T(\phi)$.
3. $\forall \phi, \psi \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\phi \dot{\wedge} \psi) \equiv T(\phi) \wedge T(\psi)$

Basic truth theories

Definition

$\mathrm{UTB}^-(\mathrm{Th})$ is the $\mathcal{L}_T := \mathcal{L}_{\mathrm{Th}} \cup \{T\}$ theory extending Th with all the axioms of the form

$$\forall x_1, \ldots, x_n \quad T(\phi(\dot{x}_1, \ldots, \dot{x}_n)) \equiv \phi(x_1, \ldots, x_n),$$

for $\phi(\bar{x}) \in \mathcal{L}_{\mathrm{Th}}$.

Definition

$\mathrm{CT}^-(\mathrm{Th})$ is the $\mathcal{L}_T$ theory extending Th with the following axioms

1. $\forall s, t \quad T(\dot{R}(s,t)) \equiv R(s^\circ, t^\circ)$, $R \in \{=, <\}$.
2. $\forall \phi \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\dot{\neg}\phi) \equiv \neg T(\phi)$.
3. $\forall \phi, \psi \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\phi \dot{\wedge} \psi) \equiv T(\phi) \wedge T(\psi)$
4. $\forall \phi(v) \in \mathrm{Form}^{\leq 1}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\dot{\exists} v \phi(v)) \equiv \exists x T(\phi(\dot{x}))$.

Some metamathematics of UTB^-

Proposition (Essentially due to Tarski)

If $\mathcal{L}_{\mathrm{PA}} \supseteq \mathrm{Th} \supseteq \mathrm{EA}$, *then* $\mathrm{UTB}^-(\mathrm{Th})$ *is proof theoretically conservative over* Th.

Some metamathematics of UTB^-

Proposition (Essentially due to Tarski)

If $\mathcal{L}_{\mathrm{PA}} \supseteq \mathrm{Th} \supseteq \mathrm{EA}$, *then* $\mathrm{UTB}^-(\mathrm{Th})$ *is proof theoretically conservative over* Th.

The proof idea is that we can interpret finitely many axioms of UTB^- in Th by changing T to the (sufficiently large) partial truth predicate.

Some metamathematics of UTB^-

Proposition (Essentially due to Tarski)

If $\mathcal{L}_{\mathrm{PA}} \supseteq \mathrm{Th} \supseteq \mathrm{EA}$, *then* $\mathrm{UTB}^-(\mathrm{Th})$ *is proof theoretically conservative over* Th.

Proposition (Folklore)

$\mathrm{UTB}^-(\mathrm{Th})$ *coincide with the following* $\mathcal{L}_T$ *theory, extending* Th *with the following axioms*

1. $\forall s, t \;\; T(\dot{R}(s, t)) \equiv R(s^\circ, t^\circ)$, $R \in \{=, <\}$.

Some metamathematics of UTB^-

Proposition (Essentially due to Tarski)

If $\mathcal{L}_{\mathrm{PA}} \supseteq \mathrm{Th} \supseteq \mathrm{EA}$, *then* $\mathrm{UTB}^-(\mathrm{Th})$ *is proof theoretically conservative over* Th.

Proposition (Folklore)

$\mathrm{UTB}^-(\mathrm{Th})$ *coincide with the following* $\mathcal{L}_T$ *theory, extending* Th *with the following axioms*

1. $\forall s, t \;\; T(\dot{R}(s,t)) \equiv R(s^\circ, t^\circ)$, $R \in \{=, <\}$.
2. $\forall \phi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \;\; T(\dot{\neg}\phi) \equiv \neg T(\phi)$.

Some metamathematics of UTB^-

Proposition (Essentially due to Tarski)

If $\mathcal{L}_{\mathrm{PA}} \supseteq \mathrm{Th} \supseteq \mathrm{EA}$, *then* $\mathrm{UTB}^-(\mathrm{Th})$ *is proof theoretically conservative over* Th.

Proposition (Folklore)

$\mathrm{UTB}^-(\mathrm{Th})$ *coincide with the following* $\mathcal{L}_T$ *theory, extending* Th *with the following axioms*

1. $\forall s, t \quad T(\dot{R}(s,t)) \equiv R(s^\circ, t^\circ)$, $R \in \{=, <\}$.
2. $\forall \phi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\dot{\neg}\phi) \equiv \neg T(\phi)$.
3. $\forall \phi, \psi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\phi \dot{\wedge} \psi) \equiv T(\phi) \wedge T(\psi)$

Some metamathematics of UTB^-

Proposition (Essentially due to Tarski)

If $\mathcal{L}_{\mathrm{PA}} \supseteq \mathrm{Th} \supseteq \mathrm{EA}$, *then* $\mathrm{UTB}^-(\mathrm{Th})$ *is proof theoretically conservative over* Th.

Proposition (Folklore)

$\mathrm{UTB}^-(\mathrm{Th})$ *coincide with the following* $\mathcal{L}_T$ *theory, extending* Th *with the following axioms*

1. $\forall s, t \quad T(\dot{R}(s,t)) \equiv R(s^\circ, t^\circ)$, $R \in \{=, <\}$.
2. $\forall \phi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\dot{\neg}\phi) \equiv \neg T(\phi)$.
3. $\forall \phi, \psi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\phi \dot{\wedge} \psi) \equiv T(\phi) \wedge T(\psi)$
4. $\forall \phi(v) \in \mathrm{Form}^{\leq 1, \mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \quad T(\dot{\exists} v \phi(v)) \equiv \exists x T(\phi(\dot{x}))$.

Some metamathematics of UTB^-

Proposition (Essentially due to Tarski)

If $\mathcal{L}_{\mathrm{PA}} \supseteq \mathrm{Th} \supseteq \mathrm{EA}$, then $\mathrm{UTB}^-(\mathrm{Th})$ is proof theoretically conservative over Th.

Proposition (Folklore)

$\mathrm{UTB}^-(\mathrm{Th})$ *coincide with the following $\mathcal{L}_T$ theory, extending* Th *with the following axioms*

1. $\forall s, t \;\; T(\dot{R}(s,t)) \equiv R(s^\circ, t^\circ)$, $R \in \{=, <\}$.
2. $\forall \phi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \;\; T(\dot{\neg}\phi) \equiv \neg T(\phi)$.
3. $\forall \phi, \psi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \;\; T(\phi \dot{\wedge} \psi) \equiv T(\phi) \wedge T(\psi)$
4. $\forall \phi(v) \in \mathrm{Form}^{\leq 1, \mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \;\; T(\dot{\exists} v \phi(v)) \equiv \exists x T(\phi(\dot{x}))$.

Some metamathematics of UTB^-

Proposition (Essentially due to Tarski)

If $\mathcal{L}_{\mathrm{PA}} \supseteq \mathrm{Th} \supseteq \mathrm{EA}$, *then* $\mathrm{UTB}^-(\mathrm{Th})$ *is proof theoretically conservative over* Th.

Proposition (Folklore)

$\mathrm{UTB}^-(\mathrm{Th})$ *coincide with the following* $\mathcal{L}_T$ *theory, extending* Th *with the following axioms*

1. $\forall s, t \;\; T(\dot{R}(s,t)) \equiv R(s^\circ, t^\circ)$, $R \in \{=, <\}$.
2. $\forall \phi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \;\; T(\dot{\neg}\phi) \equiv \neg T(\phi)$.
3. $\forall \phi, \psi \in \mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \;\; T(\phi \dot{\wedge} \psi) \equiv T(\phi) \wedge T(\psi)$
4. $\forall \phi(v) \in \mathrm{Form}^{\leq 1, \mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}} \;\; T(\dot{\exists} v \phi(v)) \equiv \exists x T(\phi(\dot{x}))$.

for $n \in \mathbb{N}$. $\mathrm{Sent}^{\mathrm{dpt}(\underline{n})}_{\mathcal{L}_{\mathrm{Th}}}(x)$ *expresses that* x *is a sentence of* $\mathcal{L}_{\mathrm{Th}}$ *of logical depth at most* n. *Let* $\mathrm{CT}^- \restriction_x$ *denote the above formula.*

Some model theory of UTB^-.

Proposition (Exercise)

If $(\mathcal{M}, T), (\mathcal{M}', T') \models \mathrm{UTB}^-(\mathrm{EA})$ *and* $(\mathcal{M}, T) \subseteq (\mathcal{M}', T')$, *then* $\mathcal{M} \preceq \mathcal{M}'$.

Some model theory of UTB^-.

Proposition (Exercise)

If $(\mathcal{M}, T), (\mathcal{M}', T') \models \mathrm{UTB}^-(\mathrm{EA})$ *and* $(\mathcal{M}, T) \subseteq (\mathcal{M}', T')$*, then* $\mathcal{M} \preceq \mathcal{M}'$*.*

Proposition (Exercise 2)

If $\mathcal{M} \preceq \mathcal{M}'$*, and* $(\mathcal{M}', T') \models \mathrm{UTB}^-(\mathrm{EA})$*, then* $(\mathcal{M}, T' \restriction_M) \models \mathrm{UTB}^-(\mathrm{EA})$*.*

Some metamathematics of CT^-

Theorem (Enayat-Visser, Leigh)

$\mathrm{CT}^-(\mathrm{Th})$ *is conservative over* Th *for all reasonable* Th.

Denote with $\mathrm{CT}_n(\mathrm{Th})$ the extension of $\mathrm{CT}^-(\mathrm{Th})$ with induction for Σ_n formulae.

Some metamathematics of CT^-

Theorem (Enayat-Visser, Leigh)

$\mathrm{CT}^-(\mathrm{Th})$ *is conservative over* Th *for all reasonable* Th.

Denote with $\mathrm{CT}_n(\mathrm{Th})$ the extension of $\mathrm{CT}^-(\mathrm{Th})$ with induction for Σ_n formulae.

Theorem (Kotlarski-Smoryński-Wcisło, Ł.)

The arithmetical part of $\mathrm{CT}_0(\mathrm{PA})$ *is* $\mathrm{REF}^\omega(\mathrm{PA})$.

Global reflection & induction

Having a truth predicate one might consider a "finitization" of the uniform reflection scheme. This principle is called Global Reflection for Th

$$\forall \phi \in \mathcal{L}_{\mathsf{Th}} \;\; \mathsf{Prov}_{\mathsf{Th}}(\phi) \rightarrow T(\phi) \tag{GR}$$

Theorem (Cieśliński-Kotlarski-Łełyk)

Over $\mathsf{CT}^-(\mathsf{PA})$ *the following principles coincide:*

Global reflection & induction

Having a truth predicate one might consider a "finitization" of the uniform reflection scheme. This principle is called Global Reflection for Th

$$\forall \phi \in \mathcal{L}_{\mathrm{Th}} \ \ \mathrm{Prov}_{\mathrm{Th}}(\phi) \rightarrow T(\phi) \tag{GR}$$

Theorem (Cieśliński-Kotlarski-Łełyk)

Over $\mathrm{CT}^{-}(\mathrm{PA})$ *the following principles coincide:*

1. Δ_0*-induction for* $\mathcal{L}_T$.

Global reflection & induction

Having a truth predicate one might consider a "finitization" of the uniform reflection scheme. This principle is called Global Reflection for Th

$$\forall \phi \in \mathcal{L}_{\mathrm{Th}} \ \ \mathrm{Prov}_{\mathrm{Th}}(\phi) \rightarrow T(\phi) \tag{GR}$$

Theorem (Cieśliński-Kotlarski-Łełyk)

Over $\mathrm{CT}^{-}(\mathrm{PA})$ *the following principles coincide:*

1. Δ_0*-induction for* $\mathcal{L}_T$*.*
2. $\mathrm{GR}(\mathrm{PA})$*.*

Global reflection & induction

Having a truth predicate one might consider a "finitization" of the uniform reflection scheme. This principle is called Global Reflection for Th

$$\forall \phi \in \mathcal{L}_{\mathrm{Th}} \;\; \mathrm{Prov}_{\mathrm{Th}}(\phi) \rightarrow T(\phi) \tag{GR}$$

Theorem (Cieśliński-Kotlarski-Łełyk)

Over $\mathrm{CT}^{-}(\mathrm{PA})$ *the following principles coincide:*

1. Δ_0*-induction for* $\mathcal{L}_T$*.*
2. $\mathrm{GR}(\mathrm{PA})$*.*
3. $\forall \phi \in \mathcal{L}_{\mathrm{PA}} \;\; \mathrm{Prov}_{\emptyset}^{T}(\phi) \rightarrow T(\phi)$*.*

The most important property of CT_0

In what follows, if $(\mathcal{M}, T) \models \mathsf{CT}_0$ and $a \in M$, then T_a denotes the restriction of T for formulae of logical depth at most a (this makes sense inside of $\mathcal{M}$).

Theorem (Ł.-Wcisło)

Suppose $(\mathcal{M}, T) \models \mathsf{CT}_0$. Then for every $b \in M$, $(\mathcal{M}, T_{b+1}) \models \mathsf{CT}^-\restriction_b + \mathrm{Ind}(\mathcal{L}_T)$.

Reflecting over the disquotational truth

Proposition (Beklemishev-Pakhomov)

$\mathsf{UTB}^-(\mathsf{EA}) + \Delta_0\text{-}\mathsf{REF}(\mathsf{UTB}^-(\mathsf{Th})) \vdash \mathsf{REF}(\mathsf{Th})$.

This is fairly obvious, since for every $\phi \in \mathcal{L}_{\mathsf{Th}}$, $T(\phi)$ is a $\Delta_0(\mathcal{L}_T)$ formula.

Reflecting over the disquotational truth

Proposition (Beklemishev-Pakhomov)

$\mathsf{UTB}^-(\mathsf{EA}) + \Delta_0\text{-}\mathsf{REF}(\mathsf{UTB}^-(\mathsf{Th})) \vdash \mathsf{REF}(\mathsf{Th})$.

This is fairly obvious, since for every $\phi \in \mathcal{L}_{\mathsf{Th}}$, $T(\phi)$ is a $\Delta_0(\mathcal{L}_T)$ formula.

Proposition (Beklemishev-Pakhomov)

$\mathsf{UTB}^-(\mathsf{EA}) + \Sigma_1\text{-}\mathsf{REF}(\mathsf{UTB}^-(\mathsf{Th})) \vdash \mathsf{CT}^-$.

Reflecting over the disquotational truth

Proposition (Beklemishev-Pakhomov)

$\mathsf{UTB}^{-}(\mathsf{EA}) + \Delta_0\text{-}\mathsf{REF}(\mathsf{UTB}^{-}(\mathsf{Th})) \vdash \mathsf{REF}(\mathsf{Th})$.

This is fairly obvious, since for every $\phi \in \mathcal{L}_{\mathsf{Th}}$, $T(\phi)$ is a $\Delta_0(\mathcal{L}_T)$ formula.

Proposition (Beklemishev-Pakhomov)

$\mathsf{UTB}^{-}(\mathsf{EA}) + \Sigma_1\text{-}\mathsf{REF}(\mathsf{UTB}^{-}(\mathsf{Th})) \vdash \mathsf{CT}^{-}$.

Reflecting over the disquotational truth

Proposition (Beklemishev-Pakhomov)

$\mathsf{UTB}^-(\mathsf{EA}) + \Delta_0\text{-}\mathsf{REF}(\mathsf{UTB}^-(\mathsf{Th})) \vdash \mathsf{REF}(\mathsf{Th})$.

This is fairly obvious, since for every $\phi \in \mathcal{L}_{\mathsf{Th}}$, $T(\phi)$ is a $\Delta_0(\mathcal{L}_T)$ formula.

Proposition (Beklemishev-Pakhomov)

$\mathsf{UTB}^-(\mathsf{EA}) + \Sigma_1\text{-}\mathsf{REF}(\mathsf{UTB}^-(\mathsf{Th})) \vdash \mathsf{CT}^-$.

This once again is fairly obvious, since $\mathsf{CT}^-\restriction_x$ is a $\Pi_2(\mathcal{L}_T)$ sentence and

$$\mathsf{EA} \vdash \forall x \mathsf{Prov}_{\mathsf{UTB}^-(\mathsf{Th})}(\mathsf{CT}^-\restriction_{\dot{x}}).$$

The curious case of Σ_1 reflection

Theorem (Beklemishev-Pakhomov)

The arithmetical consequences of CT_0 *and* $\mathsf{UTB}^-(\mathsf{EA}) + \Sigma_1(\mathcal{L}_T)\text{-}\mathsf{REF}(\mathsf{UTB}^-(\mathsf{EA}))$ *coincide.*

The curious case of Σ_1 reflection

Theorem (Beklemishev-Pakhomov)

The arithmetical consequences of CT_0 *and* $\mathrm{UTB}^-(\mathrm{EA}) + \Sigma_1(\mathcal{L}_T)\text{-REF}(\mathrm{UTB}^-(\mathrm{EA}))$ *coincide.*

Question

Does $\mathrm{CT}_0 \vdash \Sigma_1(\mathcal{L}_T)\text{-REF}(\mathrm{UTB}^-(\mathrm{EA}))$?

The curious case of Σ_1 reflection

Theorem (Beklemishev-Pakhomov)

The arithmetical consequences of CT_0 *and* $\mathrm{UTB}^-(\mathrm{EA}) + \Sigma_1(\mathcal{L}_T)\text{-REF}(\mathrm{UTB}^-(\mathrm{EA}))$ *coincide.*

Question

Does $\mathrm{CT}_0 \vdash \Sigma_1(\mathcal{L}_T)\text{-REF}(\mathrm{UTB}^-(\mathrm{EA}))$?

The curious case of Σ_1 reflection

Theorem (Beklemishev-Pakhomov)

The arithmetical consequences of CT_0 *and* $\mathrm{UTB}^-(\mathrm{EA}) + \Sigma_1(\mathcal{L}_T)\text{-}\mathrm{REF}(\mathrm{UTB}^-(\mathrm{EA}))$ *coincide.*

Question

Does $\mathrm{CT}_0 \vdash \Sigma_1(\mathcal{L}_T)\text{-}\mathrm{REF}(\mathrm{UTB}^-(\mathrm{EA}))$?

Claim: It does.

For starters: Δ_0-reflection for the disquotational truth

Theorem (Beklemishev-Pakhomov)

$\mathrm{UTB}^-(\mathrm{PA}) + \Delta_0(\mathcal{L}_T)\text{-REF}(\mathrm{UTB}^-(\mathrm{PA}))$ *is arithmetically conservative over* $\mathcal{L}_{\mathrm{PA}}\text{-REF}(\mathrm{PA})$.

Fix any $(\mathcal{M}, T) \models \mathcal{L}_{\mathrm{PA}}\text{-REF}(\mathrm{PA}) + \mathrm{UTB}(\mathrm{PA})$. For every a, let $T{\restriction}_a$ denote the restriction of T to formulae of logical depth at most a. Observe that for every $n \in \mathbb{N}$

$$(\mathcal{M}, T) \models \forall \phi(\mathrm{dpt}(\phi) \leq n \wedge \mathrm{Prov}_{\mathrm{PA}}^{T\restriction_n}(\phi) \rightarrow T(\phi)).$$

So by overspill there exists a $c > \mathbb{N}$ such that

$$(\mathcal{M}, T) \models \forall \phi(\mathrm{dpt}(\phi) \leq c \wedge \mathrm{Prov}_{\mathrm{PA}}^{T\restriction_c}(\phi) \rightarrow T(\phi)).$$

Δ_0-reflection for the disquotational truth

$$(\mathcal{M}, T) \models \forall \phi(\mathrm{dpt}(\phi) \leq c \wedge \mathrm{Prov}_{\mathrm{PA}}^{T\restriction c}(\phi) \rightarrow T(\phi)). \qquad (*)$$

Consider the following $(\mathcal{M}, T)$ definable theory

$$\mathrm{Th} := \mathrm{PA} + \{\phi \mid \mathrm{dpt}(\phi) \leq c \wedge T(\phi)\}.$$

Work in $(\mathcal{M}, T)$. $(*)$ witnesses that Th is consistent. Moreover, the trivial conservativity proof for $\mathrm{UTB}^-(\mathrm{PA})$ shows that

$$\mathrm{UTB}^-(\mathrm{Th})$$

is consistent as well. Let $\mathcal{M}' = (M', T', S')$ be an $(\mathcal{M}, T)$ definable model for $\mathrm{UTB}^-(\mathrm{Th})$, where S' is a satisfaction relation for (M', T').

Δ_0-reflection for the disquotational truth

$$\mathrm{Th} := \mathrm{PA} + \{\phi \mid \mathrm{dpt}(\phi) \leq c \wedge T(\phi)\}$$
$$\mathcal{M}' := (M', T') \models_{S'} \mathrm{UTB}^-(\mathrm{Th})$$

Claim: $\mathcal{M} \preceq_e \mathcal{M}'$
Indeed, if $\mathcal{M} \models \neg\phi(a)$ for $a \in M$, then $(\mathcal{M}, T) \models T(\neg\phi(\underline{a})) \wedge \mathrm{dpt}(\phi(\underline{a})) \leq c$. Hence $\neg\phi(\underline{a}) \in \mathrm{Th}$, hence $\mathcal{M}' \models_{S'} \neg\phi(a)$.
Claim 2: $(\mathcal{M}, T'\restriction_M) \models \mathrm{UTB}^-$. This follows, since $(M', T') \models_{S'} \mathrm{UTB}^-$ and $\mathcal{M} \preceq \mathcal{M}'$.
Claim 3: $(\mathcal{M}, T'\restriction_M) \models \Delta_0(\mathcal{L}_T)$-REF$(\mathrm{UTB}^-(\mathrm{PA}))$. Pick a $\Delta_0(\mathcal{L}_T)$ formula $\phi(x)$, $a \in M$ and, working in $(\mathcal{M}, T'\restriction_M)$ suppose $\mathrm{Prov}_{\mathrm{UTB}^-(\mathrm{PA})}(\phi(\underline{a}))$. Hence $(M', T') \models_{S'} \phi(a)$. Since

$$(\mathcal{M}, T'\restriction_M) \subseteq_e (M', T'),$$

we are done.

Outline

Ouverture: Δ_0-case

Theorem

$\mathrm{CT}_0 \vdash \Delta_0(\mathcal{L}_T)\text{-REF}(\mathrm{UTB}(\mathrm{PA}) + T)$.

Fix an arbitrary $(\mathcal{M}, T) \models \mathrm{CT}_0$, a $\Delta_0(\mathcal{L}_T)$-formula $\phi(x)$ and $a \in M$. Suppose that $\mathcal{M} \models \mathrm{Prov}^{T_{d'}}_{\mathrm{UTB}}(\phi(\underline{a}))$, for some $d' \in M$. Let $b \in M$ be big enough so that for every $T'\restriction_b = T\restriction_b$ and every $\mathcal{M}' \supseteq_e \mathcal{M}$

$$(\mathcal{M}, T\restriction_b) \models \phi(a) \iff (\mathcal{M}', T') \models \phi(a).$$

Claim: There exists $\mathcal{M}' \supseteq_e \mathcal{M}$ and $T'\restriction_b = T\restriction_b$ such that $(\mathcal{M}', T') \models \phi(a)$.

Ouverture: Δ_0-case

Claim: There exists $\mathcal{M}' \supseteq_e \mathcal{M}$ and $T'\restriction_b = T\restriction_b$ such that $(\mathcal{M}', T') \models \phi(a)$.
Let c fix $T\restriction_b$ in $\mathcal{M}$, i.e. $\phi \in c^{\mathcal{M}}$ iff

$$\phi \in T\restriction_b \vee (\phi = \neg\psi \wedge \psi < b \wedge \psi \notin T\restriction_b).$$

Let b' be big enough such that $c \subseteq T_{b'}$ and let $d = \max\{d', b'\}$.
Internally in $(\mathcal{M}, T)$ consider the following theory $\mathcal{L}_T$-definable theory

$$\mathrm{Th} := \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \mid T(\phi)\}.$$

We claim that $\mathcal{M} \models \mathrm{Con}_{\mathrm{Th}}$. This holds, since

1. The proof of conservativity of UTB over (any extension of) PA formalizes in PA.

Ouverture: Δ_0-case

Claim: There exists $\mathcal{M}' \supseteq_e \mathcal{M}$ and $T' \restriction_b = T \restriction_b$ such that $(\mathcal{M}', T') \models \phi(a)$.
Let c fix $T \restriction_b$ in $\mathcal{M}$, i.e. $\phi \in c^{\mathcal{M}}$ iff

$$\phi \in T \restriction_b \vee (\phi = \neg\psi \wedge \psi < b \wedge \psi \notin T \restriction_b).$$

Let b' be big enough such that $c \subseteq T_{b'}$ and let $d = \max\{d', b'\}$. Internally in $(\mathcal{M}, T)$ consider the following theory $\mathcal{L}_T$-definable theory

$$\mathrm{Th} := \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \ \mid \ T(\phi)\}.$$

We claim that $\mathcal{M} \models \mathrm{Con}_{\mathrm{Th}}$. This holds, since

1. The proof of conservativity of UTB over (any extension of) PA formalizes in PA.
2. $(\mathcal{M}, T) \models \forall \phi \in \mathcal{L}_{\mathrm{PA}} (\mathrm{Prov}^{T}_{\mathrm{PA}}(\phi) \rightarrow T(\phi))$.

Ouverture: Δ_0-case

Claim: There exists $\mathcal{M}' \supseteq_e \mathcal{M}$ and $T'\restriction_b = T\restriction_b$ such that $(\mathcal{M}', T') \models \phi(a)$.
Let c fix $T\restriction_b$ in $\mathcal{M}$, i.e. $\phi \in c^{\mathcal{M}}$ iff

$$\phi \in T\restriction_b \vee (\phi = \neg\psi \wedge \psi < b \wedge \psi \notin T\restriction_b).$$

Let b' be big enough such that $c \subseteq T_{b'}$ and let $d = \max\{d', b'\}$.
Internally in $(\mathcal{M}, T)$ consider the following theory $\mathcal{L}_T$-definable theory

$$\mathrm{Th} := \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \ \mid \ T(\phi)\}.$$

We claim that $\mathcal{M} \models \mathrm{Con}_{\mathrm{Th}}$. This holds, since

1. The proof of conservativity of UTB over (any extension of) PA formalizes in PA.
2. $(\mathcal{M}, T) \models \forall\phi \in \mathcal{L}_{\mathrm{PA}}(\mathrm{Prov}^T_{\mathrm{PA}}(\phi) \to T(\phi))$.

Ouverture: Δ_0-case

Claim: There exists $\mathcal{M}' \supseteq_e \mathcal{M}$ and $T'\restriction_b = T\restriction_b$ such that $(\mathcal{M}', T') \models \phi(a)$.
Let c fix $T\restriction_b$ in $\mathcal{M}$, i.e. $\phi \in c^{\mathcal{M}}$ iff

$$\phi \in T\restriction_b \vee (\phi = \neg\psi \wedge \psi < b \wedge \psi \notin T\restriction_b).$$

Let b' be big enough such that $c \subseteq T_{b'}$ and let $d = \max\{d', b'\}$.
Internally in $(\mathcal{M}, T)$ consider the following theory $\mathcal{L}_T$-definable theory

$$\mathrm{Th} := \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \mid T(\phi)\}.$$

We claim that $\mathcal{M} \models \mathrm{Con}_{\mathrm{Th}}$. This holds, since

1. The proof of conservativity of UTB over (any extension of) PA formalizes in PA.
2. $(\mathcal{M}, T) \models \forall\phi \in \mathcal{L}_{\mathrm{PA}}(\mathrm{Prov}^T_{\mathrm{PA}}(\phi) \rightarrow T(\phi))$.

Let $(\mathcal{M}', T')$ be any $(\mathcal{M}, T_d)$ definable model of Th.

Ouverture: Δ_0 - case

$$c^{\mathcal{M}} = T\restriction_b \cup \{\neg\phi \mid \phi < b \wedge \phi \notin T\restriction_b\}.$$
$$\mathrm{Th} = \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \mid T(\phi)\}.$$
$$\mathcal{M} \models "(\mathcal{M}', T') \models \mathrm{Th}".$$

We check:

- $\mathcal{M}' \supseteq_e \mathcal{M}$. This holds, since $\mathcal{M}'$ is an $(\mathcal{M}, T_d)$-definable model of Q and $(\mathcal{M}, T_d) \models \mathrm{PA}^*$.

Ouverture: Δ_0 - case

$$c^{\mathcal{M}} = T\restriction_b \cup \{\neg\phi \mid \phi < b \wedge \phi \notin T\restriction_b\}.$$

$$\mathrm{Th} = \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \mid T(\phi)\}.$$

$$\mathcal{M} \models "(\mathcal{M}', T') \models \mathrm{Th}".$$

We check:

- $\mathcal{M}' \supseteq_e \mathcal{M}$. This holds, since $\mathcal{M}'$ is an $(\mathcal{M}, T_d)$-definable model of Q and $(\mathcal{M}, T_d) \models \mathrm{PA}^*$.
- $(\mathcal{M}', T') \models \phi(a)$. This holds, since $\mathcal{M} \models "(\mathcal{M}', T') \models \mathrm{UTB} + \phi"$, for every $\phi \in \Sigma_{d'}$ and $\mathcal{M} \models \mathrm{Prov}^{T_{d'}}_{\mathrm{UTB(PA)}}(\phi(\underline{a}))$.

Ouverture: Δ_0 - case

$$c^{\mathcal{M}} = T{\restriction_b} \cup \{\neg\phi \mid \phi < b \land \phi \notin T{\restriction_b}\}.$$
$$\mathrm{Th} = \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \mid T(\phi)\}.$$
$$\mathcal{M} \models \text{"}(\mathcal{M}', T') \models \mathrm{Th}\text{"}.$$

We check:

- $\mathcal{M}' \supseteq_e \mathcal{M}$. This holds, since $\mathcal{M}'$ is an $(\mathcal{M}, T_d)$-definable model of Q and $(\mathcal{M}, T_d) \models \mathrm{PA}^*$.
- $(\mathcal{M}', T') \models \phi(a)$. This holds, since $\mathcal{M} \models \text{"}(\mathcal{M}', T') \models \mathrm{UTB} + \phi\text{"}$, for every $\phi \in \Sigma_{d'}$ and $\mathcal{M} \models \mathrm{Prov}^{T_{d'}}_{\mathrm{UTB(PA)}}(\phi(\underline{a}))$.
- $T'{\restriction_b} = T{\restriction_b}$. This holds, since for $\psi < b$

$$\psi \in T_b \Leftrightarrow \mathcal{M} \models \text{"}\mathcal{M}' \models \psi\text{"} \quad (1)$$
$$\Leftrightarrow \mathcal{M} \models \text{"}(\mathcal{M}', T') \models T(\psi)\text{"} \quad (2)$$
$$\Leftrightarrow \psi \in T'. \quad (3)$$

Ouverture: Δ_0 - case

$$c^{\mathcal{M}} = T{\restriction_b} \cup \{\neg\phi \mid \phi < b \wedge \phi \notin T{\restriction_b}\}.$$
$$\mathrm{Th} = \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \mid T(\phi)\}.$$
$$\mathcal{M} \models \text{"}(\mathcal{M}', T') \models \mathrm{Th}\text{"}.$$

We check:

- $\mathcal{M}' \supseteq_e \mathcal{M}$. This holds, since $\mathcal{M}'$ is an $(\mathcal{M}, T_d)$-definable model of Q and $(\mathcal{M}, T_d) \models \mathrm{PA}^*$.
- $(\mathcal{M}', T') \models \phi(a)$. This holds, since $\mathcal{M} \models \text{"}(\mathcal{M}', T') \models \mathrm{UTB} + \phi\text{"}$, for every $\phi \in \Sigma_{d'}$ and $\mathcal{M} \models \mathrm{Prov}^{T_{d'}}_{\mathrm{UTB(PA)}}(\phi(\underline{a}))$.
- $T'{\restriction_b} = T{\restriction_b}$. This holds, since for $\psi < b$

$$\psi \in T_b \Leftrightarrow \mathcal{M} \models \text{"}\mathcal{M}' \models \psi\text{"} \quad (1)$$
$$\Leftrightarrow \mathcal{M} \models \text{"}(\mathcal{M}', T') \models T(\psi)\text{"} \quad (2)$$
$$\Leftrightarrow \psi \in T'. \quad (3)$$

Ouverture: Δ_0 - case

$$c^{\mathcal{M}} = T{\restriction_b} \cup \{\neg\phi \mid \phi < b \wedge \phi \notin T{\restriction_b}\}.$$
$$\mathrm{Th} = \mathrm{UTB} + \mathrm{PA} + \{\phi \in \Sigma_d \mid T(\phi)\}.$$
$$\mathcal{M} \models \text{"}(\mathcal{M}', T') \models \mathrm{Th}\text{"}.$$

We check:

- $\mathcal{M}' \supseteq_e \mathcal{M}$. This holds, since $\mathcal{M}'$ is an $(\mathcal{M}, T_d)$-definable model of Q and $(\mathcal{M}, T_d) \models \mathrm{PA}^*$.
- $(\mathcal{M}', T') \models \phi(a)$. This holds, since $\mathcal{M} \models \text{"}(\mathcal{M}', T') \models \mathrm{UTB} + \phi\text{"}$, for every $\phi \in \Sigma_{d'}$ and $\mathcal{M} \models \mathrm{Prov}^{T_{d'}}_{\mathrm{UTB(PA)}}(\phi(\underline{a}))$.
- $T'{\restriction_b} = T{\restriction_b}$. This holds, since for $\psi < b$

$$\begin{aligned} \psi \in T_b &\Leftrightarrow \mathcal{M} \models \text{"}\mathcal{M}' \models \psi\text{"} && (1)\\ &\Leftrightarrow \mathcal{M} \models \text{"}(\mathcal{M}', T') \models T(\psi)\text{"} && (2)\\ &\Leftrightarrow \psi \in T'. && (3) \end{aligned}$$

And we are done.