Proof Theory in Philosophy

Carlo Nicolai

KING'S College LONDON

Slides available at https://carlonicolai.github.io

Outline

What I'm not considering

I don't consider **proof-theoretic semantics**.

I only briefly touch upon **reductive proof-theory** in the philosophy of mathematics.

Abstraction and Truth

'There never were any set-theoretic paradoxes, but the property theoretic paradoxes are still unresolved' (Gödel to Myhill)

Naïve abstraction

$$\forall x\big(x \in \{v \mid \varphi(v)\} \leftrightarrow \varphi(x)\big)$$

Naïve Truth

$$\mathtt{Tr}\ulcorner A \urcorner \leftrightarrow A$$

Here I assume that for any φ in the language there is a term $\{v \mid \varphi(v)\}$ with $\mathtt{FV}(\{v \mid \varphi(v)\}) = \mathtt{FV}(\varphi) \setminus \{v\}$. If φ is a sentence, I write $\ulcorner A \urcorner$ for 'the proposition expressed by A'.

Liar

$\Gamma, \Delta, \Theta, \Lambda, \ldots$ are multisets of formulas.

Truth rules

$$\frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow \mathtt{Tr}\ulcorner A \urcorner} \qquad \frac{A, \Gamma \Rightarrow D}{\mathtt{Tr}\ulcorner A \urcorner, \Gamma \Rightarrow D}$$

$$\lambda \Leftrightarrow \neg\mathtt{Tr}\ulcorner \lambda \urcorner \qquad \neg\lambda \Leftrightarrow \mathtt{Tr}\ulcorner \lambda \urcorner$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\lambda \Rightarrow \lambda}{\mathtt{Tr}\ulcorner \lambda \urcorner \Rightarrow \lambda}}{\mathtt{Tr}\ulcorner \lambda \urcorner, \neg\lambda \Rightarrow}}{\mathtt{Tr}\ulcorner \lambda \urcorner \Rightarrow}}{\Rightarrow \neg\mathtt{Tr}\ulcorner \lambda \urcorner}}{\Rightarrow \lambda} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\lambda \Rightarrow \lambda}{\lambda \Rightarrow \mathtt{Tr}\ulcorner \lambda \urcorner}}{\lambda, \neg\mathtt{Tr}\ulcorner \lambda \urcorner \Rightarrow}}{\lambda, \lambda \Rightarrow}}{\lambda \Rightarrow}}{\Rightarrow}$$

Liar

Truth rules

$$\frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow \mathtt{Tr}\ulcorner A \urcorner} \qquad \frac{A, \Gamma \Rightarrow D}{\mathtt{Tr}\ulcorner A \urcorner, \Gamma \Rightarrow D}$$

$$\lambda \Leftrightarrow \neg \mathtt{Tr}\ulcorner \lambda \urcorner \qquad \neg\lambda \Leftrightarrow \mathtt{Tr}\ulcorner \lambda \urcorner$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\lambda \Rightarrow \lambda}{\color{red}{\mathtt{Tr}\ulcorner \lambda \urcorner \Rightarrow \lambda}}}{\color{red}{\mathtt{Tr}\ulcorner \lambda \urcorner, \neg\lambda \Rightarrow}}}{\mathtt{Tr}\ulcorner \lambda \urcorner \Rightarrow}}{\Rightarrow \neg \mathtt{Tr}\ulcorner \lambda \urcorner}}{\Rightarrow \lambda} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\lambda \Rightarrow \lambda}{\color{red}{\lambda \Rightarrow \mathtt{Tr}\ulcorner \lambda \urcorner}}}{\color{red}{\lambda, \neg \mathtt{Tr}\ulcorner \lambda \urcorner \Rightarrow}}}{\lambda, \lambda \Rightarrow}}{\lambda \Rightarrow}}{\Rightarrow}$$

Curry

Truth rules

$$\frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow \mathtt{Tr}\ulcorner A \urcorner} \qquad\qquad \frac{A, \Gamma \Rightarrow D}{\mathtt{Tr}\ulcorner A \urcorner, \Gamma \Rightarrow D}$$

$$\kappa \Leftrightarrow \mathtt{Tr}\ulcorner \kappa \urcorner \to \bot$$

$$\frac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\kappa \Rightarrow \kappa}{\kappa \Rightarrow \mathtt{Tr}\ulcorner \kappa \urcorner} \quad \bot \Rightarrow \bot}{\kappa, \mathtt{Tr}\ulcorner \kappa \urcorner \to \bot \Rightarrow \bot}}{\kappa, \kappa \Rightarrow \bot}}{\kappa \Rightarrow \bot}}{\Rightarrow \mathtt{Tr}\ulcorner \kappa \urcorner \to \bot}}{\Rightarrow \kappa} \qquad \dfrac{\dfrac{\dfrac{\kappa \Rightarrow \kappa}{\kappa \Rightarrow \mathtt{Tr}\ulcorner \kappa \urcorner} \quad \bot \Rightarrow \bot}{\kappa, \mathtt{Tr}\ulcorner \kappa \urcorner \to \bot \Rightarrow \bot}}{\kappa \Rightarrow \bot}}{\Rightarrow \bot}$$

Curry

Truth rules

$$\frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow \mathtt{Tr}\ulcorner A \urcorner} \qquad \frac{A, \Gamma \Rightarrow D}{\mathtt{Tr}\ulcorner A \urcorner, \Gamma \Rightarrow D}$$

$$\kappa \Leftrightarrow \mathtt{Tr}\ulcorner \kappa \urcorner \to \bot$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\kappa \Rightarrow \kappa}{\boldsymbol{\kappa \Rightarrow \mathtt{Tr}\ulcorner \kappa \urcorner}} \qquad \boldsymbol{\bot \Rightarrow \bot}}{\boldsymbol{\kappa, \mathtt{Tr}\ulcorner \kappa \urcorner \to \bot \Rightarrow \bot}}}{\kappa, \kappa \Rightarrow \bot}}{\kappa \Rightarrow \bot}}{\boldsymbol{\Rightarrow \mathtt{Tr}\ulcorner \kappa \urcorner \to \bot}}}{\Rightarrow \kappa} \qquad \dfrac{\dfrac{\dfrac{\kappa \Rightarrow \kappa}{\boldsymbol{\kappa \Rightarrow \mathtt{Tr}\ulcorner \kappa \urcorner}} \qquad \boldsymbol{\bot \Rightarrow \bot}}{\boldsymbol{\kappa, \mathtt{Tr}\ulcorner \kappa \urcorner \to \bot \Rightarrow \bot}}}{\kappa \Rightarrow \bot}}{\Rightarrow \bot}$$

Internal Curry

'Consequence' predicate

$$\frac{\Gamma, A \Rightarrow B \qquad \Gamma, C \Rightarrow D}{\Gamma, A, \mathtt{C}(\ulcorner B \urcorner, \ulcorner C \urcorner) \Rightarrow D} \qquad\qquad \frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow \mathtt{C}(\ulcorner A \urcorner, \ulcorner B \urcorner)}$$

$$\nu \Leftrightarrow \mathtt{C}(\ulcorner \nu \urcorner, \ulcorner \bot \urcorner)$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\nu \Rightarrow \nu \qquad \bot \Rightarrow \bot}{\nu, \mathtt{C}(\ulcorner \nu \urcorner, \ulcorner \bot \urcorner) \Rightarrow \bot}}{\nu \Rightarrow \bot}}{\Rightarrow \mathtt{C}(\ulcorner \nu \urcorner, \ulcorner \bot \urcorner)}}{\Rightarrow \nu} \qquad \dfrac{\dfrac{\nu \Rightarrow \nu \qquad \bot \Rightarrow \bot}{\nu, \mathtt{C}(\ulcorner \nu \urcorner, \ulcorner \bot \urcorner) \Rightarrow \bot}}{\nu \Rightarrow \bot}}{\Rightarrow \bot}$$

Internal Curry

‘Consequence’ predicate

$$\dfrac{\Gamma, A \Rightarrow B \qquad \Gamma, C \Rightarrow D}{\Gamma, A, \mathtt{C}(\ulcorner B \urcorner, \ulcorner C \urcorner) \Rightarrow D} \qquad\qquad \dfrac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow \mathtt{C}(\ulcorner A \urcorner, \ulcorner B \urcorner)}$$

$$\nu \Leftrightarrow \mathtt{C}(\ulcorner \nu \urcorner, \ulcorner \bot \urcorner)$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\nu \Rightarrow \nu \qquad \bot \Rightarrow \bot}{\nu, \mathtt{C}(\ulcorner \nu \urcorner, \ulcorner \bot \urcorner) \Rightarrow \bot}}{\nu \Rightarrow \bot}}{\Rightarrow \mathtt{C}(\ulcorner \nu \urcorner, \ulcorner \bot \urcorner)}}{\Rightarrow \nu} \quad \dfrac{\dfrac{\nu \Rightarrow \nu \qquad \bot \Rightarrow \bot}{\nu, \mathtt{C}(\ulcorner \nu \urcorner, \ulcorner \bot \urcorner) \Rightarrow \bot}}{\nu \Rightarrow \bot}$$

$$\Rightarrow \bot$$

Cut-elimination for truth and abstraction

The main extension of the standard inductive strategy consists in the reduction of cuts of the following form:

Tr -rules principal in the last inferences

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{D}_0\\ \Gamma \Rightarrow \Delta, A\end{array}}{\Gamma \Rightarrow \Delta, \mathtt{Tr}\ulcorner A\urcorner} \qquad \dfrac{\begin{array}{c}\mathcal{D}_1\\ A, \Gamma \Rightarrow \Delta\end{array}}{\mathtt{Tr}\ulcorner A\urcorner, \Gamma \Rightarrow \Delta}}{\Gamma \Rightarrow \Delta}$$

…which we would like to reduce to:

$$\dfrac{\begin{array}{c}\mathcal{D}_0\\ \Gamma \Rightarrow \Delta, A\end{array} \qquad \begin{array}{c}\mathcal{D}_1\\ A, \Gamma \Rightarrow \Delta\end{array}}{\Gamma \Rightarrow \Delta}$$

This creates a problem because $\mathtt{Tr}\ulcorner A\urcorner$ is atomic whereas A is of arbitrary (logical) complexity.

Tr -measures

- I will consider two ways of keeping track of applications of the truth rules in derivations: the first applies to nodes in the derivation tree, the second applies to single formulas within derivations.
- In the first case:

$$\dfrac{\dfrac{\gamma_0 \Rightarrow \top \quad \alpha}{\gamma_0 \Rightarrow \mathtt{Tr}\ulcorner \top \urcorner \quad \alpha + 1} \qquad \gamma_1 \Rightarrow \mathtt{Tr}\ulcorner \top \urcorner \quad \beta}{\gamma_0, \gamma_1 \Rightarrow \mathtt{Tr}\ulcorner \top \urcorner \wedge \mathtt{Tr}\ulcorner \top \urcorner \quad \mathtt{max}(\alpha, \beta)}$$

- In the second case:

$$\dfrac{\dfrac{\gamma_0 \Rightarrow {}^{0}\top}{\gamma_0 \Rightarrow {}^{1}\mathtt{Tr}\ulcorner \top \urcorner} \qquad \gamma_1 \Rightarrow {}^{0}\mathtt{Tr}\ulcorner \top \urcorner}{\gamma_0, \gamma_1 \Rightarrow {}^{\mathtt{max}(1, n)}\mathtt{Tr}\ulcorner \top \urcorner \wedge \mathtt{Tr}\ulcorner \top \urcorner}$$

Contraction-Free

Systems of truth and 'set theories' can be proved to be consistent via cut elimination arguments Grišin (1982), Petersen (2000), Cantini (2003).

Truth à la Grišin GT

$$\Gamma, \mathtt{Tr}\, s \Rightarrow \mathtt{Tr}\, s, \Delta\ [0] \qquad \Gamma \Rightarrow \top, \Delta\ [0] \qquad \Gamma, \bot \Rightarrow \Delta\ [0]$$

$$\frac{A, \Gamma \Rightarrow \Delta\ [\alpha]}{\mathtt{Tr}\ulcorner A \urcorner, \Gamma \Rightarrow \Delta\ [\alpha + 1]} \qquad \frac{\Gamma \Rightarrow \Delta, A\ [\alpha]}{\Gamma \Rightarrow \Delta, \mathtt{Tr}\ulcorner A \urcorner\ [\alpha + 1]}$$

$$\frac{\Gamma \Rightarrow \Delta, A_i\ [\alpha]}{\Gamma \Rightarrow \Delta, A_0 \sqcap A_1\ [\alpha]} \qquad \frac{\Gamma \Rightarrow \Delta, A\ [\alpha] \qquad \Gamma \Rightarrow \Delta, B\ [\beta]}{\Gamma \Rightarrow \Delta, A \sqcap B\ [\mathtt{max}(\alpha, \beta)]}$$

$$\frac{A, B, \Gamma \Rightarrow \Delta\ [\alpha]}{A \star B, \Gamma \Rightarrow \Delta\ [\alpha]} \qquad \frac{\Gamma \Rightarrow \Delta, A\ [\alpha] \qquad \Theta \Rightarrow \Lambda, B\ [\beta]}{\Gamma, \Theta \Rightarrow \Delta, \Lambda, A \star B\ [\alpha + \beta]}$$

Contraction-free

Systems of truth and 'set theories' can be proved to be consistent via cut elimination arguments Grišin (1982), Petersen (2000), Cantini (2003).

Lemma

Given cut-free derivations $\mathcal{D}_0 \vdash_{GT} \Gamma \Rightarrow \Delta, A$ *and* $\mathcal{D}_1 \vdash_{GT} A, \Theta \Rightarrow \Lambda$, *there is a* $\mathcal{D} \vdash_{GT} \Gamma, \Theta \Rightarrow \Delta, \Lambda$ *with the* $\mathcal{Tr}$*-rank* ρ *of* $\mathcal{D}$ *is* $\leq \rho(\mathcal{D}_0) + \rho(\mathcal{D}_1)$.

Proof Idea.

The induction is on $(\rho(\mathcal{D}_0) + \rho(\mathcal{D}_1), |A|, |\mathcal{D}_0| + |\mathcal{D}_1|)$. □

Contraction-free

Systems of truth and 'set theories' can be proved to be consistent via cut elimination arguments Grišin (1982), Petersen (2000), Cantini (2003).

Two problems of the contraction-free approach:

- Viewed as a set theory, GS is inconsistent with extensionality, e.g defined as:

  $$s \subseteq t \star t \subseteq s, t \in r \Rightarrow s \in r$$

  This is often called **Grišin's paradox**.

- Viewed as a property theory or a truth theory, there is no known, **plausible semantics**.

However, it needs to be added that it also features a 'decent' conditional (compared, e.g. to Field (2008)).

Fixed-Point Semantics

Given our language $\mathcal{L}_{\mathtt{Tr}} := \mathcal{L}_{\cup}\{\mathtt{Tr}\}$, we start with a (classical) model $\mathcal{M}$ of $\mathcal{L}$ such that $\ulcorner\varphi\urcorner^{\mathcal{M}} = \varphi$, and set, for $X \subset |\mathcal{M}|$:

$$
\begin{aligned}
a \in \Phi(X) \Leftrightarrow\ & a = \ulcorner\top\urcorner, \text{ or}\\
& a = \ulcorner\mathtt{Tr}\ulcorner\varphi\urcorner\urcorner \text{ and } \ulcorner\varphi\urcorner \in X, \text{ or}\\
& a = \ulcorner\neg\mathtt{Tr}\ulcorner\varphi\urcorner\urcorner \text{ and } \ulcorner\neg\varphi\urcorner \in X, \text{ or}\\
& a = \ulcorner\varphi \wedge \psi\urcorner \text{ and } \ulcorner\varphi\urcorner \in X \text{ and } \ulcorner\psi\urcorner \in X, \text{ or}\\
& a = \ulcorner\neg(\varphi \wedge \psi)\urcorner \text{ and } \ulcorner\neg\varphi\urcorner \in X \text{ or } \ulcorner\neg\psi\urcorner \in X, \text{ or} \ldots
\end{aligned}
$$

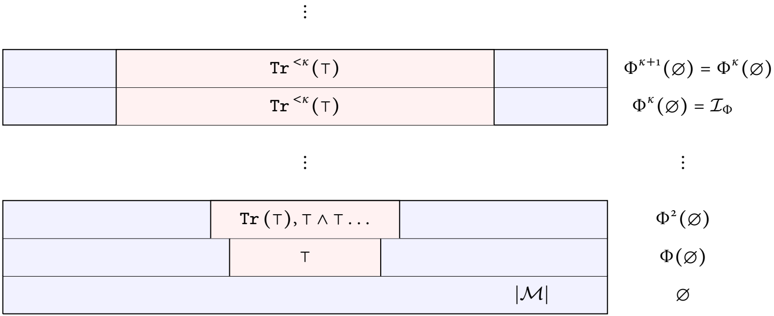
Let then $\Phi^0(X) = X$, $\Phi^{\alpha+1}(X) = \Phi(\Phi^{\alpha}(X))$, $\Phi^{\lambda}(X) = \bigcup_{\beta<\lambda}\Phi^{\beta}(X)$.

Lemma (Kripke (1975),Martin and Woodruff (1975))

If $S \subseteq |\mathcal{M}|$ is a fixed-point of Φ, then for all $\varphi \in \mathcal{L}_{Tr}$:

$$\varphi \in S \text{ iff } Tr\ulcorner\varphi\urcorner \in S$$

Fixed-Point Semantics



The structure $(\mathcal{M}, \mathcal{I}_\Phi)$ gives rise to a three-valued model for $\mathcal{L}_{\mathtt{Tr}}$ with $\mathtt{Tr}$ a 'partial' predicate. Define

$$\mathcal{M} \vDash \Gamma \Rightarrow \Delta :\Leftrightarrow (\forall \gamma \in \Gamma)\ |\gamma|^{\mathcal{M}}_{\mathcal{I}_\Phi} \neq 0 \rightarrow (\exists \delta \in \Delta)\ |\gamma|^{\mathcal{M}}_{\mathcal{I}_\Phi} = 1$$

Restricting initial sequents

Already known in other contexts Kreuger (1994); Jäger and Stärk (1998); Schroeder-Heister (2016). This is contained in Nicolai (2018a). Structural rules are absorbed.

Definition (LPT)

$$\Gamma, \bot \Rightarrow \Delta \qquad\qquad \Gamma \Rightarrow \top, \Delta$$

$$\frac{\Gamma \Rightarrow \Delta, A}{\Gamma \Rightarrow \Delta, \mathtt{Tr}\ulcorner A \urcorner} \qquad\qquad \frac{A, \Gamma \Rightarrow \Delta}{\mathtt{Tr}\ulcorner A \urcorner, \Gamma \Rightarrow \Delta}$$

$$(\neg\text{L})\ \frac{\Gamma \Rightarrow \Delta, \varphi}{\neg\varphi, \Gamma \Rightarrow \Delta} \qquad\qquad (\neg\text{R})\ \frac{\varphi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\varphi}$$

$$\vdots \qquad\qquad \vdots$$

- Now $(\mathcal{M}, S) \vDash$ LPT for S a fixed point of Φ.
- The model $(\mathcal{M}, \mathcal{I}_\Phi)$ satisfies a fully operational, paracomplete version system of naïve truth based on Strong-Kleene logic (modulo definition of consequence).

Back to cut elimination

When contraction is around, the notion of Tr -rank is not enough:

$$\dfrac{\dfrac{\begin{matrix}\mathcal{D}_{00}\\ \Gamma \Rightarrow \Delta, \mathtt{Tr}\ulcorner\psi\urcorner, \mathtt{Tr}\ulcorner\psi\urcorner \ [\alpha]\end{matrix}}{\Gamma \Rightarrow \Delta, \mathtt{Tr}\ulcorner\psi\urcorner \ [\alpha]} \qquad \begin{matrix}\mathcal{D}_{1}\\ \mathtt{Tr}\ulcorner\psi\urcorner, \Theta \Rightarrow \Lambda \ [\beta]\end{matrix}}{\Gamma, \Theta \Rightarrow \Delta, \Lambda \ [\alpha + \beta]}$$

Now the idea here would be that we transform the derivation in

$$\dfrac{\dfrac{\dfrac{\begin{matrix}\mathcal{D}^*_{00}\\ \Gamma \Rightarrow \Delta, \psi, \psi \ [\alpha]\end{matrix} \qquad \begin{matrix}\mathcal{D}^*_{1}\\ \psi, \Theta \Rightarrow \Lambda \ [\beta]\end{matrix}}{\Gamma, \Theta \Rightarrow \Delta, \Lambda, \psi \ [\alpha + \beta]} \qquad \begin{matrix}\mathcal{D}^*_{1}\\ \psi, \Theta \Rightarrow \Lambda \ [\beta]\end{matrix}}{\Gamma, \Theta, \Theta \Rightarrow \Delta, \Lambda, \Lambda \ [2\alpha + \beta]}}{\Gamma, \Theta \Rightarrow \Delta, \Lambda \ [2\alpha + \beta]}$$

$\mathtt{Tr}$ -complexity $\kappa(\cdot)$ of formulas

The ordinal $\mathtt{Tr}$ -complexity $\kappa_{\mathcal{D}}(\cdot)$ of a formula φ of $\mathcal{L}_{\mathtt{Tr}}$ in a derivation $\mathcal{D}$ is defined inductively as follows:

- formulas of $\mathcal{L}$ **have $\mathtt{Tr}$ -complexity** 0 in any $\mathcal{D}$;
- If $\mathcal{D}$ is just $\Gamma, \varphi \Rightarrow \varphi, \Delta$ with $\varphi \in \mathcal{L}$, then $\kappa_{\mathcal{D}}(\psi) = \kappa_{\mathcal{D}}(\varphi) = 0$ for all $\psi \in \Gamma, \Delta$. Similarly for $(\top)$, $(\bot)$.
- If $\mathcal{D}$ ends with

$$\frac{\Gamma \Rightarrow \Delta, A}{\Gamma \Rightarrow \Delta, \mathtt{Tr}\ulcorner A \urcorner}$$

  then the complexity of formulas in Γ, Δ is unchanged and $\kappa_{\mathcal{D}}(\mathtt{Tr}\ulcorner A \urcorner) = \kappa_{\mathcal{D}}(A) + 1$ (similarly for ($\mathtt{TR}$ -L)).
- If $\mathcal{D}$ ends with

$$\frac{\gamma^1_1, \dots, \gamma^1_n \Rightarrow \delta^1_1, \dots, \delta^1_m, \varphi \qquad \gamma^2_1, \dots, \gamma^2_n \Rightarrow \delta^2_1, \dots, \delta^2_m, \psi}{\gamma^3_1, \dots, \gamma^3_n \Rightarrow \delta^3_1, \dots, \delta^3_m, \varphi \wedge \psi}$$

  then

$$\kappa_{\mathcal{D}}(\varphi \wedge \psi) = \mathtt{max}(\kappa_{\mathcal{D}}(\varphi), \kappa_{\mathcal{D}}(\psi))$$
$$\kappa_{\mathcal{D}}(\gamma^3_i) = \mathtt{max}(\kappa_{\mathcal{D}}(\gamma^1_i), \kappa_{\mathcal{D}}(\gamma^2_i)) \quad 1 \leq i \leq n$$
$$\kappa_{\mathcal{D}}(\delta^3_j) = \mathtt{max}(\kappa_{\mathcal{D}}(\delta^1_j), \kappa_{\mathcal{D}}(\delta^2_j)) \quad 1 \leq j \leq m$$

Full cut elimination

Crucially, rules of LPT are **κ-invertible**, e.g.:
If $\mathcal{D} \vdash_{\mathrm{LPT}} \Gamma^1, \mathtt{Tr}\ulcorner A \urcorner \Rightarrow \Delta^1$, then there is $\mathcal{D}' \vdash_{\mathrm{LPT}} A, \Gamma \Rightarrow \Delta$ with $\kappa_{\mathcal{D}'}(\Gamma) \leq \kappa_{\mathcal{D}}(\Gamma^1)$, $\kappa_{\mathcal{D}'}(\Delta) \leq \kappa_{\mathcal{D}}(\Delta^1)$, and

$$\kappa_{\mathcal{D}'}(A) \leq \kappa_{\mathcal{D}}(\mathtt{Tr}\ulcorner A \urcorner), \text{ if } \kappa_{\mathcal{D}}(\mathtt{Tr}\ulcorner A \urcorner) = 0;$$
$$\kappa_{\mathcal{D}'}(A) < \kappa_{\mathcal{D}}(\mathtt{Tr}\ulcorner A \urcorner), \text{ if } \kappa_{\mathcal{D}}(\mathtt{Tr}\ulcorner A \urcorner) \neq 0.$$

Lemma

Contraction is κ-admissible and length-admissible, e.g. : If $\mathcal{D} \vdash^n_{\mathrm{LPT}} \Gamma^1, \varphi^1, \varphi^2 \Rightarrow \Delta^1$, then there is a $\mathcal{D}' \vdash^n_{\mathrm{LPT}} \Gamma, \varphi \Rightarrow \Delta$ with

$$\kappa_{\mathcal{D}'}(\Gamma^1) \leq \kappa_{\mathcal{D}'}(\Gamma); \quad \kappa_{\mathcal{D}'}(\Delta^1) \leq \kappa_{\mathcal{D}'}(\Delta)$$
$$\kappa_{\mathcal{D}'}(\varphi) \leq max(\kappa_{\mathcal{D}'}(\varphi^1), \kappa_{\mathcal{D}'}(\varphi^2)).$$

Proposition

If $\mathcal{D}_0$ is a cut-free proof of $\Gamma^1 \Rightarrow \Delta^1, \varphi^1$ in LPT, *and $\mathcal{D}_1$ is a cut-free* LPT*-proof of $\varphi^2, \Gamma^2 \Rightarrow \Delta^2$, then there is a cut-free proof $\mathcal{D}$ of $\Gamma^3 \Rightarrow \Delta^3$ with $\kappa_{\mathcal{D}}(\Gamma^3) \leq max(\kappa_{\mathcal{D}_0}(\Gamma^1), \kappa_{\mathcal{D}_1}(\Gamma^2))$ and $\kappa_{\mathcal{D}}(\Delta^3) \leq max(\kappa_{\mathcal{D}_0}(\Delta^1), \kappa_{\mathcal{D}_1}(\Delta^2))$.*

- The ideal of **semantic closure** is at odds with resources that outstrip the ones available in one's semantic theory. Cut-elimination procedures are usually formalizable in weak arithmetical systems.
- When nonlogical initial sequents are around, full cut elimination is not in general available: however by eliminating cuts on formulas of the form $\mathtt{Tr}\ulcorner A \urcorner$, one can obtain **conservativity proofs**.
- Another advantage of the approach with restricted initial sequents is that – unlike the contraction-free approaches – there are natural **infinitary systems** that arise from the logic and that give succinct presentations of Π^1_1-sets.

Truth Bearers

- There are good reasons to require an ontology of bearers of truth prior to discussing principles of truth. We want to prove **in the object language** things like:

$$\forall \varphi, \psi \, \exists \chi \, (\chi = (\varphi \dot{\wedge} \psi) \wedge \varphi \neq \chi)$$

  This is usually guaranteed by assuming a theory of finite objects (as we shall see in a moment).
- Notice that this is imposing non-trivial constraints. More 'philosophical' theories of truth are often formulated in terms of **propositions**, and not sentence types (Horwich, 1998; Soames, 1998; Jago, 2018). This rules out that propositions are coarse-grained, e.g. sets of possible worlds.

Arithmetic

- Peano arithmetic (PA) is the preferred base theory for systems of truth. It is usually formulated in $\mathcal{L}_{\mathbb{N}} = \{0, \mathtt{S}, +, \times\}$ and features equations for its primitives, e.g.

$$(x + 0) = x \qquad\qquad x + \mathtt{S}y = \mathtt{S}(x + y)$$

  and the first-order induction schema

$$\varphi(0) \wedge \forall x(\varphi(x) \to \varphi(\mathtt{S}x)) \to \forall x\, \varphi(x) \qquad \text{for } \varphi \in \mathcal{L}_{\mathbb{N}}$$

- Alternatively, one can employ a theory of strings and concatenation $\frown$ with two atoms a, b based on **Tarski's axiom**

$$a^\frown y = u^\frown v \;\leftrightarrow\; \exists w\big((x = u^\frown w \wedge v = w^\frown y) \vee (u = x^\frown w \wedge y = w^\frown v)\big)$$

  With first-order string induction, the two theories are mutually interpretable. An accessible source is Ganea (2009).
- Finite set theories are also a convenient choice. For instance
  - Kaye and Wong (2006) show that PA and KF $\setminus$ $\{\mathrm{Inf}\}$+ 'every set has a transitive closure' are **bi-interpretable**;
  - similarly, a neat set theory based by Świerczkowski (2003) based on the **adjunction operation** $x \triangleleft y \mapsto x \cup \{y\}$ is bi-interpretable with PA.

Doing with less

- Ultimately, what we require to establish the basic properties of the truth bearers are **a good notion of sequence**, and **a minimum of induction** to handle suitable forms of recursion.
- For the former the notion of a **sequential** theory is enough – see Visser (2010) for a comprehensive overview. A theory is sequential if it interprets – with no relativization of quantifiers – the theory AS given by the empyt set axiom and adjunction – which is as strong as Robinson's Q.
- As to induction, since all the relevant syntactic notions (terms, formulas, proofs) are p-time decidable, the theory S^1_2 by Buss (1986) suffices. However, many of the results that I will treat below are specific to PA (or equivalents), and it is object of current research to check which results are stable over $T \supseteq S^1_2$.

Ordinals

- A good base theory will also provide a satisfactory representation of ordinals. For our purposes it suffices to require a notation up to the Feferman-Schütte ordinal Γ_0:
  - α is principal if it cannot be expressed as $\zeta + \eta$ for $\zeta, \eta < \alpha$. Define:

$$C(0) := \text{`the class of principal ordinals'}$$
$$C(\alpha + 1) := \text{`the class of fixed points of the function enumerating } C(\alpha)\text{'}$$
$$C(\lambda) := \bigcap_{\zeta<\lambda} C(\zeta) \quad \text{for } \lambda \text{ a limit ordinal}$$

  - The *Veblen functions* φ_α are the enumerating functions of $C(\alpha)$. The class of *strongly critical* ordinals SC contains precisely the ordinals α that are themselves α-critical. Γ_ζ indicates the ζ-th strongly critical ordinal.
  - Principal ordinals α that are not themselves strongly critical are such that $\alpha = \varphi_\zeta\eta$ for $\eta, \zeta < \alpha$. Therefore, by this fact and Cantor's normal form theorem, ordinals $< \Gamma_0$ can be uniquely determined as words of the alphabet $(0, +, \varphi\cdot\cdot)$.

Ordinals

- A good base theory will also provide a satisfactory representation of ordinals. For our purposes it suffices to require a notation up to the Feferman-Schütte ordinal Γ_0
- A notation system for Γ_0 is of the form $(\mathtt{OT}, \mathtt{PT}, |\cdot|, \prec)$, with
  - $\mathtt{OT}$ the set of natural number 'codes' for ordinals $< \Gamma_0$
  - $\mathtt{OT} \subseteq \mathtt{OT}$ the set of codes of principal ordinals
  - $|\cdot|\colon \mathtt{OT} \to \mathtt{ON}$
  - $n \prec m :\leftrightarrow n \in \mathtt{OT} \wedge m \in \mathtt{OT} \wedge |n| < |m|$
- Using standard coding techniques one can show that $\mathtt{OT}, \mathtt{PT}, \prec$ are primitive recursive. Actually, Beckmann et al. (2003) show that they can be showed to be p-time and represented in S^1_2 – notice that I **do not** mean that Γ_0 can be well-founded in S^1_2!

Schemata

For ordinals $\alpha < \Gamma_0$, we denote with a the corresponding numeral in the representation of OT and we do not distinguish between ordinal functions such as the Veblen functions and their representations. The system $(\mathtt{OT}, \mathtt{PT}, \prec)$ enables us to formulate the following principles of transfinite induction:

$$(\mathrm{TI}^{\varepsilon_0}_{\mathcal{L}_{\mathtt{Tr}}}) \qquad \frac{\forall a \prec b\, \phi(a), \Gamma \Rightarrow \Delta, \phi(b)}{\Gamma \Rightarrow \Delta, \forall a \prec \varepsilon_0\, \phi(a)}$$

$$(\mathrm{TI}^{<\omega^\omega}_{\mathcal{L}_{\mathtt{Tr}}}) \qquad \frac{\forall a \prec b\, \phi(a), \Gamma \Rightarrow \Delta, \phi(b)}{\Gamma \Rightarrow \Delta, \forall a \prec c\, \phi(a)} \qquad \text{for all } \gamma (= |c|) < \omega^\omega$$

$$(\mathrm{TI}^{<\varepsilon_0}_{\mathcal{L}_{\mathtt{Tr}}}) \qquad \frac{\forall a \prec b\, \phi(a), \Gamma \Rightarrow \Delta, \phi(b)}{\Gamma \Rightarrow \Delta, \forall a \prec c\, \phi(a)} \qquad \text{for all } \gamma < \varepsilon_0$$

Truth as primitive

- Truth-theoretic deflationism holds that truth is **not a genuine property** and that its function is mainly that of a **generalizing device** (Quine, 1970; Field, 1994; Horwich, 1998).
- Unlike other notions that have been taken to be primitive for lack of consensus over a definition – e.g. knowledge, see Williamson (2000) – Tarski's theorem uncontroversially establishes this (Halbach, 2014, Ch. 1).
- Truth is a fundamental **semantic** concept. A theory of meaning for natural language expressions is not much more than a (Tarskian) theory of truth for it (Davidson, 1984).

Tarskian Truth

The theory of truth in $\mathcal{L}_{\mathtt{Tr}}$ that Davidson had in mind extends PA with the following:

Definition (CT)

$$\forall s, t(\mathtt{Tr}\,(s = t) \leftrightarrow s^{\circ} = t^{\circ})$$

$$\forall \varphi \in \mathcal{L}(\mathtt{Tr}\,(\neg\varphi) \leftrightarrow \neg\mathtt{Tr}\,\varphi)$$

$$\forall \varphi, \psi \in \mathcal{L}(\mathtt{Tr}\,(\varphi \wedge \psi) \leftrightarrow \mathtt{Tr}\,\varphi \wedge \mathtt{Tr}\,\psi)$$

$$\forall v, \forall \varphi(v) \in \mathcal{L}(\mathtt{Tr}\,(\forall v \varphi) \leftrightarrow \forall x\,\mathtt{Tr}\,\varphi(\dot{x}))$$

$$\varphi(0) \wedge \forall x(\varphi(x) \to \varphi(x+1)) \to \forall x\,\varphi(x) \qquad \text{with } \varphi(v) \in \mathcal{L}_{\mathtt{Tr}}$$

Important variations are obtained by tweaking the induction schema:

- CT $\restriction$ (a.k.a. CT^{-}) is obtained by restricting induction to $\mathcal{L}$
- CT$_{\text{int}}$ is obtained by adopting the internal induction schema

$$\forall \varphi(v)(\mathtt{Tr}\,\varphi(o/v) \wedge \forall y(\mathtt{Tr}\,\varphi(\dot{y}/v) \to \mathtt{Tr}\,\varphi(\dot{\mathrm{S}y}/v)) \to \forall x \mathtt{Tr}\,\varphi(\dot{x}/v))$$

Conservativeness

Thesis (Shapiro, 1998; Ketland, 1999)

CT proves Con(PA), therefore deflationism is untenable.

This contrasts with:

Proposition (Kotlarski et al. (1981); Visser and Enayat (2015); Leigh (2015))

$\mathrm{CT}_{\mathrm{int}}$ *is a conservative extension of* PA.

The discussion took a strong technical turn, brilliantly summarized in Cieliski (2017) – with many original contributions. It's worth mentioning:

Proposition (Enayat and Pakhomov (2018))

CT $\restriction$ *plus 'disjunctive correctness', i.e.*

$$\forall s\big(Tr\big(\bigvee_{i<s}\varphi_i\big) \leftrightarrow \exists i < s\; Tr\,\varphi_i\big)$$

is the same theory as $\mathrm{CT}[\mathrm{I}\Delta_0]$, *and therefore proves* Con(PA).

Conservativeness

- Despite the technical interest, the debate seems to be built on shaky foundations. Virtually **no deflationist** has thoroughly defended the claim that truth **has to be** conservative over the base theory.
- By contrast, it has repeatedly been argued that **truth has to be nonconservative**, but in a way that is **distinctively metalinguistic**, i.e. it does not interfere with the subject matter of the base theory over which truth is built.
- This has led to the programme of **'disentangling'** syntactic quantifiers from quantifiers over natural numbers:

Proposition (Nicolai (2015, 2016))

If one formulates $\mathrm{CT} \upharpoonright$ *as a two-sorted theory, with 'syntactic' quantifiers and 'number-theoretic' quantifiers, and truth applying only over syntactic objects, then:*

- *The theory of truth becomes* ***trivially*** *conservative over* PA*;*
- *This version of* $\mathrm{CT} \upharpoonright$ *plus 'all axioms of* PA *are true' is* ***mutually interpretable*** *with* $\mathrm{PA} + \mathrm{Con}(\mathrm{PA})$.

Feferman's project

To give a nice presentation of the **reflective closure** of PA – and possibly of further 'natural' theories: i.e. the (truth-)theory that makes explicit all that is implicit in the acceptance of PA.

First attempt: $\mathrm{CT}_{<a}$, $\alpha \leq \Gamma_0$

$$\mathcal{L}_0 := \mathcal{L}_{\mathtt{Tr}} \qquad \mathcal{L}_{<c} := \mathcal{L} \cup \{\mathtt{Tr}_b \mid b < c\}$$

With $b < c$:

$$\forall s, t(\mathtt{Tr}_b(s = t) \leftrightarrow s^\circ = t^\circ)$$
$$\forall \varphi \in \mathcal{L}_{<b}\, (\mathtt{Tr}_b(\neg\varphi) \leftrightarrow \mathtt{Tr}_b(\neg\varphi))$$
$$\forall \varphi, \psi \in \mathcal{L}_{<b}\, (\mathtt{Tr}_b(\varphi \wedge \psi) \leftrightarrow \mathtt{Tr}_b\varphi \wedge \mathtt{Tr}_b\psi)$$
$$\forall v, \forall \varphi(v) \in \mathcal{L}(\mathtt{Tr}_b(\forall v\varphi) \leftrightarrow \forall x\, \mathtt{Tr}_b\varphi(\dot{x}))$$
$$\forall \varphi \in \mathcal{L}_{<a<b}(\mathtt{Tr}_b\mathtt{Tr}_a\varphi \leftrightarrow \mathtt{Tr}_a\varphi)$$
$$\forall d \prec b, \forall \varphi \in \mathcal{L}_{<d}(\mathtt{Tr}_b\mathtt{Tr}_d\varphi \leftrightarrow \mathtt{Tr}_b\varphi)$$

Feferman's project

- The project of isolating $\mathrm{CT}_{<\varepsilon_0}$ or $\mathrm{CT}_{<\Gamma_0}$ as **natural stopping points**, that was congenial to Feferman's project, depended essentially on other results, such as Feferman's and Schütte's independent characterization of Γ_0, or the provable well-orderings of PA.
- The next step was to find an independent characterization of such theories:

Definition (KF)

$$\forall s, t(\mathtt{Tr}\,(s = t) \leftrightarrow s^\circ = t^\circ)$$
$$\forall s, t(\mathtt{Tr}\,(s \neq t) \leftrightarrow s^\circ \neq t^\circ)$$
$$\forall t(\mathtt{Tr}\,\mathtt{Tr}\,t \leftrightarrow \mathtt{Tr}\,t^\circ)$$
$$\forall t(\mathtt{Tr}\,\mathtt{Tr}\,\neg t \leftrightarrow \mathtt{Tr}\,\neg t^\circ)$$
$$\forall \varphi \in \mathcal{L}_{\mathtt{Tr}}\,(\mathtt{Tr}\,(\neg\neg\varphi) \leftrightarrow \mathtt{Tr}\,\varphi)$$
$$\forall \varphi, \psi \in \mathcal{L}_{\mathtt{Tr}}\,(\mathtt{Tr}\,(\varphi \wedge \psi) \leftrightarrow \mathtt{Tr}\,\varphi \wedge \mathtt{Tr}\,\psi)$$
$$\forall \varphi, \psi \in \mathcal{L}_{\mathtt{Tr}}\,(\mathtt{Tr}\,\neg(\varphi \wedge \psi) \leftrightarrow \mathtt{Tr}\,\neg\varphi \vee \mathtt{Tr}\,\neg\psi)$$
$$\forall v, \forall \varphi(v) \in \mathcal{L}_{\mathtt{Tr}}\,(\mathtt{Tr}\,(\forall v\varphi) \leftrightarrow \forall x\,\mathtt{Tr}\,\varphi(\dot{x}))$$
$$\forall v, \forall \varphi(v) \in \mathcal{L}_{\mathtt{Tr}}\,(\mathtt{Tr}\,(\neg\forall v\varphi) \leftrightarrow \exists x\,\mathtt{Tr}\,\neg\varphi(\dot{x}))$$

Properties of KF

- Semantically KF fits nicely with Kripke's fixed-point semantics (Kripke, 1975),

$$(\mathbb{N}, S) \models \text{KF} \;\text{ iff }\; S \text{ is a fixed point of Kripke's theory of truth}$$

- The full $\mathtt{Tr}$ -schema is available for meaningful predicates satisfying $\mathtt{D}(x) :\leftrightarrow \mathtt{Tr}\, x \vee \mathtt{Tr}\, \neg x$, i.e. for all $A \in \mathcal{L}_{\mathtt{Tr}}$:

$$\text{KF} \vdash \mathtt{D}(\ulcorner A \urcorner) \rightarrow (\mathtt{Tr}\, \ulcorner A \urcorner \leftrightarrow A)$$

Proposition (Feferman (1991); Cantini (1989))

KF *is proof-theoretically equivalent to* $(\Pi^0_1\text{–CA})_{<\varepsilon_0}$.

Proof Idea.

Lower bound: PA in $\mathcal{L}_{\mathtt{Tr}}$ proves $\text{TI}^{<\varepsilon_0}_{\mathcal{L}_{\mathtt{Tr}}}$. Now KF proves:

$$\varphi \in \mathcal{L}_{<a} \rightarrow \mathtt{D}(\varphi) \Rightarrow \varphi \in \mathcal{L}_a \rightarrow \mathtt{D}(\varphi)$$

An application of $\text{TI}^{<\varepsilon_0}_{\mathcal{L}_{\mathtt{Tr}}}$ yields an embedding of $\text{CT}_{<\varepsilon_0}$, which suffices. □

Properties of KF

- Semantically KF fits nicely with Kripke's fixed-point semantics (Kripke, 1975),

$$(\mathbb{N}, S) \models \text{KF} \text{ iff } S \text{ is a fixed point of Kripke's theory of truth}$$

- The full $\mathtt{Tr}$ -schema is available for meaningful predicates satisfying $\mathtt{D}(x) :\leftrightarrow \mathtt{Tr}\, x \vee \mathtt{Tr}\, \neg x$, i.e. for all $A \in \mathcal{L}_{\mathtt{Tr}}$:

$$\text{KF} \vdash \mathtt{D}(\ulcorner A \urcorner) \rightarrow (\mathtt{Tr} \ulcorner A \urcorner \leftrightarrow A)$$

Proposition (Feferman (1991); Cantini (1989))

KF *is proof-theoretically equivalent to* $(\Pi^0_1\text{–CA})_{<\varepsilon_0}$.

Proof Idea.

Upper bound: One formulates KF in a Tait (one-sided) infinitary calculus, and analyzes quai-normal derivations, i.e. derivations with only cuts on $\mathtt{Tr}\, t$ and $\neg\mathtt{Tr}\, t$ and proves in $\text{CT}_{<\varepsilon_0}$ that

$$\text{if KF}^\infty \vdash^\alpha \mathtt{Tr} \ulcorner A \urcorner, \text{ then } \mathtt{Tr}_{2^\alpha} \ulcorner A \urcorner$$

□

Symmetries

- One important drawback of KF is that its internal theory $\{\varphi \in \mathcal{L}_{\mathtt{Tr}} \mid \mathrm{KF} \vdash \mathtt{Tr}\ulcorner\varphi\urcorner\}$ is different from its theorems: for instance $\mathrm{KF} \vdash \lambda \vee \neg\lambda$ but $\mathrm{KF} \nvdash \mathtt{Tr}\ulcorner\lambda \vee \neg\lambda\urcorner$.
- To overcome this:

Reinhardt's thesis

One should adopt an **instrumental** reading of KF. Its conceptual core is given by its **internal theory**.

Lemma (Halbach and Horsten (2006))

There are A's in $\mathcal{L}_{Tr}$ such that $\mathrm{KF} \vdash Tr\ulcorner A\urcorner$ *but the proof essentially employs B's such that* $\mathrm{KF} \nvdash Tr\ulcorner B\urcorner$.

Kripke-Feferman in four-valued logic

- From classical logic, **remove** both negation introduction rules are replace them with, e.g.:

$$\frac{\Gamma \Rightarrow \Delta, \neg\varphi, \neg\psi}{\Gamma \Rightarrow \Delta, \neg(\varphi \wedge \psi)} \qquad \frac{\neg\varphi, \Gamma \Rightarrow \Delta \qquad \neg\psi, \Gamma \Rightarrow \Delta}{\neg(\varphi \wedge \psi), \Gamma \Rightarrow \Delta}$$

One obtains a logic in the vicinity of FDE.

- As to truth:

PKF

$$s^\circ = t^\circ \Leftrightarrow \mathtt{Tr}\,(s = t)$$
$$\neg\mathtt{Tr}\,\varphi \Leftrightarrow \mathtt{Tr}\,\neg\varphi$$
$$\mathtt{Tr}\,(\varphi \wedge \psi) \Leftrightarrow \mathtt{Tr}\,\varphi \wedge \mathtt{Tr}\,\psi$$
$$\mathtt{Tr}\,\mathtt{Tr}\,\varphi \Leftrightarrow \mathtt{Tr}\,\varphi$$

- The theories are closely related:

***ω*-categoricity**

For S a fixed point of Kripke's theory of truth,

$$(\mathbb{N}, S) \vDash \mathrm{KF} \ \text{ iff } \ (\mathbb{N}, S) \vDash_{\mathrm{FDE}} \mathrm{PKF}$$

Kripke-Feferman in four-valued logic

Unlike KF, PKF $\vdash \mathtt{Tr}\ulcorner A \urcorner$ iff PKF $\vdash A$, for all $A \in \mathcal{L}_{\mathtt{Tr}}$.

Proposition (Halbach and Horsten (2006))

PKF *proves the same arithmetical sentences as* $\mathrm{CT}_{<\omega^\omega}$.

Proposition (Feferman (1991))

KF *proves the same arithmetical sentences of* $\mathrm{CT}_{<\varepsilon_0}$.

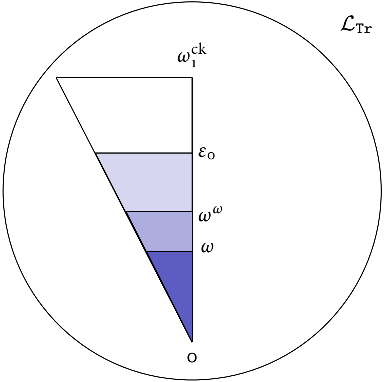
Recalling Reinhardt's thesis:

Corollary

The internal theory of KF *and* PKF *differ considerably.*

Proposition (Nicolai (2018b))

- $\mathrm{PKF} = \{A \in \mathcal{L}_{Tr} \mid \mathrm{KF}_{\mathrm{int}} \vdash Tr\ulcorner A \urcorner\}$
- $\{A \in \mathcal{L}_{Tr} \mid \mathrm{KF} \vdash Tr\ulcorner A \urcorner\} = \mathrm{PKF} + \mathrm{TI}^{<\varepsilon_0}_{\mathcal{L}_{Tr}}$



	$\mathcal{L}$	$\mathcal{L}_{\mathtt{Tr}}$
PKF, $\mathrm{KF}_{\mathrm{int}}$	$< \varphi_\omega 0$	$\varphi_0 \omega$ (or ω^ω)
KF	$< \varphi_{\varepsilon_0} 0$	$\varphi_1 0$ (or ε_0)

The costs of nonclassical logic

- This asymmetry between the amount of transfinite induction provable in PKF and KF has been taken – see Halbach (2014) and Halbach and Nicolai (2018) – as substantiating Feferman's claim that

  'nothing like sustained ordinary reasoning can be carried out [in such logic]' (Feferman, 1984)

- This seems to be supported by:

Halbach and Nicolai (2018)

$\mathrm{PKF} \upharpoonright \mathcal{L} = \{\varphi \in \mathcal{L}_{\mathtt{Tr}} \mid \mathrm{KF} \upharpoonright \mathcal{L} \vdash \mathtt{Tr} \ulcorner \varphi \urcorner\}.$

- A possible rejoinder is that transfinite induction open to **semantic notions** cannot be taken as compromising **mathematical** reasoning.

INDEC is the assertion that:

> *every countable scattered indecomposable linear ordering is either indecomposable to the left, or indecomposable to the right.*

It was proved to be true by Pierre Jullien in 1969.

Lemma (Montalbán, Friedman)

INDEC is implied by Σ^1_1-CA, which is Π^1_2-conservative over $\mathrm{ACA}_{<\varepsilon_0}$.

Proposition (Eastaugh, N.)

RCA+INDEC is proof-theoretically equivalent to KF. It follows that KF can 'nicely' interpret RCA+INDEC, but KF cannot.

- We have seen that $\mathrm{CT}_{<\varepsilon_0}$, $\mathrm{CT}_{<\Gamma_0}$, KF, and in part PKF can be seen as attempt to characterize the reflective closure of PA.
- There is a more direct strategy involving **reflection principles**, i.e. claims of the form

RFN(S)

if something is provable in a theory S, then it's true, or
$\forall x(\mathrm{Prov}_S(\ulcorner A(\dot{x})\urcorner) \to A(x))$

- It turns out that if one focuses on **classical** biconditionals

$$\mathrm{PTB} \upharpoonright \mathcal{L} = \{\mathtt{Tr}\ulcorner A\urcorner \leftrightarrow A \mid A \text{ positive of } \mathcal{L}_{\mathtt{Tr}}\}$$

one obtains:

Proposition (Horsten and Leigh (2017))

$\mathrm{RFN}^2(\mathrm{PTB} \upharpoonright \mathcal{L}) \supseteq \mathrm{KF}$.

- There seems to be a conceptual problem with this strategy:

If one formulates reflection – as it seems plausible – as:

$$\forall\varphi(\mathrm{Prov}_{\mathrm{PTB}}(\varphi) \to \mathtt{Tr}\,\varphi)$$

then there is a sentence A such that PTB with this form of reflection proves $\mathtt{Tr}\ulcorner A \land \neg A \urcorner$.

- A possible way out is to start with a set of 'biconditionals' over FDE:

$$\mathrm{TS} \restriction \mathcal{L} = \{\mathtt{Tr}\ulcorner A \urcorner \Leftrightarrow A \mid A \in \mathcal{L}_{\mathtt{Tr}}\}$$

and define reflection as

$$(\mathrm{RR})\ \frac{\mathrm{Pr}^2_S(\ulcorner \Gamma\dot{x} \Rightarrow \Delta\dot{x}\urcorner, \ulcorner \Theta\dot{x} \Rightarrow \Lambda\dot{x}\urcorner) \qquad \Gamma(x) \Rightarrow \Delta(x)}{\Theta(x) \Rightarrow \Lambda(x)}$$

- Then we can prove:

Proposition (Fischer et al. (2017))

- $\mathrm{RR}^2(\mathrm{TS} \restriction \mathcal{L}) \supseteq \mathrm{PKF}$.
- $\mathrm{RR}^\omega(\mathrm{TS} \restriction \mathcal{L}) \vdash \mathrm{TI}^{<\omega^{\omega^2}}_{\mathcal{L}_{Tr}}$

An analogy

- Solovay's theorem tells us that, for all $A \in \mathcal{L}_\Box$,

$$\mathrm{GL} \vdash A \ \text{ iff } \ \forall \star\, (\mathrm{PA} \vdash A^\star)$$

- By redefining a realization $\star\colon \mathcal{L}_\Box \to \mathcal{L}_{\mathtt{Tr}}$ as:

$$A^\star = \begin{cases} 0 = 1, & \text{if } A = \bot, \\ \text{commutes with connectives} & \\ \mathtt{Tr}\ulcorner B^\star \urcorner & \text{if } A = \Box B \end{cases}$$

one can ask:

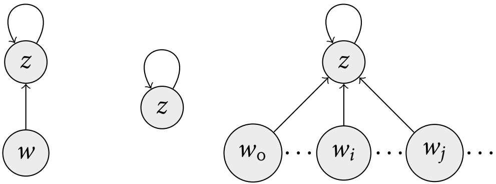
Question

$$? \vdash A \quad \Leftrightarrow \quad \forall \star\, (\mathrm{KF} \vdash A^\star)$$

Nicolai and Stern (2018)

The logic $L_\Box$

$\Box \top$	$\neg \Box \bot$
$\Box A \leftrightarrow \Box\neg\neg A$	$\Box A \wedge \neg \Box \neg A \rightarrow A$
$\Box (A \wedge B) \leftrightarrow \Box A \wedge \Box B$	$\Box \neg(A \wedge B) \leftrightarrow \Box\neg A \vee \Box\neg B$
$\Box A \leftrightarrow \Box \Box A$	$\Box \neg A \leftrightarrow \Box\neg \Box A$



Proposition

$$L_\Box \vdash A \ \textit{iff} \ \forall \star\, (\mathrm{KF} \vdash A^\star)$$

Beckmann, A., Pollett, C., and Buss, S. R. (2003). Ordinal notations and well-orderings in bounded arithmetic. *Annals of Pure and Applied Logic*, 120(1-3):197–223.

Buss, S. (1986). *Bounded arithmetic*. Bibliopolis, Napoli.

Cantini, A. (1989). Notes on formal theories of truth. *Zeitschrift für Logik un Grundlagen der Mathematik*, 35:97–130.

Cantini, A. (2003). The undecidability of Grisin's set theory. *Studia Logica*, 74(3):345–368.

Cieliski, C. (2017). *The Epistemic Lightness of Truth: Deflationism and its Logic*. Cambridge University Press.

Davidson, D. (1984). *Inquiries Into Truth And Interpretation*. Oxford University Press.

Enayat, A. and Pakhomov, F. (2018). Truth, disjunction, and induction. arXiv:1805.09890 [math.LO].

Feferman, S. (1984). Towards useful type-free theories I. *Journal of Symbolic Logic*, 49(1):75–111.

Feferman, S. (1991). Reflecting on incompleteness. *Journal of Symbolic Logic*, 56: 1–49.

Field, H. (1994). Deflationist views of meaning and content. *Mind*, 103(411):249–285.

Field, H. (2008). *Saving truth from paradox*. Oxford University Press, Oxford.

Fischer, M., Nicolai, C., and Horsten, L. (2017). Iterated reflection over full disquotational truth. *Journal of Logic and Computation*, 27(8):2631–2651.

Ganea, M. (2009). Arithmetic on semigroups. *Journal of Symbolic Logic*, 74(1):265–278.

Grišin, V. (1982). Predicate and set-theoretic calculi based on logic without contraction. *Math. Izvestija*, 18:41–59. (English Translation).

Halbach, V. (2014). *Axiomatic theories of truth. Revised edition*. Cambridge University Press.

Halbach, V. and Horsten, L. (2006). Axiomatizing Kripke's theory of truth in partial logic. *Journal of Symbolic Logic*, 71: 677–712.

Halbach, V. and Nicolai, C. (2018). On the costs of nonclassical logic. *Journal of Philosophical Logic*, 47:227–257.

Horsten, L. and Leigh, G. E. (2017). Truth is simple. *Mind*, 126(501):195–232.

Horwich, P. (1998). *Truth*. Clarendon Press.

Jäger, G. and Stärk, R. F. (1998). A proof-theoretic framework for logic programming. In Buss, S. R., editor, *Handbook of Proof-Theory*, pages 639–682. Elsevier.

Jago, M. (2018). *What Truth Is*. Oxford: Oxford University Press.

Kaye, R. and Wong, T. (2006). On interpretations of arithmetic and set theory. Online Draft.

Ketland, J. (1999). Deflationism and tarski's paradise. *Mind*, 108(429):69–94.

Kotlarski, H., Krajewski, S., and Lachlan, A. H. (1981). Construction of satisfaction classes for nonstandard models. *Canadian Mathematical Bulletin*, 24(1):283–93.

Kreuger, P. (1994). Axioms in definitional calculi. In Dyckhoff, R., editor, *Extensions of Logic Programmiing, 4th International Workshop, ELP'93, St Andrews, UK*, number 798 in Lecture Notes in Computer Science, pages 196–205.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72:690–712.

Leigh, G. (2015). Conservativity for theories of compositional truth via cut elimination. *The Journal of Symbolic Logic*, 80.

Martin, R. L. and Woodruff, P. W. (1975). On representing 'true-in-l' in l. *Philosophia*, 5(3):213–217.

Nicolai, C. (2015). Deflationary truth and the ontology of expressions. *Synthese*, 192(12):4031–4055.

Nicolai, C. (2016). A note on typed truth and consistency assertions. *Journal of Philosophical Logic*, 45(1):89–119.

Nicolai, C. (2018a). Cut-elimination for systems of naïve consequence and truth with restricted initial sequents. Draft.

Nicolai, C. (2018b). Provably true sentences across axiomatizations of Kripke's theory of truth. *Studia Logica*, 106:101–130.

Nicolai, C. and Stern, J. (2018). The modal completeness of Kripke-Feferman truth. Draft.

Petersen, U. (2000). Logic without contraction as based on inclusion and unrestricted abstraction. *Studia Logica*, 64(3):365–403.

Quine, W. V. (1970). *Philosophy of Logic*. Harvard University Press.

Schroeder-Heister, P. (2016). Restricting initial sequents: the trade-off between identity, contraction, cut. In *Advances in Proof Theory*, volume 28 of *Progress in Computer Science and Applied Logic*. Springer.

Shapiro, S. (1998). Proof and truth: Through thick and thin. *Journal of Philosophy*, 95(10):493–521.

Soames, S. (1998). *Understanding Truth*. Oxford University Press USA.

Świerczkowski, S. (2003). Finite sets and gödel's incompleteness theorems. *Dissertationes Mathematicae*.

Visser, A. (2010). What is the right notion of sequentiality? *Logic Group Preprint Series*, 288:1–24.

Visser, A. and Enayat, A. (2015). New constructions of satisfaction classes. In Fujimoto, K., Fernández, J. M., Galinon, H., and Achourioti, T., editors, *Unifying the Philosophy of Truth*. Springer Verlag.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.