INEXHAUSTIBILITY AND IMPLICIT COMMITMENT

Carlo Nicolai

KING'S College LONDON

'Tübingen', July 9th, 2021

(slides at carlonicolai.github.io)

To what extent can mathematical thought be analyzed in formal terms? Godel's theorems show the inadequacy of single formal systems for this purpose, except in relatively restricted parts of mathematics. However at the same time they point to the possibility of systematically generating larger and larger systems ***whose acceptability is implicit in the acceptance of the starting theory****. The engines for that purpose are what have come to be called reflection principles.*

Feferman, *Reflecting on Incompleteness*, p.1.

Let $T \supseteq \mathrm{EA}$ in the language $\mathcal{L} \supseteq \mathcal{L}_{\mathbb{N}}$. The following express, to different degrees of adequacy, that *all theorems of T are true*:

- Uniform Reflection:

$$\{\forall x(\mathrm{Pr}_T(\ulcorner A(\dot{x})\urcorner) \to A(x)) \mid A(v) \in \mathcal{L}\} \tag{RFN(T)}$$

- Global Reflection:

$$(\forall \varphi \in \mathcal{L})(\mathrm{Pr}_T(\varphi) \to \mathrm{Tr}(\varphi)) \tag{GRF(T)}$$

The ***Philosophical*** Problem of Reflection

*'...whose **acceptability** is implicit in **acceptance** of the starting theory...'*

- Suppose that you are **justified in believing**, say, PA
- Justification is transferred via **proof**
- By Gödel's Second Incompleteness Theorem, your justification for PA cannot be **inferentially** transferred to $\mathrm{RFN}(\mathrm{PA})$, let alone $\mathrm{GRF}(\mathrm{PA})$
- Yet, *if* $\mathrm{RFN}(\mathrm{PA})$ (resp. $\mathrm{GRF}(\mathrm{PA})$) indeed **expresses** the soundness of PA, what kind of **non-inferential** warrant supports $\mathrm{RFN}(\mathrm{PA})(\mathrm{GRF}(\mathrm{PA}))$?

A FIRST PATH: TRUTH

'Literally speaking, the intended reflection principle cannot be formulated in T itself by means of a single statement. This would require a truth definition...' (Kreisel and Levy 1968, p. 98)

- We focus on theories $T \supseteq \mathrm{EA}$ in $\mathcal{L}_{\mathbb{N}} := \{+, \cdot, 1, 0, \exp\}$.
- To formulate

$$(\forall \varphi \in \mathcal{L}_{\mathbb{N}})(\mathrm{Pr}_T(\varphi) \to \mathrm{Tr}(\varphi)), \qquad (\mathrm{GRF}_{\mathcal{L}_{\mathbb{N}}}(T))$$

  one needs to know what Tr means. A sound choice is turning Tarskian semantic clauses $(\mathrm{CT}(\cdot))$ into axioms:

$$\begin{aligned}
&\mathrm{Tr}\,(s = t) \leftrightarrow \mathrm{val}(s) = \mathrm{val}(t)\\
&(\forall \varphi \in \mathcal{L}_{\mathbb{N}})(\mathrm{Tr}\,(\neg\varphi) \leftrightarrow \neg\mathrm{Tr}\,\varphi)\\
&(\forall \varphi, \psi \in \mathcal{L}_{\mathbb{N}})(\mathrm{Tr}\,(\varphi \wedge \psi) \leftrightarrow (\mathrm{Tr}\,\varphi \wedge \mathrm{Tr}\,\psi))\\
&(\forall \varphi(v) \in \mathcal{L}_{\mathbb{N}})(\mathrm{Tr}\,(\forall v \varphi) \leftrightarrow \forall t\, \mathrm{Tr}\,(\varphi(t/v)))
\end{aligned}$$

- If non-logical schemata are extended to Tr, $\mathrm{CT}(T)$ proves $\mathrm{GRF}_{\mathcal{L}_{\mathbb{N}}}(T)$, and therefore $\mathrm{RFN}_{\mathcal{L}_{\mathbb{N}}}(T)$ because $\mathrm{Tr}\,(A\dot{x})$ and $A(x)$ are materially equivalent for $A(v) \in \mathcal{L}_{\mathbb{N}}$.

First Issue

The transition $T \mapsto \mathrm{CT}(T)$ requires an argument. Justification does not immediately transfer (and certainly not inferentially!)

Behind the adoption of $\mathrm{CT}(T)$ there's the idea that the **concept** of truth is characterized by the CT axioms.

This may be sound, but unsatisfactory:

- If we know *anything* about the concept of truth, it's that it's **self-applicable**, whereas the truth predicate of CT is not.
  Example: $\mathrm{CT}(T) \nvdash \mathrm{Tr}\ulcorner \mathrm{GRF}(T) \urcorner$), if consistent.
- This feature makes **iterations** in need of independent justification.
  Example: Justification for $\mathrm{CT}(T)$ transfers to $\mathrm{CT}_{n+1}(T)$ ($\sim (\Pi^0_1\text{-}\mathrm{CA})_{n+1}$) and even into the transfinite. This requires additional machinery (e.g. ordinals) to be justified on independent grounds.

Type-free truth can (and has been) be invoked. The **Kripke-Feferman** (KF) axioms embody the idea that the interaction between (classical) *truth and negation* is paradoxical. We now talk about *sentences* φ, ψ of $\mathcal{L}_{\mathrm{Tr}} := \mathcal{L}_{\mathbb{N}} \cup \{\mathrm{Tr}\}$.

$$\mathrm{Tr}\,(s = t) \leftrightarrow \mathrm{val}(s) = \mathrm{val}(t) \qquad \mathrm{Tr}\,\neg(s = t) \leftrightarrow \mathrm{val}(s) \neq \mathrm{val}(t)$$

$$\mathrm{Tr}\,\mathrm{Tr}\,t \leftrightarrow \mathrm{Tr}\,\mathrm{val}(t) \qquad \mathrm{Tr}\,\neg\mathrm{Tr}\,t \leftrightarrow \mathrm{Tr}\,\neg\mathrm{val}(t)$$

$$\mathrm{Tr}\,(\varphi \wedge \psi) \leftrightarrow (\mathrm{Tr}\,\varphi \wedge \mathrm{Tr}\,\psi) \qquad \mathrm{Tr}\,\neg(\varphi \wedge \psi) \leftrightarrow (\mathrm{Tr}\,\neg\varphi \vee \mathrm{Tr}\,\neg\psi)$$

$$\mathrm{Tr}\,(\forall v \varphi) \leftrightarrow \forall t\, \mathrm{Tr}\,\varphi(t/v) \qquad \exists t\, \mathrm{Tr}\,\neg\varphi(t/v) \leftrightarrow \mathrm{Tr}\,\neg(\forall v \varphi)$$

$\mathrm{KF}(T)$ not only entails $\mathrm{GRF}(T)$, but has "significant" proof-theoretic strength: $\mathrm{KF} \equiv_{\mathcal{L}_{\mathbb{N}}} (\Pi^0_1\text{-}\mathrm{CA})_{<\varepsilon_0}$, and the schematic version of KF is equivalent to ATR_0.

KF-like options have to live with the fact that

$$\{A \mid \mathrm{KF} \vdash A\} \neq \{A \mid \mathrm{KF} \vdash \mathrm{Tr}\ulcorner A \urcorner\}.$$

As a consequence:

Observation

$\mathrm{KF} + \mathrm{GRF}(\mathrm{KF})$ is internally inconsistent – and outright inconsistent if $\neg\mathrm{Tr}(\varphi \wedge \neg\varphi)$ is assumed.

However, $\mathrm{RFN}(\mathrm{KF})$ is consistent. Moreover, Graham Leigh showed that there's a tight correspondence between iterations of uniform reflection (κ) and transfinite induction over KF (ε_κ). *But what notion of soundness is* RFN *expressing, if not the one in* GRF*?*

There's a question of internal coherence: justification for $\mathrm{GRF}(T)$ (and much more!) is given by KF, but KF deems GRF *incorrect* by its own light.

Alternatively, one may take $\{A \mid \mathrm{KF} \vdash \mathrm{Tr}\ulcorner A \urcorner\}$ at face value: the resulting logic is FDE (or $\mathrm{K3}$ if internal consistency is assumed), and the system PKF. Axioms (necessarily in sequent forms) will look like:

$\mathrm{Tr}\,(s = t) \Rightarrow s = t$	$s = t \Rightarrow \mathrm{Tr}\,(s = t)$
$\mathrm{Tr}\,\mathrm{Tr}\,t \Rightarrow \mathrm{Tr}\,\mathrm{val}(t)$	$\mathrm{Tr}\,\mathrm{val}(t) \Rightarrow \mathrm{Tr}\,\mathrm{Tr}\,t$
$\mathrm{Tr}\,\neg\varphi \Rightarrow \neg\mathrm{Tr}\,\varphi$	$\neg\mathrm{Tr}\,\varphi \Rightarrow \mathrm{Tr}\,\neg\varphi$
$\mathrm{Tr}\,(\varphi \wedge \psi) \Rightarrow \mathrm{Tr}\,\varphi \wedge \mathrm{Tr}\,\psi$	$\mathrm{Tr}\,\varphi \wedge \mathrm{Tr}\,\psi \Rightarrow \mathrm{Tr}\,(\varphi \wedge \psi)$
$\mathrm{Tr}\,(\forall v \varphi) \Rightarrow \forall t\,\mathrm{Tr}\,\varphi(t/v)$	$\forall t\,\mathrm{Tr}\,\varphi(t/v) \Rightarrow \mathrm{Tr}\,(\forall v \varphi)$

Lemma (Fischer, N.)

Over $T \supseteq \mathrm{PKF}$, since $\mathrm{Tr}\ulcorner A\dot{x} \urcorner \Leftrightarrow Ax$ for any $Av \in \mathcal{L}_{\mathrm{Tr}}$, TFAE:

$$\frac{\Rightarrow \mathsf{Thm}_T(\ulcorner A\dot{x} \urcorner)}{\Rightarrow A(x)}\ (\mathsf{RFN}^R_T) \qquad \frac{\Rightarrow \mathsf{Sent}_{\mathcal{L}_{\mathrm{Tr}}}(x) \wedge \mathsf{Thm}_T(x)}{\Rightarrow \mathrm{Tr}\,x}\ (\mathsf{GRF}^R_T)$$

This may give hope, but in order to obtain significant extensions by reflection in the context of FDE $(\mathrm{K3})$, one needs the more complex rule:

$$\frac{\Rightarrow \mathsf{Prv}_S(\ulcorner \Gamma[\dot{x}] \Rightarrow \Delta[\dot{x}]\urcorner, \ulcorner \Theta[\dot{x}] \Rightarrow \Lambda[\dot{x}]\urcorner) \qquad \Gamma[x] \Rightarrow \Delta[x]}{\Theta[x] \Rightarrow \Lambda[x]} \quad (\mathrm{RR}(S))$$

To stick with finite levels, iterations give us *some* strength, although the absence of an equivalent of the Gentzen jump formula makes things slower:

Observation

$\mathrm{RR}^{\omega}(\mathrm{PKF})$ proves only all instances of transfinite induction up to ω^{ω^2}.

In sum:

- Although *some form of* coherence of the operation of extending a theory by reflection is restored, the logical stregth of GRF in its traditional form is compromised.
- One has to live with a nonclassical conditional, with all its costs.

A SECOND, TRUTH-FREE PATH

*'...whose acceptability is **implicit** in acceptance of the starting theory...'*

Question

What are principles characterizing such a notion of *implicit commitment*?

Theories τ are now **Δ_0-formulae with one free variable** that, provably in Kalmar's Elementary Arithmetic EA, define a set of sentences.

Elementary Reducibility

Suppose that τ and τ' are two theories. We say that τ is **elementarily reducible** to τ', denoted $\tau \leq_{er} \tau'$, iff there exists an EA-provably total elementary function f such that

$$\mathrm{EA} \vdash \mathrm{Proof}_{\tau}(y, x) \rightarrow \mathrm{Proof}_{\tau'}(f(y), x).$$

I consider an operator $\mathcal{I}$ on theories, which takes **a concrete axiom set** and associated proof-system and returns (a necessary part of) the implicit commitments of someone who justifiedly believes τ. $\mathcal{I}$ is characterized by the following principles:

Invariance

$$\text{if } \tau' \leq_{er} \tau, \text{ then } \mathcal{I}(\tau') \subseteq \mathcal{I}(\tau)$$

Example: the axioms of $\mathrm{Q} + \mathrm{Ind}(\mathcal{L}_{\mathbb{N}})$ and $\bigcup_{n\in\omega} \mathrm{I}\Sigma_n$ have the same implicit commitments.

Reflection

$$\text{if } \mathrm{EA} \vdash \forall x\, \tau(\ulcorner \varphi(\dot{x}) \urcorner), \text{ then } \forall x\, \varphi(x) \in \mathcal{I}(\tau).$$

Example: if one accepts PA, and EA proves that 'every number is an instance of a PA-axiom A', then $\forall x\, Ax$ is part of their implicit commitment.

Proposition (Łełyk, N.)

If τ extends EA, then $\mathrm{RFN}(\tau) \subseteq \mathcal{I}(\tau)$.

Proof.

First, one has (Feferman):

$$\mathrm{EA} \vdash \forall x \, \mathrm{Pr}_\tau(\ulcorner \mathrm{Proof}_\tau(x_1, \ulcorner \varphi(\dot{x}_2)\urcorner) \to \varphi(x_2)\urcorner) \tag{1}$$

Let

$$\tau'(x) :\leftrightarrow x \in \mathrm{EA} \vee \exists y \, x = \ulcorner \mathrm{Proof}_\tau(y_1, \ulcorner \varphi(\dot{y}_2)\urcorner) \to \varphi(y_2)\urcorner \tag{2}$$

By (1) and REFLECTION, we get $\mathrm{RFN}(\tau) \subseteq \mathcal{I}(\tau')$. Since (1) also gives us $\tau' \leq_{er} \tau$, INVARIANCE then yields $\mathrm{RFN}(\tau) \subseteq \mathcal{I}(\tau)$. □

Main Claim

Justified belief in τ is **preserved** to $\mathrm{RFN}(\tau) \in \mathcal{I}(\tau)$.

- It's plausible that elementary reducibility preserves JB.
- Therefore, REFLECTION becomes crucial. We invoke its **deductive lightness** (meta-inferential transmission of justification):
  - *Reflection* is computationally simple(r) than Uniform Reflection
  - *Reflection* mirrors τ-provability (not self-embeddable)
  - *Reflection* can be conservatively interpreted in τ (e.g. by letting $\mathcal{I}(\tau) := \{\forall x \varphi \mid \mathrm{EA} \vdash \forall x(\tau(\ulcorner \varphi(\dot{x}) \urcorner)\}$

Summing up, in two aphorisms:

If soundness means truth, truth may not be sound.

Even if there's no truth, Uniform Reflection may be justified.