A Theory of Implicit Commitment

Carlo Nicolai
(joint with Mateusz Łełyk)

KING'S College LONDON

Konstanz, February 12 2020

(slides at carlonicolai.github.io)

The Project

We defend the following thesis on novel grounds:

> IMPLICIT COMMITMENT THESIS (ICT): *Anyone who is **justified in believing** a mathematical formal system S is also implicitly committed to **various additional statements** which are expressible in the language of S but which are formally independent of its axioms.*

For a theory τ, we focus on:

$$\mathrm{RFN}(\tau) := \{\forall z \big(\mathrm{Prov}_\tau(\ulcorner \varphi(\dot{z}) \urcorner) \to \varphi(z)\big) \mid \varphi(v) \in \mathcal{L}_{\mathbb{N}}\}$$

One key contribution is a **direct axiomatization** of implicit commitment.

Principle of Invariance

Justified belief in a theory τ (and associated proof-system) commits one to theories that are reducible to τ in a **sufficiently simple way**.

Example

Let:

$$\mathrm{PA_I} := \bigcup_{n \in \omega} \mathrm{I}\Sigma_n \qquad\qquad \mathrm{PA_{II}} := \mathrm{Q1\text{-}6} + \mathrm{Ind}_{\mathcal{L}_{\mathbb{N}}}$$

Principle of Axiomatic Reflection

Justified belief in τ (and associated proof-system) commits one to universal claims whose instances are **uniformly and uncontroversially** recognized as axioms of τ.

Example

If $\mathrm{PA} \vdash$ ‘every number is an instance of a PA-axiom φ’, *then* $\mathrm{PA} \vdash \forall x\, \varphi(x)$.

The Formal Theory

- We focus on theories in $\mathcal{L}_{\mathbb{N}} := \{+, \cdot, 1, 0, \exp, \leq\}$, with $\exp(x)$ is 2^x.
- Theories τ 'are' Δ_0-formulae with one free variable that, provably in Kalmar's Elementary Arithmetic EA, define a set of sentences.

Definition

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

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

We axiomatize an operator $\mathcal{I}$ on theories, which takes a concrete axiom set and associated proof-system and returns (part of) the implicit commitments of someone who justifiedly believes τ:

Invariance

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

Reflection

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

Proposition

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

Proof.

First, one has (Feferman):

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

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 $\mathrm{RFN}(\tau) \subseteq \mathcal{I}(\tau)$. □

Individually, the principles do not force logical strength:

- Let $\mathcal{I}_{\mathrm{I}}(\tau) := \tau$. EA is arithmetically sound: the assumption $\tau' \leq_{er} \tau$ entails that $\mathcal{I}_{\mathrm{I}}(\tau') \subseteq \mathcal{I}_{\mathrm{I}}(\tau)$.

- Let
$$\mathcal{I}_{\mathrm{II}}(\tau) = \{\forall x \varphi \mid \mathrm{EA} \vdash \forall x\, \tau(\ulcorner \varphi(\dot{x}) \urcorner)\}$$
Then $\mathcal{I}_{\mathrm{II}}(\mathrm{PA})$ is deductively equivalent to PA.

Justified Belief, Stability, and Entitlement

Main Claim

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

- For INVARIANCE: since elementary reducibility preserves JB, if $\tau \mapsto \mathcal{I}(\tau)$ preserves JB (and $\tau' \leq_{er} \tau$), also $\tau' \mapsto \mathcal{I}(\tau')$ does.
- Therefore, REFLECTION becomes crucial. We invoke its **deductive lightness** (meta-inferential transmission of justification):
  - Unlike RFN(τ), it can be *conservatively interpreted* in τ.
  - Unlike RFN(τ), it involves an elementary property (i.e. membership in τ), not a *recursively enumerable* one.
  - It cannot be iterated.
  - Possibility of error (e.g. hidden ω-inconsistency) is substantially reduced.

EPISTEMICALLY STABLE THEORY*: if 'there exists a* ***coherent rationale for accepting [it]*** *which does not entail or otherwise oblige a theorist to accept statements which cannot be derived from [its] axioms' (Dean 2014).*

Observation

Having a 'coherent rationale for accepting' a theory τ entails having a notion (possibly dispositional) of what the axioms of τ are.

Claim

Once you have the general notion of axiom for a theory τ you're justifiedly believing, you're bound to have justified belief in $\mathrm{RFN}(\tau)$. This leaves untouched weaker versions, in which the 'coherent rationale' is not available to the τ-theorist.

Comparisons

- Although we cannot be justified in believing RFN(τ), we can be **entitled** to it (H& et al. 17/19, H21), and therefore we can get to know it.
- We argued that something stronger is true, however...
- Our formal framework can be used to locate the source of the entitlement in principles that are **properly weaker** than Uniform Reflection: INVARIANCE and REFLECTION.

'S accepts τ': *S believes that for every theorem φ of τ there is a normally-good-enough reason to believe that φ [in short: φ is believable]. (C17, p. 251)*

REF $\forall\varphi\big(\mathrm{Prov}_{\tau B}(\varphi) \to B(\varphi)\big)$;

MP $\forall\varphi, \psi\big(B(\varphi) \wedge B(\varphi \to \psi) \to B(\psi)\big)$;

ωR $\forall\varphi(v)\big(B(\ulcorner\forall x B(\varphi(\dot{x}))\urcorner) \to B(\forall x \varphi(x))\big)$.

$$\frac{\varphi}{B(\ulcorner\varphi\urcorner)} \qquad (\textsc{nec})$$

$\mathrm{Int}_{\mathrm{Bel}(\tau)} = \{\varphi \in \mathcal{L}_B \mid \mathrm{Bel}(\tau) \vdash B(\ulcorner\varphi\urcorner)\}$.

Cieślińki shows that $\mathrm{Int}_{\mathrm{Bel}}(\tau)$ contains ω-iterations of $\mathrm{RFN}(\tau)$.

- Believability seems rather weak to support a strong thesis such as ICT, and yet is satisfies strong rules.
- First, (ωR) and NEC entail closure under:

$$\frac{\forall x B(\ulcorner \varphi(\dot{x}) \urcorner)}{B(\ulcorner \forall x \varphi(x) \urcorner)}$$

  This seems rather strong for a 'believability' predicate.
- Moreover, NEC is unrestricted. Unlike our REFLECTION, it applies to $\mathrm{Bel}(\tau)$-proofs as well. Why?

Proposition

Let τ be Σ_1-sound. Then $\mathrm{Int}_{\mathrm{Bel}'(\tau)}$ is conservative over τ.

We presented:

- An **axiomatic** characterization of the necessary part of implicit commitments
- A **defence** of ICT based on a preservation of **justified belief**
- A contribution to the debate on entitlement to reflection principles