LOGICAL FRAMEWORKS FOR ACCEPTANCE AND COMMITMENT IN MATHEMATICS

CARLO NICOLAI
KING'S COLLEGE LONDON

1. The Question

In this talk I take a foundational theory to provide an account of how some mathematical beliefs could have been warranted in principle. In Crispin Wright's terminology, a *foundational cognitive project* can be seen as a collection of claims in some language (e.g. the arithmetical language $\mathscr{L}_{\mathbb{N}}$) about a mathematical subject matter (e.g. arithmetic), and the collection of proofs and refutations in a given formal theory T (e.g. proofs and refutations in Peano Arithmetic (PA)). The link between the subject matter and the system T is typically part of one's foundational standpoint, also providing a basic epistemic warrant for T. Proof in T can be seen as a straightforward way to provide *evidential* warrant for the relevant theorems.

By Gödel's Second Incompleteness Theorem, and for a reasonable T, one cannot have evidential warrant via proof for the formal claims 'T is consistent', 'T is sound' on the basis of the foundational project involving T – assuming, of course, that these claims actually express the relevant propositions (cf. next section!).

Yet, it appears to be *epistemically obligatory* for anyone embarked in a foundational project involving T – such that T also provides evidential warrant for its theorems – to believe that T is consistent or sound. For these (and other) reasons, such claims are often referred to as *implicit commitments*.[1]

So, and here's the problem, is there *evidential* or *non-evidential warrant* for those purported beliefs?[2]

2. Expressing Meta-Mathematical Propositions

How do we know that the the usual metamathematical sentences (or schemata), e.g.

$$\begin{aligned}
\mathrm{Con(PA)} &:\leftrightarrow \neg\exists x\, \mathrm{Proof}_{\mathrm{PA}}(x, \ulcorner 0 = 1 \urcorner) \leftrightarrow \neg\mathrm{Prov}_{\mathrm{PA}}(\ulcorner 0 = 1 \urcorner)\\
\mathrm{Rfn(PA)} &:= \{\mathrm{Prov}_{\mathrm{PA}}(\ulcorner A \urcorner) \to A \mid A \in \mathscr{L}_{\mathbb{N}}\}\\
\mathrm{RFN(PA)} &:= \{\forall x(\mathrm{Prov}_{\mathrm{PA}}(\ulcorner A(\dot{x}) \urcorner) \to A(x)) \mid A(v) \in \mathscr{L}_{\mathbb{N}}\}
\end{aligned}$$

express the relevant meta-mathematical propositions?[3]

If one cares about epistemological assumptions, some of the usual accounts won't suffice. Consider the case of *provability*:

(i) A is provable in PA iff $\mathbb{N} \vDash \mathrm{Prov}_{\mathrm{PA}}(\ulcorner A \urcorner)$: but the standard model $\mathbb{N}$ is not accessible given first-order PA;
(ii) A is provable in PA iff $\mathrm{PA} \vdash \mathrm{Prov}_{\mathrm{PA}}(\ulcorner A \urcorner)$: weak representability/Kreisel's condition does not differentiate between canonical and non-canonical provability predicates (e.g. Rosser);
(iii) A is provable in PA iff it satisfies some meaning postulates, e.g. Löb's derivability conditions: the predicate $x = x$ satisfies all those.

My way of "biting the bullet" is: we take some *natural* development of syntax formalized, hands-on, in some finite set theory based on $\varnothing$, $x, y \mapsto x \cup \{y\}$ and

$$A(\varnothing) \wedge \forall x, y(A(x) \wedge A(y) \to A(x \cup \{y\})) \to \forall x A(x),$$

where A can be taken to be of bound complexity. This theory will be *bi-interpretable*, and even *definitionally equivalent*, with a fragment of PA or other foundationally relevant theories. It's this shared, bottom-up construction process that provides the required correlation between formal and informal claims.

Date: Konstanz, September 7, 2022.

[1] Ex auctoritate: Gödel writes

> It is this theorem [the second incompleteness theorem] which makes the incompletability of mathematics particularly evident. For, it makes it impossible that someone should set up a certain well-defined system of axioms and rules and consistently make the following assertion about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics. If someone makes such a statement he contradicts himself. For if he perceives the axioms under consideration to be correct, he also perceives (with the same certainty) that they are consistent. Hence he has a mathematical insight not derivable from his axioms (Gödel, 1995, p. 309).

And here's Feferman:

> [reflection] the process of finding out what is implicit in accepting a basic system L_1, i.e., what one ought to accept, *on the same fundamental grounds*, when one accepts L_1...(Feferman, 1988, p. 131)

[2] A way of rephrasing the question: what does it mean to say that these extensions ought to be accepted 'on the same fundamental grounds'?

[3] We should note, for later use, that Rfn(PA) and RFN(PA) are informally glossed via a *truth-predicate*.

3. Non-Expansionist Strategies

3.1. **Non-Evidential Route.** Here's the setting proposed in Horsten (2021) building on Crispin Wright's non-evidentialist epistemology: Con(PA) is a *cornerstone proposition* for this cognitive project; to this proposition we are *entitled* (Wright, 2004, p. 191-2):

- We have no sufficient to believe that Con(PA) is false;
- The attempt to justify Con(PA) would involve presuppositions of in turn of no more secure and prior standing.

Potential problems of the approach: entitlements may be too generous – what about Con(ZFC)? It may be challenged the idea that the combination of entitlements and warranted beliefs can lead the agent to *claim knowledge* of their consequences.

3.2. **Evidential Route.** Fischer (2021) considers a 'tamed' version of the ω-rule with primitive recursive premisses. He proposes to add to a reasonable T a rule of the form:

$$(\omega\text{-PR}) \qquad \frac{T \vdash \forall x \mathrm{Proof}_T(\ulcorner A(\dot{x}) \urcorner, f(x)) \quad \text{for } f \in \mathrm{PR}}{\forall x A}$$

$T + \omega$-PR yields RFN(T). However, it seems as if the warrant for this rule would simply amount to a warrant for RFN(T) itself.

Łełyk and Nicolai (2022) propose a more nuanced formal framework. Given a foundational theory $T \supseteq$ EA, *they claim that* RFN(PA) *can be warranted evidentially*. The idea is to *"axiomatize" the notion of implicit commitment*. Keeping T fixed, the principles:

(INVARIANCE) If T' is proof-theoretically reducible to T in a sufficiently simple way,[4] the commitments of T' are *included* in the commitments of T.

(REFLECTION) if EA proves that $A(\overline{n})$ is an *axiom* of T for all n, then $\forall x A$ is in the commitments of T.

Proposition 1 (Łełyk and Nicolai (2022)). *T+INVARIANCE and T+REFLECTION can be conservatively interpreted in T. However, T+INVARIANCE+REFLECTION entails all instances of* RFN(T).

4. Expansionist Strategy: Truth Predicates

One may argue that the initial question should to be framed in a different setting. The very formulation of Rfn(T) and RFN(PA) call for a language expansion: a straightforward English paraphrase of Rfn(PA) and RFN(PA) is given in terms of truth.[5] However, truth is inexpressible in $\mathscr{L}_{\mathbb{N}}$ (and PA-undefinable). So, one needs to move to $\mathscr{L}_{\mathrm{Tr}} := \mathscr{L}_{\mathbb{N}} \cup \{\mathrm{Tr}\}$, where we can naturally formulate:

$$(\mathrm{GRP(PA)}) \qquad \forall \varphi (\mathrm{Prov}_{\mathrm{PA}}(\varphi) \to \mathrm{Tr}\, \varphi).$$

4.1. **Adopting a Truth Theory.** To *prove* GRP(PA), one needs suitable axioms for Tr. The usual schema 'A' is true iff A won't suffice. What suffices is the extension of PA with the formalization of the relation $\mathbb{N} \models A$ in $\mathscr{L}_{\mathrm{Tr}}$, a theory that is called CT for 'compositional truth'.[6] So the idea: warrant in PA as a foundational theory transfers to CT – roughly, because warrant in T suffices for warranting that *T is true*. Therefore, this warrant can be transferred deductively to GRP(PA). However, it could be argued that we are not using a philosophically adequate concept of truth: the CT truth-predicate is *typed*. There is also another issue: why not *iterating* CT if this comes "for free"? And if you can do this, how far?

To fix this, one could move to *type-free* theories of truth. *Kripke-Feferman* (KF) truth is a theory that axiomatizes in classical logic the class of fixed-points of the operator:

$$\Gamma(X) = \{\varphi \in \mathscr{L}_{\mathrm{Tr}} \mid (\mathbb{N}, X) \models_{\mathrm{sk}} \varphi\},$$

where $\models_{\mathrm{sk}}$ is the satisfaction relation of Strong Kleene Logic. The advantage of KF is that not only GRP(PA) is recovered, but also all *iterations* of CT or iterations of arithmetical comprehension for all (codes of) ordinals that can be well-ordered in PA. This speaks in favour of KF *instrumentally*, but what about the KF-truth predicate as a concept of truth? By itself, it features "puzzling theorems" such as:

$$(1) \qquad (\lambda \wedge \neg \mathrm{Tr}\ulcorner \lambda \urcorner) \vee (\neg \lambda \wedge \mathrm{Tr}\ulcorner \lambda \urcorner).$$

Moreover, if GRP(S) captures the soundness of S in the presence of Tr, this looks like bad news:

Fact 2. KF + GRP(KF) *is internally inconsistent.*

---

[4]This could be understood either as elementary, or even P-TIME proof-transformations.

[5]There are alternatives to this expansion, using epistemic predicates. Cieśliński (2017) contains a proposal based on a "believability" predicate.

[6]For the definition, you can consult one of Halbach (2014); Horsten (2011); Cieśliński (2017). For the proof: There are some tricky details: such as the role of induction of PA and the double role of PA as theory of syntax. But they don't matter for now.

4.2. **Entitlement again.** If CT or KF appear to require *additional* warrant to the one given for PA, is there a non-evidential route akin to the one considered in §3.1? Horsten and Leigh (2017) investigate this option. They start with simple bi-conditionals (schematic):

(2) $\mathrm{Tr}\ulcorner A \urcorner \leftrightarrow A$ for all $A \in \mathscr{L}_{\mathbb{N}}$

(3) $(\mathrm{Tr}\ulcorner A \urcorner \leftrightarrow A) \wedge (\mathrm{F}\ulcorner \overline{A} \urcorner \leftrightarrow A)$ for A in a negation-free language.

They show that CT and KF can be obtained by few iterations of $\mathrm{RFN}(\cdot)$ over (2) and (3), respectively. Leigh (2016) proves a very general theorem:

Proposition 3.
(i) *ε_α-induction together with* CT *yields an identical theory as α iterations of* $\mathrm{RFN}(\cdot)$ *over typed (uniform) disquotation;*
(ii) *ε_α-induction together with (negation-free)* KF *yields an identical theory as α iterations of* $\mathrm{RFN}(\cdot)$ *over type-free, positive (uniform) disquotation.*

However, (i) is hard to justify because of the dubious status of (2) as a collection of cornerstone propositions, and (ii), besides being formulated in an artificial language, is plagued by forms of Fact 2 above: the addition of $\mathrm{GRP}(\cdot)$ to the theories isn't as innocent as the addition of $\mathrm{RFN}(\cdot)$.

4.3. **Nonclassical Logic and Fully Disquotational Truth.** Fischer et al. (2017) and Fischer et al. (2021) show that an analogous strategy holds for *unrestricted* analogues of (2) and (3):

(TB) $$\mathrm{Tr}\ulcorner A \urcorner \Leftrightarrow A$$

where $\Leftrightarrow$ is a meta-theoretic (double)-sequent arrow. The price to pay is the *full* adoption of Strong Kleene Logic (or alternatives such as FDE, LP). However, the unrestricted, *quasi-logical* status of (TB) suggests that its instances can function as cornerstone propositions for the adoption of a truth concept.

Proposition 4. *Over $T \supseteq$ (TB),* $\mathrm{GRP}(T)$ *and* $\mathrm{RFN}(T)$ *are equivalent.*

However, to achieve some non-trivial consequences of the combination of reflection and full disquotational truth one needs to resort to more complex reflection rules:

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

The combination of (TB) and RR(S) enables one to recover stronger principles of truth – all compositional, type-free principles adapted to the nonclassical logic – and some non-trivial mathematical strength (some portions of predicative maths), albeit *way less* than what one gets in KF: ω-iterations of reflection only give us ω^{ω^2}-induction.

There is a way to do better: we can "push" the meta-theoretic biconditional in (TB) into the object language. The resulting logic is the extension of FDE with an intuitionistic conditional studied by Leitgeb (2019) – a logic called HYPE. The conrnerstone schema for the use of truth becomes:

(HTB) $$\mathrm{Tr}\ulcorner A \urcorner \rightleftharpoons A, \quad \text{for } A \text{ not containing } \rightharpoonup.$$

The rationale for accepting $\rightharpoonup$ is to see it as a non-extensional context, in which the equivalence of A and $\mathrm{Tr}\ulcorner A \urcorner$ may not be guaranteed. An analogue of Proposition 4 can be obtained in this setting. Moreover, the standard form of $\mathrm{RFN}(\cdot)$ can be preserved:

Proposition 5. *ε_α-induction over* KF*-truth in the logic* HYPE (*which is as strong as* classical KF) *is a subtheory of α-many iterations of* $\mathrm{RFN}(\cdot)$ *over* HYPE.

A corollary of the result is that there's no loss of power in moving to classical logic, with all its quirks, to HYPE. Since we appear to have a better ground for seeing (HTB) and $\mathrm{RFN}(\cdot)$ as entitlements, then the process of extending PA with truth is more convincing here.

HYPE has its quirks too, though: the version of $\mathrm{RFN}(\cdot)$ required needs to be extended to $\rightharpoonup$, and for this we *don't* have an analogue of Proposition 4.

References

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

Feferman, S. (1988). Turing in the land of o (z). In *A half-century survey on The Universal Turing Machine*, pages 113–147.

Fischer, M. (2021). Another look at reflection. *Erkenntnis*, pages 1–31.

Fischer, M., Horsten, L., and Nicolai, C. (2021). Hypatia's silence: Truth, justification, and entitlement. *Noûs*, 55(1):62–85.

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

Gödel, K. (1995). Collected works. iii: Unpublished essays and lectures. *Oxford University Press, Oxford*.

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

Horsten, L. (2011). *The Tarskian Turn*. MIT University Press, Oxford.

Horsten, L. (2021). On reflection. *The Philosophical Quarterly*, 71(4):pqaa083.

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

Leigh, G. (2016). Reflecting on Truth. *IFCoLog Journal of Logics and their Applications*, 3.

Leitgeb, H. (2019). Hype: A system of hyperintensional logic. *Journal of Philosophical Logic*, 48(2):305–405.

Łełyk, M. and Nicolai, C. (2022). A theory of implicit commitment. *Synthese*, 200(4):1–26.

Wright, C. (2004). Warrant for nothing (and foundations for free)? *Aristotelian Society Supplementary Volume*, 78(1):167–212.