ON IMPLICIT COMMITMENT FOR ARITHMETICAL THEORIES AND THE SEMANTIC CORE

CARLO NICOLAI AND MARIO PIAZZA

Abstract. According to the *implicit commitment thesis*, once accepting a mathematical formal system S, one is implicitly committed to additional resources not immediately available in S. Traditionally, this thesis has been understood as entailing that, in accepting S, we are bound to accept reflection principles for S and therefore claims in the language of S that are not derivable in S itself. It has recently become clear, however, that such reading of the implicit commitment thesis cannot be compatible with well-established positions in the foundation of mathematics which consider a specific theory S as self-justifying and doubt the legitimacy of any principle that is not derivable S: examples are Tait's finitism and the role played in it by Primitive Recursive Arithmetic, Isaacson's thesis and Peano Arithmetic, Nelson's ultrafinitism and sub-exponential arithmetical systems. This casts doubts on the very adequacy of the implicit commitment thesis for arithmetical theories. In the paper we show that such foundational standpoints are nonetheless compatible with the implicit commitment thesis. We also show that they can even be compatible with genuine soundness extensions of S with suitable forms of reflection. The analysis we propose is as follows: when accepting of system S, we are bound to accept a *fixed* set of principles extending S and expressing minimal soundness requirements for S, such as the fact that the non-logical axioms of S are true. We call this invariant component the *semantic core of implicit commitment*. But there is also a *variable* component of implicit commitment that crucially depends on the justification given for our acceptance of S in which, for instance, may or may not appear (proof-theoretic) reflection principles for S. We claim that the proposed framework regulates in a natural and uniform way our acceptance of different arithmetical theories.

1. Preamble

The acceptance of a system formalizing some portion of mathematics is the outcome of a complex justificatory process that is constrained by philosophical and ontological attitudes, influenced by pragmatical considerations (fruitfulness, generality), and also hospitable to aesthetical ones (simplicity, elegance). The acceptance of a formal system S, therefore, encompasses the possibility that some of the components of this justificatory process are not expressible or even formalizable in the language of S and that some crucial constituents of this acceptance are only left *implicit* by the process itself. Examples abound: just to remain on the formal side, for instance, soundness assertions for S involving the notion of truth are not expressible in the language of S, while most of their truth-free surrogates are not provable in S. We have come to the core idea underlying the notion of implicit commitment: when accepting a theory S, we are also bound to embrace a cluster of formal or semi-formal assertions that are not immediately available in S itself.[1]

[1]In what follows, we will always deal with systems S formulated in a language $\mathcal{L}_S$ extending the language of arithmetic $\mathcal{L} = \{0, \mathsf{S}, +, \times, \leq\}$.

Historically, the notion of implicit commitment for formal systems of arithmetic, and for mathematical theories more in general, emerged in the work of logicians and philosophers already in the 50's and 60's of the last century. Their main concern is clearly expressed in a later work by Solomon Feferman:

> To what extent can mathematical thought be analyzed in formal terms? Gödel'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 acceptance of the starting theory. The engines for that purpose are what have come to be called *reflection principles* [12, p. 1].

The articulation of this version of implicit commitment was crucial for the analysis of predicativism and, in particular, for the identification of predicatively definable sets of natural numbers.[2] In one well-known Feferman's formulation, the limit for the generation of such sets is articulated in terms of iterations of uniform reflection principles and predicative comprehension over Peano Arithmetic (see [10, 11, 31]). This project was refined and reshaped several times by Feferman in the past decades, moving from iterations of ramified analysis [11] to more succinct formulations such as the *reflective closure* of the starting theory involving a primitive notion of truth (see §3-4).

At any rate, no matter what formulation of Feferman's hierarchy of systems one chooses, the resulting picture of implicit commitment will entail that in accepting a starting theory S one is committed to statements that are not provable in S itself. Nevertheless, as we shall see later in the paper, it has recently become clear that the inclusion of reflection principles for the starting theory among the claims we are implicitly committed to when accepting it, although integral part of Feferman's foundational program, may clash with other philosophical standpoints. Therefore, we opt for a more neutral and more general formulation of the implicit commitment thesis:

> (ICT) In accepting a formal systems S one is also committed to additional resources not available in the starting theory S but whose acceptance is implicit in the acceptance of S.[3]

As it is formulated, (ICT) has the advantage of reflecting many instances of disagreement in the existing literature over what these 'additional resources' should amount to. Several authors, following Feferman, allow for conceptual resources that are not immediately available in S due to familiar Gödelian phenomena, such as consistency statements and reflection principles. Some of these authors go even further and argue that, once soundness extensions of S are admitted, they should be formulated by *explicitly* resorting to a notion of truth and not merely by implicitly referring to it via schemata [12, 45, 27, 5]. Other authors maintain instead that whatever we are committed to when endorsing the system S should already be expressible in the very language of S [48]. And finally, proponents of a more drastic view, like Jean-Yves Girard, even deny that reflection on our acceptance of S may have any epistemological value because it relies on a preexisting agreement on what axioms and rules should be believed to be true [19].

[2] For an overview of this debate, see [13].

[3] A similar formulation can be found in [6, p. 32].

It's important to notice, nonetheless, that the positions just sketched are extracted from works that are not directly concerned with a clarification of the notion of implicit commitment, but mainly with the notion of truth in the context of truth-theoretic deflationism.[4] A direct analysis of the notion of implicit commitment, however, is much needed. As Horsten and Leigh put it:

> philosophers of mathematics have hitherto largely failed to investigate the notion of implicit commitment, and have not spent much philosophical energy on analysing our warrant for reflection principles [24, p. 32].

Recently Walter Dean has claimed that what is traditionally taken as part of the principles we are committed to when accepting a system S — such as reflection principles — may clash with the justification provided for the acceptance of S itself [6]. In particular, he focuses on the well-known theses by William Tait and Daniel Isaacson, according to which finitary mathematics coincides with what can be proved in Primitive Recursive Arithmetic (PRA) and first order Peano Arithmetic (PA) respectively. Dean observes that both theses can be understood as suggesting that PRA and PA are *epistemically stable*, "in the sense that there exists a coherent rationale for accepting [these systems] which does not entail or otherwise oblige a theorist to accept statements which cannot be derived from [their] axioms"[6, p. 53]. Although we are not primarily interested in analyzing the notion of epistemic stability introduced by Dean, we will be concerned with some of its effects: the predicativist *à la* Feferman and the first-orderist *à la* Isaacson, although both assigning a privileged status to Peano Arithmetic, will do so on different grounds and this will heavily affect their stance on (ICT). For Feferman there will be a recognizable set of statements that are not derivable in PA, while being part of our implicit commitment to it; for Isaacson, by contrast, the additional resources hinted at in (ICT) will likely be non-existent. This determines a form of relativity of implicit commitment with respect to the acceptance of one's preferred arithmetical system that will be a recurring theme of this paper.

In what follows, we propose an analysis of the notion of implicit commitment for arithmetical theories. On our account, implicit commitment exhibits a variable and invariant component. We maintain, with Dean, that the set of principles defining the implicit commitment with respect to the acceptance of a theory S is relative to the justification given for this very acceptance. However, we also argue — *contra* Dean – that the acceptance of a system does involve an explicit soundness extension in the form of a *fixed* set of semantic principles, which we call 'semantic core'. The relative aspect of implicit commitment thus takes the form of different, possible extensions of this ubiquitous core. In other words, we will claim that there is a fundamental body of 'reflection' principles formulated through the notion of truth — such as the claim that all non-logical axioms of the accepted theory S are true — that are part of the implicit commitment relative to any reasonable justification offered for the acceptance of a theory S. What extends such a kernel is variable and constrained by the justification given by the idealized mathematician.

Here is a sketch of the structure of the paper. In the next section, we briefly discuss Dean's critical analysis of (ICT) carried out by relating it with Tait's and Isaacson's theses and we extend his remarks to Nelson's ultrafinitism. We claim that Dean's analysis does not lead to a dismissal of (ICT) but rather to an alternative interpretation of it. In section 3, in fact, we

[4] Girard and Feferman are obvious exceptions, although they also do not directly deal with conceptual analysis of the notion of implicit commitment.

introduce the necessary toolbox to take up this possible interpretation of (ICT), whereby the 'additional resources' we are committed to when accepting a system S do not entail statements that are *unprovable* in S. Our aim is to isolate the semantic core of an arithmetical theory, amounting to the fixed, invariable component of our commitment to it. In section 4 we defend the thesis that the distinction between semantic core and variable components of implicit commitment resolves the tension between Tait's and Isaacson's theses and (ICT); in addition, we show that the resulting picture of implicit commitment is also compatible with the traditional reading of (ICT) associated with positions such as Feferman's. Section 5 contains some concluding remarks.

2. Implicit commitment and foundational theses

As we mentioned in the previous section, Dean in [6] examines the way in which Tait and Isaacson respectively justify the acceptance of PRA and PA from a 'finitist' and a 'first-orderist' perspective. In both cases he concludes that (ICT) is incompatible with the justifications given by Tait and Isaacson. Let us now briefly recall Tait's and Isaacson's theses and how Dean employs them to criticize this strong reading of (ICT).

On the view articulated by Tait, the formal system of PRA — as standardly presented by Hilbert and Bernays ([23]) — captures precisely the finitist portion of mathematics. Inasmuch as the 'finitist portion of mathematics' is not itself a mathematical notion, Tait draws an analogy between his thesis and Church's thesis, suggesting that 'any plausible attempt to construct a finitist function that is not primitive recursive either fails to be finitist [...] or else turns out to be primitive recursive after all' [47, p. 533, p. 537]. In accordance with the spirit of Hilbert's program, Tait then investigates what counts as finitistic proof of a Π_1-sentence of $\mathcal{L}_{\mathsf{PRA}}$, namely the type of sentence capturing the *real* or *finitary* part of mathematics as opposed to its *ideal* or *transfinite* components. To this aim, he considers some finitistically acceptable principles — such as the laws of ordered pairs — governing how to manipulate concrete equations of $\mathcal{L}_{\mathsf{PRA}}$ and concludes that the proofs of expressions of the form $\forall\vec{x}\,(s(\vec{x}) = t(\vec{x}))$, for s, t terms of $\mathcal{L}_{\mathsf{PRA}}$, obtained from these principles are precisely the proofs of $s(\vec{x}) = t(\vec{x})$ in PRA. Tait's proof principles involve only a limited form of induction for finitistically acceptable types (cf. [47, p. 537]), much closer to primitive recursion than to the first-order schema of induction.[5] In fact, already the schema of induction for Σ_2-formulas, let alone the full schema of induction, would enable one to define recursive but not primitive recursive operations such as the Ackermann function. Such instances of induction are therefore not available to the finitist.

However, let us consider the so-called *uniform reflection principle* for an elementary theory S, namely the claim

$$\mathsf{RFN(S)} \qquad \forall x\,(\mathsf{Pr}_S(\ulcorner \phi(\dot{x}) \urcorner) \rightarrow \phi(x))$$

for $\phi(v)$ a formula of $\mathcal{L}_S$ with only v free, where $\mathsf{Pr}_S(\ulcorner \phi(\dot{x}) \urcorner)$ canonically expresses that the result of formally substituting the variable v with the numeral for x in $\phi(v)$ — formally the $\mathcal{L}_S$-term $\mathsf{sub}(\ulcorner \varphi \urcorner, \ulcorner v \urcorner, \mathsf{num}(x))$ — is provable in S. Letting EA be Kalmar arithmetic (cf. [2]), the following is well-known:

[5] The finitistic justification process for PRA sketched by Tait is rooted in the fundamental operation of manipulating finite sequences of objects. All operations and notions obtained by bootstrapping this operation are finistically kosher. In particular, this process of justification is not itself legitimate for the finitist because it assumes the general notion of function, which is not finitistically definable (cf., e.g., [47, pp. 531-533]).

Proposition 1 ([32]). *Over* EA, *full induction is equivalent to* RFN(EA).

Therefore, if one understands (ICT) as including RFN(EA), then the finitist should also be committed to the very induction principle of PA, which clearly isn't provable within the finitist's preferred theory PRA. Dean thus concludes that the finitist *à la* Tait cannot include RFN(EA) (and *a fortiori* RFN(PRA)) into the set of principles she is implicitly committed to when embracing PRA. If reflection principles are therefore taken to be, as in Feferman's own reading, the paradigmatic examples of 'resources not available in S', (ICT) amounts to an inadequate account of implicit commitment across reasonable arithmetical systems.

Therefore, such conclusion seems to be justified *only if* (ICT) could only be interpreted along the lines of Feferman's own account of it and , as a consequence, the reference to 'resources not available in S' in (ICT) could only be read in terms of assertions that imply, or even that are equivalent, to sentences in the language of S that are not provable in it. As we shall see later on, however, there are many sense in which a resource not available in S may fail to entail unprovable sentences in S. Indeed, the lesson that we draw from Dean's point is not that (ICT) has to be rejected given the incompatibility of the finitist's justification of PRA and RFN(PRA). On the contrary, Dean's objection points at the possibility of embracing a plausible version of (ICT) that does *not* invoke principles equivalent to or stronger than RFN(EA). Such a version of (ICT) will be articulated in the following sections.

Dean draws a similar conclusion in relation to Isaacson's thesis, according to PA captures "an intrinsic, conceptually well-defined region of arithmetical truth" [25, p. 203]. Indeed, Isaacson suggests that PA may be regarded as sound and complete with respect to a conception of arithmetical truths as "directly perceivable" by articulating "our grasp of the structure of the natural numbers" [25, p.217] [26]. Unprovable truths in PA such as Goodstein theorem and the Paris-Harrington sentence are ones that involve hidden *higher-order* (or infinitary) concepts.[6] If these claims have a clear mathematical meaning, however, it is also well-known that they are equivalent, over PA, to claims of apparent *meta*-mathematical meaning such as the Gödel sentence for PA or a canonical consistency statement Con(PA).

A similar correspondence between the mathematical and the meta-mathematical can be found at the level of the principles which are usually involved in strong readings of (ICT) such as Feferman's. Let's consider again RFN(PA). It is a classical result by Gentzen the claim that PA proves transfinite induction up to any ordinal smaller than ε_0 (henceforth TI_{ω_n}) – i.e. up to the limit of all ordinals of the form $\omega^{\omega^{\cdot^{\cdot^{\cdot^{\omega}}}}}$ for towers of order n [18]. Hence, by the properties of provability, PA proves the formalization of this fact for all n. By RFN(PA), therefore, one can conclude, within PA+RFN(PA), the claim that for all n, TI_{ω_n}, that is the schema of transfinite induction up to ε_0 ($\mathsf{TI}_{\varepsilon_0}$). Also the other direction — that is the claim that PA + RFN(PA) proves $\mathsf{TI}_{\varepsilon_0}$ — is well-known, although the proof, which can be found in [32], is definitely more involved.

As a consequence, a principle that is naturally justified by appealing to semantic or meta-mathematical considerations such as RFN(PA) on the one hand, and a principle concerning how many countable transfinite ordinals can be well-ordered on the other, are equivalent over

[6] Note that Isaacson characterization of arithmetical truth seems to entail that sentences like the Goldbach conjecture are un-arithmetical, being neither directly perceivable by grasping the structure of natural number, nor perceivable from some arithmetical truth [1]. Against the claim that a proof of *any* true PA sentence which is independent of PA needs ideas that go beyond those that are required in understanding PA, see [42].

PA. Therefore, if Isaacson's thesis on PA is to be understood in a radical way as to entail that anything that is unprovable in PA should *not* be part of the principles allowed by (ICT), both RFN(PA) and $\mathsf{TI}_{\varepsilon_0}$ should be ruled out. In a less categorical reading of Isaacson's thesis, one may still think that principles that are not provable in PA may be allowed in the set specified by (ICT); however, as stressed by Dean himself, these truths should now assume the instrumental role of confirming the theorems of PA as clear boundaries for finite mathematics (see [25, §3]).[7] However, it is not clear to us in which sense this more liberal reading of (ICT) should differ from the radical one, since the inclusion of these additional arithmetical truths in the set specified by (ICT) only reaffirms and does not characterize PA as a self-standing portion of mathematical truth. The message that Dean extracts from Isaacson's thesis looks, again, uncontroversial: if one endorses it, she is also committed to a reading of (ICT) that eschews claims that are unprovable in PA, casting serious doubts on the plausibility of (ICT) itself. And again, we will see in the next section that there are senses in which 'resources not available in PA' we might be implicitly committed to when accepting PA may fail to imply statements that are unprovable in PA itself.

Apart from the cases examined by Dean, similar questions arise in analogous foundational theses that rely on a restriction of the full induction schema of PA. For instance, one might look at the *ultrafinist* thesis advocated by Edward Nelson in [37], and echoed in several commentator's works, according to which one should mistrust the assumption of the totality of exponentiation.[8] A theory that fully meets Nelson's standards is the theory S^1_2 from [3, 4]. S^1_2 has several further advantages: In addition to being consistent with the negation of exponentiation, S^1_2 is also remarkable from a purely proof-theoretic point of view: it can be seen as a minimal theory for formalizing in a natural way the syntax of first-order theories as it is commonly done for the incompleteness theorems and as it is required for formulating reflection principles and semantic extensions of our starting theories. These notions are in fact all p-time and the functions Σ_1-definable in S^1_2 coincide with the p-time computable functions. S^1_2 is formulated in $\mathcal{L}^* = \mathcal{L} \cup \{0, \mathsf{S}, +, \times, |\cdot|, \#, \lfloor \frac{1}{2}\cdot \rfloor\}$, where $|\cdot|$ is the length function that gives the number of symbols in the binary representation of the input, $\#$ is such that $x\#y = 2^{|x|\times|y|}$ and $\lfloor \frac{1}{2}\cdot \rfloor$ gives the lower integer part of $\frac{x}{2}$. Its axioms are the defining equation of these symbols

[7] It should be noticed that we haven't made any reference to the notions of 'higher-order' or 'infinitary' in this description, and this is not by accident: it is not completely clear to us, indeed, where the boundary between finitary and infinitary should lie in the case of PA. Isaacson seems to think that such a boundary coincides with the distinction between what can be proved or not in PA: but can there be a sense in which 'higher-order' or 'infinitary' notions are not at odds with PA? To cite one simple example, consider well-orderings of order type $\alpha < \varepsilon_0$, that can be proved in PA by a well-known theorem of Gentzen; other examples that come to mind are versions of semantical reflection that, unlike RFN(PA), are conservative over PA and therefore do not lead us outside of the realm of what is acceptable by the 'first-orderist'. The next section will present and discuss examples of semantical reflection of this sort.

[8] In particular, Nelson sketches in [37, Ch 31] a foundational program under the assumption of the negation of the totality of exponentiation. Admittedly, much less clear are the reasons why Nelson advocates such position. Besides his clear nominalistic stance (cf. [37, Ch. 18], Nelson's position can be taken to hold that

> …the basic informal argument says, roughly, that the number of steps needed to terminate a recursion defining exponentiation is of the order of magnitude of exponentiation itself – a perceived circularity. [14, p. 2]

and the schema

$$\text{(PIND)} \qquad \varphi(0) \wedge \forall x\,(\varphi(\lfloor \tfrac{1}{2}x \rfloor) \to \varphi(x)) \to \forall x\,\varphi(x)$$

for φ in the class Σ_1^b, which is similar to the usual class Σ_1 formulas with the additional assumption that quantifiers in the formula have to be bounded by a term of the form $|t|$, except the outermost string of existential quantifiers that can be bounded by an arbitrary term. Crucially, S_2^1 is interpretable in Robinson arithmetic Q, witnessing its minimality, and is finitely axiomatizable [21, Ch. V].

Assuming therefore that S_2^1 is *ultrafinitistically*-acceptable, let us consider a reflection principle of the form

$$\mathsf{RFN}(\varnothing) \qquad \forall x\,(\mathsf{Pr}_{\varnothing}(\ulcorner \varphi(\dot{x}) \urcorner) \to \varphi(x))$$

where $\mathsf{Pr}_{\varnothing}(\ulcorner \varphi(\dot{x}) \urcorner)$ expresses the fact that an arbitrary numeral instance of the formula φ is provable in first-order predicate logic. However, even under these minimal assumptions, we obtain a result similar to Proposition 1.

Proposition 2. PA *is a subtheory of* S_2^1+RFN*(∅)*

Proof Sketch. It is clear that, for each $m \in \omega$,

$$(1) \qquad \varphi(\bar{0}) \wedge \forall x(\varphi(x) \to \varphi(x+1)) \to \varphi(\bar{m})$$

is provable in first-order logic by a series of modus ponens and universal instantiations starting from $\varphi(\bar{0})$. This proof, however, may not be captured in general by S_2^1. Therefore we argue as follows: assuming that φ is provably progressive in S_2^1 — that is, S_2^1 proves that it holds for 0 and that, if it holds for x, it holds for $x+1$ as well —, by employing Solovay's shortening of cuts technique (cf. again [21, Ch. V]), we downwards close φ under $\leq$ so that the resulting formula defines an initial segment of the S_2^1-numbers $\mathcal{J}$. We can safely assume $\mathcal{J}$ to be closed under multiplication and the function #.

Then we claim that

$$(2) \qquad \mathsf{S}_2^1 \vdash \forall \mathsf{x}\, \mathsf{Pr}_{\mathsf{S}_2^1}\ulcorner \mathcal{J}(\dot{\mathsf{x}}) \urcorner$$

(2) is proved by considering dyadic numerals

$$\overline{2 \times n} = (\mathsf{SS0}) \times \overline{n} \qquad\qquad \overline{2 \times n + 1} = \mathsf{S}((\mathsf{SS0}) \times \overline{n})$$

The codes of the numeral n, in this way, is of order n^c for a fixed c – therefore can be handled with # – and not 2^{cn} for fixed c, which would require exponentiation. Now reasoning in S_2^1 and starting with the proof of $\mathcal{J}(0)$, we can reason as usual to obtain a proof of $\mathcal{J}(n)$. Therefore, in S_2^1, which can be expressed as a single sentence A, plus RFN(∅),

$\forall x\, \mathsf{Pr}_{\varnothing}(\ulcorner \mathsf{A} \to \mathcal{J}(\dot{x}) \urcorner)$	
$\mathcal{J}(x)$	by A and RFN(∅)
$\mathcal{J}(x) \to \varphi(x)$	by def. of $\mathcal{J}$
$\forall x\, \varphi(x)$	logic

□

Proposition 2 strengthens the conclusion that, if one reads (ICT) as referring to 'resources not available in S' that entail claims that are not provable in S, then S cannot be taken to capture a self-standing, self-justifying portion of mathematical reality. The ultrafinitst embracing S^1_2, in fact, cannot even be committed to a reflection principle for logic, on the pain of the acceptance of the full induction schema of PA that, obviously, also entails the claim that the exponentiation function is total.

To summarize, the discussion of the theses of Tait, Isaacson, and Nelson, coupled with a strong reading of (ICT) *à la* Feferman that seems to be taken for granted by Dean, leads to at least two options: either we reject (ICT) across the board, deeming it as inadequate, or we provide a different interpretation of (ICT) equipped with an alternative reading of what the 'resources not available' in the chosen system should amount to. In the next section we set the basis for such an alternative interpretation: we will introduce in particular a wide array of semantical extensions of an arithmetical system S that, although crucially resorting to notions that are not immediately available in S — such as a truth predicate — do not entail sentences in the language of S that are not provable in S itself.

3. Soundness extensions and the semantic core

As noticed by several authors[9], resorting to schemata such as $\mathsf{RFN}(S)$ above may be plausibly seen as a surrogate for single sentences of the form

$\mathsf{GRP}(S)$ $$\forall x\,(\mathsf{Pr}_S(x) \to \mathsf{T}x)$$

where T is unary truth predicate. These surrogates only become necessary when a notion of truth is not part of the signature of the theory. Any soundness claim seems in fact to be intrinsically related to the notion of truth. If one wants to express in the object language that all non-logical axioms of S are true, for instance, one can of course resort to a schema of the form

$$\mathsf{Ax}_S(\ulcorner\varphi\urcorner) \to \varphi$$

where $\mathsf{Ax}_S(\cdot)$ is the representation of all non-logical axioms of S. Yet, this option merely highlights the fact that we are relegating the notion of truth in the meta-theory.

Clearly someone might have independent motivations to stick with the expressive limitations of the arithmetical language in asserting the soundness of a theory. Tennant in [48], for example, has made use of the well-known fact that schematic versions of reflection, such as $\mathsf{RFN}(S)$, enable us to go beyond what's provable in S to defend the possibility of a deflationary account of the notion of truth employed in these soundness claims. However, Tennant does not fully articulate a justification for these principles, although he hints at the schematic version of reflection as sufficient for fixing the norms for assertion of these soundness claims [48, p. 574]. More generally, while it is uncontroversial a soundness extension of S will contain forms of reflection such as $\mathsf{RFN}(S)$, it remains problematic whether the presence of $\mathsf{RFN}(S)$ is *sufficient* for defining a soundness extension, in the sense that its principles amount to a coherent articulation of the concepts needed to state soundness claims for S. A good illustration of how soundness claims can be derived within an adequate framework for provability and truth is provided by Feferman's reflective closure of PA ($\mathsf{Ref}(S)$), nowadays commonly known as KF from 'Kripke-Feferman'[10]:

[9] Cf. for instance, [32], [20, p. 309].

[10] See footnote 17 for a precise definition of KF.

> Which statements in the base language L of S [...] ought to be accepted if one has accepted the basic axioms and rules of S? The answer is given as an ordinary theory Ref(S) formulated in a language $L(T, F)$ [...] where T and F are partial truth and falsity predicates which are self-applicable in the sense that they apply to (codes of) statements of $L(T, F)$ [...] Thus, for example, we may reason in Ref(PA) by induction about the truth of statements which contain the notion of truth, and so arrive at statements of the form: $\forall x[Prov_{PA}(x) \to T(x)]$, and by repeating this kind of argument derive iterated reflection principles for arithmetic [12, p. 2].

Note well that we are not suggesting the impossibility of convincing arguments supporting the absence of the notion of truth from soundness extensions of a given theory; we are simply holding that given the usual way of introducing and justifying soundness claims for a theory S, the notion of truth is hard to do without. Proposals such as Tennant's, and the subsequent debate it generated [28, 5, 41], clearly show how hard it is to eradicate the intuition that reflection principles are conceptually dependent on the notion of truth. But the onus is on those who do not share this intuition to tell a principled story about soundness claims by resorting to surrogates that can play the role of semantic notions. It's hard to say what this story could amount to. Thus, throughout the paper we will stick with the widespread view and hold that soundness claims are best formulated by employing a notion of truth governed by suitable axioms.

However, this does not immediately mean that these axioms added on top of S *need to* entail GRP(S). Such a requirement, indeed, would be too strong for an arbitrary S (namely, when S also varies over, for instance, theories with restricted induction). The case made by Tait for PRA from the finitist point of view is indeed one example where one needs to be careful in calibrating the strength of the principles for the truth predicate. Similar considerations apply to Isaacson's thesis on PA and ultrafinitist's position viewed through the lens of S^1_2. Prima facie, there is not much room for the choice of the truth principles: for instance, the next proposition shows that already weak truth axioms seem to collapse the fine structure of the subsystems of PA.

Proposition 3. *The result of extending* S^1_2 *— whose language is expanded with a fresh predicate* T *— with the schema*

$$\forall x\,(\mathsf{T}\ulcorner\varphi(\dot{x})\urcorner \leftrightarrow \varphi(x)) \tag{utb}$$

for all $\mathcal{L}$*-formulas* $\varphi(v)$ *derives the full induction schema of* $\mathcal{L}$*.*

Proof. Since S^1_2 in $\mathcal{L}_\mathsf{T} := \mathcal{L} \cup \{\mathsf{T}\}$ contains $\mathsf{I}\Delta_0$ in $\mathcal{L}_\mathsf{T}$, the following is derivable in the former

$$\mathsf{T}\ulcorner\varphi(0)\urcorner \wedge \forall x\,(\mathsf{T}\ulcorner\varphi(\dot{x})\urcorner \to \mathsf{T}\ulcorner\varphi(x \dot{+} 1)\urcorner) \to \forall x\,\mathsf{T}\ulcorner\varphi(\dot{x})\urcorner \tag{3}$$

for a formula $\varphi(v)$ of $\mathcal{L}$ of arbitrary complexity, because $\mathsf{T}\ulcorner\varphi(\dot{x})\urcorner$ is a Δ_0-formula of $\mathcal{L}_\mathsf{T}$. By employing (utb), (3) yields the desired result. □

The argument employed in Proposition 3 applies equally well — with the obvious modifications — to other subsystems of PA obtained by restricting induction such as EA, PRA, of $\mathsf{I}\Sigma_n$ for

every n.[11] At any rate, Proposition 3 seems to slim our chances of finding a reasonable truth-theoretic extension of an arbitrary arithmetical theory S, that is fixing a set of reasonable truth axioms that are compatible with the principles we are implicitly committed to when we endorse S. Proposition 2 and Kreisel and Levy's result already told us that a soundness extension involving the uniform reflection principle for S may clash with the foundational standpoints — such as the ones just discussed in the previous section — advocating a restriction of the full schema of induction of PA. Proposition 3 extends these limitations to the truth-theoretic context: if even weak axioms such as (utb) are sufficient to lead us from, S^1_2, EA, or PRA to full PA, then there seems to be no hope to harmonize (ICT) and foundational positions that do not permit arithmetical consequences exceeding those of the systems associated with such positions.

Nevertheless, concluding this would simply be trading on a confusion on the meaning of 'truth axiom'. The theory of truth employed in Proposition 3 is obtained by extending the *mathematical* induction schemata of the base theory to the truth predicate. If the axioms (utb) are unequivocally truth-theoretic in character, it is natural to think of the extended induction as a *mathematical* and not as a truth-theoretic axiom. There seems to be in fact a substantial difference between metalinguistic principles declaring the truth conditions for a sentence of $\mathcal{L}$, as (utb) seems to be (partially) doing, and the extension to the truth predicate of a schema whose justification is apparently non-metalinguistic. As observed by Hartry Field, such a justification essentially depends on a 'fact about natural numbers, namely, that they are linearly ordered with each element having finitely many predecessors'[15, p. 538].

For example, the formula $\mathsf{T}\ulcorner 0 = 0\urcorner \wedge 2^x > x$ can occur into instances of the induction schema of EA formulated in $\mathcal{L}_\mathsf{T}$ (:= $\mathcal{L} \cup \{2^{\cdot}, \mathsf{T}\}$); however, it would be rather implausible to consider the corresponding instance of induction as a genuinely truth-theoretic sentence. By contrast, the truth predicate in it is merely idling and the bulk of the induction is instead a basic mathematical property of the exponential function. On the contrary, the induction instance corresponding to the $\mathcal{L}_\mathsf{T}$-formula $\mathsf{T}\ulcorner 2^{\dot{x}} > \dot{x}\urcorner$ is expressing a metalinguistic fact, namely that all substitutional instances of the formula $2^x > x$ are true. The shift in meaning between the two properties is subtle but crucial: in one case we talk about properties of a mathematical function, in the second one about formulas of $\mathcal{L}$.

Let's be clear about this point to avoid further confusion: from the internal point of view of the theory of truth, the two instances of induction corresponding to the formulas $\mathsf{T}\ulcorner 0 = 0\urcorner \wedge 2^x > x$ and $\mathsf{T}\ulcorner 2^{\dot{x}} > \dot{x}\urcorner$ are, strictly speaking, indistinguishable. However, from the external point of view of our informal metamathematical practice, they are clearly distinct. It is only because arithmetic plays a double role of theory of syntax and of object theory, that we can consider both instances as belonging to essentially the same class. This observation even led to the formulation of theories of truth that keep separate the domain of syntactic objects from the mathematical or, more generally, the object theoretic universe (see [20, 22, 38]). It's not our intention here to consider the details of this alternative framework: we will keep implicit the distinction between metalinguistic and object-linguistic instances of the induction schema. However, in what follows we *will not* extend the induction schema of S to the truth predicate to avoid any conflation between the two levels.

[11] A similar argument would even hold in the case of set theories formulated by syntactically restricting schemata.

Similar considerations can obviously be applied to the case of logical axioms: if the language of S is expanded with a truth predicate allowed to appear into logical axiom schemata, then, for example, $\mathsf{T}x \vee \neg \mathsf{T}x$ would be an instance of a logical axiom, and not a truth-theoretic axiom. No one would deny that such extension of the logical axiom schemata is a natural move: however, as we shall shortly see, the behaviour of logical axiom schemata will differ considerably from the one of mathematical induction. If in fact the extended induction schema in combination with natural truth axioms would lead us to very strong theories, the assumption of the truth of all its instances is fairly innocent: as we shall see shortly, in fact, the result of adding to a wide class of base theories S the claim 'all instances of the induction schema of S are true' is still compatible with the alternative reading of (ICT) that we suggested in the previous section and that is aimed at harmonizing (ICT) with foundational standpoints such as Tait's, Isaacson's and Nelson's. By contrast, if the extension of logical axiom schemata to the truth predicate is not only innocuous for our purposes but also desirable as a such, the formal claim of the truth of all logical axioms of S would yield a theory that is substantially stronger than S (see §3.2).

3.1. **The semantic core.** We have seen that a strong reading of (ICT) may conflict with foundational standpoints based on a form of 'arithmetical completeness' or 'epistemic stability' of some arithmetical system S. In fact, if (ICT) entails reflection principles for S and therefore claims in the arithmetical language that are not provable in S alone, then in accepting S one is also bound to accept arithmetical consequences that go beyond S, thus contradicting its alleged completeness.

In concluding §2, we envisaged the possibility of an alternative reading of (ICT) that could be immune from this problem. But how could this alternative reading look like? A hasty thought may be to let (ICT) depend exclusively on one's foundational standpoint. This is highly problematic. Let's consider, for example, someone who embraces only what's derivable or interpretable in PA: by a well-known result of Feferman, she will also accept $\neg\mathsf{Con}(\mathsf{PA})$.[12] By contrast, we have seen that there are several authors disposed to accept $\mathsf{Con}(\mathsf{PA})$ after accepting PA. Under this relativistic view of (ICT), therefore, different readings of it would not only lead to alternative sets of principles, but rather to sets of principles inconsistent with each other. In the specific case of $\neg\mathsf{Con}(\mathsf{PA})$ just mentioned, moreover, there is a clear departure from what we previously defended as a necessary condition for any plausible reading of (ICT), namely the *truth* of the principles at play. The interpretation of (ICT) that we now introduce will keep a strong link with the notion of truth, while rejecting the sort of rigidity detected in Feferman's reading of (ICT). Our approach substantiates a dynamic reading of (ICT) as displaying a fixed, semantic component — called the *semantic core of implicit commitment* — and a variable component that is relative to one's foundational standpoint.

The semantic core amounts to a set of principles of meta-theoretic nature that enable us to reflect in a natural and uniform way on our acceptance of different arithmetical theories. To introduce it, we argue in stages. In the first step, we need to expand the language of S with semantic resources, a truth predicate T in particular, and characterize it with a minimal set of principles capturing its disquotational nature. More precisely, given a suitable S, the theory $\mathsf{TB}[S]$ is obtained by expanding $\mathcal{L}_S$ with the predicate T and extending its axioms with the schema

$$\mathsf{T}\ulcorner\varphi\urcorner \leftrightarrow \varphi \tag{tb}$$

[12] See [9].

for all $\mathcal{L}_S$-sentences φ. An immediate consequence of (tb) is the truth of *each* axiom of S; it is clear therefore that if S has finitely many non-logical axioms, (tb) suffices to conclude $\forall x\,(\mathsf{Ax}_S(x) \to \mathsf{T}x)$, that is the single sentence expressing the truth of all (non-logical) axioms of S. Further claims of clear metalinguistic nature are also provable in (tb). For instance, $\mathsf{TB}[S]$ proves the claim that the global reflection principle for S entails the consistency of S. Formally:

$$\forall x\,(\mathsf{Pr}_S(x) \to \mathsf{T}x) \to \mathsf{Con}(S) \tag{4}$$

This implication is simply obtained by instantiating $\bot$ in $\mathsf{GRP}(S)$.

Already in this first step, it should be clear that we aim at *semantic extensions* of S in the sense of coherent articulations of a concept of truth over the base theory S. For instance, one could simply extend S with the sentences $\forall x\,(\mathsf{Ax}_S(x) \to \mathsf{T}x)$ or $\mathsf{GRP}(S)$ as new axioms. The sentences above clearly do not suffice to count as *axioms* for the truth predicate T: in the first case the the resulting theory is clearly interpretable in S by taking the truth predicate in question to be *defined* by $\mathsf{Ax}_S(x)$ itself; in the second case, the full schema (tb) is not necessary to derive (4), as the 'modal' axiom $\mathsf{T}\ulcorner\phi\urcorner \to \phi$ suffices. This suggests that, in these extensions of S, concepts other than truth could be employed as natural readings for the predicate T.

From the perspective of the *theorems* of S, $\mathsf{TB}[S]$ looks fairly innocent. First of all, it is conservative over S. Moreover, if S is reflexive, it is also relatively interpretable in it. This is because in any given proof in S the truth predicate T can be replaced by a S-definable truth predicate. This suffices to witness the conservativity and, by Orey's compactness theorem (see [35, §7]), the interpretability of $\mathsf{TB}[S]$ in S for reflexive S.

The disquotational principles (tb), however, fall short of many further desiderata that we would like to ascribe to the semantic core of implicit commitment. For instance the schema (tb) *cannot* enable us to establish that instances of modus ponens preserve truth because every generalization crucially involving truth provable in $\mathsf{TB}[S]$ can be reduced to a finite conjunction. This means, in particular, that $\mathsf{TB}[\mathsf{PA}]$ can only prove the weaker

$$\forall x, y\,(\mathsf{Sent}^n_{\mathcal{L}}(x) \wedge \mathsf{Sent}^n_{\mathcal{L}}(y) \wedge \mathsf{T}(x \dot{\to} y) \wedge \mathsf{T}x \to \mathsf{T}y) \tag{5}$$

where $\mathsf{Sent}^n_{\mathcal{L}}(x)$ expresses that x is a sentence of $\mathcal{L}$ of complexity $\leq n$ for any given n but not for arbitrary sentences of $\mathcal{L}$ and the expression $\dot{\to}$ (and $\dot{f}$ more generally) represents in S the *syntactic* operation of entailment (resp. f).[13]

One might think of extending $\mathsf{TB}[S]$ with further truth-theoretic principles so as to derive the non-restricted versions of (5). Obvious candidates are the so-called *compositional* truth axioms such as '$\neg\varphi$ is true if and only if φ is not true', which govern the interaction of the truth predicate and the logical constants. For instance, since we might safely assume that S is formulated in a calculus in which modus ponens is the only logical rule of inference (see, for instance, [8]), we would only need to add to S the sentence

$$\forall x, y\,(\mathsf{Sent}_{\mathcal{L}}(x) \wedge \mathsf{Sent}_{\mathcal{L}}(y) \wedge \mathsf{T}(x \dot{\to} y) \wedge \mathsf{T}x \to \mathsf{T}y) \tag{6}$$

to derive the truth-preserving character of modus ponens.

If S is finitely axiomatizable, therefore, $\mathsf{TB}(S)$+(6) enables us to prove that all non-logical axioms of S are true and that — if the logic is rightly chosen — that all rules of inferences of S preserve truth. However, there are at least two problems with this theory: in the first

[13] Here the complexity of a formula can simply be taken as the number of logical symbols in it.

place, it does not articulate a coherent semantic notion as we usually demand that the truth of a compound sentence depends on the truth of its compounds, and this theory has no such feature. In short, the theory is not compositional. Secondly, if S is not finitely axiomatizable, it cannot prove that all non-logical axioms of S are true. In fact, as the next lemma shows, it cannot do so even if we add to S a fully compositional theory of truth:

Lemma 1. *Let S be formulated in $\mathcal{L}_{\mathsf{T}}$ and assume it satisfies full induction for $\mathcal{L}_S$ – that is the truth predicate is not allowed into instance of induction. This theory extended with the sentences*

$$\mathsf{Cterm}_{\mathcal{L}_S}(x_1) \wedge \ldots \wedge \mathsf{Cterm}_{\mathcal{L}_S}(x_n) \rightarrow \left(\mathsf{T}^{\ulcorner} R(\dot{x}_1, \ldots, \dot{x}_n)^{\urcorner} \leftrightarrow R(x_1, \ldots, x_n)\right) \tag{7}$$

$$\mathsf{Sent}_{\mathcal{L}_S}(x) \rightarrow \left(\mathsf{T}(\dot{\neg} x) \leftrightarrow \neg \mathsf{T}x\right) \tag{8}$$

$$\mathsf{Sent}_{\mathcal{L}_S}(x \dot{\rightarrow} y) \rightarrow \left(\mathsf{T}(x \dot{\rightarrow} y) \leftrightarrow (\mathsf{T}x \rightarrow \mathsf{T}y)\right) \tag{9}$$

$$\mathsf{Sent}_{\mathcal{L}_S}(\dot{\forall} vx) \rightarrow \left(\mathsf{T}(\dot{\forall} vx) \leftrightarrow \forall y\, \mathsf{T}x(\dot{y}/v)\right) \tag{10}$$

cannot prove that all axioms of S are true.

In (7), R ranges over the relation symbols of $\mathcal{L}_S$.

Proof. Assume that S+(7)-(10) proves

$$\forall x\, (\mathsf{Ax}_S(x) \rightarrow \mathsf{T}x), \tag{11}$$

We can then show that the formula

$$\mathcal{K}(x) :\leftrightarrow (\forall y \leq x)\, (\mathsf{Prv}_S(y) \rightarrow \mathsf{Tend}(y)) \tag{12}$$

is progressive in it. In (12), Prv_S is a Δ_1^b predicate expressing the notion of being a proof in S and $\mathsf{end}(\cdot)$ is a Σ_1^b-function that outputs the last element of of a S-proof. Therefore, still by Solovay's result on subcuts (see Proposition 2), we find an initial segment of the S-numbers satisfying the property expressed by $\mathcal{K}(x)$ in which all logical axioms of S are true and then prove the consistency of S relative to this initial segment.[14] By a strengthening of Gödel's second incompleteness theorem due to Pudlák ([44, Cor. 3.5]), therefore, this is sufficient to show that S cannot interpret S+(7)-(10). However, S+(7)-(10) is known to be interpretable in S (see [7, §16.5]). □

The full compositional clauses (7)-(10) are without a doubt desirable features for a notion of truth. Moreover, this notion of truth is a natural component of the acceptance of S via soundness claims, and soundness claims are, in turn, an integral part of many accounts of implicit commitment. As we have seen, however, there are also limitations to which soundness claims one can assume, depending on one's foundational stance. We have considered already examples of such limitations: for example the ones related to the reflection principles RFN(EA) or RFN(PA) — and, *a fortiori*, their global versions — for positions such as finitism or first-orderism *à la* Isaacson. Nonetheless, as we shall soon point out, no such limitations occur for the compositional truth clauses. What is even more surprising is that we can allow explicit soundness claims relative to the non-logical axioms of an arbitrary theory S without trespassing into the realm of what's unprovable in S. This can be established in full generality.

Let $\mathsf{CT}[S]$ be the result of expanding the theory S with a truth predicate *not* allowed into instances of the non-logical axiom schemata, and of adding to it the principles (7)-(11).

[14] For details concerning this strategy, see [39].

Halbach in [20] attempts to prove the conservativity of $\mathsf{CT}[S] \setminus (11)$ via a cut elimination argument. His argument relies on a reformulation of $\mathsf{CT}[S]$ in a (finitary) two-sided sequent calculus with cut by rewriting (7)-(10) as rules of inference, e.g.

$$\frac{\Gamma, \mathsf{T}s \Rightarrow \Delta}{\Gamma, \mathsf{Sent}_{\mathcal{L}}(s) \Rightarrow \Delta, \mathsf{T}(\neg s)}\ (\neg\text{-R})$$

and then proceeds via an attempt to eliminate cuts on formulas of the form $\mathsf{T}s$ from derivations in this theory. Leigh in [34] shows that this strategy can only remove cuts of a provably fixed complexity (cf. [34, §3.7]). He then shows how to fix Halbach's strategy by finding suitable bounds to the complexity $\mathsf{c}(\cdot)$ of truth-cut-formulas in $\mathsf{CT}[S]$-derivations – for S interpreting EA – so that $\mathsf{CT}[S]$ can be embedded in the system resulting from replacing the full cut rule for formulas of the form $\mathsf{T}s$ with a weaker set of rules

(Cut_n)

$$\frac{\begin{array}{ccc} & & \text{Truth-free proof} \\ & & \vdots \\ \Gamma, \mathsf{T}s \Rightarrow \Delta & \Gamma \Rightarrow \Delta, \mathsf{T}s & \Gamma, \mathsf{Sent}_{\mathcal{L}}(s) \Rightarrow \mathsf{c}(s) \leq n \end{array}}{\Gamma \Rightarrow \Delta}$$

for each n and a suitably bounded version of (11). Crucially, this system enjoys a standard version of cut-elimination for cuts on truth ascriptions. Derivations of truth-free sequents of the form $\Gamma \Rightarrow \Delta$ are then regimented via the notion of *approximation* of a sequent, first considered by Kotlatski, Krajewski, and Lachlan in [29], that enables one to control such proofs in $\mathsf{CT}[S]$ and transform them into proofs of the same sequent where only applications of the modified rules are employed. Finally, one eliminates cuts on formulas of the form $\mathsf{T}s$ in a standard manner. This strategy yields the following:

Proposition 4 ([34, Thm. 2])**.** *For $S \supseteq \mathsf{EA}$, $\mathsf{CT}[S]$ is a conservative extension of S.*

Proposition 4 tells us that the semantic principles of the theory $\mathsf{CT}[S]$ can safely be included into the semantic core of the implicit commitment for S. Our main thesis is now taking shape: in accepting an arithmetical theory S, we are *always* implicitly committed to the theory $\mathsf{CT}[S]$, which amounts to the fixed, invariable component of our commitment. Whether or not $\mathsf{CT}[S]$ *exhausts* our commitments depends on the particular foundational standpoint that led us to accept a given theory S in the first place.

3.2. **Logical axioms.** The theory $\mathsf{CT}[S]$ includes the assertion of the truth of non-logical axioms of S. Hence a natural question is whether extending $\mathsf{CT}[S]$ with the claim of the truth of all *logical* axioms of S is compatible with our project.

The answer to this question is negative: $\mathsf{CT}[S]$+'all logical axioms of S are true' proves the consistency of S. Such an extension of the semantic core would thus be too strong for authors that regard S as epistemically stable. The argument is based on the formalization of the usual inductive soundness argument for S in which from the truth of all axioms of S and the truth-preserving character of the rules of inference of S one concludes the truth of all theorems of S: however, since we do not have induction with the truth predicate, one needs to find alternative means to run the argument. It turns out that the following principle suffices:

$\mathsf{I}\text{-}\mathsf{Ind}(\Gamma)$

$$\forall x\big(\mathsf{Fml}^1_{\Gamma}(x) \rightarrow \big(\mathsf{T}x(\ulcorner 0 \urcorner) \wedge \forall y\,(\mathsf{T}x(\dot{y}) \rightarrow \mathsf{T}x(\dot{y}+1)) \rightarrow \forall y\,\mathsf{T}x(\dot{y})\big)\big)$$

where Γ stands for a level Σ_n of the arithmetical hierarchy. If S is PA, Γ is simply $\mathcal{L}$. If, say, S is $\mathsf{I}\Sigma_1$, then Γ will be Σ_1. We have the following:

Proposition 5. *Let's assume S has Γ-induction. Then* $\mathsf{CT}[S]$ *proves* $\mathsf{I}\text{-}\mathsf{Ind}(\Gamma)$.

Proof sketch. Reasoning in $\mathsf{CT}[S]$, let's consider a number y such that $\mathsf{Fml}^1_\Gamma(y)$. Then the expression

$$\mathsf{sub}(y, \ulcorner 0 \urcorner) \mathbin{\dot\wedge} \dot\forall v\, (\mathsf{sub}(y, \mathsf{num}(v)) \mathbin{\dot\to} \mathsf{sub}(y, \mathsf{num}(\dot S v))) \mathbin{\dot\to} \dot\forall v\, \mathsf{sub}(y, \mathsf{num}(v))$$

would be a formal instance of the induction schema of S and therefore true by $\mathsf{CT}[S]$. By distributing the truth predicate, we obtain the desired conclusion. □

The argument for showing that $\mathsf{CT}[S]$+'all logical axioms of S are true' proves the consistency of S is now standard. By assumption, all axioms (logical and non-logical) of S are true. $\mathsf{I}\text{-}\mathsf{Ind}(\Gamma)$ suffices to show that all rules of inference preserve truth and to conclude that all theorems of S are true: one simply runs the usual soundness proof inside the scope of the truth predicate.

$\mathsf{CT}[S]$ alone was claimed to be able to prove the consistency of S by many authors, including [16]. Nevertheless, Albert Visser and Richard Heck found a gap in these arguments: essentially, the internal induction schema given by the truth of all instances of S-induction does not suffice to derive the truth of all logical axioms of S.[15] Proposition 4 gives us indirect confirmation of the status of these claims: $\mathsf{CT}[S]$ is indeed a conservative extension of S.

In what follows, therefore, we will not include the assertion of the truth of the logical axioms of S in the semantic core: despite its seemingly innocuous nature, in fact, its presence leads to a proper strengthening of the semantic core. One might of course prefer the truth of the logical axioms over the truth of the non-logical axioms of S and modify the definition of $\mathsf{CT}[S]$ accordingly. The proof of Proposition 4 may also be adapted to show the conservativeness of this new version of $\mathsf{CT}[S]$ over S. However, our project is mainly concerned with the mathematical content of implicit commitment and, since we are forced by the formal result to make a choice, we stick to the definition of the semantic core introduced in the previous section.

4. Schematic reasoning and the structure of implicit commitment

Several foundational standpoints, including the ones considered above, can be compared and distinguished by taking into account the role of the schemata of induction of the arithmetical systems associated to them. In this section we employ these different understandings of schematic reasoning to assess the effectiveness of our analysis of implicit commitment based on the distinction between the constant semantic core and its variable components.

At one end of the spectrum, we find advocates of restrictions of the arithmetical induction schema. Tait's finitism and Nelson's ultrafinitism are paradigmatic examples of this sort: in both cases claims about the totality of natural numbers can only be reached for a class of 'acceptable' predicates that are proper subclasses of the ones expressible by formulas of the language of arithmetic. The remaining instances of the induction schemata are, according to these standpoints, at least suspicious if not false. At the other end of the spectrum, we find

[15] The problem is essentially related to the axioms or rules of inference for quantifiers: to be able to show that all instances of the axiom, say, $\forall \vec{x} \varphi(\vec{x}) \to \varphi(\vec{t})$ are true, one needs to go from $\forall \vec{x}\, \mathsf{T}^\ulcorner \varphi(\dot{\vec{x}}^\urcorner)$ to $\mathsf{T}^\ulcorner \forall \vec{v} \varphi^\urcorner$. Such commutativity conditions involve a possibly nonstandard string of variables and require more than internal induction to be dealt with. This fact was personally communicated to the authors.

authors defending the view that, once accepting a system S, not only we should impose no restriction to non-logical axiom schemata, but we should also allow for extensions of these schemata to possible expansions of the starting language.

This latter view can be understood of course in different senses. On a radical reading, similar to what Vann McGee suggested in [36], the acceptance of, say, PA, should commit us to instances of induction corresponding to any subset of natural numbers. This possibility is supposed to be rooted in how mathematical language itself is learned and communicated.[16] This radical form of open-endedness of axiom schemata leads quickly to very strong theories, in fact, categorical ones. Critics of this position notice in fact that — despite McGee's efforts — it is also committed to the rich ontology of second-order logic (see [40]).

Feferman's notions of reflective closure of a theory S (see [12]) represent a less radical alternative. It comes in two versions: the reflective closure of S and the *schematic* reflective closure of S. In both cases, the interaction of semantic resources and the power of the induction of PA enable one to reach strong subsystems of second-order arithmetic. In the case of the reflective closure of PA one reaches the strength of ramified analysis up to ε_0 via the theory of self-applicable truth KF, whereas the schematic reflective closure of PA takes the form of a type-free theory of truth as strong as ramified analysis up to the Feferman-Schütte ordinal Γ_0 (i.e, roughly speaking, the theory resulting form iterating predicative comprehension α-times for $\alpha < \Gamma_0$) [11, 46].[17] Feferman's approach therefore, although clearly committed to schematic reasoning, is clearly weaker than McGee's, since it only delivers a proper subsystem of second-order arithmetic.

Among the authors that hold an intermediate position between the ones just considered we find Isaacson himself. He does not seem to impose any restriction to the class of formulas allowed to appear into instances of induction; however, he also clearly states that any further instance of induction involving extra-vocabulary would be intrinsically higher-order, inasmuch as the axioms of full PA suffice to characterize what he calls 'finite mathematics' ([25, p. 204]).

[16] As McGee writes:

> Our understanding of the language of arithmetic is such that we anticipate that the Induction Axiom Schema, like the laws of logic, will persist through all such changes. There is no single set of first-order axioms that fully expresses what we learn about the meaning of arithmetical notation when we learn the Induction Axiom Schema, since we are always capable of generating new Induction Axioms by expanding the language [36, p. 58].

[17] More precisely, such a theory amounts to an extension of the type-free theory of truth KF in $\mathcal{L}_{\mathsf{T}} \cup \{P\}$ equipped with a schematic rule of substitution $\psi(P)/\psi(\chi)$, with $\varphi(P)$ not containing truth, that replaces every subformula P of $\psi(P)$ with by χ. The axioms of KF are the axioms of PA formulated in $\mathcal{L} \cup \{\mathsf{T}\}$ and the sentences

(13) $\mathsf{Cterm}_{\mathcal{L}_{\mathsf{T}}}(\vec{x}) \rightarrow \big((\mathsf{T}^{\ulcorner} R(\dot{\vec{x}})^{\urcorner} \leftrightarrow R(\vec{x})) \wedge (\mathsf{T}^{\ulcorner} \neg R(\dot{\vec{x}})^{\urcorner} \leftrightarrow \neg R(\vec{x}))\big)$

(14) $(\mathsf{T}^{\ulcorner} \mathsf{T}(\dot{x})^{\urcorner} \leftrightarrow \mathsf{T}x) \wedge (\mathsf{T}^{\ulcorner} \neg \mathsf{T}(\dot{x})^{\urcorner} \leftrightarrow \mathsf{T}\dot{\neg} x)$

(15) $\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x) \rightarrow (\mathsf{T}\dot{\neg}\dot{\neg} x \leftrightarrow \mathsf{T}x)$

(16) $\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x \dot{\wedge} y) \rightarrow (\mathsf{T}(x \dot{\wedge} y) \leftrightarrow \mathsf{T}x \wedge \mathsf{T}y)$

(17) $\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x \dot{\wedge} y) \rightarrow (\mathsf{T}\dot{\neg}(x \dot{\wedge} y) \leftrightarrow \mathsf{T}\dot{\neg} x \vee \mathsf{T}\dot{\neg} y)$

(18) $\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(\dot{\forall} v x) \rightarrow (\mathsf{T}\dot{\forall} v x \leftrightarrow \forall y\, (\mathsf{Cterm}_{\mathcal{L}_{\mathsf{T}}}(y) \rightarrow \mathsf{T}x(y/v))$

(19) $\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(\dot{\forall} v x) \rightarrow (\mathsf{T}\dot{\neg}\dot{\forall} v x \leftrightarrow \exists y\, (\mathsf{Cterm}_{\mathcal{L}_{\mathsf{T}}}(y) \rightarrow \mathsf{T}\dot{\neg} x(y/v))$

In the next subsections, we will consider how these different positions in the spectrum interact with our account of (ICT).

4.1. **Restricted schemata and (ICT).** In §2, we have defended the claim that the notion of truth is integral to any reasonable articulation of what we are implicitly committed to when accepting a given arithmetical theory. Of course this comes as no surprise and, as we have seen, our view is shared by many authors. Our intention, however, is not to reformulate a widespread position on the role of truth in foundations, but to suggest something further. What concern us, indeed, is to examine how the notion of truth, as a device to unravel our commitments, can coexist with narrow readings of the implicit commitment thesis (ICT), namely readings which do not allow for claims that are underivable in the accepted arithmetical theory, above all uniform reflection principles.

The case studies of narrow readings of (ICT) stem Dean's analysis of Tait's and Isaacsons' theses. For instance, in the case of Tait's finitism, the uniform reflection principle for the subtheory EA of PRA was sufficient to deliver the full schema of induction of PA (see Proposition 1). If the finitist's reading of (ICT) involved principles such as RFN(PRA), she would also be committed to PA, which clearly outstrips primitive recursive reasoning. There is, therefore, a strong temptation for concluding that (ICT) is incompatible with finitism or, even more drastically, that it is false.[18] This temptation, we argue, should be resisted. *The semantic core for implicit commitment introduced in §3 gives us a way to accommodate the strong intuition that, even for the finitist's defence of* PRA*, (*ICT*) is best spelled out in terms of truth; the semantic core also tells us, however, that these additional resources, being clearly of meta-theoretic and not of object-theoretic nature, do not interfere with the arithmetical content of* PRA *that the finitist wants to preserve.*

Over PRA, which is known to be not finitely axiomatizable, the semantic core does not only involve compositional truth axioms of the form (7)-(10) on page 13, but also the single sentence stating the truth of all the infinitely many non-logical axioms of PRA. By Proposition 4, the resulting theory CT[PRA] does not yield new arithmetical consequences. *Nonetheless, it is capable of deeming true the equations for all primitive recursive functions and all instances of the induction of* PRA*, all instances of each individual propositional tautology of $\mathcal{L}_{\mathsf{T}}$, and establishing that the rules of inference of the chosen logical calculus are truth-preserving.* The first and last fact follow respectively from the assumption (11) and the axiom (9). The truth of all instances of each propositional tautology follows from a suitable instance of a logical axiom schema of CT[S] and the axioms (7)-(10): for instance, in the case of the law of excluded middle, one starts with $\mathsf{T}x \vee \neg\mathsf{T}x$ for $\mathsf{Sent}_{\mathcal{L}_S}(x)$ and concludes, by (8) and (9), $\forall x(\mathsf{Sent}_{\mathcal{L}_S}(x) \rightarrow \mathsf{T}(x \dot{\vee} \dot{\neg} x))$.

The bearing of this fact should now be clear: we have already argued that truth provides a powerful and natural tool to express one's commitment to a base theory, PRA in the case at hand. Dean cast doubts on the possibility of harmonizing a satisfactory notion of truth and the exclusive commitment to theorems of PRA that appears to be essential to Tait's standpoint. The semantic core offers a minimal sense in which this balancing process can actually succeed; we do have a notion of truth satisfying some adequacy requirements, such as the partial metalinguistic reflection available in CT[PRA] just considered, and yet we cannot go beyond what's provable in PRA.

[18] Dean seems to support something along the lines of the first claim.

Admittedly the semantic core, although satisfying many desiderata generally imposed to the truth predicate such as full compositionality, falls short of others, such as the capability of proving the base theory true. We have extensively elaborated on how this latter desideratum cannot be satisfied in the case of PRA as the finitist's base; nonetheless, one may still consider CT[PRA], and CT[S] more generally, as an insufficient soundness extension. But even for the skeptic of this kind the situation depicted above may provide useful information on the precise boundary between acceptable and non-acceptable sets of metalinguistic principles. This boundary lies exactly, in the case of PRA, in between the semantic core CT[PRA] and its extension with either the truth of all logical axioms of PRA, or with the schematic induction of PRA in which the truth predicate is allowed: in CT[S] only partial metalinguistic reflection is admitted, whereas in these extensions one can ascend to the full metalinguistic reflection expressed by GRP(PRA).

Moving to what we called ultrafinitism, in order to draw conclusions along the lines of the ones just obtaineed for PRA, we would need an analogue of Proposition 4 for all theories containing S^1_2. This claim is, unfortunately, still only a likely conjecture. At any rate, this more general version of Proposition 4 would then establish that the semantic core for implicit commitment relative to a theory S gives us a theory that does not give us new theorems in $\mathcal{L}_S$, and in particular Π_1-sentences such as the consistency of Robinson arithmetic, Con(Q), that are not available in ultrafinitistically acceptable theories.

4.2. **Full arithmetical induction and beyond.** Isaacson considers PA as specifying a self-standing portion of mathematical reality. In his view, full-induction on $\mathcal{L}$ still belongs to or even delimits the realm of finite mathematics: principles that properly extend PA, such as RFN(PA), must therefore appeal to infinitary resources. Again, the semantic core offers us the possibility of identifying a metalinguistic component of the implicit commitment to PA and by distinguishing it from the object-linguistic, or mathematical content of PA. The conservativeness of CT[PA] over PA tells us that proofs of theorems in the language of $\mathcal{L}$ in CT[PA] are not affected by the metalinguistic component embodied in the truth principles of CT[PA].

Arguably, Isaacson would regard the semantic components of CT[PA] as intrinsically infinitary, but this is not a problem for our reading of (ICT). The implicit commitment to PA, if one regards it as 'arithmetically complete', would be delimited by the semantic core, and its non-arithmetical, infinitary components do not interfere in any way with its mathematical ones in CT[PA]-proofs. This is once more an example of how the semantic core can combine the idea of a privileged access to a definite portion of mathematical reality given by a specific theory with the natural act of reflection *on* the metalinguistic aspects of this theory via semantic notions.[19]

Isaacson's position clearly contrasts with views such as Feferman's, who considers the extension of the induction schema of PA as essential to unravel the class of arithmetical assertions we are implicitly committed to when accepting PA in the first-place. In such positions, schemata are open-ended, and there is no need to stop the truth predicate to interact with the arithmetical content of PA. The semantic core CT[PA], in such view, *counts only as a class of necessary conditions that our notion of truth has to satisfy.* The theory of truth Feferman is putting forward to fully articulate our commitment to PA, namely KF, contains CT[PA] and is

[19]This separation between object-linguistic and meta-linguistic aspects of theories can be even made more drastic. Perhaps in this setting the distinction between arithmetical and syntactico/semantic content may even be more convincing for authors that stress the epistemic stability of an arithmetical theory S. We refer to [38] for an overview of such options.

spectacularly stronger than it: it corresponds in fact to ε_0-many iterations of ACA. In terms of classical ordinal analysis, KF will prove the same arithmetical theorems as PA plus transfinite induction up to $\varphi_{\varepsilon_0}(0)$.[20] According to our proposed reading of (ICT), therefore, Feferman's acceptance of PA is tied not only to the semantic core, but to a substantial amount of mathematical principles that can be measured by the big gap separating the transfinite induction schemata for $\mathcal{L}$ associated to the ordinals ϵ_0 and $\varphi_{\varepsilon_0}(0)$. In moving from Isaacson's to Feferman's position, the semantic core stayed the same, whereas the variable component, which was *empty* in the case of Isaacson, now includes a large amount of analysis.

There is, however, an unexpected bridge between Isaacson's and Feferman's positions. Once the truth predicate is not allowed into the induction schema of PA, KF becomes much closer to CT[PA]. This theory, called KF $\restriction$ in [20], is in fact conservative over PA. Any model $\mathcal{M}$ of PA can be expanded to a model $(\mathcal{M}, S)$ of KF $\restriction$ by tanking S to be a fixed point of a suitable positive inductive definition capturing the clauses of the construction of a Kripke truth set (see [33]).

Instead of being a mere curiosity, this point highlights how the difference between the view of implicit commitment associated with the first-orderist *à la* Isaacson and with the predicativist may be seen as not lying in their conception of truth, but in their understanding of schematic reasoning. If in fact our distinction between object-linguistic and metalinguistic component of a truth theory is granted, then the first-orderist can articulate a robust notion of truth and yet distinguishing between the arithmetical reality that PA is isolating and the mere metalinguistic consequences that become available once one moves to its extension CT[PA]. She might even move to a type-free notion of truth, as articulated by KF $\restriction$, for instance, without exceeding the arithmetical consequence of PA. Once the truth predicate is allowed to do *mathematical work*, however, the situation drastically changes.

This scenario reinforces the usefulness of our analysis of implicit commitment via the semantic core: the latter in fact gives us necessary conditions for soundness extensions of a mathematical theory we accept and it is compatible with both restrictive and relaxed readings of (ICT).

Of course once one has reached a satisfactory halting point, such as KF for Feferman's analysis of implicit commitment, nothing prevents one from asking herself what we are implicitly committed to when we are accepting the theory of truth. If Feferman's strategy is extended to the theory of truth, for instance, one can obtain extensions of KF via uniform reflection principles. Indeed, Horsten and Leigh in [24] have shown that extensions of KF can be obtained by starting with TB[PA] via finitely many iterations of uniform reflection.[21] However, since we are not interested in the theory of truth itself, but only in the boundary between acceptable and non-acceptable characterizations of the implicit commitment for the base theory, we do not consider further this possible extension of our analysis.

5. Conclusion

The implicit commitment thesis (ICT) prescribes that, in accepting a system S formalizing some portion of mathematics — arithmetic in our case studies — one is committed to resources not immediately available in S. Traditionally, these additional resources have been understood

[20] For a definition of the Veblen functions, see [43].

[21] A similar strategy for a nonclassical setting in which the starting point are type-free principles of the form $\mathsf{T}^{\ulcorner}\varphi^{\urcorner} \Leftrightarrow \varphi$, with $\Leftrightarrow$ a suitable non classical biconditional has been carried out by [17].

in terms of sentences in the language of S that are not provable in S already, typically reflection principles for S expressing the soundness of S.

As recently shown by Dean, however, certain foundational standpoints consider a particular arithmetical theory S as delimiting a privileged region of mathematical reality. Reflection principles for the theory S therefore, being closely related to mathematically meaningful principles that lie beyond the space of mathematics occupied by S (see §2), should be considered as incompatible with those foundational standpoints. Examples of such positions are Tait's justification of PRA, Isaacson's thesis on PA, and to some extent Nelson's strict finitism.

Starting with the observation that soundness claims of S can only be fully articulated by resorting to the notion of truth, we have proposed a dynamic and widely applicable reading of (ICT). The additional resources we are committed to when accepting S will contain principles for truth: these principles, what we called the semantic core for implicit commitment, are fixed and shared by any reasonable justification for the acceptance of a system S. They amount to compositional truth principles and include minimal soundness claims for S such as the truth of all its non-logical axioms, the truth of all instances of each propositional tautology and, in reasonably chosen cases, the truth-preserving character of its rules of inference. Further principles extending the semantic core depend on the justification for S provided by the idealized mathematician.

This analysis that we have provided is adequate with respect to the case studies considered in the first part of the paper: the semantic core, when added to S, prevents one from proving new consequences in the language of S besides the ones already available in S itself. Moreover, all natural articulations of soundness assertions of S in the form of stronger truth principles will contain the semantic core; whatever variable components one is willing to add to the semantic core, therefore, they will not be incompatible with it.[22]

References

[1] A. Arana. Logical and semantic purity. In *Philosophy of Mathematics. Set Theory, Measuring Theories, and Nominalism*, G. Preyer and G. Peter (eds.), Ontos Verlag, Frankfurt, pp. 40-52, 2008.

[2] L.D. Beklemishev. *Reflection principles and provability algebras in formal arithmetic.* Uspekhi Matematicheskikh Nauk, 60(2):3-78, 2005. In Russian. English translation in: Russian Mathematical Surveys, 60(2): 197-268, 2005.

[3] S.R. Buss. Bounded Arithmetic. Bibliopolis, Napoli, 1986.

[4] S.R. Buss. First-order Proof Theory of Arithmetic. In *Handbook of Proof Theory*, edited by S. Buss, Elsevier North-Holland, pp. 79-147, 1998.

[5] C. Cieśliński. *Truth, Conservativeness, and Provability*, Mind 119, pp. 409-422, 2010.

[6] W. Dean. *Arithmetical Reflection and the Provability of Soundness.* Philosophia Mathematica, 23(1):31-64, 2015.

[7] A. Enayat and A. Visser. New constructions of satisfaction classes. In *Unifying the Philosophy of Truth*, T. Achourioti (*et alii*) (eds.), Springer, 2015. pp. 321-335.

[8] H. B. Enderton. *A Mathematical Introduction to Logic.* Academic Press Second edition, 2001.

[9] S. Feferman. *Arithmetization of metamathematics in a general setting.* Fundamenta Mathematicae 49, pp. 35-92, 1960.

[10] S. Feferman. *Transfinite recursive progressions of axiomatic theories.* The Journal of Symbolic Logic 27, 259-316, 1962.

[11] S. Feferman. *Systems of predicative analysis.* Journal of Symbolic Logic, 29, pp. 1-30, 1964.

[12] S. Feferman. *Reflections on Incompleteness.* Journal of Symbolic Logic, 56, pp. 1-49, 1991.

[22] We notice that not only truth principles may be added, but also principles for intensional predicates such as necessity. If genuine extensions of, say, KF may be reached via the interplay of truth and modal predicates, this would represent a genuine improvement of predicativism given the natural numbers *à la* Feferman.

[13] S. Feferman. *Predicativity* In S. Shapiro, *The Oxford Handbook Of Philosophy of Mathematics and Logic*, pp. 590-624. Oxford University Press, 2005.
[14] F. Ferreira and G. Ferreira. *Interpretability in Robinson's* Q. Bulletin of Symbolic Logic, 19, pp. 289-317, 2013.
[15] H. Field. *Deflating the conservativeness argument.* Journal of Philosophy 96 (10):533-540, 1999.
[16] M. Fischer *Minimal truth and interpretability.* Review of Symbolic Logic, 2(4) pp. 799–815, 2009.
[17] M. Fischer, L. Horsten and C. Nicolai (forthcoming), *The Logic of truth.* Unpublished Manuscript.
[18] G. Gentzen. *Die Widerspruchsfreiheit der reinen Zahlentheorie.* Mathematische Annalen, 112: pp. 493-565, 1936. Translated as "The consistency of arithmetic", in M. E., Szabo, ed., *Collected Papers of Gerhard Gentzen* Amsterdam: North-Holland, 1969.
[19] J.-Y. Girard. *Proof-Theory and Logical Complexity.* Bibliopolis, Napoli, 1987.
[20] V. Halbach. *Axiomatic Theories of Truth.* Cambridge University Press, 2011.
[21] P. Hájek and P. Pudlák. Metamathematics of First-Order Arithmetic. Springer-Verlag, 1998.
[22] R. Heck. Consistency and the theory of truth. The Review of Symbolic Logic. 8 (03): pp. 424-466, 2015.
[23] D. Hilbert and P. Bernays. Grundlagen der Mathematik. 2nd edition. Berlin: Springer, 1968.
[24] L. Horsten and G. E. Leigh. Truth is Simple. Mind, forthcoming.
[25] D. Isaacson. *Arithmetical truth and hidden higher-order concepts.* In *Logic Colloquium '85.* The Paris Logic Group (ed.), Amsterdam: North-Holland, pp. 147-69, 1987. Reprinted in W.D. Hart, editor, The Philosophy of Mathematics, pp. 203-224. Oxford University Press, 1996.
[26] D. Isaacson. *Some considerations on arithmetical truth and the omega-rule.* In M. Detlefsen, ed. *Proof, Logic and Formalization*, Routledge, pp. 94-138, 1991.
[27] J. Ketland. *Deflationism and Tarski's Paradise.* Mind 108, pp. 69-94, 1999.
[28] J. Ketland. *Deflationism and the Gödel phenomena: reply to Tennant.* Mind 114, pp. 75-88, 2005.
[29] H. Kotlarski, S. Krajewski, and A. H. Lachlan, *Construction of satisfaction classes for nonstandard models.* Canadian Mathematical Bulletin 24, 283–293, 1981.
[30] G. Kreisel. Ordinal logics and the characterization of informal concepts of proof. *Proceedings of the International Congress of Mathematicians, Cambridge University Press, pp. 289-299, 1960.*
[31] G. Kreisel. Principles of proof and ordinals implicit in given concepts. In *Intuitionism and Proof Theory*, J. Myhill, and R. E. Vesley (eds.), North-Holland, Studies in Logic and the Foundations of Mathematics 60, pp. 489-516, 1970.
[32] G. Kreisel and A. Lévy. *Reflection principles and their use for establishing the complexity of axiomatic systems.* Zeitschrift für mathematische Logik und Grundlagen der Mathematik. 36: 441–454, 1968.
[33] S. Kripke. *Outline of a theory of truth.* The Journal of Philosophy, Volume 72, Issue 19, pp. 690–716, 1975.
[34] G. Leigh. *Conservativity for theories of compositional truth via cut-elimination.* The Journal of Symbolic Logic, Volume 80, Issue 3, pp. 845–865, 2015.
[35] P. Lindstrom. Aspects of Incompleteness. Springer, 1997.
[36] V. McGee. *How We Learn Mathematical Language.* The Philosophical Review 106(1), pp. 35-68, 1997.
[37] E. Nelson. Predicative Arithmetic. Princeton University Press, 1986.
[38] C. Nicolai. Deflationary truth and the ontology of expressions. *Synthese*, 192(12):4031–4055, 2015.
[39] C. Nicolai. A note on typed truth and consistency assertions. *Journal of Philosophical Logic* (45) 89, pp. 89-119, 2016.
[40] N. J. L. L. Pedersen and M. Rossberg. *Open-endedness, Schemas and Ontological Commitment.* Noûs 44 (2), pp. 329-39, 2010.
[41] M. Piazza and G. Pulcini. *A Deflationary Account of the Truth of the Gödel Sentence $\mathcal{G}$.* In *From Logic to Practice*, G. Lolli et al. (eds.), Boston Studies in the Philosophy and History of Science, Springer, pp. 71-90, 2015.
[42] M. Piazza and G. Pulcini. *What is so special about the Gödel sentence $\mathcal{G}$?.* In F. Boccuni F. and A. Sereni (eds.) *Objectivity, Realism, and Proof. FilMat Studies in the Philosophy of Mathematics*, Boston Studies in the Philosophy and History of Science, Springer, pp. 245-263.
[43] W. Pohlers. Proof Theory: The First Step into Impredicativity. Springer, 2009.
[44] P. Pudlák. *Cuts, consistency statements, and interpretations.* The Journal of Symbolic Logic, 50, 423-441, 1985.
[45] S. Shapiro. *Proof and Truth: Trough Thick and Thin.* Journal of Philosophy 95, pp. 493-521, 1998.
[46] K. Schütte. *Einge Grenze für die Beweisbarkeit der Transfiniten Induktion in der verzweigten Typenlogik.* Archiv für Mathematische Logik und Grundlagen-forschung, 7:45-60, 1965.
[47] W. Tait. *Finitism.* The Journal of Philosophy 78, pp. 524-546, 1981.

[48] N. Tennant. *Deflationism and the Gödel phenomena.* Mind 111, pp. 551-582, 2002.
[49] A.S. Troelstra and H. Schwichtenberg. Basic Proof theory. Cambridge, 2000.