J Philos Logic (2016) 45:89–119
DOI 10.1007/s10992-015-9366-6

A Note on Typed Truth and Consistency Assertions

Carlo Nicolai

Received: 11 June 2014 / Accepted: 25 February 2015 / Published online: 2 April 2015

Abstract In the paper we investigate typed (mainly compositional) axiomatizations of the truth predicate in which the axioms of truth come with a built-in, minimal and self-sufficient machinery to talk about syntactic aspects of an arbitrary base theory. Expanding previous works of the author and building on recent works of Albert Visser and Richard Heck, we give a precise characterization of these systems by investigating the strict relationships occurring between them, arithmetized model constructions in weak arithmetical systems and suitable set existence axioms. The framework considered will give rise to some methodological remarks on the construction of truth theories and provide us with a privileged point of view to analyze the notion of truth arising from compositional principles in a typed setting.

Keywords Axiomatic theories of truth · Subsystems of first-order arithmetic · Truth-theoretic deflationism

1 Introduction

Axiomatic investigations of the truth predicate appear to be pointed at at least three different kinds of goals.

The ideas and results presented in what follows heavily rely on conjectures and motivations due to Richard Heck and now contained in [24]. I would like to thank Richard for making me available his (at the time) unpublished manuscripts and for enlightening email exchanges. I also thank Ali Enayat, Volker Halbach, Graham Leigh and an anonymous referee for their feedbacks. A special thank goes to Albert Visser. This work was supported by the Arts and Humanities Research Council UK AH/H039791/1 and by the *Analysis Trust*.

C. Nicolai (✉)
Faculty of Philosophy, University of Oxford, Oxford, England
e-mail: carlo.nicolai6@gmail.com; carlo.nicolai@philosophy.ox.ac.uk

- In the first place semantic theories of truth give rise to mathematical structures that often reach a high degree of complexity; it is thus both insightful and illuminating to fix collections of principles that hold in these structures.[1]
- There are, moreover, several examples of mutual reductions of axiomatic theories of truth over a base theory B and extensions of B with set existence axioms.[2] A second way to look at these axiomatic investigations is thus to consider them as attempts to find reductions of ontological commitments to semantic commitments.[3]
- A third target is represented is evaluating the 'metaphysical' impact that axioms of truth determine on the underlying mathematical structure. The analysis of truth-theoretic deflationism, in fact, often inspires technical investigations on axiomatic theories of truth: some of the results in this area provide concrete examples of the role played in abstract reasoning by the primitive truth predicate depicted by the axioms.

The present paper constitutes a contribution to the increasing literature on axiomatic truth. To give a glimpse of its content, we briefly touch on how it relates to the three targets above.

The semantic theory that the axioms considered will capture is essentially Tarski's; frankly speaking, it is as much Tarskian as it could be. The structure under investigation accentuates the main traits of Tarski's solution to the Liar paradox. Building on [23, 33] and on unpublished work of Richard Heck, we in fact investigate a setting in which typing is not only forced by the axioms for truth but also by the way in which the syntax of the object theory is formalized (cfr. §3).

The resulting theories will be essentially typed, compositional axiomatizations of the truth predicate. They will be shown to belong to the same degree of interpretability of an intensionally correct consistency statement for their object theory over a minimal theory of syntax. By recent results of Albert Visser [46], this would in turn add a new example to the many already known of mutual reductions of compositional Tarskian truth and predicative comprehension. This suggests some partial fulfilment of the second target.

Advocates of deflationism have often proposed axiomatizations of the truth predicate based upon the celebrated T-schema. By contrast the deflationist's attitude towards the so-called compositional principles such as (*) 'a conjunction is true if and only if both conjuncts are true' remains somewhat mysterious.[4] In recent years,

[1] Nonetheless, it is still not completely clear in what sense an axiomatic theory of truth can capture a semantic construction. Fischer et al. [16] tries to fix sufficient conditions in this direction.

[2] See [21] for a survey.

[3] Or, possibly, vice versa.

[4] It has been argued, in [25] for instance, that also in the case of typed truth the diquotationalist can retrieve all instances of principles in the style of (*). This, however, cannot be equivalent to deriving the formal counterpart of (*) itself with a compact logic in the background. The reason for this was already known to Tarski: typed disquotational truth cannot establish single sentences expressing general claims involving truth. A more sophisticated account is contained in [14]. We refer to a forthcoming work of Richard Heck for the impact that the framework that we consider can have on the contents of Field's paper.

however, the expressive and deductive weakness of the disquotational truth predicate has led many authors to consider the acceptance of those principles as at least compatible with truth theoretic deflationism.[5]

Many examples of sound, compositional systems for truth have been proposed whose truth predicate is in some sense reducible to syntactic or mathematical resources.[6] For instance the theory[7] Peano Arithmetic + 'there is a truth class' (a.k.a $\mathsf{CT}\restriction$ in [23]) is known to interpretable in PA and also a conservative extension of it.[8] Similar results hold for the axiomatization of Kripke's theory of truth over PA with induction restricted to arithmetical formulas and for Friedman and Sheard's $\mathsf{FS}\restriction$. [9] Obviously the restriction on the induction schema of PA plays a crucial role in those results; nonetheless, if one focuses only on the role of compositional axioms, semantic resources — although indispensable to establish general claims — appear to be reducible to object-theoretic resources.

The results presented below will suggest that a notion of truth governed by compositional axioms of the sort considered in this work, unlike disquotational truth, is not reducible — in a sense yet to be clarified — to syntactic or mathematical resources. The equivalence of the compositional principles, modulo mutual interpretability, to assertions of syntactic nature such as intensional consistency statements for the object theory will suggest, moreover, that these axioms encapsulate enough metamathematical content to outreach the resources of the object theory.

2 Preliminaries

2.1 Bounded Arithmetic and Arithmetization

We assume some familiarity with subsystems of first-order arithmetic as introduced, for instance, in [19]. We conventionally consider only first-order, possibly many-sorted, languages. It is useful to formulate the arithmetical theories that we want to reason about, the object theories of our theories of truth, in a *relational* language extending the language $\mathcal{L}_A$ of relational arithmetic.[10] As usual, given the proviso concerning our 'official' definition of the theories, we will immediately forget about the restriction imposed to the languages when this is thought to be harmless.

[5] See, for instance, [14].

[6] We do not get into a discussion of what 'reduction' *should* precisely mean in this context. Here we are content with traditional forms of reduction such as conservativity or relative interpretability.

[7] We mention Peano Arithmetic as it is the standard syntax theory for axiomatic theories of truth. However, it is not clear, in general, whether many of the results obtained with Peano Arithmetic as base theory smoothly transfer to a setting in which the syntax is provided by a direct axiomatization of concatenation or, along similar lines, by a weak arithmetical theory.

[8] A new, very elegant proof of both claims is contained in [8].

[9] A relative interpretation $\mathsf{KF}\restriction$ in PA can be extracted form the conservativity proof of $\mathsf{KF}\restriction$ over PA given by Cantini in [5]. Moreover, each finite subsystem $\mathsf{FS}_n\restriction$ of $\mathsf{FS}\restriction$ can be shown to be embeddable in the theories $\mathsf{RT}_{<2n}$, as shown in [20]. By using well-known results on satisfaction classes, then, the conservativity of $\mathsf{FS}_n\restriction$ follows. Also, by the reflexivity of PA and Orey's compactness, $\mathsf{FS}_n\restriction$ is interpretable in PA.

[10] Cfr. [19, §I.(e)].

On occasion we will require our base theories to be *sequential*. Essentially, a theory is sequential if it can code sequences of variable length of *all* objects in the sense of the theory. More precisely, T is sequential iff it directly interprets (i.e. the interpretation is not relativized and identity is mapped into identity) *Adjunctive Set Theory*,[11] that is a theory in the a (first-order) language with $\in$ and $=$ whose axioms are:

$$\exists x\ \forall y\ y \notin x \tag{AS1}$$

$$\forall u, v\ \exists x\ \forall y (y \in x \leftrightarrow (y \in u \vee y = v)) \tag{AS2}$$

Remarkably, Q is not sequential, but many sequential theories are interpretable in it.

For many of the arguments employed in what follows the base theory will be S^1_2, which is, among other things discussed shortly, sequential. It is a theory invented by Samuel Buss in [3] to study polynomial time computability. Notably, the provably total functions of S^1_2 are exactly the p-time computable functions. S^1_2 will be employed in this work in two ways: it will be interpretable in any theory that we will consider and it will taken to be our preferred environment to formalize syntactic notions and operations. In the remaining part of this section, we will introduce S^1_2 and motivate our choice of it.

S^1_2 is formulated in the language

$$\mathcal{L}_2 = \left\{0, \mathsf{S}, +, \times, \#, |.|, \left\lfloor \frac{1}{2}. \right\rfloor, \leq\right\},$$

where $|.|, \#$ are such that

$$|x| = \lceil \mathsf{log}_2(x+1) \rceil, \qquad\qquad x\#y = 2^{|x| \times |y|};$$

that is, $|x|$ outputs the length of the binary representation of x (the upper integer part of the binary logarithm of $x + 1$) and $x\#y$ denotes the power of two with binary representation of length $|x| \times |y|$. To introduce the axioms of S^1_2, we need the notion of sharply bounded quantification. A quantifier is *sharply bounded* if it is bounded by a term of the form $|t|$, where t is a term of $\mathcal{L}_2$. A formula of $\mathcal{L}_2$ is Δ^b_0 if all its quantifiers are sharply bounded. It is Σ^b_1 iff it is of the form $\exists x < t\ \varphi$ with t a term of $\mathcal{L}_2$ and φ in Δ^b_0. The class of Δ^b_1 formulas is defined in the usual way.

The axioms of S^1_2 include the 32 sentences of BASIC (Cfr. Appendix) and the schema Σ^b_1-PIND:

$$\varphi(0) \wedge \forall x \left(\varphi\left(\left\lfloor \frac{1}{2}x \right\rfloor\right) \rightarrow \varphi(x)\right) \rightarrow \forall x\, \varphi(x) \qquad\qquad (\Sigma^b_1\text{-}\mathsf{PIND})$$

Crucially, $\#$ has polynomial growth rate: for every term $t(\bar{x})$ of $\mathcal{L}_2$ we can find a polynominal P such that for all $x_1, ..., x_n$

$$|t(x_1, ..., x_n)| \leq P(|x_1|, ..., |x_n|).$$

This makes possible a smooth, intensional development of the syntax of formal systems inside S^1_2: definitions by recursion of syntactic notions are naturally formalized

[11] This definition is due to Pavel Pudlák.

by stipulating the existence of a sequence coding the course of values of the function in question. Theories such as $\mathsf{I}\Sigma_1$ have enough induction to prove the relevant properties of these sequences.[12] By Parikh's Theorem,[13] this is not possible in theories, such as $\mathsf{I}\Delta_0$, whose provably total functions can only cope with linear increases of the length of sequences. By contrast, S^1_2 is just right to deal with polynomial increases of length of sequences.

For the details of the coding we refer to [3, 4]. Sequences are coded by strings of 0's and 1's; moreover, for any given language $\mathcal{L}_W$, we assume Δ^b_1-definitions $\mathsf{Term}_{\mathcal{L}_W}(x)$, $\mathsf{Fml}_{\mathcal{L}_W}(x)$, $\mathsf{Sent}_{\mathcal{L}_W}(x)$, $\mathsf{Prf}_{\mathcal{L}_W}(x)$, $\mathsf{Proof}_{\mathcal{L}_W}(x, y)$ of the sets of terms, formulas, sentences of the language of an arbitrary object theory, of proofs in the theory W and of the relation of being a proof in W of the formula y of $\mathcal{L}_W$. Also, we take the set of theorems of W to be defined by the $\exists\Delta^b_1$-formula $\mathsf{Prov}_U(x) :\leftrightarrow \exists y\, \mathsf{Proof}_W(x, y)$. A special status the coding procedure we are assuming is reserved to the function that assigns to each number its numeral. If $\overline{n}$ is defined as

$$\underbrace{\mathsf{S}\cdots\mathsf{S}}_{n \text{ times}}\overline{0} \tag{1}$$

then $\mathsf{NUM}(n)$, that is the code of Eq. (1), will be exponential in n and we would not be able to find bounds to the sequence defining $\mathsf{NUM}(x)$. Therefore dyadic numerals have to be employed in this procedure, that is

$$\overline{0} = \overline{0} \qquad \overline{1} = \mathsf{S}\,\overline{0}$$
$$\overline{2n} = \overline{2} \times \overline{n} \qquad \overline{2n+1} = \overline{2} \times \overline{n} + \overline{1}.$$

This definition renders the function $\mathsf{num}(x)$ that assigns to a number n the code of the corresponding dyadic numeral provably total even in $\mathsf{I}\Delta_0$, as now $\mathsf{num}(n)$ becomes of order n^a for a constant (standard) a. Moreover, S^1_2 can prove the totality (as it provides a bound for its outputs) of the function that takes codes t, r, v of terms and a variable and returns the code of the result $t(r/v)$ of replacing all occurrences of the variable coded by v in t by r. As to notational conventions, we will often abuse of Gödel corners and employ them together with Feferman's dot convention.

We conclude this section by outlining two further attractive features of S^1_2. The first links directly to the considerations just sketched: S^1_2 is synonymous (i.e. definitionally equivalent) with a theory of strings developed in [12] based on primitive concatenation and substitution. The second is represented by the finite axiomatizability of S^1_2 and explains why S^1_2 is often preferred to theories such as $\mathsf{I}\Delta_0 + \Omega_1$,[14] for

[12] Cfr. [19, §I(c),(d)]

[13] Parikh's Theorem states that if $\varphi(\bar{x}, y)$ defines a total function in $\mathsf{I}\Delta_0$, then there is a term $t(\bar{x})$ of $\mathcal{L}$ such that

$$\mathsf{I}\Delta_0 \vdash \forall \bar{x}$$
$$\exists y < t(\bar{x})\ \varphi(\bar{x}, y).$$

By Parikh's theorem therefore, any $\mathsf{I}\Delta_0$-provably total function can only increase the length sequences of 0–1 words linearly.

[14] For a proof of the finitely axiomatizability of S^1_2, we refer to [19].

which a similar result is not yet known. Essentially, one shows that all instances of Σ^b_1-PIND can be retrieved via the equivalence of every Σ^b_1-formula $\varphi(x)$ of $\mathcal{L}_2$ with a single property that encapsulates the p-time computation expressed by $\varphi(x)$.

2.2 Cuts and Interpretations

We choose to focus our attention on logical properties of compositional axioms and, on occasion, of disquotational principles. The theories of truth considered below, therefore, will not contain induction principles or related schemata in which the presence of semantic vocabulary is allowed. The possibility of performing inductive reasoning in our theory of truth, however, still appears to be important if we require our axiomatization of the truth predicate to capture portions of metamathematical reflection on the object theory. Following a standard practice initiated by Robert Solovay, the lack of induction will be compensated by the use of definable cuts.[15]

Cuts are initial segments. A formula $\varphi(x)$ is called *inductive* in $T \supseteq \mathsf{Q}$ if and only if

$$T \vdash \varphi(0) \wedge \forall y\ (\varphi(y) \rightarrow \varphi(Sy))$$

$\varphi(x)$ is a *T-cut* if and only if, additionally,

$$T \vdash \forall x, y\ (\varphi(x) \wedge y \leq x \rightarrow \varphi(y))$$

By a well-known result of Solovay, every inductive formula has a subcut:

Lemma 1 *Let $\varphi(x)$ be inductive in Q. Then there exists a Q-cut $\psi(x)$ such that*

$$\mathsf{Q} \vdash \psi(x) \rightarrow \varphi(x).$$

The ψ in Lemma 1 is obtained by closing $\varphi(\overline{n})$ for each n under transitivity of $\leq$ so that this holds for all $m \leq n$ as well. It is often useful, however, to employ a slightly modified notion of definable cut. The modifications are justified by the following Lemma:

Lemma 2 *Let T interpret S^1_2 and I be a T cut. Then we can find a subcut J of I such that T proves the following:*

$$J(x) \wedge J(y) \rightarrow J(x + y) \tag{2}$$
$$J(x) \wedge J(y) \rightarrow J(x \times y) \tag{3}$$
$$J(x) \wedge J(y) \rightarrow J(x\#y) \tag{4}$$

Therefore, in what follows, a definable cut can always be taken to be closed under addition, multiplication and the smash function. It is also convenient to assume that the cuts considered satisfy the axioms of S^1_2.

[15] Solovay's original note is unpublished. The standard reference is now [19].

Let T and W be theories containing S^1_2.[16] A *relative translation* of $\mathcal{L}_T$ into $\mathcal{L}_W$ can be described as a pair (δ, F) where δ is a $\mathcal{L}_W$-formula with one free variable — the domain of the translation — and F is a (finite) mapping that takes n-ary relation symbols of $\mathcal{L}_T$ and gives back formulas of $\mathcal{L}_W$ with n free variables. The translation extends to the mapping τ:

- $(R(x_1, ..., x_n))^\tau :\leftrightarrow F(R)(x_1, ..., x_n)$;
- τ commutes with propositional connectives;
- $(\forall x\ \varphi(x))^\tau :\leftrightarrow \forall x\ (\delta(x) \to \varphi^\tau)$ and $(\exists x\ \varphi(x))^\tau :\Leftrightarrow \exists x\ (\delta(x) \wedge \varphi^\tau)$.

An *interpretation* K is then specified by a triple (T, τ, W) such that for all sentences φ of $\mathcal{L}_T$,[17]

$$T \vdash \varphi \Rightarrow W \vdash \varphi^\tau.$$

We write $K : W \rhd T$ for 'K is an interpretation of T in W' and $W \rhd T$ for 'there is an interpretation of T in W'. Moreover, we let

$$W \equiv T :\leftrightarrow W \rhd T \wedge T \rhd W,$$

and read $T \equiv W$ as 'T and W are mutually interpretable'. T *locally* interprets W—in symbols, $T \rhd_{\mathsf{loc}} W$ — if and only if every finite subsystem of W is interpretable in T. An interpretation is *direct* if and only if it maps identity to identity and it does not relativize quantifiers. We will often not distinguish between an interpretation and the relative translation that supports it.

If $K : T \rhd W$ and $\mathcal{M}$ is any model of T, then K can be seen as a method for defining a model $\mathcal{M}^K$ of W inside $\mathcal{M}$. We state two ways to compare $\mathcal{M}^K$ and $\mathcal{M}$, depending on whether T has full induction or not.

Lemma 3 *If $K : T \rhd W$ and T has full induction, then for any $\mathcal{M} \models T$ we find a (uniformly) $\mathcal{M}$-definable embedding of $\mathcal{M}$ into an initial segment of $\mathcal{M}^K$.*

Lemma 3 is readily obtained by noticing that in $\mathcal{M}$ we define an injection $f : \mathcal{M} \to \mathcal{M}^K$ such that

$$f\left(0^{\mathcal{M}}\right) = 0^{\mathcal{M}^K}; \qquad f\left(x +^{\mathcal{M}} 1^{\mathcal{M}}\right) = f(x) +^{\mathcal{M}^K} 1^{\mathcal{M}^K}.$$

Since $\mathcal{M}$ has full induction, one can actually show that f is indeed the required isomorphism.

If, on the other hand, T does not have full-induction, the comparison between the numbers as given in $\mathcal{M}$ and $\mathcal{M}^K$ is provided by what is known as Pudlák's Lemma.

Lemma 4 (Pudlák) *If $K : T \rhd W$ and T is sequential, then there is a T-cut I that is, provably in T, an embedding of I into an initial segment of the numbers in the sense of W.*

[16] We recall that the theories are assumed to be formulated in a relational language. We owe to Albert Visser's work this description of interpretations.

[17] The equivalence of 'axioms' and 'theorems' interpretability can break down in the context of the formalization of the interpretability relation in weak theories. We refer to [44] for details.

For a more careful statement of the result we refer to [27]. A proof is contained in [38]. We just notice that in the proof we construct in T a formula $\varphi_K(x, y)$ that provably defines an isomorphism between a T-cut I and a proper initial segment J of the W-numbers as seen by K. It is important to notice, however, that there is in general no $\mathcal{L}_W$-formula $\psi(x)$ whose image under K defines the cut J in T.[18]

In what follows the next lemma will play an important role:

Lemma 5 (Wilkie, Nelson.) Q *interprets* S^1_2 *on a definable cut.*

Since Q does not prove the usual properties of $\leq$, the crucial component of the proof consists in showing that the interpretation is actually a cut. This makes possible the downwards preservation of Π_1-sentences (cf. [19, V.5(c)]).

We conclude this preliminary section with a brief description of the incompleteness phenomena determined by the techniques just described. Given a theory T containing Q and a Σ_1 provability predicate π for T satisfying the usual provability conditions, Gödel's Second Incompleteness Theorem tell us that T does not prove $\neg\pi(\ulcorner 0 = 1\urcorner)$. To obtain this, one usually shows that $\neg\pi(\ulcorner 0 = 1\urcorner)$ is actually equivalent to a Gödel sentence for T. First strengthenings of this result, when T is a reflexive extension of PA, were presented by Feferman in [10]. In particular, it was there shown that $T + \neg\pi(\ulcorner 0 = 1\urcorner)$ is not interpretable in T.[19] Moreover, since $T + \pi(\ulcorner 0 = 1\urcorner)$ is Π_1-conservative over T, then we can find an interpretation of $T + \pi(\ulcorner 0 = 1\urcorner)$ in T.[20]

Further strengthenings are essentially due to Pavel Pudlák: let T contain Q and τ be a Σ_1-definition of T in T. For any T-cut J, T does not prove

$$\forall x\,(x \in J \rightarrow \neg\,\mathsf{Proof}_\tau(x, \ulcorner 0 = 1\urcorner)),$$

where $\mathsf{Proof}_\tau(x, y)$ is a Δ_0-definition of 'x is a proof of y'. A beautiful, related form of second incompleteness theorem is the following: for any T containing Q, if T is defined by a Σ_1-formula, then T does not interpret $\mathsf{Q} + \mathsf{Con}_T$. The key observations here are that Gödel's Second Incompleteness theorems can be meaningfully stated and proved in S^1_2 and that $\mathsf{S}^1_2 + \mathsf{Con}_U$ is interpretable in $\mathsf{Q} + \mathsf{Con}_U$ on a definable cut (Lemma 5). The facts just stated depend on the coding procedure being used, e.g.

[18] Another important principle characterising definable cuts is the so-called *outside big, inside small* principle. This denomination, as far as the author knows, is due to Albert Visser. It informally states that despite that fact that a cut is an initial segment provably in a theory, the very same theory can construct, for every number of the theory, a proof that that number belongs to the cut. That is, with T interpreting S^1_2 and $\varphi(x)$ a T-cut,

$$T \vdash \forall x\ \mathsf{Prov}_T(\ulcorner\varphi(\dot{x})\urcorner), \tag{5}$$

where $\mathsf{Prov}_T(x)$ is as in §2.1. Essentially, to obtain (5) one notices that since the cut is closed under addition and we employ efficient numerals (see p. 5), for any theory-number a one can combine the standard proof of $\varphi(\overline{0})$ with $a - 1$ universal instantiations of $\forall x\,(\varphi(x) \rightarrow \varphi(2x))$ or $\forall x\,(\varphi(x) \rightarrow \varphi(2x))$.

[19] Otherwise, T formalises the statement 'if T is consistent, so is $T+\neg\pi(\ulcorner 0 = 1\urcorner)$'. But then $T+\neg\pi(\ulcorner 0 = 1\urcorner)$ would prove its own consistency.

[20] If T is sequential and has full induction, in fact, 'T interprets W' is equivalent to 'every Π_1-sentence provable in W is provable in T'.

to formalise Con_U. Versions of the second incompleteness theorem that avoid *some* specific assumptions on the formalization of the syntax — the choice of the logical calculus, in particular — can be found in [45].

The following result is due to [37] and it will be quite useful in what follows. It is a consequence of Alex Wilkie's characterization of the class of Π_1-sentences interpretable in Q as the class of Π_1-sentences σ that are provable in $\mathsf{I}\Delta_0(\mathsf{exp})$:

Proposition 1 *For any Π_1-sentences π, π':*

$$\mathsf{Q} + \pi \rhd \mathsf{Q} + \pi' \;\Leftrightarrow\; \mathsf{I}\Delta_0(\mathsf{exp}) \vdash \pi \rightarrow \pi'.$$

3 Typed Theories of Truth and Expressions

As we anticipated in the introduction, our main target is to characterize a specific set of compositional truth axioms resulting from a non conventional way of constructing axiomatic theories of truth. The idea behind the construction is essentially Tarski's, although the interest in this alternative setting was recently revived by Richard Heck and Volker Halbach. The author's works [33] and [35] represent an available starting point for the investigation of this framework. Unlike the present, those papers mainly focus on the role of inductive reasoning in the proposed setting.

3.1 A Widely Applicable Framework

To give an informal picture of the setting we are interested in, a model of our theory of truth will be a many-sorted structure compounded by three disjoint universes: the domain D of the quantifiers of $\mathcal{L}_U$ — where U is our object theory — a domain of 'syntactic' objects that is disjoint from D, and a further (still disjoint from the other two) universe of assignments, that is objects representing sequences of elements of D. The salient component of the new setting is that the extension of the truth predicate consists of pairs of objects in the 'syntactic' domain and disjoint sequences of D-objects. The objects to which truth is informally attributed are still $\mathcal{L}_U$-sentences, but their formal counterparts to which the semantic vocabulary is applied do not belong to the domain of discourse of quantifiers of $\mathcal{L}_U$ but to the domain of discourse of a disjoint, 'syntactic' universe.

More formally, we start with an arbitrary (first-order) object theory U and its language $\mathcal{L}_U$. We only require $U \rhd \mathsf{S}^1_2$. When the sequentiality of U is required, we will state this explicitly. We expand this language to a three-sorted language $\mathcal{L}_\mathsf{T}$: variables are labeled via the set of sorts $\{\mathfrak{o}, \mathfrak{s}, \mathfrak{sq}\}$;[21] $\mathcal{L}_\mathsf{T}$ will also contain, besides nonlogical constants proper of $\mathcal{L}_U$, also the constants proper of $\mathcal{L}_2$,[22] the function symbol $\cdot(\cdot)$ of type $(\mathfrak{sq}, \mathfrak{s}) \rightarrow \mathfrak{o}$ and the predicate symbol Sat of sort $(\mathfrak{s}, \mathfrak{sq})$. The former will give rise to expressions of the form $a(i) = x$, stating that the i^{th} element of a sequence

[21] For heuristic, they label respectively 'object-theoretic variables', 'syntactic variables', 'sequence variables'. Nonlogical constants, for clarity,will occasionally also be labeled by sorts.

[22] For simplicity, we may require them to be unofficial abbreviations of their relational counterparts.

Table 1 Axioms of $\mathsf{T}[U]$

A.	Axioms of U (relativized to sort $\mathfrak{o}$)
B.	Axioms of S^1_2 (relativized to sort $\mathfrak{s}$)
C.	$\exists b\ (\forall j\ (j \neq i \rightarrow a(j) = b(j)) \wedge b(i) = x)$;
D.	1. $\mathsf{Sat}(\ulcorner R(v_1, ..., v_n)\urcorner, a) \leftrightarrow R(a(1), ..., a(n))$
	2. $\mathsf{Sat}(\ulcorner \neg \varphi\urcorner, a) \leftrightarrow \neg\, \mathsf{Sat}(\ulcorner \varphi\urcorner, a)$
	3. $\mathsf{Sat}(\ulcorner \varphi \wedge \psi\urcorner, a) \leftrightarrow \mathsf{Sat}(\ulcorner \varphi\urcorner, a) \wedge \mathsf{Sat}(\ulcorner \psi\urcorner, a)$
	4. $\mathsf{Sat}(\ulcorner \forall v_i \varphi\urcorner, a) \leftrightarrow \forall b\ (\forall j (j \neq i \rightarrow a(j) = b(j)) \rightarrow \mathsf{Sat}(\ulcorner \varphi\urcorner, b))$

a is the D-object x, whereas the latter will be characterized as a satisfaction predicate. For readability, we write $x, y, z, ..., i, j, k, l, ...$ and $a, b, c, ...$ for variables of sort $\mathfrak{o}$, $\mathfrak{s}$ and $\mathfrak{sq}$ respectively. Greek letters $\varphi, \psi, ...$ are taken to range over formulas of $\mathcal{L}_U$, $A, B, C \ldots$ over formulas of $\mathcal{L}_\mathsf{T}$.

We list the axioms of the theory $\mathsf{T}[U]$ in Table 1. The notation $\mathsf{T}[U]$ is suggestive: the theory of truth is interpreted as an operator that takes an arbitrary object theory U in predicate logic and yields a compositional, Tarski-style theory of truth for it. Axiom C is essentially due to [7] and tells us that we can always manipulate an assignment a so to create the assignment b which differs only in one element from it. This axiom is essential for the formulation of the theory of truth in terms of satisfaction (cfr. D.4). It is worth noticing that in virtue of C, for instance, it is possible to consider variable assignments as functions assigning arbitrary elements of D to the first n variables of $\mathcal{L}_U$ and 0 to all the remaining variables. We will resort to this possibility below (e.g. in the proof of Theorem 1). We also assume — in C, D.1, D.4 — the notational shortcut represented by the quantification over indices of variables of sort $\mathfrak{o}$ as variables of sort $\mathfrak{s}$. The axioms of S^1_2, as we mentioned already, guarantee a smooth formalization of the syntax of U.

In [33] extensions of $\mathsf{T}[U]$ were considered. They were obtained by adding to $\mathsf{T}[U]$ a 'syntactic' induction principle

$$A(\overline{0}) \wedge \forall k(A(k) \rightarrow A(\mathsf{S}k)) \rightarrow \forall k\, A(k) \qquad \text{(S-Ind)}$$

open to all formulas A of $\mathcal{L}_\mathsf{T}$ but restricted only to variables of sort $\mathfrak{s}$; in other words, we have in Eq. (S-Ind) that $\overline{0}$ is of sort $\mathfrak{s}$ and S of type $\mathfrak{s} \rightarrow \mathfrak{s}$. $\mathsf{T}[U]$ can be shown to be, in many cases, essentially weaker than $\mathsf{T}[U]$+(S-Ind). . Nonetheless, they both prove the following:

Lemma 6 *For any $\mathcal{L}_U$-formula $\varphi(v_{i_1}, ..., v_{i_n})$ with the free variables displayed,*[23] $\mathsf{T}[U]$ *proves*

$$\forall j_1, ..., j_n\ (\mathsf{Sat}(\ulcorner \varphi\urcorner(j_1/i_1, ..., j_n/i_n), a) \leftrightarrow \varphi(a(j_1), ..., a(j_n))) \qquad (6)$$

[23] We recall that, by assumption, we can assume that there are no closed terms in the language of U apart from variables.

By a very general argument (cfr. [33], Thm. 3.11), any model of U can be shown to be expandable to a model of $\mathsf{T}[U]$ and $\mathsf{T}[U]^I$.[24] Thus:

Lemma 7 $\mathsf{T}[U]$ *is a conservative extension of* U.

Other results on extensions of $\mathsf{T}[U]$ are more sensitive to the choice of U and we refer to [33] for their description.

The study of theories of truth with 'disentangled syntax' appears to be significant in several respects. Above all, they appear to be a direct axiomatization of the Tarskian picture of the metatheory ([41]): the theory of syntax is taken to be independent from the choice of the object theory. This framework thus renders the theory of truth quite general; Tarski-style axioms of truth can be provided for a wide range of base theories, even for theories that are not expressive enough to formalize their own syntax, such as theories formalizing portions of epistemological and metaphysical discourse. In [33, 35] and [36] further applications were considered: theories in the style of $\mathsf{T}[U]$+(S-Ind). Offer in fact the possibility of formalizing informal metamathematical practice, in which expressions are clearly distinguished from the objects of the intended domain of discourse of the mathematical object theory; moreover, they enable one to examine more closely the structure of the much discussed conservativeness argument against deflationism [29, 39].

In the following sections, on the other hand, we will mostly focus on characterizing the truth predicate of theories in the style of $\mathsf{T}[U]$ by relating them to foundationally relevant arithmetical systems and principles.

4 The Henkin-Feferman Construction

Henkin's proof of the completeness theorem displays a method for exhibiting a term model for an arbitrary, consistent set of sentences $\mathcal{S}$ of a first-order language. Generalizing and refining previous work of Bernays and Wang, Feferman offered in [10] a full formalization of the construction in extensions of PA plus a Π_1-sentence expressing the consistency of $\mathcal{S}$ in a canonical way. Following Visser, we refer to the resulting arithmetized model as the 'Henkin-Feferman construction'. Suitable improvements of this method will give us significant insights on the notion of truth captured by the theories introduced in §3, therefore they deserve an in-depth treatment.

Feferman's proof required at least Δ_2-induction to be carried out. But it is possible do better. The result can be improved significantly by employing the method of cuts. In the next lemma, due to Pavel Pudlák, it is shown how to find an optimal improvement Feferman's result by replacing PA with S^1_2, and even Q. We recall some of our assumptions: U has an axiom set that is Δ^b_1 in S^1_2. $\mathsf{Prov}_U(x)$ and Con_U are then taken to be defined accordingly.

[24] This also implies that for any model $\mathcal{M}$ of U there is an elementary extension $\mathcal{N}$ of it which satisfies $\mathsf{T}[U]$.

Lemma 8 $\left(\mathsf{S}^1_2\right)$. $\mathsf{S}^1_2 + \mathsf{Con}_U \rhd U$.

Proof We follow Visser's proof given in [44] to some extent, in which it is also shown that S^1_2 suffices as metatheory for the argument. The arithmetization method employed will thus be along the lines of the one mentioned in §2.1. One can check that the idea of looking for the leftmost consistent path in the tree whose paths are completions of U is miniaturized in the proof.

Reasoning in $\mathsf{S}^1_2 + \mathsf{Con}_U$, we expand $\mathcal{L}_U$ to the language $\mathcal{L}'_U$ with a countable set of new constants: for any $\mathcal{L}'_U$-sentences of the form $\exists v\varphi$ and $\forall v\varphi$, one adds to the language corresponding constants $c_{\exists v\varphi}$ and $c_{\forall v\varphi}$. Now we relate 0-1 strings to $\mathcal{L}'_U$-sentences via the following Σ^b_1-formula:

$$\begin{aligned}\mathsf{Pc}(x, y) :\leftrightarrow \quad & x \text{ is a sequence, } y \text{ is a sentence of } \mathcal{L}'_U \ \wedge \\ & \Big((x)_y = 0 \vee (\exists u < |x|)((x)_u = 1 \wedge y = \dot{\neg} u) \vee \\ & (\exists v, w < |x|)((x)_v = 0 \wedge \mathsf{var}(w) \wedge v = \ulcorner \exists w\varphi \urcorner \wedge y = \ulcorner \varphi(c_{\exists w\varphi}) \urcorner) \ \vee \\ & (\exists v, w < |x|)((x)_v = 1 \wedge \mathsf{var}(w) \wedge v = \ulcorner \forall w\varphi \urcorner \wedge y = \ulcorner \neg\, \varphi(c_{\forall w\varphi}) \urcorner)\Big)\end{aligned}$$

for an arbitrary formula φ of $\mathcal{L}'_U$ with one free variable. $\mathsf{Pc}(x, y)$ is clearly Σ^b_1 and it should be also clear the essential role of the smash function to perform the required substitutions. Let us denote with PC_x the set of sentences y so defined and let $\mathsf{Tree}(x)$ be a predicate expressing that PC_x is consistent with U. Crucially, if $\mathsf{Tree}(x)$, either $x^\frown\langle 0\rangle$ or $x^\frown\langle 1\rangle$ hold.[25] We define next the relation among binary sequences of 'being on the left', that is

$$x <_l y :\leftrightarrow (\exists u < \min\{\mathsf{lh}(x), \mathsf{lh}(y)\})(\forall v < u)((x)_v = (y)_v \wedge (x)_u = 0 \wedge (y)_u = 1),$$

so saying that x whose PC_x is consistent with U is 'on the leftmost path' becomes:

$$\mathsf{Path}(x) :\leftrightarrow \mathsf{Tree}(x) \wedge \neg\, \exists y\, (\mathsf{Tree}(y) \wedge y <_l x) \tag{7}$$

If we further define an order $\prec_i$ such that $x \prec_i y$ means x is an initial subsequence of y, we see that $\prec_i$ defines a linear order between elements of Path; that is, $\mathsf{S}^1_2 + \mathsf{Con}_U$ proves:

$$\forall x, y\, (\mathsf{Path}(x) \wedge \mathsf{Path}(y) \rightarrow (x \prec_i y \vee x = y \vee y \prec_i x)).$$

Now let us consider the numbers defined by the formula

$$J(x) :\leftrightarrow \exists y\, (\mathsf{Path}(y) \wedge x = |y|) \tag{8}$$

We claim that $J(x)$ is inductive. Obviously, the empty sequence $\langle\rangle$ is in J. Also, by the reasoning above, if x is in J, then there is some y in Path such that $x = |y|$, then also $x + 1$ is in J. By Lemma 1, we shorten $J(x)$ to a cut $I(x)$.

[25] Here $^\frown$ denoted sequences concatenation.

The interpretation $(\cdot)^H$, corresponding to a model of U, is specified by the domain of Henkin constants

$$\delta(x) :\leftrightarrow \quad \exists y \left(I(y) \wedge \mathsf{Sent}_{\mathcal{L}'_U}(y) \wedge (y = \ulcorner \exists v \varphi \urcorner \vee y = \ulcorner \forall v \varphi \urcorner) \wedge x = \dot{c}_y \right),$$

by a substitution function

$$\mathsf{sb}(\ulcorner \varphi \urcorner, y) := \text{ the (code of the) result of substituting the elements of } y \text{ for the free variables in } \varphi, \text{ where } y \text{ is a sequence of } \delta\text{-objects;}$$

and by a predicate

$$H(x) :\leftrightarrow I(x) \wedge \mathsf{Sent}_{\mathcal{L}'_U}(x) \wedge \exists z \, (\mathsf{Path}(z) \wedge |z| \leq x \wedge x \in \mathsf{PC}_z).$$

In particular one has, by abuse of notation and denoting with $\bar{x}$ strings of objects:

$$R^H(\bar{x}) :\leftrightarrow \delta(\bar{x}) \wedge H\left(\mathsf{sb}\left(\dot{R}\bar{v}, \bar{x}\right)\right) \quad \text{for any relation symbol } R \text{ in } \mathcal{L}_U \tag{9}$$

Moreover, provably in $\mathsf{S}^1_2 + \mathsf{Con}_U$, we have:

$$\forall x \in \mathsf{Sent}^I_{\mathcal{L}'_U} \left(H(\dot{\neg} x) \leftrightarrow \neg H(x) \right)$$

$$\forall x, y \in \mathsf{Sent}^I_{\mathcal{L}'_U} \left(H(x \mathbin{\dot{\wedge}} y) \leftrightarrow H(x) \wedge H(y) \right)$$

$$\forall \ulcorner \varphi(v) \urcorner \in \mathsf{Fml}^I_{\mathcal{L}'_U} \left(H(\dot{\forall} v \ulcorner \varphi \urcorner) \leftrightarrow \forall w (\delta(w) \rightarrow H(\mathsf{sb}(\ulcorner \varphi(v) \urcorner, w))) \right)$$

In the last line, $\varphi(v)$ is a formula with only one free variable. Crucially we have on the one hand, with $\mathsf{Prov}_U(x)$ as before,

$$\forall x \left(\mathsf{Sent}^I_{\mathcal{L}_U}(x) \wedge \mathsf{Prov}_U(x) \rightarrow H(x) \right), \tag{10}$$

which is an immediate consequence of the fact that if $x \in \mathsf{Sent}_{\mathcal{L}_U}$ and $x \in I$, then either $x \in \mathsf{PC}_y$ or $\dot{\neg} x \in \mathsf{PC}_y$. On the other, we have that for all $\varphi \in \mathsf{Fml}_{\mathcal{L}'_U}$

$$\forall y \left(\delta(\bar{y}) \rightarrow \left(H(\mathsf{sb}(\ulcorner \varphi \urcorner, \bar{y})) \leftrightarrow \varphi^H(\bar{y}) \right) \right). \tag{11}$$

Now let ψ be a $\mathcal{L}_U$-sentence provable in U. By the assumption on $\mathcal{L}_U$, $\mathsf{S}^1_2 + \mathsf{Con}_U$ proves that $\ulcorner \psi \urcorner$ is in I (cf. footnote 18) and that $\mathsf{Sent}_{\mathcal{L}_U}(\ulcorner \psi \urcorner)$. Thus $H(\ulcorner \psi \urcorner)$. Thus, by Eq. 11, ψ^H.

All these conversions, as shown in [44], are available in S^1_2, although a finer grained notion of interpretability—smooth interpretabilty—has to be employed. □

Finally Eq. (10), combined with Lemma 5 and our assumptions on U, yields:

Corollary 1 $\left(\mathsf{S}^1_2\right)$ $\mathsf{Q} + \mathsf{Con}_U \rhd U$.

This completes the preliminary work needed to characterize the truth predicate of theories of the form $\mathsf{T}[U]$. We conclude the section with a brief digression on the

Table 2 $\mathsf{CT}\restriction$

CT1	$\forall x,y\ \left(\mathsf{Cterm}_{\mathcal{L}}(x) \wedge \mathsf{Cterm}_{\mathcal{L}}(y) \rightarrow (\mathsf{Tr}\ x \doteq y \leftrightarrow \mathsf{val}(x) = \mathsf{val}(y))\right)$
CT2	$\forall x\ (\mathsf{Sent}_{\mathcal{L}}(x) \rightarrow (\mathsf{Tr}\ \dot{\neg} x \leftrightarrow \neg \mathsf{Tr}\ x))$
CT3	$\forall x,y\ (\mathsf{Sent}_{\mathcal{L}}(x \dot{\wedge} y) \rightarrow (\mathsf{Tr}\ x \dot{\wedge} y \leftrightarrow (\mathsf{Tr}\ x \wedge \mathsf{Tr}\ y)))$
CT4	$\forall x \forall y\ (\mathsf{Sent}_{\mathcal{L}}(\dot{\forall} yx) \rightarrow (\mathsf{Tr}\ \dot{\forall} yx \leftrightarrow \forall z\ (\mathsf{Cterm}_{\mathcal{L}}(z) \rightarrow \mathsf{Tr}\ x(z/y))))$

analysis of the arithmetized models arising Henkin-Feferman construction when U is taken to be PA.

4.1 Classical Compositional Truth and the Henkin Feferman Construction[26]

The theory $\mathsf{CT}\restriction$ is formulated in the language $\mathcal{L}$ of arithmetic plus a unary predicate Tr not allowed to appear into instances of the schema of induction; its axioms are the axioms of PA and the sentences displayed in Table 2.

Let $\mathcal{M} \vDash \mathsf{PA}$ and $S \subseteq \mathcal{M}$. S is a *full truth class* for $\mathcal{M}$ if and only if $(\mathcal{M}, S) \vDash \mathsf{CT}\restriction$.[27] Moreover, S is a *partial*, nonstandard truth class for the nonstandard $\mathcal{M} \vDash \mathsf{PA}$ if and only if there is some $a \in \mathcal{M} - \mathbb{N}$ such that $(\mathcal{M}, S)$ satisfies $\mathsf{CT1}$-$\mathsf{4}$ for sentences whose codes are $< a$. A remarkable fact concerning truth classes for models of PA is the following:[28]

Fact 1 (Lachlan [31], Kaye [28]) . *If $\mathcal{M} \vDash \mathsf{PA}$ is countable, nonstandard and S a partial truth class for $\mathcal{M}$, then $\mathcal{M}$ is recursively saturated.*

As we have seen the Henkin-Feferman construction gives rise to models equipped with truth predicates: truth-theorists have thus recently asked for a comparison between models of PA admitting a truth class and models arising from the Henkin-Feferman construction.[29] A natural question appears to be: *given a countable model $\mathcal{M} \vDash \mathsf{Q} + \mathsf{Con}_{\mathsf{PA}}$, is its internal model $\mathcal{H} \vDash \mathsf{PA}$ given by Lemma 5 and Lemma 8 recursively saturated* ?

We briefly recall how to obtain $\mathcal{H}$ in $\mathcal{M}$: one considers an interpretation N of $\mathsf{S}^1_2 + \mathsf{Con}_{\mathsf{PA}}$ in $\mathsf{Q} + \mathsf{Con}_{\mathsf{PA}}$ on a definable cut N_0. The Henkin Feferman construction (Lemma 8) is then carried out within N_0 to construct the internal model $\mathcal{H}$ of PA. As it was shown in the proof of Lemma 8, the truth predicate of the model works for sentences in a $\mathcal{M}$-definable subcut I of N_0. Furthermore, Pudlák's lemma can be

[26]This digression has been thoroughly rewritten after receiving the comments of an anonymous referee that highlighted the many weak points of the original formulation and suggested detailed and clear improvements. I thank him for the precious and insightful advices.

[27]Cf. [23, Definition 8.7].

[28]Cf. [28, p. 228]. A brief *legenda*. Let $\mathcal{M} \vDash PA$ be nonstandard. A *type* $P(\bar{x})$ over PA is a set of formulas $\varphi(\bar{x})$ such that the set of all $\varphi(\bar{c})$ with $\varphi(\bar{x}) \in P(\bar{x})$ and $\bar{c}$ a finite sequence of new constants is consistent with PA. $P(\bar{x})$ is realized in $\mathcal{M}$ if and only if there is a sequence $\bar{a} \in |\mathcal{M}|$ such that $\mathcal{M} \vDash \varphi(\bar{a})$ for all $\varphi(\bar{x}) \in P(\bar{x})$. $P(\bar{x})$ is recursive if and only if the set of codes of formulas $\varphi(\bar{x})$ in $P(\bar{x})$ is recursive. A model $\mathcal{M}$ of PA is recursively saturated if and only if every recursive type on $\mathcal{M}$ is realized in $\mathcal{M}$.

[29]We have mainly [15] in mind.

applied as S_2^1 is available in N_0: there is thus a $\mathcal{M}$-cut I^*, a subcut of N_0, and an $\mathcal{M}$-definable embedding of I^* into an initial segment I^{**} of the $\mathcal{H}$-numbers as seen by $\mathcal{M}$ via the composition of N and the interpretation H given by the Henkin-Feferman construction. By employing the shortening techniques it is possible to identify I^* and I above: as a result, from the perspective of $\mathcal{M}$, one might construct a $\mathcal{M}$-definable subcut of N_0 shared by $\mathcal{M}$-numbers and $\mathcal{H}$-numbers. In this way $\mathcal{H}$ is coded by numbers that are actually an initial segment of the $\mathcal{H}$-numbers. Now by combining what we have just said and Fact 1, we notice that $\mathcal{H}$ will be recursively saturated unless any $I^* \subseteq N_0$ defines the standard numbers.[30] If some I^* does not define the standard numbers, in fact, the truth predicate of the Henkin-Feferman construction will give rise to a nonstandard truth class; by Fact 1, $\mathcal{H}$ would be recursively saturated.

A further question is whether the $\mathsf{Q} + \mathsf{Con}_{\mathsf{PA}}$*-subcut* I *of* N_0 *to which the truth predicate of the Henkin-Feferman construction applies is in fact a proper initial segment of the* $\mathcal{H}$*-numbers.* It turns out that there is no definite answer to this question, although it is possible to construct models of $\mathsf{Q} + \mathsf{Con}_{\mathsf{PA}}$ in which the truth predicate of the Henkin Feferman construction indeed applies only to a proper initial segment of the numbers in the internal model $\mathcal{H}$. To construct such a model it would suffice to display a model $\mathcal{N}$ of $\mathsf{Q} + \mathsf{Con}_{\mathsf{PA}}$ in which there are some $\mathcal{N}$-definable cuts that do not satisfy PA. An example, due to Albert Visser, is obtained as follows: let A be a finitely axiomatized subtheory of PA extending Q and define the theory

$$V := A + \mathsf{Con}_{\mathsf{PA}} + \{\neg\, \mathsf{Con}^J(A + \mathsf{Con}_{\mathsf{PA}}) \mid J \text{ is an} (A + \mathsf{Con}_{\mathsf{PA}})\text{-definable cut}\}$$

This theory is consistent by Pudlák's strengthening of Gödel's Second Incompleteness Theorem. Let $\mathcal{N} \vDash V$ and assume, seeking a contradiction, that PA is available in any $\mathcal{N}$-cut J. Again by downwards persistence of Π_1-sentences, $\mathsf{Con}_{\mathsf{PA}}$ is available in J together with $\neg\, \mathsf{Con}(A + \mathsf{Con}_{\mathsf{PA}})$. But PA is essentially reflexive,[31] thus $\mathsf{Con}_{\mathsf{A}}$, so $\neg\, \mathsf{Con}_{\mathsf{PA}}$ will hold in J. Since $\mathsf{Con}_{\mathsf{PA}}$ is in J, we reached the desired contradiction and displayed a model of $\mathsf{Q} + \mathsf{Con}_{\mathsf{PA}}$ in which no cuts satisfy PA.

In the following two sections, as promised, we focus on the mutual interpretability of $\mathsf{T}[U]$ — and variants thereof — and $\mathsf{Q}+\mathsf{Con}_U$ for an arbitrary choice of U. We will see that, when U is not finitely axiomatized, a further axiom stating that all axioms of U belong to our special class of true sentences is needed in order to obtain the result.

5 Truth, Reflection and Arithmetized Models

5.1 Reflection without Induction

Results akin to the ones described in this section seem to be known as folklore. The author became aware of them from (forthcoming) work by Richard Heck [24]. We

[30] As the same referee has pointed out to the author, cardinal arithmetic represents a simple example of a nonstandard model of arithmetic $\mathsf{Q} + \mathsf{Con}_{\mathsf{PA}}$ in which all definable cuts isomorphic to cuts of internal models of PA are the standard numbers. It would be interesting to find examples of models of stronger theories that enjoy this property.

[31] That is, any pure extension W of it proves $\mathsf{Con}_{\mathsf{B}}$ for any finite $B \subseteq W$.

show that if a compositional theory of truth is strong enough to prove that all axioms of the base theory are true, then the former is not interpretable in the base theory as it will prove the consistency of the base theory on a definable cut. We focus on $\mathsf{T}[U]$, but the arguments given are more generally applicable (cf. Heck's paper). Some remarks in this direction will be given at the end of the section. Let us first fix some notation.

We say that U is *finitely axiomatized* if its set of nonlogical axioms is finite. We call U *schematic*, in the sense of [11, 32], if U is axiomatized by a finite set of sentences plus the substitution instances of a finite number of *schemata*. We understand schemata as formulas of the form $\Phi(X)$ — in which X is a free predicate variable; their instances are then formulas of $\mathcal{L}_U$ resulting from replacing, in $\Phi(X)$, each occurrence of $X(t_1, .., t_n)$ by a fixed formula $\psi(t_1, ..., t_n)$ together with a suitable renaming of bound variables in Φ in order to avoid clashes.

We define

$$\mathsf{AxT}_U :\leftrightarrow \forall a\, \forall k\, (\mathsf{Ax}_U(k) \rightarrow \mathsf{Sat}(k, a))$$

where $\mathsf{Ax}_U(k)$ is again as in §2.1, suitably relativized to the syntactic sort. We will also occasionally employ the sentence

$$\mathsf{ScT}_U :\leftrightarrow \forall a\, \forall k\, (\mathsf{Sc}_U(k) \rightarrow \mathsf{Sat}(k, a))$$

where $\mathsf{Sc}_U(k)$ is a Δ_1^b predicate expressing that k is a substitution instance of one of the (finitely many) axiom schemata of U, if present.

Con_U and $\mathsf{Prov}_U(k)$ are as above. We will show:

Proposition 2

(i) *if U is finitely axiomatized,* $\mathsf{T}[U] \rhd \mathsf{Q} + \mathsf{Con}_U$*;*
(ii) $\mathsf{T}[U] + \mathsf{AxT}_U \rhd \mathsf{Q} + \mathsf{Con}_U$.

Proposition 2 is a corollary of the stronger

Lemma 9

(i) *if U is finitely axiomatized,* $\mathsf{T}[U] \rhd \mathsf{S}_2^1 + \mathsf{Con}_U$*;*
(ii) $\mathsf{T}[U] + \mathsf{AxT}_U \rhd \mathsf{S}_2^1 + \mathsf{Con}_U$.

The intuitive picture of the strategy is as follows. The crucial step is how to suitably relativize Con_U to a cut in which S_2^1 is available. Reasoning in the more general case in which U is arbitrary, we first identify a $\mathsf{T}[U]$ (resp. $\mathsf{T}[U] + \mathsf{AxT}_U$)-cut whose numbers support some lemmata of syntactic nature concerning U. We then prove a special form of global reflection for U on a suitable cut shortening the original one. This will suffice to give us consistency of U on this cut.

The lemmata of syntactic nature that we are going to prove first are needed to establish useful facts such as 'substitution of identicals is preserved under the scope of the satisfacton predicate'. Claims of this sort are usually obtained by induction on the complexity of the relevant formula φ falling under the scope of the truth or

satisfaction predicates.[32] In practice, this is not possible as we do not have such an induction principle in $\mathsf{T}[U]$. Therefore we employ $\mathsf{T}[U]$ (resp. $\mathsf{T}[U] + \mathsf{AxT}_U$)-definable cuts.

These syntactic claims are in turn needed to prove that logical axioms are true and that rules of inference of the logical calculus in which U is formulated preserve truth if formulas belonging to some suitable initial segment of our numbers are considered. We notice in fact that, by the very construction of $\mathsf{T}[U]$, no nonlogical axiom schema of U is extended to contain nonlogical vocabulary from $\mathcal{L}_\mathsf{T} - \mathcal{L}_U$. So, for instance, if U is ZFC, elements $\mathcal{L}_\mathsf{T} - \mathcal{L}_U$ are not allowed into instances of the schema of replacement. This does not mean, however, that logical rules and logical axiom schemata of U are not extended to $\mathcal{L}_\mathsf{T}$. This is for the main reason that if we restrict logic as well, it would be hard to justify $\mathsf{T}[U]$ as a *theory*, at least in the most straightforward reading of the word. If U is formulated in a Hilbert style calculus in which Modus Ponens and Generalization are the only rules of inference,[33] the lemmata in question enable one to deal with axioms and rules of the form

$$\forall v_i \varphi \rightarrow \varphi(v_j/v_i); \qquad \text{(Gen)} \quad \text{if } \Gamma \vdash \varphi(v_i), \text{ then } \Gamma \vdash \forall v_i \varphi$$

with the usual conditions on legitimate substitution and on Γ.[34] Let

$$a \overset{i}{\sim} b :\leftrightarrow \forall j\, (j \neq i \rightarrow a(j) = b(j)).$$

Moreover, let $\mathsf{lc}(k)$ be a Σ^b_1 function in S^1_2 keeping track of the logical complexity of the formula k.[35] For our purposes it suffices to define the logical complexity of a formula as the number of propositional connectives and quantifiers in it, we do not require finer grained notions. We have

Lemma 10 *The formula $K(m)$ defined as* [36]

$$\mathsf{Freev}(i,k) \wedge \mathsf{lc}(k) \leq m \wedge a \overset{i}{\sim} b \wedge a(i) = b(j) \rightarrow (\mathsf{Sat}(k,a) \leftrightarrow \mathsf{Sat}(k(j/i), b))$$

is inductive in $\mathsf{T}[U]$.

Proof $K(m)$ just says: if a formula $\varphi(v_i)$ of logical complexity less than m is satisfied by a variable assignment a, b differs from a only in what it assigns to i, and $a(i)$ is just $b(j)$, then b satisfies $\varphi(j/i)$.

$K(0)$, that is the case in which k is the code of an atomic formula, follows immediately from D.1. Assuming $K(n)$ we prove $K(n+1)$. All cases are quite

[32] Cfr. [23, § 8.6].

[33] As it is widespread practice when arguments of the kinds discussed here are involved.

[34] Here Γ is simply a finite set of formulas, not a multiset.

[35] For clarity, we notice that $\mathsf{lc}(.)$ formalizes a function of type $s \rightarrow s$.

[36] For readability, we omit quantification over parameters.

straightforward given the inductive hypothesis. For instance in the case of the universal quantifier, we have:

$$\begin{aligned}
\mathsf{Sat}(\ulcorner \forall v_k \varphi(v_i) \urcorner, a) &\leftrightarrow \forall c \overset{k}{\sim} a\ \mathsf{Sat}(\ulcorner \varphi(v_k, v_i) \urcorner, c) && \text{D.4}\\
&\leftrightarrow \forall d \overset{k}{\sim} b\ \mathsf{Sat}(\ulcorner \varphi \urcorner(j/i)), && \text{by } K(n) \text{ for suitable } d\\
&\leftrightarrow \mathsf{Sat}(\ulcorner \forall v_k \varphi(v_j) \urcorner, b) && \text{D.4}
\end{aligned}$$

□

Lemma 1 then immediately gives us

Corollary 2 *There is a* $\mathsf{T}[U]$*-cut N in which the axioms of* S^1_2 *are available and that shortens* $K(k)$*.*

By an analogous strategy, we obtain

Lemma 11 $\mathsf{T}[U]$ *proves that the formula* $M(m)$*:*[37]

$$\neg\mathsf{Freev}(i,k) \wedge \mathsf{lc}(k) \leq m \wedge a \overset{i}{\sim} b \rightarrow (\mathsf{Sat}(k,a) \leftrightarrow \mathsf{Sat}(k,b))$$

is inductive.

Corollary 3 *There is a* $\mathsf{T}[U]$*-cut I that shortens* $M(k)$ *in which the axioms of* S^1_2 *are available.*

Let us call $L(x)$ a subcut shared by the cuts N and I constructed in Corollaries 2 and 3. Let now AxLT_U the $\mathcal{L}_T$ sentence stating that all logical axioms of $\mathcal{L}_U$ are true. We have

Corollary 4 *There is a* $\mathsf{T}[U]$*-cut in which* AxLT_U *holds.*

Proof Sketch Axioms that are of the form of single sentences follow from the provability of the T-biconditionas in $\mathsf{T}[U]$ (Lemma 6). Propositional schemata follow from compositional axioms. Axioms for quantification are obtained by a crucial contribution of Corollary 2. In fact, any $\mathsf{T}[U]$-cut in which Corollary 2 holds would work, *a fortiori* we can employ $L(x)$. □

The corollaries just stated are somewhat dependent on the choices of the calculus in which U is formulated. It is worth noticing that although it is in principle possible to define AxT_U in such a way to stipulate the truth of the logical axioms U, still we would need Corollaries 2 and 3 to hold in a suitable $\mathsf{T}[U]$-definable cut to guarantee the truth-preserving character of rules of inferences involving quantification, if present, essential to obtain Lemma 12 below. A further way out would consist in formulating U in a Hilbert-style calculus in which Modus Ponens is the only rule of

[37] Again quantification over parameters is omitted.

inference.[38] We could then be dispensed from working in subcuts of L above. At any rate, the strategy just outlined is clearly superior as it renders the results below less sensitive to the choice of the calculus.

Let now J be a cut, we define

$$\mathsf{Con}_U^J :\leftrightarrow \neg\exists k\,(J(k) \wedge \mathsf{Proof}_U(k, \ulcorner 0 = 1 \urcorner))$$

Lemma 12

(i) *Let U be finitely axiomatized. Then there is a $\mathsf{T}[U]$-cut J such that*

$$\mathsf{T}[U] \vdash \mathsf{Con}_U^J;$$

(ii) *There is a $\mathsf{T}[U] + \mathsf{AxT}_U$-cut J such that*

$$\mathsf{T}[U] + \mathsf{AxT}_U \vdash \mathsf{Con}_U^J.$$

We first notice that in virtue of Lemma 6, AxT_U becomes provable in $\mathsf{T}[U]$ already if U is finitely axiomatized.

Proof of Lemma 12 We prove (i) and (ii) simultaneously. Let J_0 be a cut in which Corollaries 2 and 3 hold and in which the logical axioms of U have been shown to be true.

We recall that $\mathsf{Prf}_U(k)$ is a Δ_1^b-formula in S_2^1 expressing that k is the code of a U-proof, and $\mathsf{Ist}(k)$ a Σ_1^b function yielding, when applied to a sequence k, the last element of k. One first notices that the following formula $B(k)$ is inductive

$$\forall l \leq k\ \Big((\forall i \leq \mathsf{lh}(l))(\mathsf{Ic}((l)_i) \in J_0) \wedge \mathsf{Prf}_U(l) \rightarrow \forall a\ \mathsf{Sat}(\mathsf{Ist}(l), a)\Big) \tag{12}$$

The crucial step is to obtain $B(k+1)$. If $\mathsf{Ist}(k+1)$ is a logical axiom, then its truth follows by assumption and Corollary 4. If it is a nonlogical axiom and U is finitely axiomatized, then $\mathsf{Ist}(k+1)$ is satisfied by all sequences by Lemma 6; otherwise we employ the assumption AxT_U. If on the other hand $\mathsf{Ist}(k+1)$ is obtained via Modus Ponens or (Gen), then compositional axioms D.1-D.4, together with Corollary 3, suffice to obtain the result. By Lemma 1 we shorten B to a cut J.

In other words, we have reflection in the cut J:

$$\forall k\ \big(\mathsf{Fml}_{\mathcal{L}_U}(k) \wedge \exists m\,(J(m) \wedge \mathsf{Proof}_U(m, k)) \rightarrow \forall a\,\mathsf{Sat}(k, a)\big) \tag{13}$$

We notice that we tacitly assumed a monotone coding here as in the proof of Lemma 8: if k is a sequence, then $(k)_i < k$ with $i < \mathsf{lh}(x)$. We have thus $\mathsf{Fml}_U^J(k)$ in Eq. (13).

Given Eq. 13, we can conclude Con_U^J in the usual way, that is with the help of Lemma 6. □

Lemma 9 is now readily obtained by taking J as the domain of the interpretation. We thus also have Proposition 2.

As observed by Heck, this holds for *any* Tarski-style typed theory of truth proving the truth of all axioms of the base theory, even in the case of theories constructed

[38] Cfr. for instance [9].

in the usual, 'entangled' way. For our present purposes, however, it was sufficient to formulate our arguments in terms of the particular setting under examination. The proofs above can be easily adapted to the usual setting. We just mention few examples. We write $\mathsf{CT}\restriction[U]$ as the results of adding a full truth class to an arbitrary model of U.

Observation 1 *Let U be a finitely axiomatized extension of S^1_2, then $\mathsf{CT}\restriction[U]$ proves the consistency of U on a cut.*

Observation 2 *Let U be a schematic extension of S^1_2, then $\mathsf{CT}\restriction[U]+$'all axioms of U are true' proves the consistency of U on a cut.*

As a consequence, $\mathsf{CT}\restriction[U]$ (resp. $\mathsf{CT}\restriction[U]+$'all axioms of U are true') interprets $\mathsf{S}^1_2 + \mathsf{Con}_U$. To give some familiar examples, we have for instance,

(i) $\mathsf{CT}\restriction[\mathsf{I}\Sigma_n]$ proves the consistency of $\mathsf{I}\Sigma_n$ for each n and thus $\mathsf{CT}\restriction[\mathsf{I}\Sigma_n]$ interprets $\mathsf{S}^1_2 + \mathsf{Con}_{\mathsf{I}\Sigma_n}$;
(ii) $\mathsf{CT}\restriction+\mathsf{AxT}_{\mathsf{PA}}$ proves the consistency of PA on a cut and thus $\mathsf{CT}\restriction+\mathsf{AxT}_{\mathsf{PA}}$ interprets $\mathsf{S}^1_2 + \mathsf{Con}_{\mathsf{PA}}$.

As far as the author knows, it is an open problem whether $\mathsf{CT}\restriction[\mathsf{I}\Sigma_n]$ interprets $\mathsf{I}\Sigma_1 + \mathsf{Con}_{\mathsf{I}\Sigma_n}$ for all n, or whether $\mathsf{CT}\restriction+\mathsf{AxT}_{\mathsf{PA}}$ interprets $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$.

Observation 3 *Let U contain S^1_2. If $\mathsf{CT}\restriction[U]$ is interpretable in U, then $\mathsf{CT}\restriction[U]$ does not prove that all axioms of U are true.*

In the light of Observation 3 and of the fact that $\mathsf{CT}\restriction[\mathsf{PA}]$ and $\mathsf{CT}\restriction[\mathsf{ZF}]$ are interpretable in PA and ZF respectively, we have, modulo a suitable formalization of the syntax,

(i) $\mathsf{CT}\restriction[\mathsf{PA}]$ does not prove that all instances of induction of PA are true;
(ii) $\mathsf{CT}\restriction[\mathsf{ZF}]$ does not prove that all instances of replacement are true.

More on these arguments applied to the usual typed setting can be found in [24]. In the next section we focus on the opposite direction. We show how the Henkin-Feferman construction can give us an interpretation of the truth predicate of $\mathsf{T}[U]$.

5.2 Truth with Disentagled Syntax and The Henkin-Feferman Construction

The idea that inspires this section is that, unlike the case of $\mathsf{CT}\restriction[U]$, the special satisfaction class defined by $\mathsf{T}[U]$ and variants thereof is indeed uniformly comparable with the truth predicate of the Henkin-Feferman construction as it is always interpretable, loosely speaking, as a truth predicate for 'small numbers'. We first consider for simplicity the case in which U is finitely axiomatized.

Theorem 1 (S^1_2) *Let U be finitely axiomatized. Then $\mathsf{S}^1_2 + \mathsf{Con}_U \rhd \mathsf{T}[U]$.*

Proof The idea of the proof is based upon the observation that the Henkin-Feferman construction, as presented in §4, gives us the means to interpret the truth predicate for the language of U given by $\mathsf{T}[U]$.

We define a translation $(.)^F$ specified by a triple of domains $(\delta^{\mathfrak{s}}, \delta^{\mathfrak{o}}, \delta^{\mathfrak{sq}})$ and a mapping of symbols of $\mathcal{L}_T$ into formulas of $\mathcal{L}_U$ that for simplicity we won't distinguish from $(.)^F$ itself. We recall the proof of Lemma 8: the domain of our interpretation was represented by Henkin constants in the $\mathsf{S}^1_2 + \mathsf{Con}_U$-cut I of numbers associated with the leftmost consistent — with the completion of U — path in the full binary tree. Constants in this cut will represent also the domain of quantifiers of the 'object-theoretic' part of $\mathcal{L}_T$:

$$\delta^{\mathfrak{o}}(x) :\leftrightarrow \quad \exists y \left(I(y) \wedge \mathsf{Sent}_{\mathcal{L}'_U}(y) \wedge (y = \ulcorner \exists v \varphi \urcorner \vee \right.$$
$$\left. y = \ulcorner \forall v \varphi \urcorner) \wedge x = \dot{c}_y \right),$$

Quantifiers ranging over variable assignments will be relativized to finite sequences of elements of $\delta^{\mathfrak{o}}$.

$$\delta^{\mathfrak{sq}}(x) :\leftrightarrow x \text{ is a (finite) sequence and } (\forall y \in x)(\delta^{\mathfrak{o}}(y))$$

where $x \in y$ is defined as in [3, §2.5]. As we have mentioned on page 10, the axiom C governing sequences of variable assignments does not prevent one to interpret them as finite sequences. All syntactic objects are relativized to the cut I; in other words, $\delta^{\mathfrak{s}} = I$.

Nonlogical constants are mapped by $(\cdot)^F$ to corresponding formulas of $\mathcal{L}_A$ in the following way, where $F(\cdot)$ and $\mathsf{sb}(\cdot, \cdot)$ are the truth predicate and the substitution function given in the proof of Lemma 8. We also assume a suitable machinery for renaming bound variables to avoid clashes. We thus have:

$$\begin{aligned}
(R)^F(x_1, ..., x_n) &:= F(\mathsf{sb}(\ulcorner R(\bar{v}) \urcorner, \langle x_1, ..., x_n \rangle)) \text{ for } n\text{-ary relations } \mathrm{R} \in \mathcal{L}_U \\
(P)^F(x_1, ..., x_n) &:= P(x_1, ..., x_n) \text{ for any relation symbol } P \text{ of sort } (\mathfrak{s}, ..., \mathfrak{s}); \\
(\mathsf{Sat})^F(x, y) &:= F(\mathsf{sb}(x, y)); \\
((\cdot)\cdot)^F(x, y, z) &:= ((x)_y = z \wedge y < \mathsf{lh}(x)) \vee (y \geq \mathsf{lh}(x) \wedge z = 0)
\end{aligned}$$

in the last line, $(x)_y$ outputs the y^{th} element of the finite sequence x.[39]

We check that $\mathsf{S}^1_2 + \mathsf{Con}_U$ satisfies the translation of the axioms of $\mathsf{T}[U]$. Axioms of U are obtained by Lemma 8. Axioms of S^1_2 are assumed to be available in the cut I. The translation of C. (cfr. Table 1) follows from the sequentiality of S^1_2. To verify the translations of the axioms for Sat, we focus on the instructive cases of atomic formulas and quantified formulas.

Ad D.1: $\mathsf{Sat}(\ulcorner R(v_1, ..., v_n) \urcorner, a) \leftrightarrow R(a(1), ..., a(n))$

[39] Again a full definition can be found in [3, 4].

Let us assume, without loss of generality, that R is of sort $(\mathfrak{o}, \mathfrak{o})$. We have, for arbitrary $x, y \in I$ and $z \in \delta^{\mathfrak{sq}}$, $F(\mathsf{sb}(\dot{R}xy, z))$. But by Eq. (9) in the proof of Lemma 8, this is just $R^F((z)_x, (z)_y)$.

Ad D.4: $\mathsf{Sat}(\ulcorner \forall v_i \varphi \urcorner, a) \leftrightarrow \forall b(\forall j(j \neq i \rightarrow a(j) = b(j)) \rightarrow \mathsf{Sat}(\ulcorner \varphi \urcorner, b))$

Let x be the code of a formula of $\mathcal{L}'_U$ with only the variable coded by y free and let x be in I. Thus $y \in I$. Moreover, let z be in $\delta^{\mathfrak{sq}}$. We notice that if $\mathsf{Sent}^I_{\mathcal{L}_U}(\dot{\forall} vx)$, then $F(\mathsf{sb}(x, \langle y \rangle)) \leftrightarrow F(x(y/v))$, where $r(s/w)$ the standard substitution function — provably total in S^1_2 — replacing free variables with other terms in formulas.

For the left to right direction, if $F(\mathsf{sb}(\dot{\forall} yx, z))$, then obviously $F(\dot{\forall} yx)$. Thus by the properties of F for all $u \in \delta^{\mathfrak{o}}$, $F(x(u/y))$. If we now consider an arbitrary sequence $v \in \delta^{\mathfrak{sq}}$ that differs from z only in what it assigns to the variable coded by y, then we have $F(\mathsf{sb}(x, v))$ as what v assigns to y can be taken to be arbitrary by assumption.

For the other direction, given that for all $v \in \delta^{\mathfrak{sq}}$ that differ from an arbitrary $z \in \delta^{\mathfrak{sq}}$ only in what they assign to the variable coded by y we have $F(\mathsf{sb}(x, v))$, we assume that there is $w \in \delta^{\mathfrak{o}}$ such that $\neg F(x(w/y))$. Now consider the sequence s that is exactly like z but it assigns w to y. We thus have $F(\mathsf{sb}(x, s))$, contradicting $\neg F(x(w/y))$. □

By Lemma 5 and our assumption on U we also have

Corollary 5 $\left(\mathsf{S}^1_2\right)$ *Let U be finitely axiomatized. Then* $\mathsf{Q} + \mathsf{Con}_U \rhd \mathsf{T}[U]$.

Moreover, since U is finitely axiomatized by assumption, we have $\mathsf{T}[U] \vdash \mathsf{AxT}_U$. Therefore by Propositions 1 and 2, we can characterize the canonical consistency statement Con_U for U as the unique solution, modulo provability in $\mathsf{I}\Delta_0(\mathsf{exp})$, to the equation between $\mathsf{T}[U]$ and the result of adding to Q a Π^0_1-sentence of the language of U.

Proposition 3 *Let U be finitely axiomatized.* Con_U *is the unique Π_1-sentence π, modulo provable equivalence in* $\mathsf{I}\Delta_0(\mathsf{exp})$*, such that*

$$\mathsf{T}[U] \equiv \mathsf{Q} + \pi.$$

To obtain the analogue of Proposition 3 for the case in which U is not finitely axiomatizable, we focus instead on the theory $\mathsf{T}[U] + \mathsf{AxT}_U$. The reason was made clear in the previous section: if we don't have AxT_U, we don't know how to prove the consistency of U on a cut, and this was shown to be essential to define the interpretation of $\mathsf{Q} + \mathsf{Con}_U$ in a uniform way.[40]

[40] Of course there may be cases in which $\mathsf{T}[U]$, with U not finitely axiomatizable, but still it can prove the consistency of U on a cut.

Corollary 6 $\mathsf{S}_2^1 + \mathsf{Con}_U \rhd \mathsf{T}[U] + \mathsf{AxT}_U$.

Proof To obtain the result, it then suffices to show in $\mathsf{S}_2^1 + \mathsf{Con}_U$ the translation of

$$\forall k\ (\mathsf{Ax}_U(k) \rightarrow \forall a\ \mathsf{Sat}(a, k)), \tag{14}$$

However, from the proof of Lemma 8 we know that reflection holds in the cut I (cfr. 10), that is in the cut defined by shortening (8) in Lemma 8. Thus we have

$$\forall x\ (\mathsf{Sent}_{\mathcal{L}_U}(x) \wedge I(x) \wedge \mathsf{Ax}_U(x) \rightarrow F(x)) \tag{15}$$

Therefore, it suffices to obtain in $\mathsf{S}_2^1 + \mathsf{Con}_U$ that,

$$\text{for all } x \in \mathsf{Sent}^I_{\mathcal{L}_U}, \mathsf{Ax}^F_U(x) \rightarrow \mathsf{Ax}_U(x). \tag{16}$$

That is, if a number in the cut I is recognised as an axiom in the arithmetization of the syntax relativized to the cut I, then it is also recognised as an axiom in the unrelativized arithmetization in $\mathsf{S}_2^1 + \mathsf{Con}_U$. This, however, is unproblematic. □

Thus by Lemma 5:

Corollary 7 $\mathsf{Q} + \mathsf{Con}_U \rhd \mathsf{T}[U] + \mathsf{AxT}_U$.

Again by Proposition 2 and Lemma 1, we finally have

Corollary 8 Con_U *is the unique* Π_1*-sentence* π*, modulo provable equivalence in* $\mathsf{I}\Delta_0(\mathsf{exp})$*, such that*

$$\mathsf{T}[U] + \mathsf{AxT}_U \equiv \mathsf{Q} + \pi.$$

By inspection of the proofs of Proposition 2 and Corollary 7 it follows that we can replace in these results AxT_U with ScT_U, when U is schematic in the sense explained in §5.1.

6 Reductions and Truth-Theoretic Content

In the introduction we emphasized three conceptual areas that have motivated and inspired axiomatic investigations of the truth predicate. Using the *fil rouge* provided by those categories, we look at what we achieved and what we could not achieve in the previous sections.

There is no need to spend many words on the mathematical interest of the structure of the models of theories in the style of $\mathsf{T}[U]$. As far as the author can see, the motivation for investigating theories of truth of this sort certainly does not reside in the complexity of the model theoretic constructions that they force. As we briefly sketched in §3 and how it is argued more extensively in [33], these theories give us instead the possibility of formalizing informal metamathematical practice and of discerning patterns of reasoning involved in the provability of general claims involving the notion of truth. It thus seems to be a better idea to consider to what degree our work can be relevant for the other customary areas of application of axiomatic truth.

I now discuss whether axioms of truth with built-in syntax help us in discovering new connections between truth and set existence axioms or in providing new insights on the notion of truth arising from the axioms under consideration.

6.1 Predicative Comprehension

A well-known example of how axiomatic truth theories are related to subsystems of second-order arithmetic is represented by the strict connection between the theory CT — that is PA + 'there is a full inductive truth class' — and the system ACA: the truth predicate of CT can be defined in ACA and there is a relative interpretation of ACA in CT.[41] As we shall see in a moment, Theorem 1 above offers a new example of reductions of this sort.

According to the Vicious Circle Principle, we cannot accept definitions $\varphi(x)$ in which the definiendum $\{x : \varphi(x)\}$ belongs to the range of quantifiers in $\varphi(x)$. If one restricts the full comprehension schema to accommodate the vicious circle principle, predicative comprehension is obtained.

Given a one-sorted, sequential theory U, there are several methods for adding predicative comprehension to it.[42] A natural choice would be to expand $\mathcal{L}_U$ with a new sort and extend U with the new scheme. Alternatively, one can use a *flattened* version of this two-sorted theory. This will be the method we employ for obtaining predicative comprehension. It works as follows: starting with $\mathcal{L}_U$, we expand it to the language $\mathcal{L}_U^2$ by means of predicates O, S corresponding to the sorts $\mathfrak{ob}$, $\mathfrak{cl}$ of objects and classes of them respectively and with a binary predicate E intended to express membership of objects in sets. All quantifiers of $\mathcal{L}_U$ are relativized to O; moreover, we need an axiom forcing obvious conditions on E:

$$E(x,y) \rightarrow O(x) \wedge S(y); \tag{17}$$

Following [45], we define φ to be *sorted* if and only if there is a function s that assigns the right sort to variables in φ, that is for all $x, y \in \mathsf{Freev}(\varphi)$,

(i) if $R(x_1, ..., x_n)$ is a subformula of φ, then $s(x_i)$ is $\mathfrak{ob}$ for $1 \leq i \leq n$;
(ii) if $E(x,y)$ is a subformula of φ, then $s(x)$ is $\mathfrak{ob}$ and $s(y)$ is $\mathfrak{cl}$;
(iii) if $x = y$ is a subformula of φ, then $s(x)$ is identical to $s(y)$;
(iv) quantifiers in φ are relativized to the sort $s(x)$ for all variables x in φ.

Following again Visser's notation, we call $\mathsf{PC}(U)$ the theory formulated in $\mathcal{L}_U^2$ whose axioms are the axioms of U in which all quantifiers are relativized to O and the axiom of *predicative comprehension*

$$\forall \bar{x} \exists y \forall u (E(u,y) \leftrightarrow \varphi(u, \bar{x})) \tag{PC}$$

where $s(x)$ is $\mathfrak{ob}$ in the sorted formula φ and quantifiers of φ are of sort $\mathfrak{ob}$ as well. We have:

[41] A detailed proof of both claims can be found in [23].
[42] We follow [45, 46].

Theorem 2 (Visser) *If U is finitely axiomatized and sequential, then* $\mathsf{PC}(U) \equiv \mathsf{Q} + \mathsf{Con}_U$.

We can thus extend our characterization of Con_U to $\mathsf{PC}(\cdot)$, when U is finite and sequential: We have that $\mathsf{T}[U]$ and $\mathsf{PC}(U)$ belong to the same degree of interpretability. That is:

Corollary 9
Let U be finitely axiomatized and sequential. Then $\mathsf{PC}[U] \equiv \mathsf{T}[U]$*;*

Moreover, we have:

Corollary 10 *Let U be sequential and finitely axiomatized. Then* Con_U *is the unique* Π^0_1*-sentence* π *— modulo* $\mathsf{I}\Delta_0(\mathsf{exp})$ *provable equivalence — such that*

$$\mathsf{PC}(U) \equiv \mathsf{Q} + \pi \equiv \mathsf{T}[U]$$

In order to obtain a similar result for an arbitrary, sequential U, we move to a slightly more general framework.[43] We first employ the fact, due to Vaught ([43]), that for any sequential theory W one can find a provably identical theory $\widetilde{W}$ that is axiomatized by a single schema. If τ is our intensional, fixed Δ^b_1 representation of W, $\widetilde{W}$ will be axiomatized by Σ_τ.[44] Given a theory U given by a single schema Σ_υ, the functor $\mathsf{PC}^+(U)$ is defined exactly as $\mathsf{PC}(U)$ with the extra conditions that schematic variables in Σ_υ are treated like variables of sort $\mathfrak{cl}$ and that its universal closure is considered.

Theorem 3 (Visser [46]) *Let U be sequential and axiomatized by a schema. Then* $\mathsf{PC}^+(U) \equiv \mathsf{Q} + \mathsf{Con}_U$.

We can thus extend our characterization of the links between $\mathsf{T}[\cdot]$ and $\mathsf{PC}[\cdot]$ to arbitrary, sequential theories:

Corollary 11 *Let U be sequential. Then* $\mathsf{PC}^+(\widetilde{U}) \equiv \mathsf{T}[U] + \mathsf{AxT}_U$.

Finally, we have:

Corollary 12 *Let U be sequential. Then* Con_U *is the unique* Π^0_1*-sentence* π *— modulo* $\mathsf{I}\Delta_0(\mathsf{exp})$ *provable equivalence — such that*

$$\mathsf{PC}^+(\widetilde{U}) \equiv \mathsf{Q} + \pi \equiv \mathsf{T}[U] + \mathsf{AxT}_U$$

[43] We still follow [45, 46].

[44] Vaught considers weaker assumptions on U than sequentiality. Moreover Vaught result, as noticed in [45], can be verified in S^1_2.

We notice that $\widetilde{U}$ crucially depends on Σ_υ. Different choices of the formula representing U may lead to non equivalent results. We refer to [45] for more details and examples.

The possibility of getting rid of exponentiation in results of this sort is an open problem.

6.2 Base Theories and Truth Theoretic Content

According to several authors,[45] the conservativeness or the relative interpretability of systems of truth over the base theory suggest at least the possibility of a reduction of semantic resources to mathematical or syntactic resources. We have seen that theories in the style of $\mathsf{T}[U]$ and $\mathsf{T}[U]+\mathsf{AxT}_U$ are conservative over and non interpretable in U. These features, combined together, seem to be appealing for some variants of deflationism that accept the requirement of conservativeness of the theory of truth over the mathematical base theory but hold that truth still serves an indispensable expressive role.[46] In addition, model-theoretic conservativeness seem to be more appealing than proof-theoretic conservativeness for the deflationist as the theory of truth does not restrict the class of models of the object theories (Cfr. [34]). In this respect the theories in the style of $\mathsf{T}[U]$ display a similar status as Fischer's PT^-.[47]

From a methodological point of view, results such as Observations 1 and 2 and the subsequent observations indicate that the choice of the base theory is highly relevant to draw conceptual conclusions — even for theories constructed in the usual, 'entangled' way. Let us be explicit and consider the well-known case of the theories $\mathsf{CT}\restriction[\cdot]$. $\mathsf{CT}\restriction$ — that is $\mathsf{CT}\restriction[\mathsf{PA}]$ — is conservative and interpretable in the base theory PA.[48] If we move to finitely axiomatized base theories and consider theories such as $\mathsf{CT}\restriction[\mathsf{I}\Sigma_n]$, for instance, the techniques of the shortening of cuts will entail its non interpretability in $\mathsf{I}\Sigma_n$.[49] If our focus is on the notion of truth forced by the truth axioms, it appears to be disappointing to know that inessential variations in the base theories — inessential with respect to the way in which the theories fulfil the task of mimicking structural properties of sentences — may impinge on our conclusions. In other words, we may want *either* our 'operator' $\mathsf{CT}\restriction[\cdot]$ to perform in a uniform way across theories interpreting a sufficient amount of syntax, *or* to have strong motivations for preferring one base theory over another. Now PA is usually taken to be the standard choice. ZF is sometimes considered in alternative. But why? Do we have strong philosophical reasons for these choices?

Surely there are practical motivations. PA or ZF represent safe environments in which it is possible to reproduce and establish structural properties of the intended bearers of truth. Moreover, they display a fundamental foundational role and much

[45] Again we have in mind [6, 13, 26, 29, 39].

[46] Such versions may be found in [13, 26]. A discussion of this form of deflationism in relation to the results considered can be found in [36].

[47] For the presentation of PT^- we refer to Fischer's paper [15]. For a discussion of its role, see [26].

[48] Notice, we are reasoning in the usual setting.

[49] Although the theory of truth will remain conservative over $\mathsf{I}\Sigma_n$.

is known about their metamathematical properties. But do we have more than this? What about possible alternatives? If we remain in the arithmetical realm, some finitely axiomatized theories, such as $\mathsf{I}\Sigma_n$ for some n, might be a questionable choice: in general we lack motivations for restricting the induction schema of PA to Σ_n-induction; moreover, we know that there are theories that formalize the syntax of formal languages in a natural way that are much weaker than these. Why not considering, for instance, a theory such as S^1_2 or its synonymous theory of strings due to Ferreira that is calibrated to give us an efficient treatment of syntactic notions and it is arguably close to the theoretical optimality? If one employs Buss' S^1_2 as base theory for our theory of truth, which is in a sense 'minimal' as it is interpretable in Q and also known to be finitely axiomatizable, the addition of axioms stating the existence of a truth class to it would yield a theory that is not interpretable in — and most likely conservative over — S^1_2, whereas the addition of axioms forcing the existence of a truth class to PA determine a conservative and interpretable theory. Thus if conservativity and non interpretability in the mathematical object theory are then desirable features for the deflationist, does she also have reasons for supporting a specific choice of the mathematical or syntactic base theory? It seems that more philosophical work needs to be done and we defer the answer to these questions to a forthcoming work.

Let us now move to a closer analysis of the technical material presented above. If one aims at gaining some insights concerning the metaphysical status of the truth predicate by comparing the theory of truth and the base theory, it is very likely that she will have in mind a scenario in which there is at least a somewhat discernible separation between truth axioms on the one hand and axioms of the object theory on the other. We know, however, that things are slightly more complicated. As already noticed by Heck and by [23], the theory of syntax represents a further variable that has to be taken into account. Its role becomes apparent when schemata of the base theory such as induction principles are extended to contain semantic vocabulary. There are *mathematical* and there are *syntactic* uses of those extended schemata; as shown in [33], moreover, this diagnosis can be extended also to typed theories of compositional truth constructed over set-theoretic base theories.[50] We refer to the references for details but the idea is straightforward: inductive reasoning involving truth, such as proofs by induction on the complexity of formulas, capture essentially truth-theoretic and syntactic patterns of reasoning and should be distinguished, at least at the level of philosophical reflection, from instances of those principles that refer to the subject matter of base theory.

The framework investigated in the present work seem to suggest that similar considerations can be applied to *compositional* truth axioms. On the one hand, in fact, we know that also very weak truth axioms are in a sense deeply intertwined with assumptions on the ontology of expressions. Already the compositional axiom for negation over pure predicate logic is not in some sense ontologically neutral as it entails the

[50]There it is also shown that the different roles played by different instances schemata involving the truth predicate are much more evident in the case of set theoretic base theories.

existence of two objects [22]. A truth axioms is always a *truth-theoretic and syntactic axiom*. In the usual setting with entangled syntax, it seems, this amounts to saying that we do not have definite criteria to distinguish between mathematical and truth-theoretic content.[51] Theories such as $\mathsf{T}[U]$ above, on the contrary, are based upon the possibility of forcedly distinguishing between truth-theoretic and syntactic content on the one side, and mathematical content on the other. If one looks for crystal clear criteria for distinguishing between areas of metamathematical reflection and focus on the axioms of truth *qua* principles of semantic nature, therefore, theories of truth and expressions such as $\mathsf{T}[U]$ appear to be an inviting option; mathematical objects belong to one sort and syntactic objects interacting with semantic machinery to another. An evaluation of the impact that axioms of truth bear on the underlying mathematical structure, if meaningful, seems thus to find a more comfortable environment here than in the usual constructions.

The price to pay is of course the apparent rigidity of the framework at issue. Also, the conservativeness of the theory of truth becomes an almost trivial property and interpretability acquires a predominant role. If this is tolerated, however,[52] it seems that some conceptual gains are reached. Corollaries 6 and 8 tell us that, over an arbitrary base theory U, *Tarski-style truth axioms bundled together with enough syntax to make them meaningful can be characterized precisely as an intensionally correct consistency statement for* U. For arbitrary U, the truth-theoretic conglomerate considered has to contain also the claim ascertaining the truth of all instances of the schemata of U. These facts seem thus to deliver a clear message: if one considers mutual interpretability as a trustworthy method of comparison, axioms of truth-theoretic and syntactic content, when added to a mathematical base theory, correspond to a metamathematical claim that, given Pudlák's beautiful version of Gödel's Second Incompleteness Theorem, is out of reach for the base theory itself.

We end the paper with a countercheck for our latter remarks. For them to count as a (partial) characterization of compositionality in a typed theory of truth, in fact, one needs to ascertain that similar characterizations are not available for typed axiomatizations of the truth predicate that allegedly reflect different intuitions on the truth predicate, such as disquotational theories. The following claim tells us that in fact disquotational truth axioms *cannot* be equated to an intensionally correct consistency statement for the object theory, at least when the latter is sequential.

Let $\mathsf{ut}[U]$ be the theory in $\mathcal{L}_{\mathsf{T}}$ whose axioms are the axioms of U, the axioms of S^1_2, C. in Table 1, and the schema

$$\forall j_1, ..., j_n \ (\mathsf{Sat}(\ulcorner\varphi\urcorner(j_1/i_1, ..., j_n/i_n), a) \leftrightarrow \varphi(a(j_1), ..., a(j_n))) \qquad \text{(utb)}$$

[51] Cfr. also [23, §21.2].

[52] And we judge that *some* rigidity is mandatory when one wants to oppose truth-theoretic contents to other components of our systems.

for any $\mathcal{L}_U$-formula $\varphi(v_{i_1}, ..., v_{i_n})$ with the free variables displayed. Moreover, let t[U] like ut[U] but with Eq. (utb) replaced by

$$\forall a\ \mathsf{Sat}(a, \ulcorner\varphi\urcorner) \leftrightarrow \varphi \tag{tb}$$

for all $\mathcal{L}_U$-sentences φ.

We have:

Lemma 13 *Let U be sequential. Then $U \rhd_{\mathsf{loc}} \mathsf{ut}[U]$.*

Proof Let us consider an arbitrary finite subsystem $\mathsf{ut}[U]_0$ of $\mathsf{ut}[U]$. We interpret $\mathsf{ut}[U]_0$ in U. By our overall assumptions on U, $U \rhd \mathsf{S}^1_2$: this takes care of the syntactic part of $\mathsf{ut}[U]_0$. The object theoretic part is straightforwardly interpreted in U. The domain of sequences is relativized to sequences in U; The function $(\cdot)\cdot$ to a formula $\beta(x, y, z)$ characterizing the unique object x corresponding to the y^{th} element of z, existing by the sequentiality of U. Moreover, we have

$$\begin{aligned}(\mathsf{Sat})^{\kappa}(x, y) :\leftrightarrow\ & (x = \ulcorner\varphi_1(v_1, ..., v_{n_1})\urcorner \wedge \varphi_1((y)_1, ..., (y)_{n_1}))\ \vee \\ & \vdots \\ & (x = \ulcorner\varphi_k(v_1, ..., v_{n_k})\urcorner \wedge \varphi_k((y)_1, ..., (y)_{n_k}))\end{aligned}$$

where $\varphi_1(v_1, ..., v_{n_1}), ..., \varphi_k(v_1, ..., v_{n_k})$ are the finitely many $\mathcal{L}_U$-formulas occurring in the disquotational axioms of $\mathsf{ut}[U]_0$ with exactly the free variables displayed. It is almost immediate to verify that the translations of formulas of the finitely many instances of Eq. (utb) occurring in $\mathsf{ut}[U]_0$ become provable in U. □

Proposition 4 *Let U be as in Lemma 13. Then $\mathsf{ut}[U]$ does not interpret $\mathsf{Q} + \mathsf{Con}_U$.*

Proof Assume that $\mathsf{ut}[U]$ interprets $\mathsf{Q} + \mathsf{Con}_U$. Thus for an appropriate finite subsystem of $\mathsf{ut}[U]$, let us call it $\mathsf{ut}[U]_1$, we have

$$\mathsf{ut}[U]_1 \rhd \mathsf{Q} + \mathsf{Con}_U$$

By Lemma 13, $U \rhd \mathsf{Q} + \mathsf{Con}_U$, which is impossible by Pudlák's result. □

A fortiori, we obtain

Corollary 13 *Let U be sequential.* t[U] *does not interpret $\mathsf{Q} + \mathsf{Con}_U$.*

We notice that Proposition 4 would not hold when $\mathsf{ut}[U]$ proves already the consistency of U on a cut. This is for instance the case when U is predicate logic Pred: S^1_2 proves in fact $\mathsf{Con}_{\mathsf{Pred}}$, by formalizing the construction of a one-element model.

Appendix

Definition 1 (Basic Axioms of S_2^1)

$a \leq b \rightarrow a \leq Sb$
$a \neq Sa$
$0 \leq a$
$a \leq b \wedge a \neq b \leftrightarrow Sa \leq b$
$a \neq 0 \rightarrow 2 \times a \neq 0$
$a \leq \vee b \leq a$
$a \leq b \wedge b \leq a \rightarrow a = b$
$a \leq b \wedge b \leq c \rightarrow a \leq c$
$|0| = 0$
$|S0| = S0$
$a \neq 0 \rightarrow |2 \times a| = S(|a|) \wedge |S(2 \times a)| = S(|a|)$
$a \leq b \rightarrow |a| \leq |b|$
$|a\#b| = S(|a| \times |b|)$
$0\#a = S0$
$a \neq 0 \rightarrow 1\#(2 \times a) = 2 \times (1\#a) \wedge$
$1\#(S(2 \times a)) = 2 \times (1\#a)$
$a\#b = b\#a$
$|a| = |b| \rightarrow a\#c = b\#c$
$|a| = |b| + |c| \rightarrow a\#d = (b\#d) \times (c\#d)$
$a \leq a + b$
$a \leq b \wedge a \neq b \rightarrow S(2 \times a) \leq 2 \times b \wedge S(2 \times a) \neq 2 \times b$
$a + b = b + a$
$a + 0 = a$
$a + Sb = S(a + b)$
$(a + b) + c = a + (b + c)$
$a + b \leq a + c \leftrightarrow b \leq c$
$a \times 0 = 0$
$a \times Sb = (a \times b) + a$
$a \times b = b \times a$
$a \times (b + c) = (a \times b) + (a \times c)$
$S0 \leq a \rightarrow (a \times b \leq a \times c \leftrightarrow b \leq c)$
$a \neq 0 \rightarrow |a| = S(|\lfloor \frac{1}{2} a \rfloor|)$
$a = \lfloor \frac{1}{2} b \rfloor \leftrightarrow 2 \times a = b \vee S(2 \times a) = b$

References

1. Bar-On, D., & Simmons, K. (2007). The use of force against deflationism: Assertion and truth'. In Greimann, D., & Siegwart, G. (Eds.), *Truth and Speech Acts: Studies in the Philosophy of Language* (pp. 61–89). New York: Routledge.
2. Bennet, J.H. (1962). On Spectra. PhD Thesis, Princeton.
3. Buss, S. (1986). *Bounded Arithmetic*. Naples: Bibliopolis.
4. Buss, S. (ed.) (1998). Handbook of Proof Theory. Elsevier.
5. Cantini, A. (1989). Notes on formal theories of truth. *Archives for Mathematical Logic*, *35*, 97–139.
6. Cieśliński, C. (2010). Truth, Conservativeness, Provability. *Mind*, *119*(474).
7. Craig, W., & Vaught, W. (1958). Finite axiomatizability using additional predicates. *The Journal of Symbolic Logic*, *23*, 289–308.
8. Enayat, A., & Visser, A. (2013). *New constructions of satisfaction classes*. Logic Preprints Series: University of Utrecht.
9. Enderton, H. (2001). A mathematical introduction to logic Harcourt Press.
10. Feferman, S. (1960). Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, *49*, 35–91.
11. Feferman, S. (1991). Reflecting on incompleteness. *Journal of Symbolic Logic*, *56*, 1–49.
12. Ferreira, F. (1988). Polynomial Time Computable Arithmetic and Conservative Extensions. Ph.D.thesis, The Pennsylvania State University, State College.
13. Field, H. (1999). Deflating the conservativeness argument. *Journal of Philosophy*, *96*(10), 533–540.
14. Field, H. (2005). Compositional principles vs. schematic reasoning. *The Monist*, *89*, 9–27.
15. Fischer, M. (2009). Minimal Truth and Interpretability. *The Review of symbolic logic*, *2*, 799–815.
16. Fischer, M., Halbach, V., Speck, J., & Stern, J. Axiomatizing Semantic Theories of Truth?. Unpublished Manuscript.
17. Fujimoto, K. (2012). Classes and truths in set theory. *Annals of Pure and Applied Logic*, *164*(11), 1484–1523.
18. Gödel, K. (1931). Über formal unentscheidbare sätze der principia mathematica und verwandter systeme, I. *Monatshefte für Mathematik und Physik*, *38*, 173–98.
19. Hájek, P., & Pudlák, P. (1993). *Metamathematics of first-order arithmetic*. Berlin: Springer.
20. Halbach, V. (1994). A system of complete and consistent truth. *Notre Dame Journal of Formal Logic*, *35*, 311–327.
21. Halbach, V. (2000). Truth and reduction. *Erkenntnis*, *53*, 197–126.

22. Halbach, V. (2001a). How innocent is deflationism?. *Synthese*, *126*, 167–194.
23. Halbach, V. (2014). *Axiomatic theories of truth*. Revised Edition: Cambridge University Press.
24. Heck, R. Consistency and the Theory of Truth. forthcoming in The Review of Symbolic Logic.
25. Horwich, P. (1998). Truth, 2nd ed. Blackwell.
26. Horsten, L. (2011). The Tarskian Turn. Deflationism and Axiomatic Truth. Princeton University Press.
27. Joosten, J., & Visser, A. (2004). Visser (2004), 'Characterizations of Interpretability', in Logic Preprint Group Series Utrecht.
28. Kaye, R. (1991). Models of peano arithmetic. *Oxford University Press*, 1991.
29. Ketland, J. (1999). Deflationism and Tarski's paradise. *Mind*, *108*(429), 69–94.
30. Kikuchi, M., & Tanaka, K. (1994). On formalizations of model-theoretic proofs of Gödel's theorems'. *Notre Dame Journal of Formal Logic*, *35*(3), 403–412.
31. Lachlan, A. (1981). Full satisfaction classes and recursive saturation. *Canadian Mathematical Bulletin*, *24*, 295–297.
32. Lavine, S. (1999). Skolem was wrong, unpublished manuscript.
33. Leigh, G., & Nicolai, C. (2013). Axiomatic truth, syntax and metatheoretic reasoning. *The Review of Symbolic Logic*, *6*(4), 613–636.
34. McGee, V. (2006). In praise of a free lunch: why disquotationalists should embrace compositional semantics In: Bolander et alii (eds.), Self Reference, CSL Publications.
35. Nicolai, C. (2014). *Truth, deflationism and the ontology of expressions An Axiomatic Study*. Oxford: DPhil Thesis.
36. Nicolai, C. (2014). Deflationary truth and the ontology of expressions. To appear in *Synthese*.
37. Paris, J.B., & Wilkie, A. (1987). On the scheme of induction for bounded formulas. *Annals of Pure and Applied Logic*, *35*, 261–302.
38. Pudlák, P. (1985). Cuts, consistency statements and interpretations. *The Journal of Symbolic Logic*, *50*, 423–441.
39. Shapiro, S. (1998). Truth and Proof: through thick and thin. *Journal of Philosophy*.
40. Simpson, S. (2009). *Subsystems of Second-Order Arithmetic*. Second edition: Cambridge University press–ASL.
41. Tarski, A. (1956a). The concept of truth in formalized languages. In Woodger, H.J. (Ed.), *Logic, Semantic, Metamathematics: papers of Alfred Tarski from, 1922-1938* (pp. 152–278). Oxford: Clarendon Press.
42. Tennant, N. (2002). Deflationism and the Gödel phenomena. *Mind*, *111*, 551–82.
43. Vaught, R.A. (1967). Axiomatizability by a schema. *The Journal of Symbolic Logic*, *32*, 473–479.
44. Visser, A. (1991). The formalization of interpretability. *Studia Logica*, *50*(1), 81–106.
45. Visser, A. (2009). Can we make the second incompleteness theorem coordinate free?. *Journal of Logic and Computation*, *21*(4), 543–560.
46. Visser, A. (2009). The predicative frege hierarchy. *Annals of Pure and Applied Logic 3c*, *160*, 129–153.