MORE ON SYSTEMS OF TRUTH AND PREDICATIVE COMPREHENSION

CARLO NICOLAI

Abstract. In the paper we survey the known connections between theories that extend a common base theory with typed truth axioms on the one hand and predicative set-existence assumptions on the other. How general can the mutual reductions between truth and comprehension be taken to be? In trying to address this question, we consider (typed) classical, positive truth and predicative comprehension as operations on theories.

Keywords: Predicative Comprehension. Axiomatic Theories of Truth. Relative Interpretability.

1. Introduction

The proof-theoretic analysis of axiomatic truth theories has uncovered several connections holding between systems extending a base theory with set existence axioms and systems extending the same base theory with axioms characterizing a primitive truth predicate. Taken at face value, these results suggest that assumptions on the existence of certain sets can be replaced by suitable semantic assumptions and vice versa. As we shall see shortly, for instance, the extension of Peano Arithmetic with specific assumptions on arithmetically definable sets is inter-reducible with a full, Tarskian truth theory over PA.

These connections have played an important role in the foundation of mathematics (e.g. in the analysis of the limits of predicativity Feferman (1991)), in the analysis of different solutions to the semantic paradoxes (proving more truth-theoretic iterations has been generally considered a virtue of a theory of truth), in the debate concerning the nature of truth (is truth a light or a substantial property?). They have been also suggested as a tool to carry out ontological reductions.[1]

At any rate, these mutual reductions between truth and set existence axioms involve theories built on a fixed base theory, usually Peano Arithmetic. In this work we set the basis for a different approach: truth and predicative comprehension will be taken as functors applying to arbitrary object theories satisfying some minimal requirements. Crucially, the results of applying these functors to the base theory will turn out to be equivalent tools for uncovering the first step of our commitment implicit in the acceptance of the base theory. Furthermore, the generality given by the proposed approach will also be reflected in a more uniform correspondence between set-theoretic and semantic assumptions, as it will be exemplified in §5.

The paper has two main parts. In the first, we survey several results connecting extensions Peano Arithmetic with truth axioms or predicative comprehension axioms. In the second, we introduce the more general setting just announced.

As this is more congenial to the original, second part of the work we will mostly focus, in the first part, on systems of *typed truth* — that is, theories formulated in languages in

[1] Cf. Halbach (2014, Ch. 23) and §5 of the present paper.

which the truth predicate applies to sentences not containing the truth predicate itself — and subsystems of ACA. The core argument, which gives rise to several variations, is the folklore mutual reduction between the system of Tarskian, compositional truth CT and ACA itself. We will also take into account the system PT of positive, typed truth, which is a notational variant of CT: its subsystems are not so easily comparable with their Tarskian counterparts but still enjoy several reductions to subsystems of ACA. All of this will be surveyed in §3.

The second part, starting with §4, will investigate a generalization of the systems considered in §3. We define three main operations on recursively enumerable (RE) theories: (i) $\mathsf{T}[\cdot]$ results in a Tarskian truth theory; (ii) $\mathsf{Tp}[\cdot]$ in a typed theory of truth simulating positive inductive defintions; (iii) $\mathsf{PC}[\cdot]$ adds predicative comprehension to the object theory. In order to study $\mathsf{PC}[\cdot]$ and relate it in full generality to the truth theories also studied, a variant of it, $\mathsf{PCS}[\cdot]$, has to be taken into account: it applies to theories axiomatized by schemata in which schematic variables are replaced by 'second-order' variables. The upshot of §4 will be that, modulo mutual interpretability, the extension of arbitrary RE base theories via these three operations yields equivalent results. In §5 we reflect on the significance of the technical work carried out in the previous sections and consider possible extensions.

2. Preliminaries

2.1. **Theories and Arithmetization.** Unless otherwise specified, arbitrary theories will be formulated in many-sorted, first-order logic. We assume they have a Δ^b_1 specification[2] — provably in a weak syntax theory to be defined — and that they are formulated in a Hilbert-style calculus in with Modus Ponens as its only rule of inference.[3]

On occasion, we consider *sequential* theories, namely theories that have a nice coding of sequences, which is in fact a safe tool to formalize syntax.[4] We employ $T, U, V, W, \ldots$ to range over arbitrary theories; $A, B, C, \ldots$ are taken to range over finitely axiomatized theories.

Whereas in §3 we work with PA as base theory, so all the usual representations of syntactic notions and operations can be assumed, in §4 we work in the theory S^1_2. S^1_2 is a theory introduced by Sam Buss in Buss (1986) to study polynomial time computability; the functions that are provably recursive in S^1_2 are exactly the p-time computable functions. Remarkably, S^1_2 is finitely axiomatizable and mutually interpretable with the induction free fragment of first-order arithmetic Q. For details concerning coding in S^1_2, we refer to Buss (1986),Buss (1998). In particular, for any given language $\mathcal{L}_W$, we assume Δ^b_1-definitions $\mathsf{Term}_{\mathcal{L}_W}(x)$, $\mathsf{CTerm}_{\mathcal{L}_W}(x)$, $\mathsf{Fml}_{\mathcal{L}_W}(x)$, $\mathsf{Sent}_{\mathcal{L}_W}(x)$, $\mathsf{Prf}_{\mathcal{L}_W}(x)$, $\mathsf{Proof}_{\mathcal{L}_W}(x, y)$ of the sets of terms, closed terms, formulas, sentences of the language of W, of proofs in the theory W and of the relation of being a proof in W of the $\mathcal{L}_W$-formula y. Also, we take the set of theorems of W to be

[2]That is, their axiom set is specified by a formula provably equivalent to formulas containing sharply bounded quantifiers (Cf. Hájek &Pudlák (1993, V.4. Def. 4.2)) and one bounded existential or universal quantifier only.

[3]Cf. Enderton (2001) for an example of such axioms system for first-order logic.

[4]More precisely, T is sequential iff it directly interprets (i.e. identity is mapped into identity and quantifiers are not relativized) *Adjunctive Set Theory*, that is a theory in a (first-order) language with $\in$ and $=$ whose axioms are:

(AS1) $\exists x\ \forall y\ y \notin x$

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

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

defined by the $\exists\Delta^b_1$-formula $\mathsf{Pr}_W(x) :\leftrightarrow \exists y\, \mathsf{Proof}_W(x,y)$.[5] As to notational conventions, we will often abuse of Gödel corners and interchangeably employ them and Feferman's dot convention.

2.2. **Reductions.** The preferred means of reduction throughout the paper will be *many sorted, —possibly non direct— relative interpretability.*

Let T and W be theories containing S^1_2.[6] 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.[7] The translation extends to the mapping τ:

- $(R)^\tau(x_1,...,x_n) :\leftrightarrow F(R)(x_1,...,x_n)$;
- τ commutes with the 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 $\mathcal{L}_T$,

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

W *locally* interprets T if and only if every finite subsystem of T is interpretable in W. 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.

On occasion, we will employ the notion of *truth definability*, extensively studied by Fujimoto (2010). Truth definability will be mainly applied in the first part of the paper, as it assumes a fixed base theory (at least in its original formulation in Fujimoto (2010)). Let U, V be theories extending a syntactic base theory B: it is assumed, moreover, that U, but not necessarily also V, results from the extension of B via truth axioms. U is *relatively truth definable* in V if there is a $\mathcal{L}_B$-conservative relative interpretation of U into V. In other words, a truth definition of U in V is a relative interpretation τ_0 that behaves like the identity mapping when applied to truth-free formulas of $\mathcal{L}_U$.

2.3. **Definable Cuts and Incompleteness.** Cuts are initial segments of the natural numbers. We will in particular be interested in initial segments definable in theories T containing S^1_2. More precisely, a formula $\varphi(x)$ is called *inductive* in T containing S^1_2 if and only if

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

$\varphi(x)$ is a T-*cut* if and only if, additionally, it defines an initial segment of the T-numbers. In other words,

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

Here it is important to notice that $0, S, \leq$ are understood in T via the assumed interpretation of S^1_2 in T. By a well-know result of Solovay,[8] every inductive formula has a subcut:

[5] The class of $\exists\Delta^b_1$ formulas is obtained from the class of Δ^b_1-formulas by closing it under unbounded existential quantification.

[6] The relation of containment can be understood as subtheory relation or relative interpretability relation.

[7] The definition generalizes naturally to the many-sorted case by considering multiple domains.

[8] Although contained in an unpublished note (cf. Hájek &Pudlák (1993)).

Lemma 1. Let $\varphi(x)$ be inductive in T containing S^1_2. Then there exists a T-cut $\psi(x)$ such that

$$T \vdash \psi(x) \to \varphi(x).$$

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 $\mathcal{I}$ be a T-cut. Then we can find a subcut $\mathcal{J}$ of $\mathcal{I}$ such that T proves the following:

(1) $$\mathcal{J}(x) \wedge \mathcal{J}(y) \to \mathcal{J}(x + y)$$

(2) $$\mathcal{J}(x) \wedge \mathcal{J}(y) \to \mathcal{J}(x \times y)$$

(3) $$\mathcal{J}(x) \wedge \mathcal{J}(y) \to \mathcal{J}(x \# y).$$

where the smash function $\#$ is such that $x \# y = 2^{|x| \times |y|}$, and $|x| = \lceil \log_2(x + 1) \rceil$ — that is, the upper integer part of the binary logarithm of $x + 1$.

Therefore, in what follows, a definable cut can always be taken to be closed under addition, multiplication and the smash function. Lemma 2 is particularly important as it guarantees that, given a T-cut $\mathcal{I}$ — with T again containing S^1_2, we can always restrict our attention to a subcut $\mathcal{J}$ of $\mathcal{I}$ satisfying the axioms of S^1_2. This will be extensively employed in §4: having S^1_2 available on a cut enables one to have a meaningful and smooth arithmetization of the syntax in the cut. This will prove to be crucial in many of the interpretations defined later on.

We will extensively make use of the following:

Lemma 3 (Wilkie, Nelson)**.** Q interprets S^1_2 on a definable cut.

Q does not have any induction: it cannot thus prove the usual properties of $\leq$. An important part of the proof of Lemma 3 consists in showing how the interpretation is indeed defined on a initial segment of the Q-numbers. This makes possible the downwards preservation of Π_1-sentences (cf. Hájek &Pudlák (1993, V.5(c))) so that, for instance, $\mathsf{S}^1_2 + \mathsf{Con}(U)$ is interpretable in $\mathsf{Q} + \mathsf{Con}(U)$.

Another remarkable fact concerning S^1_2 is that Gödel's Second Incompleteness theorems can be meaningfully stated and proved in it. This makes possible a generalization of the Gödel's second incompleteness theorem, due to Pavel Pudlák in its original form. The beautiful version that we present is due to Albert Visser.

Lemma 4. Let U be given by a Σ_1-formula. Then U does not interpret $\mathsf{S}^1_2 + \mathsf{Con}(U)$.

Proof. Assume it does. There must be, therefore, A be a finite subsystem of U that interprets $\mathsf{S}^1_2 + \mathsf{Con}(U)$. By Σ_1-completeness, S^1_2 proves the formalization of the fact that $A \subset U$. Thus A interprets $\mathsf{S}^1_2 + \mathsf{Con}(A)$. Since A and $\mathsf{S}^1_2 + \mathsf{Con}(A)$ are finite, we have

$$\mathsf{S}^1_2 \vdash \mathsf{Con}(A) \to \mathsf{Con}(\mathsf{S}^1_2 + \mathsf{Con}(A)).$$

□

3. Arithmetical Comprehension

In this section we survey the mutual reductions between truth axioms and predicative comprehension in the standard setting, that is when PA is taken as object theory. After a few useful definitions and lemmata, in §3.1 we consider how Tarskian and positive truth can define the comprehension schema of ACA and variants thereof. We then consider the converse direction in §3.2. §3.3 hints at how to compare iterations of Tarskian truth and predicative comprehension to each other and to systems of type-free truth.

The language $\mathcal{L}_2$ extends the language $\mathcal{L} := \{0, S, +, \times, \leq\}$ of arithmetic with an additional sort for 'sets of natural numbers', or 'reals'.

Definition 1.

(i) ACA is formulated in $\mathcal{L}_2$. It extends PA with

$$\exists Y \forall u (u \in X \leftrightarrow \varphi(u)) \tag{CA}$$

where φ does not contain bound set-variables (it may contain set parameters); and

$$\varphi(0) \wedge \forall y(\varphi(y) \to \varphi(Sy)) \to \forall y\, \varphi(y) \tag{S-Ind^2}$$

for $\varphi \in \mathcal{L}_2$.

(ii) To obtain ACA_0 one replaces (S-Ind^2) with the single $\mathcal{L}_2$-sentence

$$0 \in X \wedge \forall y\, (y \in X \to Sy \in X) \to \forall y\, y \in X \tag{A-Ind^2}$$

Lemma 5. ACA_0 is not interpretable in PA.

Proof. If ACA_0 were interpretable in PA, then ACA_0 would be interpretable in a suitable finite subsystem A of PA.

Thus PA would be interpretable in A. But PA is reflexive — that is, it proves the consistency of all its finite subtheories. Therefore,

$$\mathsf{PA} \vdash \mathsf{Con}(A) \to \mathsf{Con}(\mathsf{ACA}_0)$$
$$\mathsf{PA} \vdash \mathsf{Con}(\mathsf{ACA}_0)$$

But $\mathsf{PA} \subset \mathsf{ACA}_0$, so the last line contradicts Gödel's second incompleteness theorem. □

3.1. Arithmetical Comprehension from Truth.

Definition 2 (CT)**.**

(i) CT extends PAT (i.e. PA in $\mathcal{L} \cup \{\mathsf{Tr}\}$) with the universal closure of

(CT1) $\mathsf{CTerm}_{\mathcal{L}}(x) \wedge \mathsf{CTerm}_{\mathcal{L}}(y) \to (\mathsf{Tr}\, x \dot{\circ} y \leftrightarrow \mathsf{val}(x) \circ \mathsf{val}(y))$ with $\circ \in \{=, \leq\}$

(CT2) $\mathsf{Sent}_{\mathcal{L}}(x) \to (\mathsf{Tr}_{\dot\neg} x \leftrightarrow \neg \mathsf{Tr}\, x)$

(CT3) $\mathsf{Sent}_{\mathcal{L}}(x \dot\wedge y) \to (\mathsf{Tr}(x \dot\wedge y) \leftrightarrow \mathsf{Tr}\, x \wedge \mathsf{Tr}\, y)$

(CT4) $\mathsf{Sent}_{\mathcal{L}}(\dot\forall v x) \to (\mathsf{Tr} \dot\forall v x \leftrightarrow \forall y(\mathsf{CTerm}(y) \to \mathsf{Tr}\, \mathsf{sub}(x, v, y)))$

In (CT4), v *codes* a variable.

(ii) $\mathsf{CT}\restriction$ is obtained by allowing only $\mathcal{L}$-formulas as instances of the induction scheme of CT;

(iii) CT^- := (CT1) – (CT4) plus:

(IInd) $\quad \mathsf{tot}(x) \to \mathsf{Tr}x(\ulcorner \mathsf{o} \urcorner) \wedge \forall y\,(\mathsf{Tr}x(\dot{y}) \to \mathsf{Tr}x(\mathsf{S}\dot{y})) \to \forall y\,\mathsf{Tr}x(\dot{y})$

where

(tot) $\quad \mathsf{tot}(x) :\leftrightarrow \mathsf{Fml}^1_{\mathcal{L}}(x) \to \forall y\,(\mathsf{Tr}x(\dot{y}) \vee \mathsf{Tr}_{\dot{\neg}}x\,(\dot{y}))$

and $\mathsf{Fml}^1_{\mathcal{L}}(x)$ expresses that x is a formula of $\mathcal{L}$ with only the first variable free.

CT^- is considered in Fischer (2009). It is finitely axiomatized and thus weaker than CT: the latter is in fact reflexive.

We summarize some useful facts about CT and its restricted variants. Most of them, in slightly different form, can also be found for instance in Cantini (1989), Feferman (1991), Halbach (2014).

Lemma 6.

(i) for all $\varphi(v) \in \mathcal{L}$, $\mathsf{CT}\restriction \vdash \forall x\,(\mathsf{CTerm}_{\mathcal{L}}(x) \to (\mathsf{Tr}\,\mathsf{sub}(\ulcorner \varphi \urcorner, v, x) \leftrightarrow \varphi(\mathsf{val}(x))))$;
(ii) $\forall x\,(\mathsf{Sent}_{\mathcal{L}}(x) \wedge \mathsf{Pr}_{\mathsf{PA}}(x) \to \mathsf{Tr}x)$, where $\mathsf{Pr}_{\mathsf{PA}}(\cdot)$ expresses canonical provability in PA;
(iii) $\mathsf{CT}\restriction \vdash \forall x(\mathsf{Fml}^1_{\mathcal{L}}(x) \to \mathsf{tot}(x))$
(iv) $\mathsf{CT}^- \vdash \mathsf{Con}(\mathsf{PA})$
(v) $\mathsf{CT} \vdash \mathsf{Sent}_{\mathcal{L}}(\dot{\forall} vz) \wedge \mathsf{CTerm}_{\mathcal{L}}(x) \wedge \mathsf{CTerm}_{\mathcal{L}}(y) \to$
$(\mathsf{val}(x) = \mathsf{val}(y) \;\to\; (\mathsf{Tr}(\mathsf{sub}(z, v, x)) \;\leftrightarrow\; \mathsf{Tr}(\mathsf{sub}(z, v, y)))))$

Proof. (i) is obtained by external induction on φ. For (ii) one employs the induction axioms of CT. (iii) follows from CT2. (iv) is obtained by combining (IInd) and (i) to mimic in CT^- the CT-proof of (ii). (v) also follows from a formal induction on the complexity of the formula z. □

By extending PAT with the schema of Lemma 6(i) taken as axiom, we obtain the theory UTB. In $\mathsf{UTB}\restriction$, like in $\mathsf{CT}\restriction$, the truth predicate is not allowed into instances of the induction scheme. The next lemma summarizes some well-known facts concerning the interpretability of typed truth in PA.

Lemma 7.

(i) PA locally interprets $\mathsf{TB}\restriction$, $\mathsf{UTB}\restriction$, TB, UTB;
(ii) PA interprets $\mathsf{TB}\restriction$, $\mathsf{UTB}\restriction$, TB, UTB;
(iii) PA interprets $\mathsf{CT}\restriction$;
(iv) PA does not interpret CT^-,CT.

Proof. Ad (i), in a finite subsystem S of UTB instances of the uniform disquotation scheme involve sentences $\varphi \in \mathcal{L}$ such that $\mathsf{lcx}(\varphi) \leq n$ for a standard n, where $\mathsf{lcx}(x)$ a primitive recursive function that, applied to a $\mathcal{L}$-formula φ, yields the number of its logical symbols. We interpret the truth predicate of S with the partial truth predicate, definable in PA, for sentences of complexity up to n. (ii) follows from (i) and Orey's compactness theorem, according to which if W is locally interpretable in T and T is reflexive, then T interprets W. The proof of (iii) contained in Fischer (2009) works if instead of the cut-elimination argument of Halbach (1999),[9] one employs the proof of the conservativity of $\mathsf{CT}\restriction$ over PA

[9]It has been shown to contain a mistake by Kentaro Fujimoto.

contained in Leigh (2014). A new, direct proof is contained in Enayat &Visser (2014, Theorem 5.1). (iv) is straightforward from Lemma 6(iv). □

Proposition 1. CT interprets ACA. Moreover, the interpretation behaves like the identity mapping on the arithmetical vocabulary.

Proof. A full proof is contained in Halbach (2014): we present some of its details as they will be useful for later proofs.

We define a substitution function for elements of $\mathsf{Fml}^1_{\mathcal{L}}$:

$$\mathsf{sb}(\ulcorner \varphi(v_1) \urcorner, \ulcorner t \urcorner) := \ulcorner \varphi(t) \urcorner$$

We let $\mathsf{sb}(x, y)$ take a default value, e.g. $\ulcorner 0 = 0 \urcorner$, when it is not applied to an element of $\mathsf{Fml}^1_{\mathcal{L}}$ and to a term. The translation ι is then crucially defined by the following clauses:

$(x \in y)^\iota :\leftrightarrow \mathsf{Tr}(\mathsf{sb}(x, y))$ $\quad$ y is satisfied by x

ι maps the arithmetical nonlogical vocabulary of $\mathcal{L}_2$ into itself;

ι commutes with propositional quantifiers;

$(\forall X\, \varphi)^\iota :\leftrightarrow \forall X^\iota\, (\mathsf{Fml}^1_{\mathcal{L}}(X^\iota) \rightarrow (\varphi)^\iota)$

Set variables and number variables are kept distinct to avoid clashes: this is emphasized by denoting with X^ι the $\mathcal{L}$-variable resulting from the translation. Strictly speaking, X^ι is a variable of the only available sort.

To verify that ι behaves as required, we consider the crucial case of comprehension axioms, i.e. we prove

$$(\exists X\, \forall x\, (x \in X \leftrightarrow \varphi(x))^\iota \tag{4}$$

Notice that $\varphi(x)$ *may contain first- and second-order parameters.* Then, in CT:

$$\forall x(\varphi^\iota(x, \mathsf{Tr}(\mathsf{sb}(w, X^\iota))) \leftrightarrow \varphi^\iota(x, \mathsf{Tr}(\mathsf{sb}(w, X^\iota)))) \quad \text{where } w \text{ is a closed term} \tag{5}$$

$$\forall x(\mathsf{Tr}^\ulcorner \varphi^\iota(\dot{x}, \mathsf{sb}(w, X^\iota))^\urcorner \leftrightarrow \varphi^\iota(x, \mathsf{Trsb}(w, X^\iota))) \quad \text{Lemma 6(i) and CT1-4} \tag{6}$$

$$\exists Y^\iota\, \forall x(\mathsf{Tr}(\mathsf{sb}(\dot{x}, Y^\iota)) \leftrightarrow \varphi^\iota(x, \mathsf{Tr}(\mathsf{sb}(w, X^\iota)))) \tag{7}$$

Essentially, the last line results from the application of the compositional axioms of CT to move the occurrences of the truth predicate in $\varphi^\iota(x, \mathsf{Tsb}(w, X^\iota))$ in front of the formula. In line (6), we reason for an arbitrary φ: strictly speaking, an external induction on the complexity of the formula is needed in which the different compositional axioms are employed. The last line follows from the properties of substitution. □

In the proof of Lemma 1, the compositional axioms CT1-4 are only employed in moving from (5) to (6) and, in particular, to deal with parameters occurring into the comprehension scheme. We then let $\mathsf{ACA}^{\mathsf{pf}}$ to be exactly as ACA but with no parameters allowed to appear into the comprehension scheme, and $\mathsf{ACA}\restriction$ and $\mathsf{ACA}^{\mathsf{pf}}\restriction$ their counterparts with induction involving only $\mathcal{L}$-formulas. From the proof of Proposition 1 we can thus extract:

Proposition 2.

(i) $\mathsf{UTB}\restriction$ interprets $\mathsf{ACA}^{\mathsf{pf}}\restriction$

(ii) UTB interprets $\mathsf{ACA}^{\mathsf{pf}}$.

In both cases, the translation behaves like id for arithmetical vocabulary.

Proposition 2 tells us that the admission of second-order parameters in the comprehension schema corresponds, on the truth-theoretic side, to a substantial jump from a disquotational theory of truth to a compositional one. The substantiality of this jump can be measured in terms of proof-theoretic strength — and thus in terms of interpretability strength — as Lemma 6 shows. In other words, the presence of parameters in the comprehension schema enables one to go from a theory that is conservative over PA to a theory that proves Con(PA).

On the negative side, we have

Proposition 3.

(i) $\mathsf{TB}\restriction$, $\mathsf{UTB}\restriction$, $\mathsf{CT}\restriction$ do not interpret ACA_0;
(ii) CT^- interprets $\mathsf{ACA}_{\Pi^1_1}$

Proof. (i) If one of $\mathsf{TB}\restriction$, $\mathsf{UTB}\restriction$, $\mathsf{CT}\restriction$ interpreted ACA_0, by Lemma 7, PA would interpret ACA_0, quod non. (ii) follows from Cantini (1989, Proposition 3.6). □

Positive inductive definitions can be employed to define the set of arithmetical truths Moschovakis (1974). The idea is to identify this set as the fixed point of a suitable operator on sets of natural numbers extracted by the inductive definition Halbach (2014, §8.7). The theory PT captures the clauses of this inductive definition. PT and its subsystems are studied in Fischer (2009). The peculiarity of positive inductive definitions is that in their clauses the truth predicate can only appear in the scope of an even number of negation symbols.

Definition 3.

(i) PT extends PAT with the axioms:

(PT1) $\mathsf{CTerm}_{\mathcal{L}}(x) \wedge \mathsf{CTerm}_{\mathcal{L}}(y) \rightarrow \big(\mathsf{Tr}\, x \dot{\circ} y \leftrightarrow \mathsf{val}(x) \circ \mathsf{val}(y)\big)$

(PT2) $\mathsf{CTerm}_{\mathcal{L}}(x) \wedge \mathsf{CTerm}_{\mathcal{L}}(y) \rightarrow \big(\mathsf{Tr}\, \dot{\neg} x \dot{\circ} y \leftrightarrow \neg\, (\mathsf{val}(x) \circ \mathsf{val}(y))$

(PT3) $\mathsf{Sent}_{\mathcal{L}}(x) \rightarrow (\mathsf{Tr}_{\dot{\neg}\dot{\neg}} x \leftrightarrow \mathsf{Tr}\, x)$

(PT4) $\mathsf{Sent}_{\mathcal{L}}(x \dot{\wedge} y) \rightarrow (\mathsf{Tr}\, x \dot{\wedge} y \leftrightarrow (\mathsf{Tr}\, x \wedge \mathsf{Tr}\, y))$

(PT5) $\mathsf{Sent}_{\mathcal{L}}(x \dot{\wedge} y) \leftrightarrow (\mathsf{Tr}\, \dot{\neg}(x \dot{\wedge} y) \leftrightarrow (\mathsf{Tr}\, \dot{\neg} x \vee \mathsf{Tr}_{\dot{\neg}} y)$

(PT6) $\mathsf{Sent}_{\mathcal{L}}(\dot{\forall} v x) \leftrightarrow (\mathsf{Tr}\, \dot{\forall} v x \leftrightarrow\ \forall y(\mathsf{Cterm}(y) \rightarrow \mathsf{Tr}(x(y/v))))$

(PT7) $\mathsf{Sent}_{\mathcal{L}}(\dot{\neg}\dot{\forall} v x) \leftrightarrow (\mathsf{Tr}_{\dot{\neg}\dot{\forall}} v x \leftrightarrow\ \exists y(\mathsf{Cterm}(y) \wedge \mathsf{Tr}_{\dot{\neg}} x(y/v)))$

(ii) $\mathsf{PT}\restriction$ restricts the induction scheme to $\mathcal{L}$–formulas
(iii) In PT^- the induction scheme of PAT is replaced by (IInd).

PT is identical to CT, as PT proves (CT2) by formal induction on the complexity of sentences. The proof of Proposition 1 carries over without modifications: PT thus interprets ACA in the manner described.

$\mathsf{PT}\restriction$ on the other hand, that is when induction is restricted to $\mathcal{L}$-sentences, is a *proper* subtheory of $\mathsf{CT}\restriction$. To see this, one notices that the general theory of inductive definitions entails the existence of fixed points of positive inductive definitions: this fact can be used to define a subset $S \subseteq \mathcal{M}$ for any $\mathcal{M} \vDash \mathsf{PA}$ such that $(\mathcal{M}, S) \vDash \mathsf{PT}\restriction$ (cf. Cantini (1989)). By a theorem of Lachlan (Lachlan (1981)), by contrast, not any model of PA can be expanded

to a model of $\mathsf{CT}\restriction$. This explains the properness of the subtheory relation. Therefore, by Lemma 7, PA interprets $\mathsf{PT}\restriction$.[10]

To get closer to our interests and consider subsystems of second-order arithmetic, we have:

Proposition 4.

(i) PT interprets ACA;
(ii) PT^- interprets ACA_0.

Proof. (i) follows from the identity of CT and PT and Lemma 1. (ii) is proved by Fischer (2009): the comprehension axioms require external induction on the complexity of the formula involved. Compare (5)-(7) in the proof of Proposition 1: in (6), the hidden induction on the complexity of formulas is replaced here by an induction on its *positive* complexity. Moreover, PT^- proves the translation of $\mathsf{tot}(x)$ where $x \in \mathsf{Fml}^1_{\mathcal{L}}$. □

3.2. **Truth from Comprehension.**

Proposition 5. ACA defines the truth predicate of CT.

Proof. The proof is contained in Takeuti (1987). Let $\mathsf{lcx}(\cdot)$ be as above, and $\mathcal{T}(X, x)$ mean 'X is a truth set for sentences of complexity $\leq x$', that is

$$\begin{aligned}\mathcal{T}(X,x) :\leftrightarrow\ & \forall u,v\,(\mathsf{Cterm}_{\mathcal{L}}(u) \wedge \mathsf{Cterm}_{\mathcal{L}}(v) \to (u \dot{\circ} v \in X \leftrightarrow \mathsf{val}(u) \circ \mathsf{val}(v))) \wedge \\ & \forall u\,(\mathsf{Sent}_{\mathcal{L}}(\dot{\neg} u) \wedge \mathsf{lcx}(\dot{\neg} u) \leq x \to (\dot{\neg} u \in X \leftrightarrow u \notin X)) \wedge \\ & \forall u,v\,(\mathsf{Sent}_{\mathcal{L}}(u \dot{\wedge} v) \wedge \mathsf{lcx}(u \dot{\wedge} v) \leq x \to (u \dot{\wedge} v \in X \leftrightarrow (u \in X \wedge v \in X))) \wedge \\ & \forall u,v\,(\mathsf{Sent}_{\mathcal{L}}(\dot{\forall} uv) \wedge \mathsf{lcx}(\dot{\forall} uv) \leq x \to (\dot{\forall} uv \in X \leftrightarrow \forall y(\mathsf{CTerm}_{\mathcal{L}}(y) \to \mathsf{sb}(y,v) \in X)))\end{aligned}$$

It is first noticed that $\mathsf{ACA}^{\mathsf{pf}}$ with the induction schema restricted to $\mathcal{L}_2$-formulas with no bound set variables proves, by induction on x, that truth definitions for sentences of restricted complexity are unique:

$$\mathcal{T}(X,x) \wedge \mathcal{T}(Y,x) \wedge \mathsf{Sent}_{\mathcal{L}}(y) \wedge \mathsf{lcx}(y) \leq x \to\ (y \in X \leftrightarrow y \in Y) \tag{8}$$

By arithmetical comprehension, one obtains, in $\mathsf{ACA}^{\mathsf{pf}}\restriction$,

$$\exists X\, \mathcal{T}(X, 0), \tag{9}$$

because sentences with logical complexity ≤ 0 are atomic statements. Moreover, in ACA_0, we have, still by arithmetical comprehension, that

$$\forall x\,(\exists X(\mathcal{T}(X,x)) \to \exists X(\mathcal{T}(X, x+1))) \tag{10}$$

An instance of the induction schema of ACA (indeed $\mathsf{ACA}_{\Sigma^1_1}$ would suffice, as we only need an instance of Σ^1_1-induction), gives us

$$\forall x\, \exists X\, \mathcal{T}(X,x) \tag{11}$$

The required truth definition is given by the formula

$$\tau(x) :\leftrightarrow \exists X(\mathcal{T}(X, \mathsf{lcx}(x)) \wedge x \in X) \tag{12}$$

□

Corollary 1.

[10] There is something more that holds: any model of PA can be expanded to a model of PT^-.

(i) $\mathsf{ACA}^{\mathsf{pf}}\restriction$ defines the truth predicate of $\mathsf{UTB}\restriction$;
(ii) $\mathsf{ACA}^{\mathsf{pf}}$ defines the truth predicate of UTB;
(iii) ACA_0 interprets PT^-;
(iv) $\mathsf{ACA}_{\Pi^1_1}$ defines the truth predicate of CT^-;
(v) ACA defines the truth predicate of PT.

Proof. (i) and (ii) are immediate by inspection of the proof of Proposition 5, by induction on the complexity of the formula involved in the uniform diquotation. A detailed proof of (iii) can be found in Fischer (2009): one replaces $\mathsf{lcx}(\cdot)$ with a measure of the *positive complexity* $\mathsf{pcx}(\cdot)$ in an analog of (12). Then it is shown that ACA_0 proves (IInd) under the translation: to this end, the induction schema for arithmetical formulas of $\mathcal{L}_2$, provable in ACA_0, is employed. (iv), in addition, requires Π^1_1 induction to prove CT2. (v) is immediate by Proposition (5) and the fact that PT is a subtheory of CT. □

3.3. **Extensions and Hierarchies.** The processes of extending PA to CT and ACA can be iterated. To this end, one first assumes a suitable notation for ordinals below Γ_0 (cf. Pohlers (2009, Ch. 2)): in other words, any ordinal $\alpha < \Gamma_0$ can be coded by a natural number and thus represented in PA by a unique term $\overline{\alpha}$.

To define the theories $\mathsf{RT}_{<\theta}$ for $\theta \leq \Gamma_0$ —RT standing for 'ramified truth'— one considers languages $\mathcal{L}_{<\theta} := \mathcal{L}_{\mathsf{PA}} \cup \{\mathsf{T}_0, ..., \mathsf{T}_\eta\}$ for $\eta < \theta$, where $\mathsf{T}_\alpha := (\mathsf{T}, \alpha)$. The axioms of $\mathsf{RT}_{<\theta}$ thus stipulate how the truth predicates T_η work for sentences of $\mathcal{L}_{<\eta}$, for $\eta < \theta$ (cf. Halbach (2014, p. 113)).

The definition of iterations of ACA to $\theta \leq \Gamma_0$, called $\mathsf{RA}_{<\theta}$ for 'ramified analysis up to θ', enjoys different variations. Since we are not so much interested in iterations of predicative comprehension, we refer to Feferman (1964) for the definition of $\mathsf{RA}_{<\theta}$.

The fundamental fact that links ramified truth and analysis is the following, whose proof is sketched in Feferman (1991):

Proposition 6 (Feferman)**.** For $\alpha \leq \Gamma_0$, $\mathsf{RT}_{<\alpha}$ and $\mathsf{RA}_{<\alpha}$ are mutually interpretable.

In the light of Proposition 6, one may for instance obtain a reduction of the classical axiomatization of Kripke's theory of truth KF (cf. Feferman (1991), Halbach (2014)) to $\mathsf{RT}_{<\epsilon_0}$: in particular, KF defines all truth predicates of $\mathsf{RT}_{<\epsilon_0}$. Also, from Cantini (1989, §9) one can extract an interpretation of KF in $\mathsf{RT}_{<\epsilon_0}$. Given the mutual truth definability between the remarkable disquotational system PUTB of positive uniform (type free) disquotation and KF (cf. Halbach (2009)), Proposition 6 closely relates PUTB and iterations of predicative comprehension. It also shows the close relationships between KF, PUTB, and $\widehat{\mathsf{ID}}_1$.[11]

Many other results and reductions between type-free truth and second order-arithmetic have been investigated in recent years. As we anticipated in the introduction, however, the strategy put forward in the next section does not enjoy a plausible extension to the type-free case yet: we lack natural ways of extending the strategy proposed in the next section allow self-applications of the operations on theories to be defined. Therefore we decided to omit a systematic survey of the relationships between type-free systems and subsystems of second-order arithmetic.

[11] $\widehat{\mathsf{ID}}_1$ is the theory postulating the existence of fixed-points for arbitrary positive arithmetical operators. For a detailed definition of the full ID_1, cf. Pohlers (2009, Ch. 9, Def. 9.1).

4. Truth and Predicative Comprehension as Functors

As anticipated in the introduction, we will now consider the reductions surveyed in §3 from a different angle. This approach finds its roots in a philosophical attempt to distinguish patterns of reasoning belonging to the syntactico/truth-theoretic component of a theory of truth on one side and its mathematical/object-theoretic component on the other. This separation has been used, also by the author, to analyze the connections between axiomatic truth and truth-theoretic deflationism (cf. Halbach (2014),Heck (2015), Nicolai (2015a)). The fundamental insight offered by this approach is thus that, unlike what happens in the standard construction, the assumption of a full, Tarskian theory of truth does not immediately lead to new object-theoretic, or mathematical insights, although it may do so via a suitably defined interpretation.

In §4.1 we define and introduce the basic properties of $\mathsf{T}[\cdot]$ and $\mathsf{Tp}[\cdot]$ that operate on theories U by applying Tarskian truth and positive, typed truth to them. Although inductive reasoning involving truth-theoretic and syntactic notions is not directly available in $\mathsf{T}[\cdot]$ (and $\mathsf{Tp}[\cdot]$), it can be mimicked in it by exploiting the properties of definable cuts introduced in §2: this will be the core of §4.2. The arithmetized model corresponding to the formalization —in weak arithmetical context— of the construction of the term model used in Henkin's proof of the completeness theorem, comes with a truth predicate. This truth predicate is used in §4.3 to interpret the truth predicates of $\mathsf{T}[\cdot]$ and $\mathsf{Tp}[\cdot]$. In §4.4 we turn to predicative comprehension and investigate the functors $\mathsf{PC}[\cdot]$ and $\mathsf{PCS}[\cdot]$; in §4.5 we finally relate them to $\mathsf{T}[\cdot]$ and $\mathsf{Tp}[\cdot]$.

In the following sections the object theory U will be assumed to be formulated in a relational language. On occasion we will require sequentiality or the capability of interpreting S^1_2.

4.1. Typed-Truth as an Operation on Theories.

The functor $\mathsf{T}[\cdot]$ applies to an arbitrary RE theory U and yields the three-sorted theory $\mathsf{T}[U]$ in a language $\mathcal{L}_T$ with sorts in $\{\mathfrak{s}, \mathfrak{o}, \mathfrak{sq}\}$ (for 'syntax', 'object-theory', 'sequences/variable assignments'), constants proper of the language $\mathcal{L}_\#$ of S^1_2,[12] the function symbol $\cdot(\cdot)$ of type $(\mathfrak{sq}, \mathfrak{s}) \to \mathfrak{o}$ and the predicate symbol Sat of sort $(\mathfrak{s}, \mathfrak{sq})$. The former will give rise to expressions of the form $v_i^{\mathfrak{sq}}(v_i^{\mathfrak{s}}) = v_i^{\mathfrak{o}}$, stating that the $v_i^{\mathfrak{s}}$-th element of a variable assignment $v_i^{\mathfrak{sq}}$ is the U-object $v_i^{\mathfrak{o}}$, whereas the latter will be characterized as a satisfaction predicate. Greek letters $\varphi, \psi, \ldots$ are taken to range over formulas of $\mathcal{L}_U$.

The axioms of $\mathsf{T}[U]$, besides the axioms of U, are

- (S^1_2) axioms of S^1_2
- (sq) $\exists v_j^{\mathfrak{sq}} \left(\forall v_k^{\mathfrak{s}} (v_k^{\mathfrak{s}} \neq v_i^{\mathfrak{s}} \to v_i^{\mathfrak{sq}}(v_k^{\mathfrak{s}}) = v_j^{\mathfrak{sq}}(v_k^{\mathfrak{s}})) \wedge v_j^{\mathfrak{sq}}(v_i^{\mathfrak{s}}) = v_i^{\mathfrak{o}}\right)$
- (tat) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner R(v_1^{\mathfrak{o}}, \ldots, v_n^{\mathfrak{o}})\urcorner) \leftrightarrow R(v_i^{\mathfrak{sq}}(\ulcorner v_1^{\mathfrak{o}}\urcorner), \ldots, v_i^{\mathfrak{sq}}(\ulcorner v_n^{\mathfrak{o}}\urcorner))$ for every relation symbol R in $\mathcal{L}_U$
- (t$\neg$) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg}\ulcorner\varphi\urcorner) \leftrightarrow \neg\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\varphi\urcorner)$
- (t$\vee$) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\varphi \vee \psi\urcorner) \leftrightarrow \mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\varphi\urcorner) \vee \mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\psi\urcorner)$

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

(tq) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner \exists v_i^{\mathfrak{o}} \varphi \urcorner) \leftrightarrow (\exists v_j^{\mathfrak{sq}} \overset{\ulcorner v_i^{\mathfrak{o}} \urcorner}{\sim} v_i^{\mathfrak{sq}})(\mathsf{Sat}(v_j^{\mathfrak{sq}}, \ulcorner \varphi \urcorner))$

where

$$v_j^{\mathfrak{sq}} \overset{v_i^{\mathfrak{s}}}{\sim} v_i^{\mathfrak{sq}} \;\leftrightarrow\; \forall v_j^{\mathfrak{s}} (v_j^{\mathfrak{s}} \neq v_i^{\mathfrak{s}} \rightarrow v_j^{\mathfrak{sq}}(v_j^{\mathfrak{s}}) = v_i^{\mathfrak{sq}}(v_j^{\mathfrak{s}}))$$

(sq) states minimal conditions on the existence and on the behaviour of sequences. (tq) simply formalizes Tarski's clause on quantified formulas: $\exists v_i \varphi$ is satisfied by a variable assignment if and only if there is a second variable assignment, differing from the former only in what they assign to v_i (this is captured by the abbreviation $v_j^{\mathfrak{sq}} \overset{v_i^{\mathfrak{s}}}{\sim} v_i^{\mathfrak{sq}}$), satisfying φ.

Intuitively, by applying the operator $\mathsf{T}[\cdot]$ to U, one expands a model of U with a universe of syntactic objects —here numbers in the sense of S^1_2— and a universe of 'mixed' objects, sequences of $\mathcal{L}_U$-variable assignments living in a disjoint domain.

We define an alternative functor: $\mathsf{Tp}[U]$ adds *positive* compositional axioms similar to the axioms of $\mathsf{PT}\restriction$ in Definition 3 but in a many-sorted way compatible with the idea behind the construction of $\mathsf{T}[U]$. Again $\mathsf{Tp}[U]$ is formulated in $\mathcal{L}_T$; its axioms, besides the axioms of U, S^1_2, (sq), (tat), (t$\vee$), are

(fat) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg} \ulcorner R(v_1^{\mathfrak{o}}, \ldots, v_n^{\mathfrak{o}}) \urcorner) \leftrightarrow \neg(R(v_i^{\mathfrak{sq}}(\ulcorner v_1^{\mathfrak{o}} \urcorner), \ldots, v_i^{\mathfrak{sq}}(\ulcorner v_n^{\mathfrak{o}} \urcorner)))$
for every relation symbol R in $\mathcal{L}_U$

(tdn) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg}\dot{\neg} \ulcorner \varphi \urcorner) \leftrightarrow \mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner \varphi \urcorner)$

(f$\vee$) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg} \ulcorner \varphi \vee \psi \urcorner) \leftrightarrow \big(\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg} \ulcorner \varphi \urcorner) \wedge \mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg} \ulcorner \psi \urcorner)\big)$

(fq) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg} \ulcorner \exists v_i^{\mathfrak{o}} \varphi \urcorner) \leftrightarrow (\forall v_j^{\mathfrak{sq}} \overset{\ulcorner v_i^{\mathfrak{o}} \urcorner}{\sim} v_i^{\mathfrak{sq}})(\mathsf{Sat}(v_j^{\mathfrak{sq}}, \dot{\neg} \ulcorner \varphi \urcorner))$

By external induction on the positive complexity of $\varphi(\vec{v}_i^{\mathfrak{o}})$, we have:

Lemma 8. $\mathsf{Tp}[U]$ (and thus $\mathsf{T}[U]$) proves, for all $\varphi(\vec{v}_i^{\mathfrak{o}})$ in $\mathcal{L}_U$,

(i) $\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner \varphi(v_{i_1}^{\mathfrak{o}}, \ldots, v_{i_n}^{\mathfrak{o}}) \urcorner) \leftrightarrow \varphi(v_i^{\mathfrak{sq}}(\ulcorner v_{i_1}^{\mathfrak{o}} \urcorner), \ldots, v_i^{\mathfrak{sq}}(\ulcorner v_{i_n}^{\mathfrak{o}} \urcorner))$

We will often refer to variables of sort $\mathfrak{o}, \mathfrak{s}, \mathfrak{sq}$ as $u, v, w, x, y, z \ldots$, $i, j, k, l, m, n \ldots$, and $a, b, c, \ldots$ respectively.

4.2. **Consistency from Truth.** A remarkable fact concerning the method of shortening of cuts, introduced in §2, is that it enables one to prove the consistency of the object theory U relativized to a definable cut, once we are able to prove, in the theory of truth, that all axioms of U are true.[13]

Therefore we define

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

where $\mathsf{AxL}_U(k)$ and $\mathsf{Ax}_U(k)$ are Δ^b_1-representations of the logical and the nonlogical axioms of U in S^1_2. AxT_U thus states that all logical and nonlogical axioms of U are true, i.e. satisfied by all sequences. This definition of AxT_U enables us to simplify the proof of a similar result given in Nicolai (2015); to this extent, we recall that U is taken to be formulated in a Hilbert-style calculus in which Modus Ponens is the only rule of inference. It should be noticed, however, that the result is not dependent on the choice of a specific logical calculus: this

[13] This fact has been brought to my attention by Richard Heck (cf. Heck (2015)), although already well-known in other contexts.

only simplifies the description. Thus we omit details on the specific formulation of U in the statements of the results.

To simplify the notation, from now on we set:

$$\mathsf{T}^+[U] := \mathsf{T}[U] + \mathsf{AxT}_U$$
$$\mathsf{Tp}^+[U] := \mathsf{Tp}[U] + \mathsf{AxT}_U$$

Lemma 9. $\mathsf{T}[U]^+$ proves the consistency of U on a cut.

Proof. We recall that $\mathsf{Prf}_U(k)$ is taken to be a Δ_1^b-formula in S_2^1 expressing that k is the code of a U-proof, and $\mathsf{lst}(k)$ a Σ_1^b function yielding, when applied to a sequence k, the last element of k.

We break down the proof in several steps.

We first consider the following formula with only n free.

$$\mathcal{K}(n) :\leftrightarrow \ (\forall m \leq n)\left(\mathsf{Prf}_U(m) \to \forall a\, \mathsf{Sat}(a, \mathsf{lst}(m))\right) \tag{13}$$

It expresses that any proof smaller than n has a true conclusion. We show that it is inductive. The claim clearly holds for $n = 0$, by assumption. Assuming it holds for arbitrary proofs $k, l \leq n$ and sequence a, we want to prove it for proofs $m \leq n+1$. In the interesting case, we have

$$\mathsf{lst}(m) = \mathsf{mp}(\mathsf{lst}(k), \mathsf{lst}(l)),$$

where $\mathsf{mp}(m, n)$ is a Σ_1^b-function yielding the result of applying Modus Ponens to m and n. In particular, here we require $\mathsf{lst}(k)$ of the form $\mathsf{lst}(l) \dot{\to} \mathsf{lst}(m)$. The claim is thus obtained by applying (t$\neg$) and (t$\vee$) to $\mathsf{Sat}(a, \mathsf{lst}(k))$ and combining it with $\mathsf{Sat}(a, \mathsf{lst}(l))$.

We shorten $\mathcal{K}$ to a cut $\mathcal{J}$ such that, for formulas j, we have

$$\exists m\, (\mathcal{J}(m) \wedge \mathsf{Proof}_U(m, j)) \to \forall a\, \mathsf{Sat}(a, j)$$

We can safely assume that S_2^1 is available in $\mathcal{J}$.

Assuming the monotonicity of the coding — that is codes of elements of sequences are smaller of the code of the whole sequence —- if $m \in \mathcal{J}$, also j will be. Finally we reason in the standard way: let $\bot$ code, in the S_2^1-numbers, an absurdity in the sense of U. If $\mathsf{Pr}_U^{\mathcal{J}}(\bot)$, then $\mathsf{Sat}(a, \bot)$. By Lemma 8, we obtain $\neg\mathsf{Pr}_U^{\mathcal{J}}(\bot)$, that is $\neg\, \exists m\, (\mathcal{J}(m) \wedge \mathsf{Proof}_U(m, \bot))$, that is $\mathsf{Con}^{\mathcal{J}}(U)$.

□

By Lemma 8, the parametrized Tarski-biconditionals are provable in $\mathsf{Tp}[U]$. To prove that all logical axioms of A are true on a cut, we need extra-care. In a word, we have to make sure that syntax and truth interact well on this cut. The problem is essentially that, in order to prove the truth of a logical axiom schema, say $\forall v_0 \varphi \to \varphi(y/v_0)$, one has to employ an instance of the very same principle, e.g

$$\forall m \mathsf{Sat}(a, k(m/{}^{\ulcorner}v_0{}^{\urcorner})) \to \mathsf{Sat}(a, k(t/{}^{\ulcorner}v_0{}^{\urcorner})) \tag{14}$$

to be able to apply (t$\neg$),(t$\vee$), (tq) and obtain

$$\mathsf{Sat}(a, {}^{\ulcorner}\forall v_0 \varphi \to \varphi(y/v_0){}^{\urcorner}) \tag{15}$$

The step from (14) to (15) requires, for instance, that variable assignments and formal substitution behave as required under the scope of Sat. Since these lemmata are usually proved by induction, we need to resort to a definable cut in which this behaviour is preserved. For

details, we refer to Heck (2015) and Nicolai (2015). Once this is done, we have (recall that $A, B, C\ldots$ range over finitely axiomatized theories)

Corollary 2. $\mathsf{T}[A]$ proves the consistency of A on a cut.

Surprisingly enough, a similar results can be transferred to $\mathsf{Tp}^+[U]$. We have already seen that CT and PT, unlike their variants featuring arithmetical induction only, are extensionally identical. Induction extended to the truth predicate is in fact needed to prove the full axiom for negation. The technology of shortening of cuts enables us to mimic the role played by the extended induction schema of PT in deriving (CT2).

Lemma 10. There is a $\mathsf{Tp}[U]$-definable cut in which (t¬) holds.

Proof. We reason in $\mathsf{Tp}[U]$. Let $\mathsf{lc}(x)$ a Σ^b_1-function that, applied to a $\mathcal{L}_U$-formula φ, yields the number of its logical symbols. By Lemma 1 the claim is obtained once we show that the following is inductive:

$$\mathcal{K}_0(n) := \forall k \forall a(\mathsf{Fml}_{\mathcal{L}_U}(k) \wedge \mathsf{lc}(k) \leq n \to (\mathsf{Sat}(a, \dot{\neg} k) \leftrightarrow \neg \mathsf{Sat}(a, k))$$

If $n = 0$, we use (tat) and (fat). Assuming the claim is true for formulas k with $\mathsf{lc}(k) \leq n$, we obtain the claim by the truth/falsity clauses for connectives and quantifiers and the 'inductive' assumption. For instance, if $k = \dot{\neg} k_0$, we have:

$$\begin{aligned} \neg \mathsf{Sat}(a, \dot{\neg} k_0) &\leftrightarrow \neg\neg \mathsf{Sat}(a, k_0) && \text{by assumption} \\ &\leftrightarrow \mathsf{Sat}(a, k_0) \\ &\leftrightarrow \mathsf{Sat}(a, \dot{\neg}\dot{\neg} k_0) && \text{(tdn)} \end{aligned}$$

□

Corollary 3. $\mathsf{Tp}[U]$ interprets $\mathsf{T}[U]$ on a definable cut.

More generally, Lemma 10 gives a strategy to interpret $\mathsf{CT}\restriction$ into $\mathsf{PT}\restriction$, the standard systems of Tarskian and positive truth considered in §3.

Once we have (t¬) available on suitable initial segment $\mathcal{K}_0$, we can run the proof of Lemma 9 paying attention to work in $\mathsf{Tp}^+[U]$-definable cuts that shorten $\mathcal{K}_0$ and in which the axioms of S^1_2 are satisfied. We thus have:

Proposition 7. $\mathsf{Tp}[U]^+$ proves the consistency of U on a cut.

Proof. Instead of $\mathcal{K}(n)$ in (13), we can start with the formula

$$\mathcal{K}_1(n) :\leftrightarrow (\forall m \leq n)(\mathsf{Prf}_U(m) \wedge \mathcal{K}_0(\mathsf{lst}(m)) \to \forall a\, \mathsf{Sat}(a, \mathsf{lst}(m))) \tag{16}$$

By the same argument used in the case of $\mathcal{K}(n)$ in the proof of Lemma 9, it can be shown that $\mathcal{K}_1(n)$ is inductive (only here the properties of $\mathcal{K}_0$ are used in the inductive step). The proof then proceeds unchanged. □

By Lemma 8, and taking care of truth of all logical axioms of A on a suitable definable cut as suggested above, we have

Corollary 4. $\mathsf{Tp}[A]$ proves the consistency of A on a cut.

Finally, we state the desired property of the result of applying $\mathsf{T}^+[\cdot]$ and $\mathsf{Tp}^+[\cdot]$ to arbitrary base theories:

Corollary 5.

(i) $\mathsf{T}^+[U]$ and $\mathsf{Tp}^+[U]$ interpret $\mathsf{S}^1_2 + \mathsf{Con}(U)$, and thus $\mathsf{Q} + \mathsf{Con}(U)$;
(ii) $\mathsf{T}[A]$ and $\mathsf{Tp}[A]$ interpret $\mathsf{S}^1_2 + \mathsf{Con}(A)$, and thus $\mathsf{Q} + \mathsf{Con}(A)$.

Proof. Ad (i), we simply take the cut $\mathcal{J}$ and the cut resulting from $\mathcal{K}_1$ from Proposition 7 as domains of our interpretations. We notice that, by Lemma 2, S^1_2 is holds in them. (ii) is a weaker claim than (i). □

4.3. **Truth from Consistency.** We now turn to the question of how to recover the semantic machinery of $\mathsf{T}[\cdot]$ and $\mathsf{Tp}[\cdot]$ via the assumption of the consistency of the base theory. Feferman offered in Feferman (1960) a full formalization of Henkin's term model construction for a first-order, recursive, set of sentences S in extensions of PA plus a Π_1 assertion of the consistency of S. Following Visser (1991) and Visser (2009b), we refer to the resulting arithmetized model as the 'Henkin-Feferman construction'. The technology of definable cuts enables us to formalize this term model in very weak subsystems of arithmetic complemented with the consistency statement for S.

Proposition 8. $\mathsf{S}^1_2 + \mathsf{Con}(U)$ interprets U.

Proof Sketch. A full proof can be found in Visser (1991). As usual, we start with $\mathcal{L}_U$ and extend it to $\mathcal{L}^h_U$ by adding a constant $c_{\exists v_i \varphi}$ for every $\exists v_i \varphi$ in the extended language: notice that we have enough induction in $\mathsf{S}^1_2 + \mathsf{Con}(U)$ to formalize this inductive argument. The Henkin theory, from which the arithmetized model for U is read off, is also defined in stages:

$$U^h_0 := U$$

$$U^h_{n+1} := \begin{cases} U^h_n \cup \{\varphi\}, & \text{if } n = \ulcorner \varphi \urcorner \text{ (with } \varphi \in \mathcal{L}^h_U) \text{ and } \mathsf{Con}(U^h_n \cup \{\varphi\}), \\ U^h_n \cup \{\varphi\} \cup \{\psi(c_\varphi)\}, & \text{if } \varphi \text{ is } \exists v \psi \text{ and } \mathsf{Con}(U^h_n \cup \{\varphi\}) \\ U^h_n & \text{otherwise} \end{cases}$$

Unlike the definition of $\mathcal{L}^h_U$, Σ^b_1-induction is not sufficient to prove the existence of the union of all U^h_n's. However, it turns out that the formula defining the U^h_n's is inductive, so it can be shortened to a cut $\mathcal{I}$. The required, complete theory in $\mathcal{L}^h_U$, for sentences in $\mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}^h_U}$ can thus be taken to be $\mathcal{F} = \bigcup_{n \in \mathcal{I}} U^h_n$.

From $\mathcal{F}$, we define the required interpretation $\mathfrak{F}$, whose domain $\delta_{\mathfrak{F}}$ is the set of codes of Henkin constants of $\mathcal{L}^h_U$ in $\mathcal{I}$. To the theory $\mathcal{F}$ there corresponds a truth predicate S. We set

$$R^{\mathfrak{F}}(v^0) :\leftrightarrow \delta_{\mathfrak{F}}(x) \wedge R(x) \in \mathsf{S} \qquad \text{for any } R \in \mathcal{L}_U \tag{17}$$

We notice that in (17) and in what follows, we write $\varphi(x) \in S$ to express that the result of formally substituting v_1 for x in φ falls into the extension of S. Crucially, $\mathsf{S}^1_2 + \mathsf{Con}(U)$ proves:

(a) for $\varphi \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}^h_U}$, $(\mathsf{S}\ulcorner\neg\varphi\urcorner) \leftrightarrow \neg(\mathsf{S}\ulcorner\varphi\urcorner)$;
(b) for $\varphi, \psi \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}^h_U}$, $\mathsf{S}(\ulcorner\varphi \vee \psi\urcorner) \leftrightarrow (\mathsf{S}(\ulcorner\varphi\urcorner) \vee \mathsf{S}(\ulcorner\psi\urcorner))$;
(c) for $\exists v \varphi(v) \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}^h_U}$ and $v^0 \in \mathsf{Var}^{\mathcal{I}}$, $\mathsf{S}(\ulcorner\exists v^0 \varphi\urcorner) \leftrightarrow \exists x\,(\delta_{\mathfrak{F}}(x) \wedge (\varphi(x) \in \mathsf{S}))$;
(d) for all $\varphi \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}_U}$, $(\mathsf{Pr}_U(\ulcorner\varphi\urcorner) \to \mathsf{S}(\ulcorner\varphi\urcorner))$

Notice, in (d), that reflection holds for sentences of the *non extended language* $\mathcal{L}_U$ in the cut $\mathcal{I}$. □

By the interpretability of $\mathsf{S}^1_2 + \mathsf{Con}(U)$ in $\mathsf{Q} + \mathsf{Con}(U)$ on a definable cut, we have:

Corollary 6. $\mathsf{Q} + \mathsf{Con}(U)$ interprets U.

Now we show that the truth predicate $\mathsf{S}(\cdot)$ defined in the proof of Proposition 8 enables us to interpret the satisfaction predicate of $\mathsf{T}^+[U]$, and thus of $\mathsf{Tp}^+[U]$. Therefore we obtain Tarskian and positive, typed truth for the base theory form the assertion of the consistency of U.

Proposition 9. $\mathsf{S}^1_2 + \mathsf{Con}(U)$ interprets $\mathsf{T}^+[U]$.

Proof. We define the translation $\mathfrak{H}$. It is assume a suitable renaming of bound variables omitted for readability; S is the truth predicate defined in the proof of Proposition 8.

$$(R)^{\mathfrak{H}}(u) :\leftrightarrow R(u) \in \mathsf{S} \text{ for predicates } R \text{ of sort } \mathfrak{o} \tag{18}$$

$$(P)^{\mathfrak{H}}(u) :\leftrightarrow P(u) \text{ for predicates } P \text{ of sort } \mathfrak{s}; \tag{19}$$

$$(\mathsf{Sat})^{\mathfrak{H}}(u,v) :\leftrightarrow \mathsf{S}(\mathsf{sb}(v,u)); \text{ where sb substitutes elements of } u \text{ in } v \tag{20}$$

$$((\cdot)\cdot)^{\mathfrak{H}}(u,v,w) :\leftrightarrow ((u)_v = w \wedge v < \mathsf{lh}(u)) \vee (v \geq \mathsf{lh}(u) \wedge w = 0) \tag{21}$$

$$(\exists x\, A)^{\mathfrak{H}} :\leftrightarrow \exists x\, (\delta_{\mathfrak{F}}(x) \wedge A) \text{ where } \delta_{\mathfrak{F}} \text{ is as in Prop. 8;} \tag{22}$$

$$(\exists k\, A)^{\mathfrak{H}} :\leftrightarrow \exists k\, (\mathcal{I}(k) \wedge A) \text{ with } \mathcal{I} \text{ again as in Prop. 8} \tag{23}$$

$$(\exists a\, A)^{\mathfrak{H}} :\leftrightarrow \exists a, a \text{ is a (finite) sequence and } (\forall y \in a)(\delta_{\mathfrak{H}}(y)) \tag{24}$$

In (21), $(x)_y$ is an efficient version of the β function that outputs the y^{th} element of the finite sequence x (cf. **?**). In (22)-(24) we have kept the suggestive notation concerning variables even though, strictly speaking, we are not considering the language $\mathcal{L}_T$ anymore.

A full verification that $\mathfrak{H}$ is indeed a relative interpretation can be found in Nicolai (2015). Crucially, for AxT_U, one notices that, by Prop. 8(d),

$$\forall u\, (\mathsf{Sent}_{\mathcal{L}_U}(u) \wedge \mathcal{I}(u) \wedge \mathsf{Ax}_U(u) \rightarrow \mathsf{S}(u)) \tag{25}$$

Therefore, it suffices to notice that in $\mathsf{S}^1_2 + \mathsf{Con}(U)$

$$\text{for all } u \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}_U}, \mathsf{Ax}^{\mathfrak{H}}_U(u) \rightarrow \mathsf{Ax}_U(u). \tag{26}$$

□

Corollary 7.

(i) $\mathsf{S}^1_2 + \mathsf{Con}(U)$ interprets $\mathsf{Tp}^+[U]$;
(ii) $\mathsf{Q} + \mathsf{Con}(U)$ interprets $\mathsf{T}^+[U]$, and thus $\mathsf{Tp}^+[U]$.

Proof. (i) is immediate from the previous Proposition. (ii) follows from Lemma 3. □

Corollary 7 clearly holds for $\mathsf{T}[A]$ and $\mathsf{Tp}[A]$.

4.4. **The functors $\mathsf{PC}[\cdot]$ and $\mathsf{PCS}[\cdot]$.** The facts reported in this subsection — except the connections between these facts and the results above — are essentially due to Visser (2009b). We show that, given an arbitrary theory U,

The functor $\mathsf{PC}[\cdot]$ adds predicative comprehension to an arbitrary U. In order to be completely general, as our overall program requires, we need to be able to act on arbitrary U. To mention a famous example, we should be able to generalize the transformation of PA into ACA_0 to arbitrary theories. This can be done in the following way.

We first render the language two-sorted: objects over which quantifiers of $\mathcal{L}_U$ range are taken to be of sort $\mathfrak{o}$. We add to the language variables X of sort $\mathfrak{c}$, standing for sets (or concepts, or unary predicates) of objects of sort $\mathfrak{o}$ and a binary predicate A of type $(\mathfrak{c}, \mathfrak{o})$ such that $Xx :\leftrightarrow \mathsf{A}(X, x)$. To obtain $\mathsf{PC}[U]$, the axioms of U formulated in the new language are extended with the scheme of predicative comprehension that, in a pedantic version, reads

$$\exists u^{\mathfrak{c}}\, \forall v^{\mathfrak{o}}\, (\mathsf{A}(u^{\mathfrak{c}}, v^{\mathfrak{o}}) \leftrightarrow \varphi(v^{\mathfrak{o}}, \vec{w}^{\mathfrak{o}}, \vec{y}^{\mathfrak{c}})) \tag{ca}$$

If U is sequential, then $\mathsf{PC}[U]$ can be finitely axiomatized (Visser (2009b, Thm 3.4)).

In order to describe the next results in full generality, a slight modification of $\mathsf{PC}[\cdot]$ has to be considered. The required parallelism between consistency and predicative comprehension will break down if we considered arbitrary representations of the axiom set of U. A powerful theorem of Vaught states that any RE, sequential theory U can be axiomatized by a scheme. This scheme can be taken to be of the form $\Psi(\vec{B})$, where the B's are *schematic variables* for formulas of $\mathcal{L}_U$. Vaught's result states that U and the theory U_{V} resulting from the process just described are extensionally identical (cf. Vaught (1967)).

To define the theory $\mathsf{PCS}[U]$, we formulate U_{V} in a language with sorts in $\{\mathfrak{o}, \mathfrak{c}\}$ and translate $\Psi(\vec{B})$ into $\Psi(\vec{X})$: we call this 'translation' of U_{V} as given by $\Psi(\vec{X})$ the *v-class form* of U and denote it with U^{c}. To obtain $\mathsf{PCS}[U]$, one then adds to U^{c} both (ca) and the universal closure of $\Psi(\vec{X})$. A modification of the strategy required in Lemma 1 shows that we can find a definable $\mathsf{PCS}[U]$-cut on which there are no proofs of U-contradictions. In other words:

Lemma 11. Let U_{V} be sequential. Then $\mathsf{PCS}[U]$ proves the consistency of U_{V} on a cut.

Proof Sketch. We recall the main steps of Theorem 7.1 in Visser (2009b), and reason in PCS[U]. Let $\mathsf{lc}(\cdot)$ be as above. By the sequentiality of U_{V}, syntax is formalized in a standard way (e.g. by the natural numbers). We consider a reformulation modification of the definition of $\mathcal{T}(X, x)$ in Proposition 5, which we call $\mathcal{T}^*(X, x)$.

Now in Lemma 5 we were able to use (ca) and the induction schema of ACA that the uniqueness condition for truth sets. In $\mathsf{PCS}[U]$ we have no such induction. We have to use again the shortening techniques. In particular, we can show that the (pseudo-) class

$$\mathcal{M} := \{x \mid \exists! X\, \mathcal{T}^*(X, x)\}, \tag{27}$$

that is the class of all numbers which bound the complexity of formulas occurring in our truth sets, is indeed inductive. For any x, this unique X is denoted with X^x. We shorten $\mathcal{M}$ to a cut $\mathcal{N}$.

A satisfaction predicate is now extracted from $\mathcal{N}$. $\mathsf{Fml}^{1,\mathcal{N}}_{\mathcal{L}_U}$ is the class of formulas of $\mathcal{L}_U$, lying in $\mathcal{N}$, with one free variable. It is

$$\mathsf{St}(s, \ulcorner \varphi \urcorner) :\leftrightarrow (\exists x \in \mathcal{N})((s, \ulcorner \varphi \urcorner) \in X^x) \tag{28}$$

St enjoys the nice property that, for any φ in $\mathsf{Fml}^{1,\mathcal{J}}_{\mathcal{L}_U}$ and assignment s, one can find —again by (ca)— a Y_φ such that, for all y and v_i,

$$Y_\varphi y \text{ if and only if } \mathsf{St}(s(y/v_i), \ulcorner \varphi \urcorner) \tag{29}$$

In fact this Y_φ can simply be taken to be the set of y that satisfy φ in the sense of $X^{\mathsf{lc}(\varphi)}$. Since $\varphi \mathsf{Fml}^{1,\mathcal{J}}_{\mathcal{L}_U}$ we also know that Y_φ exists.

By employing an argument similar to the one employed in Lemma 9, we consider the set of U_{V}-proofs $\mathcal{P}$ whose members $\pi \in \mathcal{P}$ are such that $\mathsf{lh}(\pi) \leq x$, $\forall i \leq \mathsf{lh}(\pi)$, $(\pi)_i \in \mathsf{Fml}^{1,\mathcal{N}}_{\mathcal{L}_U}$ and such that $\forall s\ \mathsf{St}(s, \mathsf{lst}(\pi))$. The set $\mathcal{A}$ of such x is inductive: crucially, for the axiom $\Psi(\vec{\psi})$ of U_{V} with $\vec{\psi} \in \mathsf{Fml}^{1,\mathcal{N}}_{\mathcal{L}_U}$, we require

$$\mathsf{St}(s(y/v_i), \ulcorner \Psi(\vec{\psi}) \urcorner) \tag{30}$$

By definition of $\mathsf{PCS}[\mathrm{U}]$, we have $\forall \vec{Y}\ \Psi(\vec{Y})$ and thus $\Psi(\vec{\psi})$ for an arbitrary $\psi \in \mathsf{Fml}^{1,\mathcal{N}}_{\mathcal{L}_U}$. By (29), we associate to each formula in $\vec{\psi}$ the appropriate set Y_{ψ_i}. Thus $\Psi(\vec{Y}_{\psi})$. By the properties of St, we obtain (30). A similar argument holds for the logical axiom schemata. The inductive step is unproblematic and it is similar to the inductive step in the proof of Lemma 9.

We thus shorten $\mathcal{A}$ to a cut $\mathcal{A}_0 \subset \mathcal{A}$.

It follows that, as in Lemma 9, no proof of contradiction can be reached in $\mathcal{A}_0$: i.e. we have $\mathsf{Con}^{\mathcal{A}_0}(U_{\mathsf{V}})$. □

Corollary 8. Let U_{V} be sequential. Then $\mathsf{PCS}[U]$ interprets $\mathsf{S}^1_2 + \mathsf{Con}(U_{\mathsf{V}})$.

We now show, with the help of Proposition 8 that also the converse direction holds. In other words, that we can obtain predicative comprehension from consistency.

Lemma 12. $\mathsf{S}^1_2 + \mathsf{Con}(U_{\mathsf{V}})$ interprets $\mathsf{PCS}[U]$.

Proof. We first define the domain(s) of our interpretation $\mathfrak{I}$. The domain of quantifiers ranging over variables of sort $\mathfrak{o}$, $\delta_{\mathfrak{I}}$, is just $\delta_{\mathfrak{S}}$ (cf. Proposition 8), that is the domain of Henkin constant in a $\mathsf{S}^1_2 + \mathsf{Con}(U)$-definable initial segment of the natural numbers. The domain of class variables, with $\mathcal{L}^{\mathfrak{c}}_U$ denoting again the extension of $\mathcal{L}_U$ with a countable set of Henkin constants, is

$$\delta^{\mathfrak{c}}_{\mathfrak{I}} := \{x \mid x \in \mathsf{Fml}^1_{\mathcal{L}^{\mathfrak{c}}_U} \cap \mathcal{I} \text{ and the free variable of } x \text{ is of sort } \mathfrak{o}\}$$

We recall that $\mathcal{I}$ is the cut defined in the proof of Proposition 8. In the official definition of the language of $\mathsf{PCS}[U]$, Xx reads $\mathsf{A}(X, x)$. Therefore we translate:

$$(\mathsf{A}(X, x))^{\mathfrak{I}} :\leftrightarrow \mathsf{S}(\ulcorner \varphi \urcorner (x/v)), \qquad \text{if } X \text{ is mapped into } \varphi(v)$$

Unsurprisingly, the clause just states that Xx is translated as 'φ is satisfied by x'. We call φ_X the element of $\delta^{\mathfrak{c}}_{\mathfrak{I}}$ associated to X by the translation.

Crucially, $\mathsf{S}^1_2 + \mathsf{Con}(U)$ proves, by meta-induction on the complexity of $\Phi(\vec{Y})$,[14] and by the crucial contribution of Proposition 8(d):

$$(\forall \vec{Y} \in \delta^{\mathfrak{c}}_{\mathfrak{I}})\big((\Phi(\vec{Y}))^{\mathfrak{I}} \leftrightarrow \mathsf{S}(\ulcorner \Phi(\vec{\varphi_Y}) \urcorner)\big) \tag{31}$$

The base case is basically given by the definition of the translation. The other cases are easy by employing the properties of S.

To verify that $\mathfrak{I}$ is indeed a relative interpretation, we prove in $\mathsf{S}^1_2 + \mathsf{Con}(U)$ the translation of the axioms of $\mathsf{PCS}[U]$: in particular (i) axioms of the form $\forall \vec{Y}\ \Phi(\vec{Y})$ and (ii) all instances of comprehension.

(i) for $\Phi(\vec{X})$ with $\vec{X}$ arbitrary, by Lemma 8 we know that $\mathsf{Pr}_{U_{\mathsf{V}}}(\ulcorner \Phi(\vec{\varphi_X}) \urcorner)$. Thus by reflection (Lemma 8(d)), $\mathsf{S}(\ulcorner \Phi(\vec{\varphi_X}) \urcorner)$. By (31), $(\Phi(\vec{X}))^{\mathfrak{I}}$. By the properties of interpretations,, since $\vec{X}$ is arbitrary, $(\forall \vec{X} \Phi(\vec{X}))^{\mathfrak{I}}$.

[14] $\Phi(\vec{Y})$ does not contain any $\mathfrak{c}$-quantifiers.

Ad (ii), to any $\varphi(u, \vec{Z})$ with no bound set variables, we associate to it $\varphi_X(v_i, \vec{\varphi_Z})$ by the definition of $\mathfrak{I}$. Since φ is a standard formula, we know $\varphi_X(v_i, \vec{\varphi_Z})$ will be in $\delta^{\mathfrak{c}}_{\mathfrak{I}}$. But $\varphi_X(u, \vec{\varphi_Z})$ is just $(Xu)^{\mathfrak{I}}$, and it is equivalent to $\mathsf{S}(\varphi_X(u, \vec{\varphi_Z}))$ by the properties of $\mathfrak{I}$. By (31), we conclude $(\varphi(u, \vec{Z}))^{\mathfrak{I}}$. This shows that, given any $\varphi(u, \vec{Z})$, the translation of the comprehension axiom of $\mathsf{PCS}[U]$ is satisfied. □

Thus, by Lemma 3,

Corollary 9. Let U be as above. $\mathsf{Q} + \mathsf{Con}(U_{\mathsf{V}})$ interprets $\mathsf{PCS}[U]$.

4.5. **Truth and Comprehension, via Consistency.** Finally we can combine the claims introduced in the last two sections.

We start with predicative comprehension. By Lemmata 11 and 12, we have

Corollary 10. Let U_{V} be sequential. Then $\mathsf{PCS}[U]$ is mutually interpretable with $\mathsf{Q} + \mathsf{Con}(U_{\mathsf{V}})$.

Corollary 11. If U is finitely axiomatized and sequential, then $\mathsf{PC}[U]$ is mutually interpretable with $\mathsf{Q} + \mathsf{Con}(U)$.

Proof. If U is finitely axiomatized, we don't need to resort to the schematic version of U to prove Lemmata 11 and 12. □

Now we move to Tarskian and positive typed truth. By Proposition 9, Corollary 7, and Corollary 5,

Proposition 10.

(i) Let U be finitely axiomatized, then $\mathsf{T}[U]$ and $\mathsf{Tp}[U]$ are mutually interpretable with $\mathsf{Q} + \mathsf{Con}(U)$;
(ii) Let U be arbitrary, then $\mathsf{T}^+[U]$ and $\mathsf{Tp}^+[U]$ are mutually interpretable with $\mathsf{Q} + \mathsf{Con}(U)$.

Predicative comprehension and Tarskian truth, via consistency, enjoy a mutual reduction for a wide range of choices of the object theory:

Proposition 11.

(i) Let U be finitely axiomatized and sequential. Then $\mathsf{PC}[U]$ is mutually interpretable with $\mathsf{T}[U]$ and $\mathsf{Tp}[U]$;
(ii) Let U_{V} be sequential. Then $\mathsf{PCS}[U]$ is mutually interpretable with $\mathsf{T}^+[U]$ and $\mathsf{Tp}^+[U]$.

5. Conclusion

Proof-theoretic reductions, such as the relative interpretation of PA in ZF, have often been considered as examples of ontological reductions — of natural numbers to sets, in the case at issue. Ideally, we would require a reduction of ACA to CT to replace our commitment to arithmetically definable sets with a semantic commitment to a Tarskian, compositional truth predicate. It might be thought, for instance, that any plausible criterion of ontological reduction of what is implicit in the acceptance of T to what is implicit in the acceptance of W should at least require T to be relatively interpretable in W. In such scenario, the reductions investigated in this work — and summarized in §4.5 — seem to offer the possibility of freely moving from one's ontological commitment to subsets of the domain of the object theory to

ideological commitment to concepts giving more structure to but not enlarging the original domain, such as truth and satisfaction.

These claims may be contrasted by reflecting on the combinatorial nature of a relative interpretation, which appears to be primarily a syntactical reduction: there seems to be no need to resort to a world-language relation to make sense of the interpretation of ACA in CT, for instance. It might simply be taken to be no more than a complex relation between syntactically individuated entities, that is languages and theories. The significance of the results presented in this work, however, is largely independent from one's stance on the relationships between proof-theoretic and ontological reductions. Even from a more neutral stance our study seems to reinforce the belief that set existence principles and principles governing primitive predicates for truth and satisfaction are deeply intertwined and that a general criterion of theory choice — once we accept the base theory U, how much comprehension should we add to it? Which truth axioms for it are preferable? — should consider them as interdependent. The analysis of the operations on theories $\mathsf{T}^+[\cdot]$, $\mathsf{Tp}^+[\cdot]$ and $\mathsf{PCS}[\cdot]$ in §4 generalizes in fact the standard approach surveyed in §3 in several respects and strengthens our conviction that predicative comprehension and suitable truth axioms are indeed parallel logico-mathematical devices.

One might in fact suspect that the reductions between ACA and CT rely on principles that have not much to do with truth and predicative comprehension: by considering for instance ACA_0 and $\mathsf{CT}\restriction$, resulting from ACA and CT by restricting induction, the symmetry between truth and predicative comprehension suggested by the case with full induction is lost. ACA_0 is not interpretable in $\mathsf{CT}\restriction$. Again part of the reason for this lies with the reflexivity of PA, which is a peculiarity of the base theory and has nothing to do with truth or with predicative comprehension. $\mathsf{T}^+[\cdot]$ and $\mathsf{PCS}[\cdot]$, in this respect, fare reasonably better. Their construction is rooted in the idea that $\mathsf{T}^+[U]$ genuinely applies a syntactico/truth-theoretic package to the base theory, and that $\mathsf{PCS}[U]$ generalizes to arbitrary theories the method for obtaining ACA_0 from PA: this approach tries to minimize and control peculiar features of the object theory, such as the reflexivity of PA, for instance. Furthermore, $\mathsf{T}^+[\mathsf{PA}]$ and $\mathsf{PCS}[\mathsf{PA}]$ are comparable, as Proposition 11 shows, whereas a similar characterization for $\mathsf{CT}\restriction$+'all axioms of PA are true' still seems to be missing.[15] This suggests that the operations on the theories considered here offer a more uniform, smoother comparison of semantic and set existence assumptions.[16]

Another attractive feature of the proposed results is that they represent different but equivalent ways to uncover the 'lower bound' of our implicit commitment to the acceptance of a base theory U. Corollaries 10-11 and Propositions 10-11 tell us that the application of $\mathsf{T}^+[\cdot]$, $\mathsf{Tp}^+[\cdot]$ and $\mathsf{PCS}[\cdot]$ correspond to an intensional consistency statement for U, which is not reducible — neither provable nor interpretable — to U by the incompleteness phenomena. Considered in this instrumental fashion, positive, typed truth and Tarskian truth on one side and and predicative comprehension on the other represent equivalent routes for making explicit our commitment to U via the assertion of its consistency.

[15] In particular, it is known that $\mathsf{CT}\restriction$+'all axioms of PA are true' interprets ACA_0, but the converse claim seems to be an open problem.

[16] As we have seen, more generality requires different means of reduction: conservativeness for suitable class of formulas (e.g. Π^0_2 or formulas of the base language), for instance, becomes a trivial requirement as all the theories resulting from the application of our functors will be conservative over the base theory. Relative interpretability becomes the preferred means of reduction.

Further technical and philosophical developments are of course possible. On the philosophical side, a unified account of theory choice for systems of truth is much needed, given the explosion the research on axiomatic truth has had in recent years. From the technical standpoint, an obvious extension of the work would have to take into account hierarchies of applications of the operators: the difficulty of the hierarchical approach, in this context, seems to be represented by the fact that methods for a hierarchical analysis of $\mathsf{T}[\cdot]$, for instance, are foreseeable only when a single object theory is fixed. This would lead to a weakening of the approach that draws much of its strength from its general applicability. At any rate this is only tentative and we defer a treatment of these extensions to forthcoming works.

References

Buss, S. (1986), *Bounded Arithmetic*, Bibliopolis, Naples.

Buss, S. (ed.) (1998). *Handbook of Proof Theory*. Elsevier.

Cantini, A. (1989). Notes on Formal Theories of Truth. *Archives for Mathematical Logic* 35: 97–130.

Enayat, A. and A. Visser. (2014). New Constructions of Satisfaction Classes. Logic Preprint Groups Series.

Enderton, H. (2001). *A Mathematical Introduction to Logic*. Harcourt Academic Press.

Feferman, S. (1960), Arithmetization of metamathematics in a general setting, *Fundamenta Mathematicae*, 49, 35-91.

Feferman, S. (1964). Systems of Predicative Analysis. *Journal of Symbolic Logic* 27: 1–30.

Feferman, S. (1991). Reflecting on Incompleteness. *The Journal of Symbolic Logic* 56: 1–49.

Fischer, M. (2009). Minimal Truth and Interpretability. *The Review of Symbolic Logic* 4(2), 2009: 799-815.

Fujimoto, K. (2010). Relative Truth Definability of Axiomatic Truth Theories. *The Bulletin of Symbolic Logic* 16(3): 305–344.

Hájek, P. and P. Pudlák (1993). *Metamathematics of First-Order Arithmetic*. Springer.

Halbach, V. (1999). Conservative Theories of Classical Truth. *Studia Logica* 62: 353–370.

Halbach, V. (2009). Reducing Compositional to Disquotational Truth. *Review of Symbolic Logic* 2: 786-798.

Halbach, V. (2014). *Axiomatic Theories of Truth (Revised Edition)*. Cambridge University Press 2014.

Heck, R. (2015). Consistency and the Theory of Truth. to appear in *The Review of Symbolic Logic*.

Lachlan, A. (1981). 'Full Satisfaction Classes and Recursive Saturation'. *Canadian Mathemtical Bulletin* 24: 295-297.

Leigh, G. (2014). Conservativity for Theories of Compositional Truth via Cut-Elimination. To appear in *The Journal of Symbolic Logic*.

Mostowski, A. (1950), Some impredicative definitions in the axiomatic set-theory. *Fundamentae Mathematicae* 37: 111–124.

Moschovakis, Y. (1974). *Elementary Induction in Abstract Structures*. American Elsevier.

Nicolai, C. (2015). A Note on Typed Truth and Consistency Assertions. *The Journal of Philosophical Logic* (online first version). DOI: 10.1007/s10992-015-9366-6.

Nicolai, C. (2015a). Deflationary Truth and the Ontology of Expressions. *Synthese* (online first version), DOI: 10.1007/s11229-015-0729-x.

Pohlers, W. (2009). *Proof Theory. The First Step into Impredicativity*. Springer.

Takeuti, G. (1987). *Proof Theory*. North-Holland, Amsterdam.

Vaught, R. A. (1967). Axiomatizability by a Schema. *The Journal of Symbolic Logic* 32: 473–479.

Visser, A. (1991), 'The Formalization of Interpretability', *Studia Logica* 50(1): 81-106, 1991.

Visser, A. (2009), 'Can we make the second Incompleteness Theorem Coordinate Free?', *Journal of Logic and Computation* 21(4), pp.: 543-560.

Visser, A. (2009). The predicative Frege hierarchy. *Annals of Pure and Applied Logic* 160: 129-153.

Visser, A. (2012). Vaught's Theorem on Axiomatizability by a Scheme. *Bulletin of Symbolic Logic* 18: 382–402.