Notes on bounded induction for the compositional truth predicate

Bartosz Wcisło, Mateusz Łełyk

September 29, 2016

Abstract

We prove that the theory of the extensional compositional truth predicate for the language of arithmetic with Δ_0-induction scheme for the truth predicate and the full arithmetical induction scheme is not conservative over Peano Arithmetic. In addition, we show that a slightly modified theory of truth actually proves the global reflection principle over the base theory.

1 Introduction

This paper concerns conservativeness of the compositional truth predicate with bounded induction over Peano Arithmetic. We say that a theory Th_1 is **conservative** over a theory Th_2 with respect to a class of formulae $\Gamma \subseteq \mathrm{Sent}_{\mathcal{L}_{Th_2}}$ (or simply Γ-conservative) iff for every sentence $\phi \in \Gamma$ if $Th_1 \vdash \phi$, then $Th_2 \vdash \phi$. If Γ is the whole class of first-order formulae over the language of the theory Th_2, then we simply say that Th_1 is **conservative** over Th_2. If Γ happens to be the class of all formulae in the language of Peano Arithmetic, then we say that Th_1 is **arithmetically conservative** over Th_2. Verifying various conservativeness results for theories of interest forms an established line of research. It is important from both philosophical and purely logical point of view. From a philosophical point of view, it might be argued that conservativeness of Th_1 over Th_2 assures that accepting the axioms of the former theory does not force us to make any new commitments as to what is actually the case than accepting the latter. This motivation is particularly important in case of the truth predicate, whose triviality is extensively discussed in contemporary philosophy. Namely, the adherents of the **deflationary theory of truth** claim that the truth predicate does not have any actual content but is rather a purely logical device. This vague and imprecise claim has been explicated by some critics (most notably Horsten in [7], Shapiro in [14] and Ketland in [9]) in terms of conservativeness. Were the truth predicate void of substantial content, whatever that would exactly mean, the axioms governing it should not allow us to derive more arithmetical theorems than our theory of arithmetic alone. Thus conservativeness results help to clarify the picture and to distinguish the principles of truth that account for it having a substantial content. From the point of view of pure logic, conservativeness is one of the very basic relations between the theories of interest. It shows which principles may be added to a given theory without the risk of inconsistency or inadequacy. In this context it is important that in practice most conservativeness results are possible to obtain in weak theories like PRA or $I\Sigma_1$.[1] Moreover verifying that one theory is conservative over another might deliver new information about the strength of the latter, especially when the first one allows us to employ notions and principles which are *prima facie* not available in the second one.

The situation was particularly interesting in the case of the compositional truth predicate over PA. Let us first define what we precisely mean by this notion. Before that, let us introduce a few notational conventions. We assume that the reader is familiar with the arithmetization of syntax as explained, e.g. in [8].

Convention 1.

- We assume that we have some fixed Gödel coding. By $\ulcorner x \urcorner$ we denote the Gödel code of x. We will also use $\ulcorner x \urcorner$ to represent the numeral for the Gödel code of x.
- "$y = \underline{x}$" is an arithmetical formula representing a natural primitive recursive function assigning to an element x the code of its numeral, i.e. $\ulcorner SS \ldots S(0) \urcorner$, where the successor function symbol S occurs x times;
- Term(x) is some fixed arithmetical formula representing the set of (Gödel codes of) arithmetical terms;
- the arithmetical formula Form(x) represents the set of (Gödel codes of) arithmetical formulae, the formula Sent(x) represents the set of

[1] Theorem 5.2 in [2] states that the conservativity proof for CT^- may be carried out in PRA. The same argument is valid in the case of CT^- with the internal induction for the arithmetical formulae. In [13] (Theorem 2) it is shown that CT^- and CT^- with the internal induction for the arithmetical formulae are conservative over PA provably in $I\Delta_0 + \exp_1$, where $\exp_1$ denotes the hyper-exponential function.

(Gödel codes of) arithmetical sentences, i.e. formulae with no free variables;

- the arithmetical formula $\mathrm{Ax}_{\mathrm{PA}}(x)$ represents the set of axioms of Peano Arithmetic;

- by a proof in the sequent calculus from some set of axioms Γ we mean a proof in the sequent calculus, where the additional initial sequents "$\longrightarrow \phi$" are allowed, where $\phi \in \Gamma$. In particular, by a proof of ϕ from PA in the sequent calculus we mean an ordinary proof of ϕ from axioms of PA in sequent calculus for first-order logic;[2]

- by $\mathrm{Pr}_{\mathrm{Th}}(x)$ we mean an arithmetical formula formalising the unary relation: "there exists (a code of) a proof d of the sentence x in the sequent calculus from the axioms of the theory Th";

- if $\tau(y)$ is some formula, then by $\mathrm{Pr}_{\tau}(x)$ we mean an arithmetical formula formalising the relation: "there exists (a code of) a proof d of the sentence x in the sequent calculus, where the initial sequents of the form '$\longrightarrow \phi$' are allowed for ϕ such that $\tau(\phi)$". Since in our applications $\tau(y)$ will be thought of as some form of a truth predicate, $\mathrm{Pr}_{\tau}(x)$ reads as: "there exists a proof of x from true premises";

- Subst$(\phi(v), t)$ is an arithmetical formula representing the primitive recursive substitution function, which assigns to a (code of a) formula $\phi(v)$ with at most one free variable and (a code of a) term t the unique (code of the) sentence resulting from substituting t for every free occurrence of v in ϕ. Additionally we assume that Subst is the identity function whenever applied to sentences (i.e., whenever there is no free variable in ϕ);

- the arithmetical formula "$y = t^{\circ}$" represents a natural primitive recursive function assigning to each (code of a) term t its value (and undefined if t is not a code of a term);

- "$x \in y$" is an arithmetical formula expressing that x-th bit of the binary expansion of y is 1 and we write that $x \notin y$ iff either it is 0 or $y < 2^{x}$;

[2] Rather than a proof of ϕ from the empty set of premises in sequent calculus with the additional induction rule or with the ω-rule.

- let M be an arbitrary model of a signature expanding the language of PA. Let $I \subset M$ be an arbitrary set. We say that I is **coded** in M iff there exists an element c such that $x \in I$ iff $M \models x \in c$. We call that c the **code of the set** I. We will sometimes identify the subset I with its code c.[3]

We will usually suppress the distinction between syntactical objects and their codes, e.g. we will sometimes write, e.g. $\mathrm{Pr}_{\mathrm{Th}}(0 = 0)$ instead of $\mathrm{Pr}_{\mathrm{Th}}(\ulcorner 0 = 0 \urcorner)$. In order to avoid any confusion let us introduce a few more notational conventions.

Convention 2.

- By using variables ϕ, ψ we implicitly restrict quantification to (Gödel codes of) arithmetical sentences. I.e. by $\forall \phi \ \Psi(\phi)$ we mean $\forall x \ \mathrm{Sent}(x) \rightarrow \Psi(x)$ and by $\exists \phi \ \ \Psi(\phi)$ we mean $\exists x \ \ \mathrm{Sent}(x) \wedge \Psi(x)$. For brevity we will sometimes also use variables ϕ, ψ to run over arithmetical formulae, whenever it is clear from the context which one we mean;

- similarly, $\phi(v), \psi(v)$ run over arithmetical formulae with at most one indicated free variable (i.e. $\phi(v)$ is either a formula with exactly one free variable or a sentence);

- s, t run over codes of closed arithmetical terms;

- $v, v_1, v_2, \ldots, w, w_1, w_2, \ldots$ run over codes of variables;

- to enhance readability we suppress the formulae representing the syntactic operations. We will usually be writing the results of these operations instead, e.g. $\Phi(\psi \wedge \eta)$ instead of $\Phi(x) \wedge$ "x is the conjunction of ψ and η" similarly, we write $\Phi(\psi(t))$ instead of $\Phi(x) \wedge \mathrm{Subst}(\psi, t)$;

- Let M be an arbitrary model of PA. If $x \in M$ is greater than k for any number $k \in \omega$, then we call it **nonstandard**. We call it **standard** otherwise. Since we often ignore the difference between syntactical objects and their Gödel codes, we will often say that a sentence, formula or term is nonstandard meaning that its code x satisfies in M all the sentences $x > k$ for natural numbers k.

[3] There is a difference between the above definition of a coded set, and the notion of a *coded subset of* ω. Usually we say that a subset $A \subseteq \omega$ is coded in a model $M \supset \omega$ iff there exists c such that $\{x \in M \mid M \models x \in c\} \cap \omega = A$. The set A will not, in general, be coded in any model of PA in our sense of coding.

Definition 3. By CT^- we mean the theory obtained by extending the arithmetical signature with an additional predicate $T(x)$ (with the intended reading "x is a Gödel code of a true sentence") and extending axioms of PA with the following ones:

1. $\forall s, t \ \Big(T(s = t) \equiv s^\circ = t^\circ\Big)$.
2. $\forall \phi, \psi \ \Big(T(\phi \otimes \psi) \equiv T\phi \otimes T\psi\Big)$.
3. $\forall \phi \ \Big(T(\neg\phi) \equiv \neg T\phi\Big)$.
4. $\forall v, \phi(v) \ \Big(T(Qv \ \phi(v)) \equiv Qx \ T(\phi(\underline{x}))\Big)$.

Here $\otimes \in \{\wedge, \vee\}$ and $Q \in \{\forall, \exists\}$.

Note that we in CT^- *do not assume* that any formulae with the truth predicate satisfy the induction scheme.

Remark 4. Note that the compositionality axioms for quantifiers which we have listed above are *not* exactly the ones typically assumed in the definition of compositional theories of truth with no induction. For simplicity, let us discuss this difference in a specific case of the universal quantifier, although analogous remarks apply to the existential quantifier as well. A standard axiom as given e.g. in [5] has the following form:

$$\forall v, \phi(v) \ \ T(\forall v \ \phi(v)) \equiv \forall t \ T(\phi(t)). \tag{$*$}$$

According to the above axiom a universal formula $\forall v \ \phi(v)$ is true only if for arbitrary term t its substitutional instance $\phi(t)$ is true, which is not the same as to say that for arbitrary *numeral* $\underline{x}$ the formula $\phi(\underline{x})$ is true. In the presence of Σ_1-induction for the formulae containing truth predicate both versions of the quantifier axioms are equivalent, since in such a case we may prove the principle of extensionality (or, as we will also sometimes call it, **regularity**), i.e. the following sentence:

$$\forall\phi\forall t, s \ \ \big(s^\circ = t^\circ \to T\phi(t) \equiv T\phi(s)\big). \tag{REG}$$

The principle states that the truth-value of a formula does not depend on specific terms which occur in the formula, but rather on their values (and PA proves that for every term there exists a numeral with the same value). It is arguable however, whether the axiom $(*)$ as stated above is really intuitive under the assumption that we lack any induction for the extended language whatsoever. Presumably the initial intuition is that $\forall x\ \phi(x)$ is true iff the formula $\phi(x)$ is satisfied by all elements. If we want to work with the truth predicate rather than the satisfaction relation, we have to rethink what it means to be satisfied by all elements. One way to say it is that for *an arbitrary* element x the sentence $\phi(t)$ is true, where t is *some* term denoting x. And since in systems whose intended domain are natural numbers every element x is denoted by a numeral $SS\ldots S0$ (with S repeated x times), this intuition seems to be perfectly addressed by the axiom we opt for, i.e.:

$$\forall v, \phi(v)\ \ T(\forall v\ \phi(v)) \equiv \forall x\, T(\phi(\underline{x})).$$

On the other hand, there is the intuition that universal sentences should satisfy *dictum de omni* principle, i.e. whenever $\forall v\ \phi(v)$ is true, we expect $\phi(t)$ to be true for arbitrary terms. These two intuitions for the truth of universal sentences may diverge when we lack induction for the formulae containing the truth predicate, but we see no obvious reason why substitution principle should be regarded more essential than the intuition that all natural numbers can be named by numerals. Therefore, we do not think that the compositional axioms for quantifiers which we assume are less natural than the ones that are typically formulated. Namely, we assume that to infer that a universal sentence $\forall v\ \phi(v)$ is true, we only need to assume that $\phi(\underline{x})$ holds for all numerals $\underline{x}$ rather than require the stronger hypothesis to hold, namely that $\phi(t)$ holds for all *terms*. Probably the most satisfactory solution, when working with such weak theories as CT^-, would be to embrace both intuitions and accept asymmetric compositional axioms for the universal quantifier, i.e.:

1. $\forall v, \phi(v)\ \ T\big(\forall v\ \phi(v)\big) \longrightarrow \forall t\, T\big(\phi(t)\big).$
2. $\forall v, \phi(v)\ \ \forall x\, T\big(\phi(\underline{x})\big) \longrightarrow T\big(\forall v\ \phi(v)\big).$

Of course, in this case, we should also introduce an analogous pair of axioms for the existential quantifier. We remark that, since a variant of CT^- using the above asymmetric axioms for quantifiers is stronger than the one chosen by us, our non-conservativity results apply also to the extensions of CT^- with the compositional axioms for quantifiers defined in such way.[4]

Let us make one more comment at the end of this section. Taking into account that the choice of the precise formulation for the quantifier axioms seems problematic, one may wonder, whether it would not be easier to

[4] We thank the anonymous referee for the remark that led to this discussion.

prove our results for the compositional *satisfaction* predicate rather than the truth predicate, since then there seems to be a canonical choice of quantifier axioms.

The basic reason why we have decided to work with the truth predicate is rather trivial. We wanted to conform to the standard conventions in the field of the axiomatic truth theories, especially that in the context of weak theories results obtained for the satisfaction predicate do not in general carry over to the truth-theoretic framework quite automatically precisely because of the extensionality issues.

2 Known Results on Conservativity of Extensions of CT^-

Let us list a few important theories obtained via augmenting CT^- with some induction.

Definition 5. By CT we mean the theory obtained by adding to CT^- all the instances of the induction scheme (i.e. including the ones for formulae containing the truth predicate). By CT_1 we mean CT^- with induction for Π_1-formulae containing the truth predicate. By CT_0 we mean CT^- with induction for Δ_0 formulae containing the truth predicate.

One of the most important questions of theory of truth may be now rephrased: which reasonable extensions of CT^- are conservative over PA? This question is indeed nontrivial thanks to the following theorem:[5]

Theorem 6 (Krajewski–Kotlarski–Lachlan, Enayat–Visser, Leigh)**.** CT^- *is conservative over* PA.

This result may be still improved in a substantial way.

Definition 7. By the principle of **internal induction** we mean the following axiom:

$$\forall\phi(v)\ \Big(\forall x\ \Big(T(\phi(\underline{x}))\to T(\phi(\underline{Sx}))\Big)\longrightarrow\Big(T(\phi(\underline{0}))\to\forall x\, T(\phi(\underline{x}))\Big)\Big).\quad\text{(INT)}$$

[5] Let us explain ourselves for this complicated attribution. A related result has been obtained by Kotlarski, Krajewski and Lachlan in [12], who essentially proved Theorem 6 for a variant of CT^- with *satisfaction* relation in place of the truth predicate. However, it is by no means obvious to us how to modify the proof of Kotlarski, Krajewski and Lachlan so that it works for the *truth* predicate. In order to do this we would apparently have to prove that the extensionality principle for the satisfaction relation is conservative over PA, which does not seem trivial. Theorem 6 for the version of CT^- axiomatised in *purely relational language* was proved by Enayat and Visser in [2]. The full-blown result was shown by proof-theoretic methods by Leigh in [13].

The name of this axiom is introduced in analogy to the distinction between internal and external induction axioms in subsystems of second-order arithmetic.

The proof from [2] (as well as the one from [13]) of conservativeness of CT^- over PA , gives as a corollary the following theorem.[6]

Theorem 8 (Enayat–Visser, Leigh)**.** $\mathrm{CT}^- +$ (INT) *is conservative over* PA.

Let us recall the regularity principle (REG). Instead of considering an unusual variant of CT^- we could have added the above principle to the standard list of axioms, because it yields both forms of the compositional axiom for the quantifiers equivalent. Fortunately, this would not trivialise our work, thanks to the following theorem, which may be read directly off the Enayat–Visser construction.

Theorem 9 (Enayat–Visser)**.** $\mathrm{CT}^- +$ (INT) + (REG) *is conservative over* PA.

It is easily observed that the internal induction may be proved in a system CT_0 that is in CT^- with Δ_0-induction for the truth predicate. Then, as we shall briefly indicate, it follows that CT^- with Π_1 induction for the truth predicate is enough to prove the following principle:

Definition 10. By the **Global Reflection Principle** we mean the following axiom:

$$\forall \phi \left(\mathrm{Pr}_T(\phi) \longrightarrow T(\phi)\right), \tag{GRP}$$

where $\mathrm{Pr}_T(x)$ is a special case of the predicate $\mathrm{Pr}_\tau(x)$ defined in Convention 1 with $\tau(y) = T(y)$. Hence the intuitive reading of $\mathrm{Pr}_T(x)$ is: "there exists a proof d of the sentence d in sequent calculus from the initial sequents '$\longrightarrow \phi$', where we have $T(\phi)$, i.e. a proof in sequent calculus from true premises".

Note that, speaking informally, (GRP) says that the set of true sentences is closed under reasonings in First-Order Logic.

Definition 11. By the **Axiom Soundness Property** for Th we mean the following principle:

$$\forall \phi \left(\mathrm{Ax}_{\mathrm{Th}}(\phi) \longrightarrow T(\phi)\right). \tag{ASP}$$

[6] Actually in the both cited papers a much more general theorem has been presented from which Theorem 8 follows as a direct corollary, see [2], remarks in Section 6 and [13], Theorem 3.

Note that if a theory of truth proves both the axiom soundness of Th and satisfies the global reflection principle, then it proves the following statement:

$$\forall \phi \left(\mathrm{Pr}_{\mathrm{Th}}(\phi) \longrightarrow T(\phi) \right).$$

Then a crude lower-bound for the strength of non-conservative truth-theoretic principles is given by the following theorem:

Theorem 12. CT_1 *proves the global reflection principle and the axiom soundness property for* PA *and thus the consistency of* PA. *In effect, this theory is not conservative over* PA.

The above result is obtained as follows: working in CT_1, by a straightforward induction on length of proofs in the sequent calculus we show that for every substitution of closed terms for free variables in the sequent, if every formula in the antecedent is true, then some formula in the succedent of the sequent arrow is true. This is obviously a Π_1 statement, since the quantifiers "every formula in the antecedent", "every formula in the succedent" may be assumed to be bounded by the size of the proof in question (under a reasonable coding of syntax and of sets). The consistency of PA follows, since we may prove in CT_0 that the parameter-free variant of induction holds and then, by Π_1 induction on the length of the block of universal quantifiers, that any universal closure of an instance of the induction scheme is true as well.

Then a natural question arises of how to improve bounds on minimal truth-theoretic principles which, combined with CT^-, are non-conservative over PA. The first natural candidate to consider is the theory CT_0. Indeed, Kotlarski in his paper [11] has presented an alleged proof that the theory CT_0 proves the global reflection principle.[7] Unfortunately, as observed independently by Albert Visser and Richard Heck, the proof contained a gap which seemed to require an essentially new approach to surmount. Actually, after the gap has been revealed, it has been completely unclear whether or not the theorem is true at all. We shall discuss the erroneous proof in the appendix. In the present paper we provide a non-conservativeness proof for our variant of CT_0. Moreover, we prove that the global reflection principle is arithmetically conservative over CT_0. In addition we show that a very natural modification of CT_0 actually proves the principle.

[7] More precisely, he considered a theory of satisfaction rather than truth, and only aimed to show that every formula derivable in PA is satisfied under all valuations.

3 The main result

In the paper [3] Fujimoto has argued that the right way to compare the conceptual strength of theories of truth Th_1, Th_2 over the same base theory B (in our case $B =$ PA) is to check whether Th_1 defines a truth predicate satisfying axioms of Th_2. In such case we say that Th_2 is **B-relatively definable**[8] in Th_1. The relative definability is a very natural and strong relation between two theories. In particular, if Th_2 is relatively definable in Th_1, it implies the following other relations[9] (on the other hand none of the these relations implies that Th_2 is relatively interpretable in Th_1):

1. Th_2 is arithmetically conservative over Th_1.
2. Th_2 is interpretable is Th_1.
3. Every model of Th_1 expands to a model of Th_2.

Let us briefly comment upon the first of the three points. Suppose that Th_1 relatively interprets Th_2, i.e. there is a formula $\tau(x)$ which provably satisfies the truth axioms of Th_2. Consider any proof in Th_2 with an arithmetical consequence ϕ. Then replace all occurrences of the truth predicate in that proof with the formula $\tau(x)$ and precede all the uses of Th_2's truth-theoretic axioms with a proof in Th_1 that $\tau(x)$ indeed satisfies the axioms that we used. In such a way we can obtain a proof of ϕ in Th_1.

We show that in this sense CT_0 with the global reflection principle and the axiom soundness property is as strong as CT_0 alone.

Theorem 13. CT_0 *with the global reflection axiom and the axiom soundness property for* PA *is* PA*-relatively definable in* CT_0. *In particular* CT_0 *is not conservative over* PA.

In other words, there is a formula $T'(x)$ such that in CT_0 one can prove that $T'(x)$ satisfies compositional conditions for arithmetical sentences, global reflection principle and axiom soundness property for PA. Moreover, we can prove in CT_0 every instance of Δ_0-induction scheme for the predicate $T'(x)$.

Before we present the proof of the above theorem in full detail, let us give a sketch of our argument. This outline will be imprecise and not fully

[8] Fujimoto calls this relation relative *truth* interpretability (and keeps the parameter B implicit).

[9] All the three points have been observed in the original paper [3], see p. 324 for the first two of them and Proposition 28 (1) for the third one.

correct, but the rest of the section will be generally devoted to spelling out all the details in a proper manner.

One of the main tools in metamathematics are various forms of partial truth predicates. Probably the best known of them are arithmetical truth predicates for the classes Σ_n.[10] Consider a partial arithmetical truth predicate $\tau(x)$. Its main feature is that for some formulae we have:

$$\tau(\ulcorner \phi \urcorner) \equiv \phi.$$

If $\tau(x)$ is an arithmetical formula and the truth predicate $T(x)$ satisfies CT^-, then this entails that for standard sentences ϕ we have:

$$T\tau(\underline{\phi}) \equiv T\phi.$$

Now, it may be hoped that we can define some families of formulae formulae $T_c(v)$ (or, more precisely, their codes, so that the parameter c may be nonstandard), which behave like truth predicates in the sense that CT_0 may prove:

$$TT_c(\underline{\phi}) \equiv T\phi$$

for all arithmetical formulae ϕ with codes smaller than c. We could wish to do still better and to require that the newly defined truth predicates are compositional for small enough formulae in the sense that we have e.g.

$$TT_c(\underline{\phi \wedge \psi}) \equiv TT_c(\underline{\phi}) \wedge TT_c(\underline{\psi}).$$

Now, the key idea is that in the presence of Δ_0 induction for our original truth predicate $T(x)$ we may hope for a full induction for the "truth predicates" $TT_c(\underline{x})$. Before we explain why this should be case, let us introduce some notation, which will also be used in the proof proper.

First of all, we define a formula $T'_c(x)$ as $TT_c(\underline{x})$. The next notational convention deserves a separate definition.

Definition 14. Let $P(u)$ be a fresh[11] unary predicate with the only variable u, $\delta(x)$ an arbitrary formula and ψ another formula with exactly one free variable, possibly with parameters. Then by

$$\delta[\psi/P](x)$$

[10] For a definition see e.g. [4] Definitions 1.71 and 1.74. See [4] pp. 50–61 for a general discussion.

[11] That is, neither from the arithmetical signature, nor $T(u)$.

we mean the effect of formally substituting the formula $\psi(t)$ for all occurences of $P(t)$, where t is an arbitrary term, possibly changing the names of the bounded variables so as to avoid clashes.

Probably, the above definition is best understood via an example.

Example 15. Let $\delta(x) = \exists z, w \ \ (P(z+w) \wedge \forall z < w \neg P(z)) \wedge (x = x) \wedge P(x)$. Then

$$\delta[(u > u)/P](x) = \exists z, w \ (z+w > z+w \wedge \forall z < w \, (\neg z > z)) \wedge (x = x) \wedge x > x.$$

It is clear that the formula substitution operation may be formalized in PA.

Let us sketch, why the predicates $T'_c(x) := TT_c(\underline{x})$ satisfy full induction scheme. First, one can check that the principle of internal induction (INT) may be proved by a straightforward application of Δ_0-induction. Then we would like to show that for arbitrary arithmetical ϕ we have:

$$\Big(\forall x \ \big(\phi[T'_c/P](x) \rightarrow \phi[T'_c/P](Sx)\big)\Big) \longrightarrow \Big(\phi[T'_c/P](0) \rightarrow \forall x \ \phi[T'_c/P](x)\Big).$$

So let us fix any arithmetical formula ϕ and consider the formula $T(\phi[T_c/P])(x)$. Now, since T is compositional we can push it down the syntactic tree of the (nonstandard, but arithmetical) formula $\phi[T_c/P]$ finitely many levels, until it meets a partial truth predicate T_c, but not any further. Since we assumed that ϕ is of standard syntactic shape we do not need any induction to do that. We obtain the following equivalence, which we call **the generalised commutativity** principle (strictly speaking, we only get something resembling it — see the further comments):

$$T(\phi[T_c/P](x)) \equiv \phi[TT_c/P](x). \tag{GC}$$

Now, by the internal induction principle we have:

$$\forall x \ \Bigg(T\Big(\phi[T_c/P](\underline{x})\Big) \rightarrow T\Big(\phi[T_c/P](\underline{Sx})\Big)\Bigg) \longrightarrow$$
$$\Bigg(T\Big(\phi[T_c/P](\underline{0})\Big) \rightarrow \forall x \, T\Big(\phi[T_c/P](\underline{x})\Big)\Bigg).$$

If the (GC) principle were really the case, we could conclude that:

$$\Big(\forall x\ \big(\phi[T'_c/P](x) \rightarrow \phi[T'_c/P](Sx)\big)\Big) \longrightarrow \Big(\phi[T'_c/P](0) \rightarrow \forall x\ \phi[T'_c/P](x)\Big).$$

Unfortunately, (GC) is strictly speaking not true, since we *do not have* for example:

$$T(\exists x\ T_c(x)) \equiv \exists x\ T(T_c(x)),$$

but rather

$$T(\exists x\ T_c(x)) \equiv \exists x\ T(T_c(\underline{x})).$$

Most probably, reformulating (GC) in a proper way would be quite messy. Instead of it, we use a different statement suggested to us by Cezary Cieśliński as a way to bypass the mentioned difficulty, which will appear as Lemma 21.

So we may hope to construct a family of predicates $T'_c(\underline{\phi})$ which are compositional for formulae ϕ with the complexity smaller than c (under an appropriate choice of the complexity measure) and satisfy the full induction scheme. Now we are very close to construct a compositional truth predicate satisfying the global reflection scheme. Namely, pick any model (M, T) of CT_0 and choose some formula ψ and a set of sentences Γ both in the domain of M such that $M \models \mathrm{Pr}_{\Gamma}(\psi)$ with a proof d in the domain of M. Now, one can check that the standard argument using induction on the structure of the proof allows us to show that if all the sentences γ in Γ satisfy $T'_d(\gamma)$, then also $T'_d(\psi)$ holds.

The last step of our construction is simply to take the sum over $c \in M$ of the predicates $T'_c(\underline{x})$. If we can show that the predicates $T'_c(\underline{x})$ and $T'_d(\underline{x})$ agree on the formulae of complexity smaller than both c and d, then the predicate $T'(x) = \bigcup_{c\in M} T'_c(x)$ should be fully compositional and satisfy the global reflection principle. Moreover, $T'(x)$ should satisfy Δ_0-induction, since for any a its restriction to the interval $[0, a]$ is equal to $T'_a \cap [0, a]$ and therefore fully inductive.

Now, our task is essentially twofold: first, we have to define the family of arithmetical predicates $(T_c(x))$ such that provably in CT_0 the predicate $T_c(x)$ is compositional for formulae of complexity smaller than c. Second, we have to spell out all the details of the above argument. The rest of this section will be devoted to these issues. Let us start with some definitions, which will play a crucial role in constructing our family of arithmetical truth predicates with nice properties.

Definition 16. Let $M \models \mathrm{PA}$, $c \in M$ and $(\phi_i)_{i \leq c} \in M$ be a coded sequence of sentences. Then $\bigvee_{i \leq c} \phi_i$ denotes the code of the disjunction of all ϕ_i-s with parentheses grouped to the left (for the sake of determinateness only — provably in CT_0 the truth of a sentence does not depend on how we parenthesize blocks of disjunctions and conjunctions).

Lemma 17 (Disjunctive correctness). CT_0 *proves that for all x and all indexed families of sentences*[12] $(\phi_i)_{i \leq x}$

$$T\left(\bigvee_{i \leq x} \phi_i\right) \equiv \exists i \leq x \ \ T(\phi_i). \tag{DC}$$

Lemma 18 (The internal induction). CT_0 *proves the internal induction axiom* (INT).

Proofs of both lemmata are carried out by a completely straightforward application of Δ_0-induction. Let us show as an example the proof of disjunctive correctness.

Proof of Lemma 17. We show by induction on x that the disjunctive correctness holds for all indexed families of sentences $a < x$. In particular we will assume that under our coding, every subfamily of a is also smaller than x. We can assume that by convention a disjunction over empty sequence is some fixed *falsum*, say $0 \neq 0$ and a disjunction $\bigvee_{i \leq 0} \phi_i$ over a sequence of length one is simply ϕ_0.

Assuming the above conventions, the claim is trivial, when a is the empty sequence or when a has exactly one element. Suppose now that $a = \bigvee_{i \leq l} \phi_i$ and that $a < x+1$. If $a < x$, then our claim holds by the induction hypothesis, so we may assume that $a = x$. We will focus on the left-to-right direction in our equivalence. Let

$$a = \bigvee_{i \leq l} \phi_i.$$

Without loss of generality we may assume that $l > 0$. Then by definition we have:

[12] Note that the quantification over indexed families of sentences is expressed here with an arithmetical formula "for all y, if y is a (code of a) sequence of arithmetical sentences of length x". In particular, such a sequence will always have, according to M, the number of its elements equal to some $x \in M$. Therefore, from the point of view of M it will be finite (but not necessarily equal to some $k \in \omega$).

$$T\Big(\bigvee_{i\leq l}\phi_i\Big)\equiv T\Big((\bigvee_{i\leq l-1}\phi_i)\vee\phi_l\Big).$$

By compositionality this implies:

$$T\Big(\bigvee_{i\leq l}\phi_i\Big)\equiv T\Big(\bigvee_{i\leq l-1}\phi_i\Big)\vee T(\phi_l).$$

Which by induction hypothesis is equivalent to:

$$\Big(\exists i\leq(l-1)T(\phi_i)\Big)\vee T(\phi_l).$$

The claim follows. The right-to-left direction is proved in a similar fashion.

□

Before we proceed to the proof of our theorem, let us discuss a few preparatory steps.

Let us introduce our main technical tool — a particular class of partial truth predicates. Let A_n be the set of arithmetical sentences whose logical complexity is exactly n, where by the logical complexity we mean the maximal depth of nesting of logical symbols, i.e. quantifiers and connectives and each symbol is counted separately (e.g. prefixing a formula with a block of five universal quantifiers, raises its complexity by five). The binary relation $x\in A_y$ is clearly primitive recursive and thus represented in PA. Its precise definition is as follows:

$$\begin{aligned}
x\in A_0 &\equiv \exists s,t\ \ x=(s=t)\\
x\in A_{y+1} &\equiv \exists v,\phi(v)\ \ x=(\exists v\ \phi(v))\wedge(\phi(0))\in A_y\\
&\vee\ \ \exists v,\phi(v)\ \ x=(\forall v\ \phi(v))\wedge(\phi(0))\in A_y\\
&\vee\ \ \exists\phi,\psi\ \ x=(\phi\vee\psi)\wedge\bigvee_{v,w:\max(v,w)=y}\Big(\phi\in A_v\wedge\psi\in A_w\Big)\\
&\vee\ \ \exists\phi,\psi\ \ x=(\phi\wedge\psi)\wedge\bigvee_{v,w:\max(v,w)=y}\Big(\phi\in A_v\wedge\psi\in A_w\Big)\\
&\vee\ \ \exists\phi\ \ x=(\neg\phi)\wedge(\phi\in A_y).
\end{aligned}$$

Let us define a family of arithmetical predicates $(\Theta_n)_{n\in\omega}$ in the following way:

$$\begin{array}{rcl}
\Theta_0(x) & = & \exists s,t \;\; x = (s = t) \wedge s^\circ = t^\circ \\
\Theta_{n+1}(x) & = & \exists v, \phi(v) \;\; x = (\exists v \; \phi(v)) \wedge \exists y \; \Theta_n(\phi(\underline{y})) \\
& \vee & \exists v, \phi(v) \;\; x = (\forall v \; \phi(v)) \wedge \forall y \; \Theta_n(\phi(\underline{y})) \\
& \vee & \exists \phi, \psi \;\; x = (\phi \vee \psi) \wedge \bigvee_{k,l \leq n} \Big(\phi \in A_k \wedge \psi \in A_l \wedge (\Theta_k(\phi) \vee \Theta_l(\psi))\Big) \\
& \vee & \exists \phi, \psi \;\; x = (\phi \wedge \psi) \wedge \bigvee_{k,l \leq n} \Big(\phi \in A_k \wedge \psi \in A_l \wedge (\Theta_k(\phi) \wedge \Theta_l(\psi))\Big) \\
& \vee & \exists \phi \;\; x = (\neg \phi) \wedge \neg \Theta_n(\phi).
\end{array}$$

Clearly, the functions $n \mapsto \ulcorner A_n \urcorner$ and $n \mapsto \ulcorner \Theta_n \urcorner$ are primitive recursive. Following our conventions we will write A_x and Θ_x for the arithmetical formulae representing these functions as well as for their values. We will sometimes write $a \in A_c$ meaning that an element a satisfies the formula $A_c(x)$ (where c is a parameter, possibly nonstandard). Note, that $A_c(x)$ may indeed be expressed with an arithmetical formula with a parameter.

Let a simplistic partial arithmetical truth predicate $T_n(x)$ be defined in the following way:

$$T_n(x) = \bigvee_{j \leq n} x \in A_j \wedge \Theta_j(x).$$

As before, we shall write simply $T_x(y)$ to denote the function $x, y \mapsto \ulcorner T_x(y) \urcorner$. Note that the definition of the predicates T_n closely parallels that of the arithmetical satisfaction predicates for Σ_n-classes, only it is much simpler. Namely: we assume that *every single* quantifier or connective increases the complexity of a formula and do not make a distinction between bounded and unbounded quantifiers.

The key fact needed in the proof of our theorem is that if $T(x)$ satisfies the axioms of CT_0, then partial truth predicates defined for a parameter c as

$$T'_c(x) = T(T_c(\underline{x})),$$

or, in more detail, as

$$T'_c(x) = \exists y, z, w \; \Big(y = \underline{x} \wedge z = \ulcorner T_c(v) \urcorner \wedge w = \mathrm{Subst}(z, y) \wedge T(w)\Big)$$

enjoy remarkably good properties: they are compositional for formulae in the respective classes A_c and they are fully inductive.

Lemma 19 (Compositionality of T'_c)**.** CT_0 *proves that for every y the following conditions hold:*

1. $\forall s, t\ \ T'_y(s = t) \equiv s^\circ = t^\circ$.
2. $\forall \phi, \psi\ \ \Big((\phi \otimes \psi) \in A_z \wedge y > z \to \Big(T'_y(\phi \otimes \psi) \equiv T'_y\phi \otimes T'_y\psi\Big)\Big)$.
3. $\forall \phi\ \ \Big((\neg\phi) \in A_z \wedge y > z \to \Big(T'_y(\neg\phi) \equiv \neg T'_y\phi\Big)\Big)$.
4. $\forall v, \phi\ \ \Big((Qv\ \phi) \in A_z \wedge y > z \to \Big(T'_y(Qv\ \phi) \equiv Qx\ T'_y(\phi(\underline{x}))\Big)\Big)$.

Where $\otimes \in \{\wedge, \vee\}$ *and* $Q \in \{\forall, \exists\}$.

Proof. Let us fix arbitrary c. We will prove that the above conditions hold for T'_c and we will focus on the case of existential quantifier. Let $(\exists v\ \ \phi) \in A_j$ for some $j < c$ and suppose that

$$T\Big(T_c(\underline{\exists v\ \ \phi})\Big).$$

Then, by the disjunctive correctness of T, we must have $T(\eta)$ for some of the disjuncts η in $T_c(\underline{\exists v\ \phi})$. Fix any y and note that, since $(\exists v\ \ \phi) \in A_y$ is an arithmetical formula of standard syntactic structure with (possibly non-standard) parameters, we have:

$$T\Big((\exists v\ \phi) \in A_{\underline{y}}\Big) \equiv (y = j).$$

Therefore the only possible candidate for our true disjunct η is:

$$T\Big((\exists v\ \phi) \in A_{\underline{j}} \wedge \Theta_j(\underline{\exists v\ \ \phi})\Big).$$

This implies that in particular we have:

$$T\Big(\Theta_j(\underline{\exists v\ \phi})\Big).$$

Which entails:

$$T\Big(\exists y\ \ \Theta_{j-1}(\underline{\phi(\underline{y})})\Big).$$

This by definition equals to the following statement:

$$T\Big(\exists y, z, \psi(\ \ z = \underline{y} \wedge \psi = \mathrm{Subst}(\underline{\phi(v)}, z) \wedge \Theta_{j-1}(\psi))\Big).$$

Since the compositional truth predicate T satisfies Tarski biconditionals for standard sentences with possibly nonstandard numerals, this implies:

$$\exists y, z, \psi \left((z = \underline{y}) \wedge (\psi = \text{Subst}(\phi(v), z)) \wedge T(\Theta_{j-1}(\underline{\psi}))\right),$$

which may be again abbreviated as:

$$\exists x \ T\Big(\Theta_{j-1}(\underline{\phi(\underline{x})})\Big).$$

This, again by the disjunctive correctness property, entails:

$$\exists x \ T\Big(T_c(\underline{\phi(\underline{x})})\Big).$$

Finally, by definition this is simply:

$$\exists x \ T'_c(\phi(\underline{x}))$$

The other cases are analogous.

□

The next lemma we will use is the key ingredient of our non-conservativeness result. It states that for arbitrary c (possibly nonstandard) the predicates $T'_c(x)$ satisfy the full induction scheme.

Lemma 20. *Let $\phi(v)$ be any formula of the arithmetical language expanded with a fresh unary predicate $P(v)$. Then* CT_0 *proves that for every c*

$$\Big(\forall x \ (\phi[T'_c/P](x) \rightarrow \phi[T'_c/P](Sx))\Big) \longrightarrow \Big(\phi[T'_c/P](0) \rightarrow \forall x \ \phi[T'_c/P](x)\Big).$$

Note that in the above lemma by $\phi[T'_c/P](x)$ we mean an actual formula, rather than its formalised version. Thus, in effect, the above lemma states that the formulae T'_c are really inductive for an arbitrary choice of the parameter c.

As outlined at the beginning of the current section, the above lemma would be very easily proved, if the generalized commutativity principle (GC) were true, which is unfortunately not the case. We shall bypass the difficulty with the following lemma.[13]

[13] We are grateful to Cezary Cieśliński for formulating this fact and suggesting it as a way to prove inductiveness of T'_c in a proper manner.

Lemma 21. *Let $\phi(x_1,\ldots,x_n)$ be an arbitrary formula, in the arithmetical language augmented with a fresh predicate $P(x)$. Then* CT_0 *proves that for every c there exists an arithmetical formula $\psi(x_1,\ldots,x_n)$ such that*

$$\forall x_1,\ldots,x_n \left(\phi[T'_c/P](x_1,\ldots,x_n) \equiv T(\psi(\underline{x_1},\ldots,\underline{x_n}))\right).$$

Proof. We proceed by (meta-)induction on the complexity of a formula ϕ. If ϕ is an arithmetical atomic formula $s(x_1,\ldots,x_n) = t(x_1,\ldots,x_n)$ then:

$$\phi[T'_c/P](x_1,\ldots,x_n) = \phi(x_1,\ldots,x_n) \equiv T(\phi(\underline{x_1},\ldots,\underline{x_n})).$$

If $\phi(x) = P(t(x_1,\ldots,x_n))$, then:

$$\phi[T'_c/P](x_1,\ldots,x_n) = T'_c(t(x_1,\ldots,x_n)).$$

By definition the last formula may be expanded to:

$$\exists x,y,z \ \left(x = t(x_1,\ldots,x_n) \wedge y = \underline{x} \wedge z = \text{Subst}(T_c,y) \wedge T(z)\right).$$

Let

$$\psi(x_1,\ldots,x_n) = \exists x \ \left(x = t(x_1,\ldots,x_n) \wedge T_c(x)\right).$$

Now, by compositional clauses in CT^- we have the following equivalences:

$$\begin{aligned} T\left(\exists x \ \left(x = t(\underline{x_1},\ldots,\underline{x_n}) \wedge T_c(x)\right)\right) &\equiv \exists x\, T\left(\underline{x} = t(\underline{x_1},\ldots,\underline{x_n}) \wedge T_c(\underline{x})\right) \\ &\equiv \exists x \ \left(\underline{x}^\circ = t(\underline{x_1},\ldots,\underline{x_n})^\circ \wedge T(T_c(\underline{x}))\right) \\ &\equiv \exists x \ \left(x = t(x_1,\ldots,x_n) \wedge T(T_c(\underline{x}))\right). \end{aligned}$$

Note, that the term t above is standard and, consequently, it may be written down explicitly in the last step. Now, the last formula in the above equivalences is precisely the abbreviation of the following one:

$$\exists x,y,z \ \left(x = t(x_1,\ldots,x_n) \wedge y = \underline{x} \wedge z = \text{Subst}(T_c,y) \wedge T(z)\right).$$

So the claim follows with ψ as above.

If ϕ is a boolean combination of some formulae ξ, η, then the induction step is straightforward, so suppose that $\phi = \exists y \;\; \eta(x, y)$. Then we have:

$$\phi[T'_c/P](x_1, \ldots, x_n) = \exists\, y(\eta[T'_c/P](x_1, \ldots, x_n, y)).$$

By induction hypothesis, there exists $\psi'(x_1, \ldots, x_n, y)$ such that

$$\forall x_1, \ldots, x_n, y \,\Big(\eta[T'_c/P](x_1, \ldots, x_n, y) \equiv T(\psi'(\underline{x_1}, \ldots, \underline{x_n}, \underline{y}))\Big).$$

Using the compositional clauses for T, we have:

$$\forall x_1, \ldots, x_n \Big(\phi[T'_c/P](x_1, \ldots, x_n) \equiv T(\exists y \psi'(\underline{x_1}, \ldots, \underline{x_n}, y))\Big).$$

So, the claim holds with $\psi = \exists y \psi'$. The universal quantifier case is analogous.

□

Proof of Lemma 20. Fix $\phi(v)$ as in the claim of the lemma. Working in CT_0, fix an arbitrary c. By the previous lemma, there exists an arithmetical formula $\psi(x)$ such that

$$\forall x \,\Big(\phi[T'_c/P](x) \equiv T\psi(\underline{x})\Big).$$

By the internal induction principle we have:

$$\Big(\forall x \,\big(T(\psi(\underline{x})) \rightarrow T(\psi(\underline{Sx}))\big)\Big) \longrightarrow \Big(T(\psi(\underline{0}) \rightarrow \forall x\, T(\psi(\underline{x}))\Big).$$

By lemma 21 this entails:

$$\Big(\forall x \,\big(\phi[T'_c/P](x) \rightarrow \phi[T'_c/P](Sx)\big)\Big) \longrightarrow \Big(\phi[T'_c/P](0) \rightarrow \forall x\, \phi[T'_c/P](x)\Big).$$

Hence Lemma 20 follows. □

Let us stress that we did not put any restrictions on the complexity of ϕ, so that what we obtain for the predicates T'_c is the full induction, rather than Δ_0-induction as one could possibly expect.

Now let us state another important lemma. It states that the predicates of the form T'_c are compatible.

Lemma 22. *Provably in* CT_0 *the predicates* T'_d *have the following properties:*

1. *If* $x \notin A_c$ *for all* $c \leq d$*, then we have* $\neg T'_d(x)$*.*
2. *For arbitrary* $d < e$ *and* $\phi \in A_d$ *we have* $T'_d(\phi) \equiv T'_e(\phi)$*.*

Proof. Both claims are a straightforward application of the disjunctive correctness and the fact that the formula "$x \in A_y$" is standard, so we have:

$$\forall x, y\Big(T(\underline{x} \in A_{\underline{y}}) \equiv (x \in A_y)\Big).$$

□

Now we introduce one more truth predicate, whose properties will almost immediately imply our theorem.

Definition 23. Let the formula $T'(x)$ be defined in the following way:

$$T'(x) = \exists v\ \big((x \in A_v) \wedge TT_v(\underline{x})\big).$$

Intuitively, the extension of the predicate $T'(x)$ is the sum of the extensions of predicates $T'_c(x)$. We should expect that it behaves reasonably, since by Lemma 22 the predicate $T'_c(\phi)$ is an extension $T'_d(\phi)$ to arithmetical sentences in $A_c \setminus A_d$ whenever $c > d$. Provably in CT_0 the newly introduced predicate $T'(x)$ satisfies the following four properties:

1. Δ_0-induction scheme.
2. Compositionality.
3. Regularity.
4. Global reflection principle.
5. Axiom soundness property for PA.

Now we will spell out the listed properties in the series of lemmata.

Lemma 24 (Bounded induction for T')**.** *For any* Δ_0 *formula* ϕ *in the arithmetical language enriched with the fresh unary predicate* $P(x)$*,* CT_0 *proves that:*

$$\forall x\ \Big(\phi[T'/P](x) \rightarrow \phi[T'/P](Sx)\Big) \longrightarrow \Big(\phi[T'/P](0) \rightarrow \forall x\ \phi[T'/P](x)\Big).$$

In the proof of the above lemma we will use the following characterisation of Δ_0-induction:[14]

Fact 25. *Let $A(x)$ be an arbitrary formula in a language $\mathcal{L}$ extending the language of the arithmetic. Then the following conditions are equivalent for an arbitrary $\mathcal{L}$-structure M satisfying* PA*:*

1. *$M \models \Delta_0$-induction for the formula $A(x)$.*
2. *For every b there exists an element (a coded set) c such that*

$$M \models \forall x < b \ \Big(A(x) \equiv x \in c\Big),$$

in which case we say that the set of elements below b satisfying $A(x)$ in M is coded.

Proof of Lemma 24. We will show that for an arbitrary model M of CT_0 and arbitrary element $b \in M$ the set of elements below b satisfying $T'(x)$ is coded in M. Then the claim of the lemma will follow by the above Fact.

In any model of CT_0 the extension of the predicate T' is the sum over a's of the extensions of the predicates T'_a. Now observe that for arbitrary b and arbitrary $x < b$ the following conditions are equivalent by Lemma 22 (if we assume, as we may, that the complexity of every formula is no greater than its code, i.e. for arbitrary x we have $x \in A_y$ for some $y \leq x$):

1. $T'(x)$.
2. $T'_b(x)$.

Now, there clearly exists an element c such that:

$$\forall x < b \ \Big(T'_b(x) \equiv x \in c\Big),$$

since the predicate $T'_b(x)$ is fully inductive by Lemma 20. □

Lemma 26 (Compositionality of T')**.** *Provably in* CT_0 *the following conditions hold:*

1. $\forall s, t \ \Big(T'(s = t) \equiv s^\circ = t^\circ\Big)$.
2. $\forall \phi, \psi \ \Big(T'(\phi \otimes \psi) \equiv T'\phi \otimes T'\psi\Big)$.

[14] See [10], Proposition 1.4.2.

3. $\forall \phi \left(T'(\neg\phi) \equiv \neg T'\phi\right)$.

4. $\forall v, \phi(v) \left(T'(Qv\ \phi(v)) \equiv Qx\ T'(\phi(\underline{x}))\right)$.

Where $\otimes \in \{\wedge, \vee\}$ *and* $Q \in \{\forall, \exists\}$.

Proof. The lemma follows immediately from the definition of $T'(x)$, Lemma 22 and Lemma 19. □

The fact that the predicate $T'(x)$ may be presented as a sum of the predicates T'_c, which are fully inductive and compositional for formulae of logical complexity no greater than c guarantees that T' enjoys all sorts of good properties. The next lemma states that it is fully extensional.

Lemma 27 (Regularity of T')**.** *The predicate T' satisfies the regularity axiom, i.e.*

$$\forall\phi\forall t, s\ \left(s^\circ = t^\circ \to T'\phi(t) \equiv T'\phi(s)\right).$$

The same hold for predicates T'_b for arbitrary b

Proof. We know that for all $d \leq b$ and $x \in A_d$ we have

$$T'(x) \equiv T'_b(x).$$

Thus it is enough to prove that for arbitrary b the following equivalence holds:

$$\forall d \leq b \forall \phi(v) \in A_d \forall t, s\ \left(s^\circ = t^\circ \to T'_b\phi(t) \equiv T'_b\phi(s)\right).$$

This however may be shown by a straightforward Π_1-induction on d. □

Lemma 28 (Quantifier axioms for T')**.** *Provably in* CT_0 *The predicate T' satisfies the compositional axioms for quantifiers in the term formulation, i.e.*

1. $\forall v, \phi(v)\ \ T'(\forall v\ \phi(v)) \equiv \forall t\ T'(\phi(t))$.

2. $\forall v, \phi(v)\ \ T'(\exists v\ \phi(v)) \equiv \exists t\ T'(\phi(t))$.

Proof. This follows immediately from Lemma 27, since provably in PA (and in fact in much weaker theories) for every term t there exists a unique numeral $\underline{a}$ with $t^\circ = (\underline{a})^\circ$. □

Recall that by the global reflection principle for T' we mean the axiom:

$$\forall\phi \left(\mathrm{Pr}_{T'}\phi \longrightarrow T'\phi\right),$$

where $\mathrm{Pr}_{T'}$ is an instance of Pr_τ defined in Convention 1 with $\tau = T'$. Intuitively $\mathrm{Pr}_{T'}$ means first-order provability from some premises ϕ satisfying $T'(\phi)$.

Lemma 29 (Global reflection for T'). CT_0 *proves the global reflection principle for the formula* T'.

Proof. Working in CT_0, take an arbitrary (code of a) proof d with the conclusion ϕ, all of whose premises ψ satisfy $T'(\psi)$. Then there exists a b such that for every formula η in d there exists $y < b$ such that:

$$\eta \in A_y.$$

Then for each $x < b$ we have:

$$T'(x) \equiv T'_b(x).$$

So it is enough to prove that $T'_b(\phi)$ holds. But this may be easily proved by Lemmata 20 and 28 analogously to the usual proof of the global reflection principle for CT with compositional clauses for quantifiers in term formulation and with the full induction for the compositional truth predicate, since the formulae $T'_b(x)$ are fully inductive by Lemma 20 and compositional for formulae of logical complexity at most b by Lemma 19. They satisfy compositional axioms for quantifiers in the term formulation, since T' does by Lemma 28 and as already noted for all $x < b$ we have $T'(x) \equiv T'_b(x)$. □

Analogously, we obtain the following lemma:

Lemma 30 (Axiom soundness property). CT_0 *proves the following sentence:*

$$\forall \phi \left(\mathrm{Ax}_{\mathrm{PA}}(\phi) \longrightarrow T'(\phi) \right).$$

Now we are ready to prove the main theorem of our paper.

Proof of Theorem 13. We want show that CT_0 PA-relatively defines CT_0 extended with the global reflection principle and the axiom soundness property for PA. It is enough to observe that by the lemmata 24, 26, 29 and 30 the formula $T'(x)$ satisfies the axioms of the latter theory, so it gives us the interpretation of the latter theory in CT_0 which fixes the arithmetical vocabulary. □

Let us observe that to establish the non-conservativity result for CT_0 one actually needs only two properties of the compositional truth predicate: disjunctive correctness and internal induction. By Theorem 8 of Enayat–Visser and Leigh[15] the theory CT^- augmented with the latter principle is still conservative. We do not know yet whether this is true also for the former.

[15] See [2], remarks in Section 6 and [13], Theorem 3.

Let us end this section by a corollary we have referred to in the abstract. Recall the usual version of CT^- with compositional axioms for quantifiers of the form

$$\forall \phi(v)\ \ T(Qv\, \phi(v)) \equiv Qt\, T(\phi(t)).$$

Let us denote by CT^-_t this usual version of CT^-. Since the regularity principle implies that the theories CT^- and CT^-_t are equivalent, our results actually imply:

Corollary 31. CT^-_t *extended with* Δ_0*-induction for the full language and the regularity principle is not conservative over* PA.

4 A variation of the main result

In this section we shall prove that CT_0 is actually very close to proving the global reflection principle for its truth predicate. We will show that a very natural modification of this theory accomplishes this aim. First we shall formulate CT_0 over a theory PA^+ which is simply PA formalized in an enriched language with additional function symbols for some primitive recursive functions and extended with axioms determining the meaning of new symbols. Observe that the axioms of CT_0 remain unchanged, but the notion of *a term* is substantially enriched. To such a theory we add an axiom which generalizes the regularity principle (REG) to substitutions of boundedly-many terms at once.

Now we introduce a short sequence of definitions:

Definition 32.

1. Let $\mathcal{L}^+ = \mathcal{L}_{\mathrm{PA}} \cup \{(x)_y\}$, where $(x)_y$ is a two-argument function (with the intended reading "the result of the projection of x to its y-th component").

2. Let $\pi(x, y, z)$ be an $\mathcal{L}_{\mathrm{PA}}$ formula expressing that z is the y-th element of the sequence x. PA^+ is the theory theory in $\mathcal{L}^+$ containing the usual axioms of PA (we allow formulae of $\mathcal{L}^+$ in the induction axioms) and additionally the following axiom for $(x)_y$

$$\forall x, y, z\ \ \big((x)_y = z \equiv \pi(x, y, z)\big)$$

Convention 33. For each n, $v = \langle x_1, \ldots, x_n\rangle$ is the arithmetical formula representing in PA the relation "v is the code of the sequence of length n, containing as its first element x_1, as its second x_2,..., and as its n-th x_n".

We extend the arithmetization so that it embraces the two-argument symbol $(x)_y$. We assume that in PA$^+$ the definable syntactical relations apply to the enriched language, hence Term(x), Form(x) mean "x is a (code of a) $\mathcal{L}^+$ term", "x is a (code of a) $\mathcal{L}^+$ formula" respectively. Additionally, we need the following formulae [16]:

Definition 34 (PA$^+$)**.**

1. TermSeq(x) which says "x is a sequence of closed terms."

2. We extend the function $(\cdot)^\circ$ to terms of $\mathcal{L}^+$ by putting

$$((t)_s)^\circ = ((t)^\circ)_{(s)^\circ}$$

   where t, s are arbitrary terms. In PA$^+$ we define a generalized function of valuation which, apart from terms, is applicable also to *sequences* of terms, yielding the sequence of values of those terms. Since we will be interested only in values of sequences of terms, we will denote this function by $(\cdot)^\circ$, in the same way in which the standard function of valuation was denoted.

3. Subst(x, y) is now a *generalized* function of substitution with the following properties

   (a) y must be a sequence of closed terms such that if a is the number of distinct variables occurring in x then the length of y is at least a,

   (b) Subst(x, τ) is the effect of formal substitution in x for free variable x_z the term coded by the z-th element of sequence τ (for every z).

   As indicated in the previous section, in order not to complicate the formulae, we shall be writing $\phi(\tau)$ instead of Subst(ϕ, τ), even in the case when τ is a sequence of terms.

4. Let τ be any sequence. By τ^* we denote the unique sequence y such that

   (a) len(τ) = len(y)

[16] Mind that *sequence*, *term* etc. in the definition below mean *a code of sequence*, *a code of term* rather than truly finite sequence, true term, defined externally.

(b) y has on its $i-$th place ($i < \text{len}(y) = \text{len}(\tau)$) the Gödel code of the term $(\underline{\tau})_{\underline{i}}$.[17]

Example 35. Let $n = \langle \ulcorner SS(0) \urcorner, \ulcorner S(0) + 0 \urcorner, \ulcorner S(S(0) + S(0)) \urcorner \rangle$. Then

$$(n)^\circ = \langle 2, 1, 3 \rangle$$

and

$$n^* = \langle \ulcorner (S^n 0)_{S0} \urcorner, \ulcorner (S^n 0)_{SS0} \urcorner, \ulcorner (S^n 0)_{SSS0} \urcorner \rangle$$

and $(\ulcorner (S^n 0)_{S(0)+S(0)} \urcorner)^\circ = (n)_2 = \ulcorner S(0) + 0 \urcorner$.

Remark 36 (PA$^+$)**.** Let τ be any sequence. Then τ^* is a sequence of closed terms and $\tau = (\tau^*)^\circ$.

Remark 37. Mind the difference between $\phi(\tau^*)$ and $\phi(\tau)$. For example if

$$\phi = \ulcorner x_1 + x_2 = x_3 \urcorner$$

and $\tau = \langle \ulcorner 0 \urcorner, \ulcorner S0 + SS0 \urcorner, \ulcorner SS0 \urcorner \rangle$, then (provably in PA$^+$)

$$\phi(\tau) =_{def} \text{Subst}(\phi, \tau) = \ulcorner 0 + (S0 + SS0) = SS0 \urcorner$$

and

$$\phi(\tau^*) =_{def} \text{Subst}(\phi, \tau^*) = \ulcorner (\underline{\tau})_{S0} + (\underline{\tau})_{SS0} = (\underline{\tau})_{SSS0} \urcorner$$

Note that in the previous sections it was irrelevant for our argumentation whether the above mentioned functions are primitive or defined symbols. Now it becomes crucial to our argument, and the reason for extending the language will become apparent when proving the main theorem of this section.

Definition 38. Let ϕ be an $\mathcal{L}^+$ formula. We say that an occurrence of term t in ϕ is **bounded** if and only if it contains a bounded occurrence of a variable. An occurrence of a term t is **free** if it is not bounded.

Example 39. $\ulcorner SS(v) \urcorner$ and $\ulcorner SS(v) + S(y) \urcorner$ have bounded occurrences in

$$\phi = \ulcorner \exists v \;\; SS(v) + S(y) = SS(z) \urcorner$$

but $\ulcorner S(y) \urcorner$ and $\ulcorner SS(z) \urcorner$ don't.

[17] Formally: $\forall i < \text{len}(y) \;\; (y)_i = \ulcorner (\urcorner {}^\frown \underline{\tau} {}^\frown \ulcorner) \urcorner {}^\frown \underline{i}$.

We will need a more refined measure of complexity of formulae than the one defined in the previous section (called *logical complexity* there). To define precisely the conditions which it should meet, let us introduce one more notion, borrowed from [13]:

Definition 40 (PA$^+$)**.** Let ϕ be an $\mathcal{L}^+$-formula and w be any number which is not a code of any $\mathcal{L}^+$ symbol. Let $\mathcal{L}^w$ be a language resulting by adding w to $\mathcal{L}^+$. We treat w as an additional free variable for marking places for terms in a formula. Formally: the formula ϕ' results from ϕ by formally substituting w for every free variable of ϕ. The formula $\bar{\phi}$ results from ϕ' by formally substituting w for every term t such that every variable occurring in t is equal to w. If ψ is any $\mathcal{L}^+$- formula, we put

$$\phi \sim \psi$$

iff $\bar{\phi} = \bar{\psi}$.

The above definition serves for formalizing the relation of "being the same up to substitution of terms for free occurrences of terms". We demand that our measure of complexity have two properties:[18]

FIN Provably in PA$^+$, for every n, k, there are only finitely many $\sim$-equivalence classes of formulae of complexity less than n which use variables (either as bounded or free ones) with indices smaller than k.

COM The measure of $\phi \otimes \psi$ and $\neg\psi$ is greater than the measure of ϕ, ψ (for $\otimes \in \{\wedge, \vee\}$). The measure of $Qx\phi(x)$ (for $Q \in \{\forall, \exists\}$) is greater than the measure of $\phi(t)$ for every closed term t.

The measure given by the logical complexity of a formula, as used in the previous section, does not satisfy property FIN (at least for formulae of language $\mathcal{L}$ (and consequently $\mathcal{L}^+$) which contains infinitely many closed terms[19]). If we tried to use the Gödel number of ϕ as its size, then the resulting measure would not satisfy COM. The next definition supplies us with an appropriate measure.

Definition 41. Let ϕ be an $\mathcal{L}^w$ formula.

1. The **syntactic tree** of an $\mathcal{L}^w$ formula ϕ is defined as usual except for the fact that we unravel also terms occurring in ϕ. In consequence the only $\mathcal{L}^w$ symbols that are allowed to occur in leaves of the syntactic tree of ϕ are individual constants and variables.

[18] "FIN" is a short for "FINite" and "COM" for "COMpositional".

[19] For example there are infinitely many sentences of the form $t_1 = t_2$ and syntactic tree for every such formula has logical complexity 0.

2. We say that a formula ϕ has **complexity at most** n if and only if the height of the syntactic tree of $\bar{\phi}$ (i.e. the largest number of vertices on a maximal path) is at most n. The **complexity** of ϕ is the least n such that ϕ is of complexity at most n.

It is straightforward to check that our definition of complexity measure satisfies FIN and COM as stated above. $\bar{\phi}$ will serve us also to define the *template* of ϕ. In fact $\bar{\phi}$ is very close to match our objectives: the only improvement we need to introduce is to number the occurrences of w in $\bar{\phi}$. The next definition accomplishes this aim:

Definition 42 (PA$^+$)**.**

1. Let $\{e_i\}$ be an injective enumeration of a set of numbers which are not codes of any $\mathcal{L}^+$ symbols (starting from e_1 for simplicity). We will treat them as additional free-variable symbols (we will allow substituting closed terms for them but they are not part of $\mathcal{L}^+$). Let $\mathcal{L}^*$ be the language resulting from $\mathcal{L}^+$ by adding $\{e_i\}$ as new variable symbols.[20]

2. Let ϕ be an $\mathcal{L}^w$ formula and x_1, x_2 be two occurrences of free variables in ϕ. We put
$$x_1 \preccurlyeq_\phi x_2$$
if and only if x_1 is more to the left in the syntactic tree of ϕ than x_2.

3. Suppose now ψ is an $\mathcal{L}^+$ formula. ϕ^* is the formula resulting from $\bar{\phi}$ by substituting $e_0, e_1, e_2, \ldots$ in $\bar{\phi}$ for the first, the second, the third, ... occurrence of w respectively, (where the respective ordering of occurrences of free variables is $\preccurlyeq_\phi$). ϕ^* is called **the template of** ϕ. Let us note that the only variables which occurs freely in ϕ^* are the e_i's. In particular no variables from $\mathcal{L}$ occurs freely in ϕ^*. Let us note also that for every $\phi \in \mathcal{L}^+$
$$\phi \sim \phi^*$$

4. Provably in PA$^+$, for every $\phi \in \mathcal{L}^+$ there exists a coded set of those indices i such that e_i occurs in ϕ^*. For a given ϕ, such a set will be denoted by $E(\phi)$.

5. We define a natural extension of the function Subst so that it can operate on templates (and denote with the same symbol and use the same conventions): we assume that if ϕ^* is the template of an $\mathcal{L}^+$ formula

[20] In particular we assume that provably in PA$^+$ $\mathcal{L}^+$ is a sublanguage of $\mathcal{L}^*$.

and τ is a sequence of $\mathcal{L}^+$ closed terms, then Subst(ϕ^*, τ) is the result of the following substitution in ϕ^*: for every $i \in E(\phi)$[21].

$$\text{the } i\text{-th element of } \tau \text{ is substituted for } e_i$$

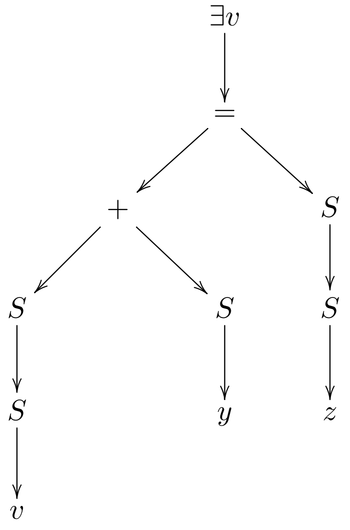
6. Provably in PA$^+$ for every $\phi \in \mathcal{L}^+$ the template of ϕ is uniquely determined. Moreover there is only one sequence of terms τ such that

   (a) $\phi = \phi^*(\tau)$

   (b) $\text{len}(\tau) = \max\{i \mid e_i \in E(\phi)\}$

   The last condition is added only to guarantee uniqueness of such τ, and it won't play any important role in our proof. Such a sequence of terms will be denoted by τ_ϕ.

Example 43. The syntactic tree of $\ulcorner \exists v \;\; SS(v) + S(y) = SS(z) \urcorner$ is the (code of the) following:



The ordering on the set of occurrences of free variables in this formula is simply:

$$\ulcorner y \urcorner \preccurlyeq_\phi \ulcorner z \urcorner$$

Example 44. The following formula

$$\phi = \ulcorner \exists z \;\; SSS(z) + v = SSSSSSSS(y) \urcorner$$

[21] For simplicity we do not assume that Subst is defined for all formulae of $\mathcal{L}^*$, but only for formulae of $\mathcal{L}^+$ and templates (which do not contain any free variables from $\mathcal{L}^+$.)

has complexity 7. $\bar{\phi}$ is equal to $\ulcorner \exists z \;\; SSS(z) + w = w \urcorner$. The template of ϕ is

$$\phi^* = \ulcorner \exists z \;\; SSS(z) + e_1 = e_2 \urcorner$$

and the corresponding sequence of terms is

$$\tau_\phi = \langle \ulcorner v \urcorner, \ulcorner SSSSSSSS(y) \urcorner \rangle$$

To give another example: if

$$\psi = \ulcorner SSS(z) + v = SSSSSSSS(y) \urcorner$$

then $\bar{\psi} = \ulcorner w = w \urcorner$ and

$$\psi^* = \ulcorner e_1 = e_2 \urcorner.$$

The corresponding sequence of terms is

$$\tau_\psi = \langle \ulcorner SSS(z) + v \urcorner, \ulcorner SSSSSSSS(y) \urcorner \rangle$$

So far we introduced four different languages: $\mathcal{L}$, $\mathcal{L}^+$, $\mathcal{L}^w$ and $\mathcal{L}^*$. Let us stress that the last two play only auxiliary roles in our reasoning and the first one was relevant only in the previous chapter and now is replaced by $\mathcal{L}^+$ as the "basic" language. Formulae Term(x), Sent(x) etc. should be read as "x is a $\mathcal{L}^+$ term" and "x is a $\mathcal{L}^+$ sentence" (respectively). In particular by writing $\forall \phi$ we implicitly quantify over $\mathcal{L}^+$ sentences (compare Convention 1).

Definition 45. CT_0^+ is the theory containing the following axioms

- PA^+
- compositional axioms for T for the language $\mathcal{L}^+$, as in Definition 3.
- Δ_0-induction for formulae of the language $\mathcal{L}^+ \cup \{T\}$.
- **Generalized regularity principle**:

$$\forall \phi \forall x, y \; \big(\mathrm{TermSeq}(x) \wedge \mathrm{TermSeq}(y) \wedge (x)^\circ = (y)^\circ \rightarrow T(\phi(x)) \equiv T(\phi(y))\big) \tag{GREG}$$

By an adaptation of methods of Enayat–Visser from their proof of conservativity of CT^- + (INT), one can show that CT^- + (INT) + (GREG) is conservative over PA.

Lemma 46. $\mathrm{CT}_0^+ \vdash \forall \phi \forall x, y \; \big(\mathrm{TermSeq}(x) \wedge y = (x)^\circ \rightarrow T(\phi(x)) \equiv T(\phi(y^*))\big)$

Proof. Using the definitions introduced above and Remark 36 one shows that provably in PA^+ for every sequence σ and every sequence of terms τ such that $\sigma = (\tau)^\circ$,

$$(\tau)^\circ = \sigma = (\sigma^*)^\circ$$

hence by the Generalized regularity principle

$$T(\phi(\tau)) \equiv T(\phi(\sigma^*))$$

which ends the proof. □

Definition 47 (PA^+)**.** Let ϕ be any $\mathcal{L}^*$ formula and let x be any $\mathcal{L}^+$ variable not occurring in ϕ (either as free or a bounded one). Let τ_x be the sequence of terms satisfying the following conditions:

1. $\mathrm{len}(\tau_x) = \mathrm{len}(\tau_\phi)$
2. for every $i \in E(\phi)$, $(\tau_x)_i = \ulcorner (x)_i \urcorner$

We define $\phi_x = \mathrm{Subst}(\phi^*, \tau_x)$.

Example 48. Let $\phi = \ulcorner \exists z \;\; SSS(z) + v = SSSSSSSS(y) \urcorner$ be the formula from Example 44. Then $\phi^* = \ulcorner \exists z \;\; SSS(z) + e_1 = e_2 \urcorner$ and

$$\phi_x = \ulcorner \exists z \;\; SSS(z) + (x)_0 = (x)_1 \urcorner$$

Definition 49 (PA^+)**.** The **index** of a formula ϕ is the maximum of its complexity and the greatest index of a variable occurring in it (either as free or a bounded one).

Example 50. Templates of sentences of index ≤ 3 are precisely (codes of) the following ones:

1. $e_1 = e_2$, $\neg(e_1 = e_2)$
2. $(e_1 = e_2) \wedge (e_3 = e_4)$, $(e_1 = e_2) \vee (e_3 = e_4)$
3. $Qx_i(x_i = e_1)$, $Qx_i(e_2 = x_i)$, $Qx_i(e_1 = e_2)$, $Qx_i(x_i = x_i)$, for $i \leq 3$ and $Q \in \{\forall, \exists\}$.

The following is the last technical definition in our paper.

Definition 51 (PA^+)**.** Let ϕ^* be a template. Two sequences of terms τ, σ are said to be ϕ^*-**equivalent** if for all i such that e_i occurs in ϕ^*, $(\tau)_i = (\sigma)_i$. If τ and σ are ϕ^* equivalent, we denote it by $\tau \sim_{\phi^*} \sigma$.

Example 52. Let $\phi = \ulcorner \exists x (S(x) = 0 \wedge y < z)\urcorner$. Then $\phi^* = \ulcorner \exists x \;\; S(x) = 0 \wedge e_1 < e_2 \urcorner$ and the following two sequences of terms are *not* ϕ^* equivalent:

1. $\langle \ulcorner S0 \urcorner, \ulcorner 0 \urcorner \rangle$
2. $\langle \ulcorner 0 \urcorner, \ulcorner 0 \urcorner \rangle$

Remark 53 (PA$^+$)**.** If $\tau \sim_{\phi^*} \sigma$ then $\phi^*(\tau) = \phi^*(\sigma)$.

Remark 54. Let us observe that provably in PA$^+$ we have: for any (arithmetical) formula ϕ and any sequence of terms τ it holds that

$$\phi_x(\underline{\tau}) =_{df} \text{Subst}(\phi_x, \underline{\tau}) = \text{Subst}(\phi^*, \tau^*) =_{df} \phi^*(\tau^*) \tag{1}$$

Hence, in CT_0^+ we have: for every $\mathcal{L}^+$ sentence ϕ, every sequence σ and every sequence of terms τ such that $\tau \sim_{\phi^*} \tau_\phi$ (for the definition of τ_ϕ see Definition 42 point 6) and $\sigma = (\tau)^\circ$

$$\begin{aligned} T\phi &\equiv T\phi^*(\tau) && \text{by definitions of } \phi^* \text{ and } t_\phi \text{ and Remark 53} \\ &\equiv T\phi^*(\sigma^*) && \text{by Lemma 46} \\ &\equiv T\phi_x(\underline{\sigma}) && \text{by (1)} \end{aligned}$$

In particular for any $\tau \sim_{\phi^*} \tau'$ if $\sigma = (\tau)^\circ$ and $\sigma' = (\tau')^\circ$ we have

$$T\phi_x(\underline{\sigma}) \equiv T\phi_x(\underline{\sigma'})$$

After these comments we are ready to state and prove our second theorem.

Theorem 55. CT_0^+ *proves the Global Reflection Principle and the Axiom Soundness Property for* PA.

Proof. We work in CT_0^+. The fact that CT_0^+ proves that all axioms of PA$^+$ are true is obvious. Let us show the Global Reflection Principle. Note that, provably in PA$^+$, for every c, d there are only finitely many templates for formulae of index less than c. Each such formula can be obtained from one of those finitely many templates by the procedure of substituting terms for the e_i's. Let $\gamma(c)$ be a (code of a) set of all templates for sentences of index at most c. Let y, z be any $\mathcal{L}^+$ variables which do not occur in those sentences. As in the main theorem we shall make use of simplistic truth predicates. But this time working in extended language enables us to make them even more simplistic. Let us put:

$$T_c(x) := \bigvee_{\phi \in \gamma(c)} \big(\exists y, z \;\; (\text{TermSeq}(y) \wedge x = \text{Subst}(\phi, y) \wedge z = (y)^\circ \wedge \phi_z)\big)$$

For example the disjuncts of $T_3(x)$ corresponding to the templates $\ulcorner e_1 = e_2 \urcorner$ and $\ulcorner \exists x_2(x_2 = e_1) \urcorner$ are the following

$$\big(\exists y, z \ \ (\text{TermSeq}(y) \wedge x = \text{Subst}(\ulcorner e_1 = e_2 \urcorner, y) \wedge z = (y)^\circ \wedge (z)_0 = (z)_1)\big)$$

$$\big(\exists y, z \ \ (\text{TermSeq}(y) \wedge x = \text{Subst}(\ulcorner \exists x_2(x_2 = e_1) \urcorner, y) \wedge z = (y)^\circ \wedge \exists x_2(x_2 = (z)_0))\big)$$

Observe that having projection functions as primitive symbols in the language makes it possible to use finitely many quantifiers at the beginning of each disjunct of $T_c(x)$. In case of their absence we would have to add one quantifier for each variable e_i of a template ϕ, which in case of non-standard formulae would result in a block of quantifiers of non-standard length. As in the proof of the main theorem the function $c \mapsto T_c$ is primitive recursive. As in the proof of our main theorem, define

$$T'_c(x) = T\big(T_c(\underline{x})\big)$$

Arguing exactly as in the proof of Lemma 20 we show that for every c every formula with T'_c satisfies the induction scheme. This time our simplistic predicates have an additional feature: for any sentence ϕ of index less or equal to c we have

$$T(\phi) \equiv T'_c(\phi)$$

Note that it follows from the above, that TT_c are compositional on the sentences of index less than c. To prove the above assertion let us fix a sentence ϕ of index less than or equal to c and observe that for $\sigma = (\tau_\phi)^\circ$ we have

$$\begin{aligned} T\big(T_c(\underline{\phi})\big) &\equiv T\Bigg(\bigvee_{\psi \in \gamma(c)} \big(\exists y, z \ \ (\text{TermSeq}(y) \wedge \phi = \text{Subst}(\psi, y) \wedge z = (y)^\circ \wedge \psi_z)\big)\Bigg) \\ &\equiv T(\phi_z(\underline{\sigma})) \\ &\equiv T(\phi) \end{aligned}$$

Indeed, the first equivalence holds by definition, and the third is obtained by Remark 54. Let us focus on the second one. From left to right: if

$$T\Bigg(\bigvee_{\psi \in \gamma(c)} \big(\exists y, z \ \ (\text{TermSeq}(y) \wedge z = (y)^\circ \wedge \phi = \text{Subst}(\psi, y) \wedge \psi_z)\big)\Bigg)$$

holds then by Disjunctive Correctness of T we get that for some template ψ it holds that

$$T\big(\exists y, z \ \ (\text{TermSeq}(y) \wedge z = (y)^\circ \wedge \phi = \text{Subst}(\psi, y) \wedge \psi_z)\big)$$

Hence for some τ', σ' such that $\sigma' = (\tau')^\circ$ we have by compositionality of T

$$\text{TermSeq}(t) \wedge \phi = \text{Subst}(\psi, \tau') \wedge T\psi_z(\underline{\sigma'})$$

But for $\psi \neq \phi^*$ this sentence can be easily disproved in CT^- (because the middle conjunct is disprovable already in PA^+). Moreover it must be the case that $\tau' \sim_{\phi^*} \tau_\phi$. Hence, for $\sigma = (\tau_\phi)^\circ$ we get $T(\phi_z(\underline{\sigma}))$ (invoking Remark 54). From right to left it is easier: if, for $\sigma = (t_\phi)^\circ$ we have $T(\phi(\underline{\sigma}))$ then also

$$\text{TermSeq}(\tau_\phi) \wedge \phi = \text{Subst}(\phi^*, \tau_\phi) \wedge \sigma = (\tau_\phi)^\circ \wedge T\phi_z(\underline{\sigma})$$

Hence $T(\exists y, z\ \ \text{TermSeq}(y) \wedge \phi = \text{Subst}(\phi^*, y) \wedge z = (y)^\circ \wedge \phi_z)$ and once again by Disjunctive Correctness we get

$$T\Bigg(\bigvee_{\psi \in \tau(c)} \big(\exists y, z\ \ (\text{TermSeq}(y) \wedge \phi = \text{Subst}(\psi, y) \wedge z = (y)^\circ \wedge \psi_z)\big)\Bigg)$$

Now we proceed as in the proof of our main theorem. Let d be a proof in First Order Logic of a sentence $\phi \in \mathcal{L}^+$ from true premises. There is a c such that each formula ψ occurring in d is of index at most c. Then by induction on the length of d we check that in each sequent $\Gamma \longrightarrow \Delta$ in d if for every ψ in Γ we have $T'_c(\psi)$, then for some θ in Δ we have $T'_c(\psi)$. This is legitimate, since $T'_c(x)$ is inductive and compositional on formulae occurring in d. Hence $T'_c(\phi)$ and by the above considerations also $T(\phi)$. □

5 Appendix

In the following section we shall discuss the proof by Kotlarski and the mistake that has been pointed out by Richard Heck and Albert Visser. We decided to include this discussion, since the alleged result has been cited, sometimes with repeating the erroneous proof, in at least three different papers, some of them written already after the gap in the proof has been observed. Let us present Kotlarski's argument.

We would like to show that CT_0 proves that all theorems of PA (axiomatised with a parameter-free induction scheme) are true. We know that CT_0 proves that any instance of the parameter-free induction scheme for arithmetical formulae under any substitution of closed terms for free variables is true. It suffices to show that for any proof d formalized in Hilbert system if all its premises are true, then the conclusion is true. But the only rule of Hilbert calculus, namely Modus Ponens, is clearly truth-preserving. In the Hilbert system, we assume only finitely many axiom schemes for first-order logic and arbitrary propositional tautologies. We may also easily check in CT^- that all these finitely many axiom schemes are true for arbitrary sentences. By Δ_0-induction we may check that all sentences which are propositional tautologies are true. Thus every provable sentence is true.

However Kotlarski overlooked a crucial detail in the formulation of Hilbert calculus. We can assume that metavariables ϕ, ψ occurring in the axiom schemes for this calculus represent arbitrary arithmetical formulae, not necessarily sentences. In such a case we apparently have to add a generalization rule to Hilbert calculus. Then if we want to show by induction that first-order derivations preserve truth, the actual induction thesis should state something like: "for any substitution of terms for free variables the resulting sentence is true" which is a Π_1 statement. In another possible formulation of Hilbert calculus we may assume that metavariables ϕ, ψ occurring in its axiom schemes represent only arithmetical *sentences*, but then we have to take as logical axioms arbitrary *universal closures* of propositional tautologies rather than tautologies themselves. But then in turn we have to prove the following statement: universal closures of propositional tautologies are true. So consider any sentence of the form:

$$\forall x_1 \forall x_2 \ldots \forall x_c \ \ \phi(x_1, \ldots, x_c),$$

where $\phi(x_1, \ldots, x_c)$ is a propositional tautology. Using Δ_0-induction for the truth predicate we may indeed prove that for arbitrary numerals $\underline{a_1}, \ldots, \underline{a_c}$ the following sentence is true:

$$\phi(\underline{a_1}, \ldots, \underline{a_c}).$$

Then by compositional axioms we may even prove for any fixed standard k (i.e. an element of true ω, viewed externally) that the following is also true, where the block of the universal quantifiers is of standard length:

$$\forall x_{c-k} \ldots \forall x_c \ \ \phi(\underline{a_1}, \ldots, \underline{a_{c-k-1}}, x_{c-k}, \ldots, x_c).$$

But to prove that the whole universal closure is true we apparently need to use induction over the following statement: "For all $y < c$ and all numerals $\underline{a_1}, \ldots, \underline{a_{c-y}}$ the sentence $\forall x_{c-y} \ldots \forall x_c \ \ \phi(\underline{a_1}, \ldots, \underline{a_{c-y-1}}, x_{c-y}, \ldots, x_c)$ is true". And to do this we would need Π_1 induction, since we quantify over arbitrary numerals.

Moreover, it seems that the problem of nonstandard blocks of universal quantifiers is not a mere technicality. Note that taking universal closures of propositional tautologies in Hilbert calculus basically allows us to dispense of eigenvariables of sequent calculus and it seems reasonable that to prove that reasoning in sequent calculus preserves truth, we really need to quantify over all possible substitutions of terms for eigenvariables and thus basically we are forced to employ Π_1 induction for the truth predicate.

6 Acknowledgements

We are extremely grateful to our supervisor, Cezary Cieśliński for his invaluable support and for introducing us to the problem of the conservativity of Δ_0-induction for the compositional truth predicate. We would also like to express our sincere gratitude to the anonymous referee, who spotted a substantial but subtle mistake in the first version of the paper. We are also grateful to Ali Enayat and Albert Visser for extremely interesting discussions.

The research presented in this paper was supported by the National Science Centre, Poland (NCN), grant number 2014/13/B/HS1/02892.

References

[1] Cezary Cieśliński. Truth, conservativeness and provability. *Mind*, (119):409–422, 2010.

[2] Ali Enayat and Albert Visser. New constructions of satisfaction classes. *Logic Group Preprint Series*, (303), 2013.

[3] Kentaro Fujimoto. Relative truth definability in axiomatic truth theories. *Bulletin of Symbolic Logic*, (16):305–344, 2010.

[4] Petr Hájek and Pavel Pudlák. *Metamathematics of First-Order Arithmetic*. Springer-Verlag, 1993.

[5] Volker Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, 2011.

[6] Richard Heck. Consistency and the theory of truth. *Review of Symbolic Logic*, to appear.

[7] Leon Horsten. *The semantical paradoxes, the neutrality of truth and the neutrality of minimalist*, pages 173–87. Studies in the General Philosophy of Science. Tilburg University Press.

[8] Richard Kaye. *Models of Peano Arithmetic*. Oxford University Press.

[9] Jeffrey Ketland. Deflationism and Tarski's paradise. *Mind*, (108):69–94, 1999.

[10] Roman Kossak and James Schmerl. *The Structure of Models of Peano Arithmetic*. Clarendon Press.

[11] Henryk Kotlarski. Bounded induction and satisfaction classes. *Zeitschrift für matematische Logik und Grundlagen der Mathematik*, (32):531–544, 1986.

[12] Henryk Kotlarski, Stanisław Krajewski, and Alistair Lachlan. Construction of satisfaction classes for nonstandard models. *Canadian Mathematical Bulletin*, (24):283–93, 1981.

[13] Graham Leigh. Conservativity for theories of compositional truth via cut elimination. *The Journal of Symbolic Logic*, 2015.

[14] Stewart Shapiro. Proof and truth: Through thick and thin. *Jornal of Philosophy*, (95):493–521, 1998.