Models of Positive Truth

Mateusz Łełyk, Bartosz Wcisło

July 24, 2018

Abstract

This paper is a follow-up to [4]. We give a strenghtening of the main result on the semantical non-conservativity of the theory of PT^- with internal induction for total formulae ($PT^- + INT(tot)$). We show that if to PT^- the axiom of internal induction for all arithmetical formulae is added ($PT^- + INT$), then this theory is semantically stronger than $PT^- + INT(tot)$. In particular the latter is not relatively truth definable (in the sense of [9]) in the former. Last but not least we provide an axiomatic theory of truth which meets the requirements put forward by Fischer and Horsten in [8].

1 Introduction

1.1 Axiomatic Theories of Truth

Axiomatic theories of truth is a branch of mathematical logic and philosophy which studies the properties of formal theories generated in the following way:

1. We take a *base theory* B which we demand to be sufficiently strong to (strongly) represent basic syntactical operations.

2. We extend the language of B by adding one new unary predicate T and some axioms for it so that the resulting theory Th prove all sentences of the form

$$T(\ulcorner \phi \urcorner) \equiv \phi$$

for ϕ *in the language of our base theory B.*

For a brief introduction to the subject see [11] and for a more complete one—[10]. The big question that this paper answers in a tiny part is the following: how various axioms for the truth predicate influence its *strength*. For the purpose of investigating this question we focus on the truth theories with *Peano Arithmetic* as a base theory. The notion of *strength* may enjoy many different

1

explications. For example, the simplest one is given by inclusion of sets of consequences: we might say that Th_1 is not weaker than Th_2 if and only if Th_1 proves all the axioms of Th_2. For many applications this is too fine-grained: many theories, intuitively differing in strength, become incomparable out of not-that-important reasons (obviously this is not a formal notion). Better adjusted notion was introduced by Kentaro Fujimoto in [9] and is a special kind of interpretability. We recall the definition:

Definition 1. 1. Let Th_1 and Th_2 be axiomatic truth theories and let T_{Θ_2} be the truth predicate of Th_2. For any sentence Θ of $\mathcal{L}_{\text{Th}_2}$ and a formula $\phi(x) \in \mathcal{L}_{\text{Th}_1}$ with precisely one free variable let

$$\Theta^{\phi(x)}$$

denote the $\mathcal{L}_{\text{Th}_1}$ sentence which results from Θ by substituting $\phi(x)$ for every occurrence $T_{\text{Th}_2}(x)$ (and renaming variables, if necessary).

2. We say that Th_1 *relatively truth defines* Th_2 if and only if there exists a formula $\phi(x) \in \mathcal{L}_{\text{Th}_1}$ such that for any axiom Θ of Th_2

$$\text{Th}_1 \vdash \Theta^{\phi(x)}$$

If Th_2 relatively truth defines Th_1 we will denote it by $\text{Th}_1 \leq_F \text{Th}_2$[1].

In terms of interpretations, relative truth definability is a $\mathcal{L}_{\text{PA}}$-conservative interpretation between truth theories (for the terminology related to interpretations see e.g. [?]). It was argued in [9] that relative truth definability provides a good explication of *epistemological reduction* between truth theories. We may treat it as an explication of a notion of strength: Th_1 is *Fujimoto-stronger* than Th_2 if and only if Th_1 relatively truth defines Th_2 but not *vice versa*. This relation will be denoted by $\lneq_F$.

1.2   Strength relative to PA

In some philosophical debates, especially the ones related to the deflationism, the need for a differently oriented formal explication of strength seems to emerge. It has been claimed (most importantly in [16], [15] and [12]) that deflationary thesis that truth is a "simple" (aka "light", "metaphysically thin") notion implies that the deflationary theory of truth should be *conservative* over PA.[2] Let us recall that a theory of truth can be conservative over PA in two senses:

Definition 2. Let Th be a theory of truth.

---

[1] "F" stands for "Fujimoto."
[2] This thesis however has been recently criticised at length in [2]

1. We say that Th is *proof-theoretically conservative* over PA if and only if for every $\phi \in \mathcal{L}_{\mathrm{PA}}$, if Th $\vdash \phi$, then PA $\vdash \phi$.[3].

2. We say that Th is *model-theoretically conservative* over PA if and only if every model $\mathcal{M}$ of PA admits an *expansion* to a model of Th.[4].

Remark 3. Note that, in the definition of model-theoretical conservativity, we do not merely demand every model to have an *extension* to a model of Th, in which case both notion of conservativity would be the same. We say that $\mathcal{M}'$ is an *expansion* of $\mathcal{M}$ if $\mathcal{M}$ and $\mathcal{M}'$ are the same model, except that $\mathcal{M}'$ carries interpretation of additional function, relation and constant symbols.

The two notions lead in the natural way to the following generalisations:

Definition 4. Let Th_1 and Th_2 be two truth theories.

1. We say that Th_1 is *proof-theoretically not stronger* than Th_2 if every $\mathcal{L}_{\mathrm{PA}}$ sentence provable in Th_1 is provable in Th_2. If Th_1 is proof-theoretically not stronger than Th_1, we will denote it with $\mathrm{Th}_1 \leq_P \mathrm{Th}_2$.[5]

2. We say that Th_1 is *model-theoretically not stronger* than Th_2 if every model which can be expanded to a model of Th_2, can be expanded to a model of Th_1. If Th_1 is syntactically not stronger than Th_1, we will denote it with $\mathrm{Th}_1 \leq_M \mathrm{Th}_2$.[6]

Obviously we say that Th_2 is proof-theoretically (model-theoretically) stronger than Th_1 if $\mathrm{Th}_1 \leq_P \mathrm{Th}_2$ but $\mathrm{Th}_2 \not\leq_P \mathrm{Th}_1$ (respectively, $\mathrm{Th}_1 \leq_M \mathrm{Th}_2$ but $\mathrm{Th}_2 \not\leq_M \mathrm{Th}_1$). This relation will be denoted $\lneq_P$ ($\lneq_M$ respectively).

Let us observe that the three notions of strength introduced above can be ordered with respect to their "granularity". Indeed, for any theories Th_1 and Th_2 we have:

$$\mathrm{Th}_1 \leq_F \mathrm{Th}_2 \implies \mathrm{Th}_1 \leq_M \mathrm{Th}_2 \implies \mathrm{Th}_1 \leq_P \mathrm{Th}_2. \qquad (FMP)$$

Hence also

$$\mathrm{Th}_2 \not\leq_P \mathrm{Th}_1 \implies \mathrm{Th}_2 \not\leq_M \mathrm{Th}_1 \implies \mathrm{Th}_2 \not\leq_P \mathrm{Th}_2 \qquad (\neg PMF)$$

Having three different notions of strength makes it possible to decide not only whether one theory of truth is stronger than another one, but also *how much* stronger it is.

---

[3]This property is also called *syntactical conservativity*
[4]This relation is also called *semantical conservativity*
[5]P is meant to abbreviate "Proof".
[6]M is meant to abbreviate "Model".

1.3 Compositional Positive Truth and its Extensions

Before continuing let us introduce some handy notational conventions:

Convention 1.

1. By using variables ϕ, ψ we implicitly restrict quantification to (Gödel codes of) arithmetical sentences. For example, by $\forall \phi \ \Psi(\phi)$ we mean $\forall x \ \big(\mathrm{Sent}(x) \to \Psi(x)\big)$ and by $\exists \phi \ \Psi(\phi)$ we mean $\exists x \ \big(\mathrm{Sent}(x) \wedge \Psi(x)\big)$. 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

   (a) $\phi(v), \psi(v)$ 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); $\phi(\bar{x}), \psi(\bar{x}) \dots$ run over arbitrary arithmetical formulae.

   (b) s, t run over codes of closed arithmetical terms;

   (c) $v, v_1, v_2, \dots, w, w_1, w_2, \dots$ run over codes of variables;

2. $\mathrm{Form}_{\mathcal{L}_{\mathrm{PA}}}(x), \mathrm{Form}_{\mathcal{L}_{\mathrm{PA}}}^{\leq 1}(x), \mathrm{Sent}_{\mathcal{L}_{\mathrm{PA}}}(x)$ are natural arithmetical formulae strongly representing in PA the sets of (Gödel codes of) formulae of $\mathcal{L}_{\mathrm{PA}}$, formulae of $\mathcal{L}_{\mathrm{PA}}$ with at most one free variable, sentences of $\mathcal{L}_{\mathrm{PA}}$, respectively.

3. if ϕ is a $\mathcal{L}_{\mathrm{PA}}$ formula, then $\ulcorner \phi \urcorner$ denotes either its Gödel code or the numeral denoting the Gödel code of ϕ (context-dependently). This is the unique way of using $\ulcorner \cdot \urcorner$ in this paper.

4. to enhance readability we suppress the formulae representing the syntactic operations. For example we write $\Phi(\psi \wedge \eta)$ instead of $\Phi(x) \wedge$ "x is the conjunction of ψ and η", similarly, we write $\Phi(\psi(t))$ instead of $\Phi(x) \wedge x = \mathrm{Subst}(\psi, t)$;

5. $\underline{x}$ denotes the (Gödel code of) standard numeral for x, i.e. $\ulcorner \underbrace{S \dots S(0)}_{x \text{ times } S} \urcorner$

6. y° is the standard arithmetically definable function representing the value of term (coded by) y.

The main objective of this study is to measure the strength of theories that are compositional, but do not enjoy the global axiom for commutativity with the negation, i.e.

$$\forall \phi \ \big(T(\neg \phi) \equiv \neg T(\phi)\big) \tag{NEG}$$

Let us formulate the theories which will be of the main interest.

Definition 5. PT^- is the axiomatic truth theory with the following axioms for the truth predicate:

1. (a) $\forall s,t \ \left(T(s=t) \equiv (s^\circ = t^\circ)\right)$

    (b) $\forall s,t \ \left(T(\neg s=t) \equiv (s^\circ \neq t^\circ)\right)$

2. (a) $\forall \phi, \psi \ \left(T(\phi \vee \psi) \equiv T(\phi) \vee T(\psi)\right)$

    (b) $\forall \phi, \psi \ \left(T(\neg(\phi \vee \psi)) \equiv T(\neg\phi) \wedge T(\neg\psi)\right)$

3. (a) $\forall v \forall \phi(v) \ \left(T(\exists v \phi) \equiv \exists x \ T(\phi(\underline{x}))\right)$

    (b) $\forall v \forall \phi(v) \ \left(T(\neg\exists v \phi) \equiv \forall x \ T(\neg\phi(\underline{x}))\right)$

4. $\forall \phi \ \left(T(\neg\neg\phi) \equiv T(\phi)\right)$

5. $\forall \phi \forall s,t \left(s^\circ = t^\circ \rightarrow \left(T(\phi(s)) \equiv T(\phi(t))\right)\right)$

In the arithmetized language, we treat $\wedge$ and $\forall$ as symbols defined contextually, i.e. $\phi \wedge \psi = \neg(\neg\phi \vee \neg\psi)$ and $\forall v \phi = \neg\exists v \neg\phi$. Then it is easy to check that the following sentences are provable in PT^-:

1. $\forall \phi, \psi \ (T(\phi \wedge \psi) \equiv (T(\phi) \wedge T(\psi)))$.

2. $\forall \phi, \psi \ (T \ (\neg \ (\phi \wedge \psi)) \equiv (T \ (\neg\phi) \vee T \ (\ulcorner \neg\psi \urcorner)))$.

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

4. $\forall v \forall \phi(v) \ (T(\neg\forall v \phi) \equiv (\exists x \ T(\neg\phi(\underline{x}))))$

In PT^- the internal logic (i.e. the logic of all *true* sentences) is modelled after the Strong Kleene Scheme. Let us observe that axioms of PT^- make it possible to accept a disjunction $\phi \vee \psi$ as true simply on the basis of the truth of one of ϕ and ψ and regardless of whether the second one has its truth value determined. The second theory we will study is more cautious in this respect. Let us define

$$\text{tot}(\phi(v)) := \text{Form}^{\leq 1}(\phi(v)) \wedge \forall x \left(T(\phi(\underline{x})) \vee T(\neg\phi(\underline{x}))\right)$$

In particular if ψ is a sentence, then

$$\text{PAT}^- \vdash \text{tot}(\psi) \equiv \left(T(\psi) \vee T(\neg\psi)\right).$$

where PAT^- is the extension of PA in $\mathcal{L}_{\text{PA}} \cup \{T\}$, with no non-logical axioms for T.

Definition 6. WPT^- is the axiomatic truth theory with the following axioms for the truth predicate:

1. (a) $\forall s,t \ \left(T(s=t) \equiv (s^\circ = t^\circ)\right)$

    (b) $\forall s,t \ \left(T(\neg s=t) \equiv (s^\circ \neq t^\circ)\right)$

2. (a) $\forall \phi, \psi \; \big(T(\phi \vee \psi) \equiv \big(\mathrm{tot}(\phi) \wedge \mathrm{tot}(\psi) \wedge \big(T(\phi) \vee T(\psi)\big)\big)\big)$

   (b) $\forall \phi, \psi \; \big(T(\neg(\phi \vee \psi)) \equiv T(\neg\phi) \wedge T(\neg\psi)\big)$

3. (a) $\forall v \forall \phi(v) \; \big(T(\exists v \phi) \equiv \mathrm{tot}(\phi(v)) \wedge \exists x T(\phi(\underline{x}))\big)$

   (b) $\forall v \forall \phi(v) \; \big(T(\neg\exists v \phi) \equiv \forall x \; T(\neg\phi(\underline{x}))\big)$

4. $\forall \phi \; \big(T(\neg\neg\phi) \equiv T(\phi)\big)$

5. $\forall \phi \forall s, t \big(s^\circ = t^\circ \to T(\phi(s)) \equiv T(\phi(t))\big)$

Using the above mentioned conventions regarding $\wedge$ and $\forall$, it is an easy exercise to show that the following sentences are provable in WPT^-:

1. $\forall \phi, \psi \, (T(\phi \wedge \psi) \equiv (T(\phi) \wedge T(\psi)))$.

2. $\forall \phi, \psi \, (T\,(\neg\,(\phi \wedge \psi)) \equiv (\mathrm{tot}(\phi) \wedge \mathrm{tot}(\psi) \wedge (T\,(\neg\phi) \vee T\,(\neg\psi))))$.

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

4. $\forall v \forall \phi(v) \, (T(\neg\forall v \phi) \equiv (\mathrm{tot}(\phi(v)) \wedge \exists x \; T(\neg\phi(\underline{x}))))$

In WPT^- the internal logic is modelled after the *Weak Kleene Scheme*. $(\mathrm{W})\mathrm{PT}^-$ can be seen as a natural stratified counterpart of $(\mathrm{W})\mathrm{KF}^-$[7] Since in particular $(\mathrm{W})\mathrm{PT}^-$ is a subtheory of $(\mathrm{W})\mathrm{KF}^-$ and the latter are well known to be model-theoretically conservative over PA (see [1]; we will outline direct proof of model-theoretical conservativity of PT^- in Section 3), we have

Proposition 7. PT^- *and* WPT^- *are model-theoretically conservative.*

In particular we see that the axiom (NEG) may contribute to the strength of truth theories: it is easy to see that $(W)\mathrm{PT}^- + (\mathrm{NEG})$ is deductively equivalent to the theory CT^-, hence in particular by the well-known theorem of Lachlan (see [10],[14]) it is not semantically conservative.

For the sake of convenience let us isolate one easily noticeable feature of PT^- and WPT^-:

Definition 8 (UTB). Let $\phi(x_0, \ldots, x_n)$ be any arithmetical formula. $\mathrm{UTB}^-(\phi)$ is the following $\mathcal{L}_T$ sentence

$$\forall t_0 \ldots t_n \; \big(T(\ulcorner \phi(t_0, \ldots, t_n) \urcorner) \equiv \phi(t_0{}^\circ, \ldots, t_n{}^\circ)\big) \qquad (\mathrm{UTB}^-(\phi))$$

Define

$$\mathrm{UTB}^- := \{\mathrm{UTB}^-(\phi(x_0, \ldots, x_n)) \mid \phi(x_0, \ldots, x_n) \in \mathcal{L}_{\mathrm{PA}}\}$$

And UTB to be the extensions of UTB^- with all instantiations of induction scheme with $\mathcal{L}_T$ formulae.

---

[7]For the definition of all mentioned theories not defined in this paper, consult [10] or [9] (for WKF).

Fact 9. *Both* PT⁻ *and* WPT⁻ *prove* UTB⁻.

In [6],[7] and [8] (this last philosophical motivation was summarized also in [4]) authors motivated the need for a weak theory of truth which would be able to prove in a single sentence the fact that every arithmetical formula satisfy the induction scheme. Such a fact can be naturally expressed by an $\mathcal{L}_T$ sentence

$$\forall \phi(x) \ \left[\big(\forall x\big(T(\phi(\underline{x})) \to T(\phi(\underline{x+1}))\big)\big) \longrightarrow \big(T(\phi(0)) \to \forall x T(\phi(\underline{x}))\big)\right] \quad \text{(INT)}$$

For further usage let us abbreviate the formula

$$\big(\forall x\big(T(\phi(\underline{x})) \to T(\phi(\underline{x+1}))\big)\big) \longrightarrow \big(T(\phi(0)) \to \forall x T(\phi(\underline{x}))\big)$$

by $\text{INT}(\phi(x))$. Using Fact 9, we see that both PT⁻ + INT and WPT⁻ + INT can prove any arithmetical instance of the induction schema in a uniform way, for each formula using the same finitely many axioms[8]. In particular, it can be finitely axiomatised by taking $I\Sigma_1$ together with axioms for the truth predicate from PT⁻ and (INT). To achieve this goal, however, none of the discussed theories uses the full strength of (INT). By UTB⁻(ϕ) every standard formula is total, provably in WPT⁻. Hence it makes good sense to consider a version of (INT) restricted to total formulae, i.e.

$$\forall \phi(v) \ \left[\textbf{tot}(\phi(v)) \longrightarrow \text{INT}(\phi(v))\right] \qquad \text{(INT} \upharpoonright_{\text{tot}})$$

The theory PT⁻ + INT $\upharpoonright_{\text{tot}}$ was claimed to be model-theoretically conservative in [5] (and then used in [6],[7] and [8] as such). However, as shown in [4], the proof of its conservativity contained an essential gap and no prime model of PA[9] admits an expansion to the model of PT⁻ + INT $\upharpoonright_{\text{tot}}$. Moreover, it was shown that every recursively saturated model of PA can be expanded to a model of this theory. In particular, PT⁻ + INT $\upharpoonright_{\text{tot}}$ is model-theoretically stronger than PT⁻ and weaker than UTB and CT⁻.

In the current study, we further approximate the class of models expandable to PT⁻ + INT $\upharpoonright_{\text{tot}}$ and compare the strength of UTB with the strength of PT⁻ + INT. Moreover we show that WPT⁻ + INT is model theoretically conservative and meets the requirements posed in [8]. Our results jointly with some well-known facts from the literature give the following picture of interdependencies between proof-theoretically conservative theories of truth:

---

[8]The proof is really easy: we fix $\phi(x)$ (with parameters), prove the instantiation of the UTB⁻(ϕ) scheme for ϕ and substitute $\phi(x)$ for $T(\phi(\underline{x}))$ in INT

[9]For the definition of all notions from the model theory of PA see [14].
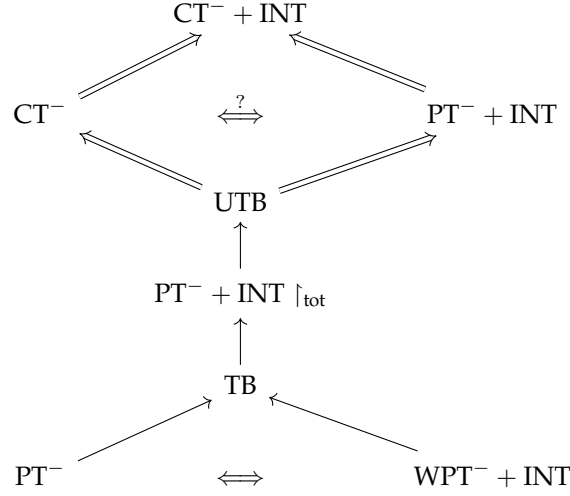
where $\longrightarrow$ stands for $\preceq_M$ and $\Longrightarrow$ for $\leq_M$. The question whether any of $\Longrightarrow$ arrows is in fact a $\longrightarrow$ arrow is open. Similarly, the relation between classes of models of CT^- and $PT^- + INT$ is unknown.

2 Models of $PT^- + INT \upharpoonright_{tot}$

In the paper [4], it has been shown that $PT^- + INT \upharpoonright_{tot}$ is not semantically conservative over PA and, moreover, any (not necessarily countable) recursively saturated model of PA admits an expansion to a model of $PT^- + INT \upharpoonright_{tot}$. The non-conservativity result has been obtained by demonstrating that no prime model of PA can be expanded in such a way. Now, we will show a strengthening of that result. Let us first recall one definition.

Definition 10. Let M be a model of PA. We say that M is **short recursively saturated** if any recursive type (with finitely many parametres from M) of the form $p(x) = \{x < a \wedge \phi_i(x) \mid i \in \mathbb{N}\}$ is realised in M where a is some fixed parameter from M.

In other words, a model is short recursively saturated if it realises all types which are finitely realised *below* some fixed element. This notion is strictly weaker than full recursive saturation. For example, the standard model $\mathbb{N}$ is short recursively saturated but not recursively saturated. More generally, a countable model is short recursively saturated if and only if it has a recursively saturated elementary end extension, see [18], Theorem 2.8.

Theorem 11. *Let $M \models$ PA and suppose that M has an expansion $(M, T) \models PT^- + INT \upharpoonright_{tot}$. Then M is short recursively saturated.*

The proof of our theorem will closely parallel the proof of Theorem 4 from [4]. In particular, we will again use a propositional construction invented by Smith.

Definition 12. Let $(\alpha_i)_{i \leq c}, (\beta_i)_{i \leq c}$ be any sequences of sentences. We define the **alternative with stopping condition** (α_i)

$$\bigvee_{i=i_0}^{c,\alpha} \beta_i$$

by backwards induction on i_0 as follows:

1. $\bigvee_{i=c}^{c,\alpha} \beta_i = \alpha_c \wedge \beta_c.$

2. $\bigvee_{i=k}^{c,\alpha} \beta_i = \neg(\alpha_k \wedge \neg\beta_k) \wedge \left((\alpha_k \wedge \beta_k) \vee \left(\neg\alpha_k \wedge \bigvee_{i=k+1}^{c,\alpha} \beta_i \right) \right).$

We may think that this is a formalisation in propositional logic of the following instruction: for i from i_0 up to c, search for the first number j such that α_j holds and then check whether also β_j holds. *Then stop your search.* The whole formula is true if this β_j is true and is false if either β_j is false or there is no j such that α_j holds. It turns out that this intuition may be partially recovered in theories of truth, even if one does not assume that the truth predicate satisfies induction axioms.

Lemma 13. *Fix* $(M, T) \models \mathrm{PT}^-$. *Suppose that* $(\alpha_i)_{i \leq c}$ *is a sequence of arithmetical sentences coded in* M. *Suppose that the least* j *such that* $T(\alpha_j)$ *holds is standard, say* $j = j_0$, *and that for any* $k \leq j_0$ *either* $T(\beta_k)$ *or* $T(\neg\beta_k)$ *holds. Then*

1. $(M, T) \models T\left(\bigvee_{i=c}^{c,\alpha} \beta_i \right) \equiv T(\beta_{j_0}).$

2. $(M, T) \models T\left(\neg \bigvee_{i=c}^{c,\alpha} \beta_i \right) \equiv T(\neg\beta_{j_0}).$

For a proof, see [4], Lemma 2.3. Now we are ready to prove that any model of PA expandable to a model of $\mathrm{PT}^- + \mathrm{INT} \restriction_{\mathrm{tot}}$ is short recursively saturated.

Proof. Fix any recursive type $p(x) = (\phi_i(x) \wedge x < a)_{i \in \omega}$ (with a parameter a) and suppose that for any finite set $\phi_0, \ldots, \phi_k$ there is some $b_k < a$ such that $M \models \phi_0(b_k) \wedge \ldots \wedge \phi_k(b_k)$. Let

$$
\begin{aligned}
\alpha_0(x) &= \neg x < \underline{a} \vee \neg\phi_0(x) \\
\alpha_{j+1}(x) &= x < \underline{a} \wedge \phi_0(x) \wedge \ldots \wedge \phi_j(x) \wedge \neg\phi_{j+1}(x).
\end{aligned}
$$

In a sense, formulae $\alpha_j(x)$ measure how much of the type p is realised by x. Now, if the type p is ommited in the model M, then for any x, there exists a standard j such that $(M, T) \models T\alpha_j(\underline{x})$. Let $\beta_j(y)$ be defined as

$$\beta_j(y) = y < \underline{a} \wedge \phi_0(y) \wedge \ldots \wedge \phi_{j+1}(y).$$

9

Now, fix any nonstandard c and consider the (nonstandard) formula

$$\phi(x, y) = \bigvee_{i=0}^{c, \alpha(x)} \beta_i(y).$$

By Lemma 13 and our assumption that the type p is omitted in M, the sentence $\phi(\underline{x}, \underline{y})$ is either true or false for any fixed $x, y \in M$. But this means that the formula $\phi(x, y)$ is total. One can check that then a formula

$$\psi(z) = \exists y \forall x < z \phi(x, y)$$

is also total. Note that this formula intuitively says that there is a y which satisfies more of a type p than any of the elements of M up to z. Now, we will show that $\psi(z)$ is progressive, i.e.,

$$(M, T) \models \forall z \Big(T\psi(\underline{z}) \to T\psi(\underline{z+1}) \Big).$$

Fix any z and suppose that $T\psi(\underline{z})$ holds. Then there exists a y such that $T(\ulcorner \forall x < z\ \phi(x, \underline{y}) \urcorner)$. Now, let j be the least number such that $T\alpha_j(\underline{z})$. Since j is a standard number and p is a type, there exists y' such that $\phi_0(y') \wedge \ldots \wedge \phi_{j+1}(y')$ holds in M, i.e. $(M, T) \models T\beta_j(\underline{y'})$. Let $y'' = y$ if also

$$T\Big(\phi_0(\underline{y}) \wedge \ldots \wedge \phi_{j+1}(\underline{y}) \Big)$$

and $y'' = y'$ otherwise. In other hand, we fix either y or y', whichever satisfies "more" formulae ϕ_i. One readily checks that then

$$(M, T) \models \forall x < z + 1\ \phi(\underline{x}, \underline{y''}).$$

We have shown that the formula $\psi(z)$ is total and progressive. By the internal induction for total formulae this means that

$$(M, T) \models \forall z\ T\psi(\underline{z}).$$

In particular, we have $T\psi(\underline{a})$, where a is the parameter used as a bound in the type p. But then for some d, we have

$$(M, T) \models \forall x < a\ T\phi(\underline{x}, \underline{d}).$$

Now, since p is a type, for an arbitrary $k \in \omega$, there exists some $x < d$ such that $\phi_0(x) \wedge \ldots \wedge \phi_k(x)$. Since $(M, T) \models T\phi(\underline{x}, \underline{d})$, it follows that $d < a \wedge \phi_0(d) \wedge \ldots \wedge \phi_{k+1}(d)$. As we have chosen an arbitrary k, we see that actually d satisfies the type p. We conclude that M is short recursively saturated. $\square$

Let us summarize our findings from [4] and this paper:

- Any recursively saturated model of PA (possibly uncountable) admits an expansion to a model of $PT^- + INT \restriction_{tot}$.

- If a model M expands to a model of $PT^- + INT \restriction_{tot}$, then it is short recursively saturated.

Unfortunately, we do not know whether any of the implications reverses.

Cieśliński and Engström have (independetly) found the following characterisation the class of models of PA which admit an expansion to a model of TB, i.e., the truth theory axiomatised with the induction scheme for the whole language and the following scheme of Tarski's biconditionals:

$$T(\ulcorner\phi\urcorner) \equiv \phi,$$

where ϕ is an arithmetical sentence.

Theorem 14 (Cieśliński, Engström). [10] *Let M be a nonstandard model of* PA. *Then the following are equivalent:*

1. *M admits an expansion to a model $(M, T) \models$ TB.*

2. *There exists an element $c \in M$ such that for all (standard) arithmetical sentences ϕ, $M \models \ulcorner\phi\urcorner \in c$ iff $M \models \phi$, i.e., M codes its own theory.*

It can be easily shown that every nonstandard short recursively saturated model $M \models$ PA satisfies the second item of the above characterisation. Hence, every short recursively saturated model of PA admits an expansion to a model of TB. Thus we obtain the following corollary:

Corollary 15. TB $\leq_M PT^- + INT \restriction_{tot}$, *i.e., every model $M \models$ PA which admits an expansion to a model of* $PT^- + INT \restriction_{tot}$, *also admits an expansion to a model of* TB.

2.1 A non-result

We are going to show that the method used to prove that every recursively saturated model of PA admits an expansion to a model of $PT^- + INT \restriction_{tot}$ cannot be used to obtain a stricter upper bound on the class of models expandable to this theory (if such a stricter upper bound exists). Let us recall that in [4], it was shown that every recursively saturated model of PA can be expanded to a model of $PT^- + INT \restriction_{tot}$. Let us recall the standard definition of a function which generates possible extensions for the PT^- truth predicate.

---

[10] See [3], Theorem 7.

Convention 2. If $\mathcal{M} \models$ PA, then $\text{Sent}_{\mathcal{M}}$, $\text{Form}^{\leq 1}_{\mathcal{M}}$, $\text{Term}_{\mathcal{M}}$ denote the set of sentences of $\mathcal{L}_{\text{PA}}$, the set of formulae with at most one free variable and the set of terms, respectively, *in the sense of* $\mathcal{M}$.

Definition 16. Let $\mathcal{M} \models$ PA and $A \subseteq \mathcal{M}$. Define:

$$
\begin{aligned}
\Theta_{\mathcal{M}}(\phi, A) \quad := \quad & \mathcal{M} \models \exists s, t \; [\phi = (s = t) \wedge s^\circ = t^\circ] \\
\vee \quad & \mathcal{M} \models \exists s, t \; [\phi = \neg(s = t) \wedge s^\circ \neq t^\circ] \\
\vee \quad & \exists \psi \in \text{Sent}_{\mathcal{M}}[\mathcal{M} \models \phi = \neg\neg\psi] \wedge \psi \in A \\
\vee \quad & \exists \psi_1, \psi_2 \in \text{Sent}_{\mathcal{M}}[\mathcal{M} \models \phi = (\psi_1 \vee \psi_2)] \wedge \left((\psi_1 \in A) \vee (\psi_2 \in A)\right) \\
\vee \quad & \exists \psi_1, \psi_2 \in \text{Sent}_{\mathcal{M}}[\mathcal{M} \models \phi = \neg(\psi_1 \vee \psi_2)] \wedge \left((\neg\psi_1 \in A) \wedge (\neg\psi_2 \in A)\right) \\
\vee \quad & \exists \psi(x) \in \text{Form}^1_{\mathcal{M}}[\mathcal{M} \models \phi = \exists x \psi] \wedge \exists x \in \mathcal{M} \; (\psi(\underline{x}) \in A) \\
\vee \quad & \exists \psi(x) \in \text{Form}^1_{\mathcal{M}}[\mathcal{M} \models \phi = \neg\exists x \psi] \wedge \forall x \in \mathcal{M} \; (\neg\psi(\underline{x}) \in A)
\end{aligned}
$$

Let $\Gamma^{\mathcal{M}} : \mathcal{P}(M) \to \mathcal{P}(M)$ be the function defined:

$$\Gamma^{\mathcal{M}}(A) = \{\phi \in \mathcal{M} \mid \Theta_{\mathcal{M}}(\phi, A)\}. \tag{Γ}$$

Let us now define:

$$
\begin{aligned}
\Gamma^{\mathcal{M}}_0 \quad &= \quad \Gamma^{\mathcal{M}}(\emptyset) \\
\Gamma^{\mathcal{M}}_{\alpha+1} \quad &= \quad \Gamma^{\mathcal{M}}(\Gamma^{\mathcal{M}}_\alpha) \\
\Gamma^{\mathcal{M}}_\beta \quad &= \quad \bigcup_{\alpha < \beta} \Gamma^{\mathcal{M}}_\alpha, \text{ for } \beta \text{ a limit ordinal.}
\end{aligned}
$$

It can be checked that for some ordinal α we must get $\Gamma^{\mathcal{M}}_{\alpha+1} = \Gamma^{\mathcal{M}}_\alpha$, i.e., $\Gamma^{\mathcal{M}}_\alpha$ is a fixpoint of $\Gamma^{\mathcal{M}}$. In general, if A is any fixpoint of $\Gamma^{\mathcal{M}}$, then

$$(\mathcal{M}, A) \models \text{PT}^-.$$

Let $\alpha_{\mathcal{M}}$ denote the least ordinal α such that $\Gamma^{\mathcal{M}}_\alpha$ is a fixpoint of $\Gamma^{\mathcal{M}}$.

In [4], the following lemmata were proved:

Lemma 17. *If* $\mathcal{M} \models$ PA *is recursively saturated, then* $\alpha_{\mathcal{M}} = \omega$.

Lemma 18. *If* $\mathcal{M} \models$ PA *and* $\alpha_{\mathcal{M}} = \omega$, *then* $(\mathcal{M}, \Gamma^{\mathcal{M}}_\omega) \models \text{PT}^- + \text{INT} \restriction_{\text{tot}}$.

Now we shall show that the converse to 17 holds. In particular, our method of finding extensions for $\text{PT}^- + \text{INT} \restriction_{\text{tot}}$ works only for recursively saturated models.

Lemma 19. *For every non-standard* $\mathcal{M} \models$ PA. *If* $\alpha_{\mathcal{M}} = \omega$, *then* $\mathcal{M}$ *is recursively saturated.*

Proof. We prove the contraposition: suppose that a non-standard model $\mathcal{M}$ is not recursively saturated. Let $p(x)$ be a recursive type using parameters from $\bar{a}$ which is omitted in $\mathcal{M}$. Let $(\phi_i(x, \bar{y}))_i$ be an arithmetically representable enumeration of formulae in $p(x)$. Without loss of generality, assume that $\phi_0(x, \bar{y}) = (x = x)$. Let

$$\psi_i(x, \bar{y}) = \bigwedge_{j<i} \phi_j(x, \bar{y}) \wedge \neg \phi_i(x, \bar{y})$$

Then every $b \in M$ satisfies exactly one of $\psi_i(x, \bar{a})$ (since $p(x)$ is omitted). Now, for every $n \in \omega$ we shall define formulae $\theta_n(x)$ as follows:

$$
\begin{aligned}
\theta_n^0(x) &= (x \neq x) \\
\theta_n^{k+1}(x, \bar{y}) &= \psi_{n-(k+1)}(x, \bar{y}) \vee \theta_n^k(x, \bar{y}) \\
\theta_n(x, \bar{y}) &= \theta_n^n(x, \bar{y})
\end{aligned}
$$

Let us observe that the above construction can be arithmetized and therefore for some $b \in M \setminus \mathbb{N}$ there exists a (code of a) formula $\theta_b(x, \bar{y})$, which is of the following form:

$$(\psi_0(x, \bar{y}) \vee (\psi_1(x, \bar{y}) \vee (\psi_2(x, \bar{y}) \vee (\ldots (\psi_{b-1}(x, \bar{y}) \vee x \neq x) \ldots)$$

Then for each $c \in M$, there exists $n \in \omega$ such that $\theta_b(\underline{c}, \underline{\bar{a}}) \in \Gamma_n^{\mathcal{M}}$, since each c satisfies some $\psi_i(x, \bar{a})$. But also for every $i \in \omega$, there exists $c \in M$ such that the least n for which $\psi_n(c, \bar{a})$ is greater than i. Consequently, there is no $k \in \omega$ for which

$$\theta_b(c, \underline{\bar{a}}) \in \Gamma_k^{\mathcal{M}}$$

for every $c \in M$. In particular, $\forall v \theta(v, \underline{\bar{a}}) \notin \Gamma_\omega^{\mathcal{M}}$ and consequently the PT^- axiom

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

is not satisfied in $(\mathcal{M}, \Gamma_\omega^{\mathcal{M}})$. Hence $\alpha_{\mathcal{M}} \neq \omega$. $\qquad \square$

3   Models of $\mathrm{PT}^- + \mathrm{INT}$

Since we have shown that $\mathrm{PT}^- + \mathrm{INT} \upharpoonright_{\text{tot}}$ is not a model-theoretically weak theory, as was originally hoped, one could start wondering whether it differs in some significant respect from $\mathrm{PT}^- + \mathrm{INT}$. In this section, we will show that actually this is the case. Namely, it turns out that $\mathrm{PT}^- + \mathrm{INT}$ is still model-theoretically stronger than $\mathrm{PT}^- + \mathrm{INT} \upharpoonright_{\text{tot}}$. As we shall see, any model of PA expandable to a model of $\mathrm{PT}^- + \mathrm{INT}$, is also expandable to a model of UTB. We know that any model of PA expandable to a model of UTB is recursively saturated and that this containment is strict, i.e., not every recursively saturated

13

model of PA admits an expansion to a model of UTB.[11] On the other hand, it has been shown in [4], Theorem 3.3 that any recursively saturated model of PA admits an expansion to a model of $\mathrm{PT}^- + \mathrm{INT} \restriction_{\mathrm{tot}}$.

Theorem 20. *Suppose that (M, T) is a model of $\mathrm{PT}^- + \mathrm{INT}$. Then there exists a T' such that $(M, T') \models \mathrm{UTB}$.*

Proof. Let $(M, T) \models \mathrm{PT}^- + \mathrm{INT}$. We will find T' such that $(M, T') \models \mathrm{UTB}$. Without loss of generality we may assume that M is nonstandard. As in the previous section, we will use Lemma 13. Let us fix any primitive recursive enumeration $(\phi_i)_{i=0}^\infty$ of arithmetical formulae. Then let

$$\alpha_i'(\phi, t)$$

be defined as the (formalised version of the) formula "t is a (finite) sequence of terms $(t_1, \ldots, t_n)$ and $\phi = \phi_i(t_1, \ldots, t_n)$" and let

$$\alpha_i(\phi, t, b) = \alpha_i'(\phi, t) \vee \underline{i} > b.$$

Let

$$\beta_i'(t)$$

be defined as "t is a (finite) sequence of terms $t_1, \ldots, t_n$ and $\phi_i(t_1, \ldots, t_n)$." Let

$$\beta_i(t, b)$$

be $\beta_i'(t) \wedge \underline{i} \leq b$. Note that ϕ is not a free variable of the formula β_i. Let us fix any nonstandard $c \in M$ and let

$$\tau(\phi, t, b) = \bigvee_{i=0}^{\alpha(\phi, t, b), c} \beta_i(t).$$

Note that for any standard c the predicate τ is equivalent to the very simple arithmetical truth predicate:

$$\tau_n(\phi, t, b) = \bigvee_{i=0}^{n} \phi = \phi_i(t) \wedge \phi_i(t) \wedge \underline{i} < b.$$

At this point one may wonder, what is the role of the variable b. It is indeed technical. We artificially truncate our truth predicates so that they work only

---

[11]We know that there exist rather classless recursively saturated models of PA, i.e., recursively saturated models $M \models \mathrm{PA}$ with the following property: for every $X \subseteq M$ such that every initial segment of X is coded M, the set X is definable in M with an arithmetical formula (with parametres). Since no subset of M definable with an arithmetical formula can satisfy UTB^-, we see that no such model M can admit an expansion to a model of UTB. The existence of recursively saturated, rather classless models has been shown by Kaufmann [13] under an additional set-theoretic assumption $\diamond$. The assumption has been dropped by Shelah, [17], Application C, p. 74.

for the first b formulae. This is to some extent controlled by the parameter c in the definition of τ, since whenever c is standard, the formula τ works like a truth predicate only for the first c sentences. However, c is not a variable in the formula τ, but rather a parameter describing the syntactic shape of τ, whereas we need this truncation to be expressed with a variable for reasons which will shortly become clear.

It turns out that for some parameter b the formula given by

$$T'(\phi) = \exists t\, T(\tau(\underline{\phi}, \underline{t}, \underline{b}))$$

satisfies UTB. We will prove this claim in a series of lemmata. This will obviously conclude our proof. $\qquad\square$

Lemma 21. *Let $\tau'(\phi, t) = \tau(\phi, t, b)$ for some fixed nonstandard b. Then for an arbitrary standard arithmetical formula $\phi(v_1, \dots, v_n)$ and an arbitrary sequence of terms $t = (t_1, \dots, t_n)$, possibly nonstandard (the length of the sequence is assumed to be standard).*

$$(M, T) \models T\tau'(\underline{\phi(t_1, \dots, t_n)}, \underline{t}) \equiv \phi(t_1, \dots, t_n).$$

Proof. If ϕ is standard, then $\phi = \phi_i$ for some standard i. So by Lemma 13

$$(M, T) \models T\tau'(\underline{\phi(t_1, \dots, t_n)}, \underline{t}) \equiv \beta_i(t) \equiv \phi_i(t_1, \dots, t_n),$$

which is exactly the claim of the lemma. $\qquad\square$

Note that the above lemma is true in pure PT^-. We have used no induction at all. Now we only need to check that for some parameter b the predicate $T'(\phi, t)$ defined as $T\tau(\underline{\phi}, \underline{t}, \underline{b})$ is fully inductive.

Lemma 22. *Let T' be defined as in the above proof. Then for some b, the formula $\tau'(\phi, t) = \tau(\phi, t, \underline{b})$ is total and consistent i.e. for all ϕ and t, exactly one of $T\tau'(\underline{\phi}, \underline{t})$, $T\neg\tau'(\underline{\phi}, \underline{t})$ holds.*

Proof. Note first that for any standard b, the formula $\tau(\phi, t, \underline{b})$ is total and consistent. Namely, since $\alpha_i(\phi, t, n)$ is true for any $i > n$, we see that for any ϕ, t the least i such that $\alpha_i(\phi, t, n)$ holds is standard (it is at most $n + 1$) and then the assumptions of Lemma 13 are satisfied. This implies that for any fixed ξ the formula $T\tau(\underline{\xi}, \underline{t}, \underline{n})$ is equivalent to some $\phi_i(t^\circ) \wedge \underline{i} \leq \underline{n}$, which is a standard formula. This implies that for any t, exactly one of $T\tau(\underline{\xi}, \underline{t}, \underline{n}), T\neg\tau(\underline{\xi}, \underline{t}, \underline{n})$ holds.

Now, consider the formula

$$\psi(b) = \forall \phi, t\, \Big(\tau(\phi, t, b) \vee \neg\tau(\phi, t, b) \Big).$$

We have just shown that for an arbitrary standard n we have $T \forall b < \underline{n}\, \psi(b)$. So by internal induction we have for some nonstandard d_1

$$T\left(\forall b \leq \underline{d_1} \forall \phi, t\, \Big(\tau(\phi, t, b) \vee \neg\tau(\phi, t, b) \Big) \right),$$

which gives

$$\forall b \leq d_1 \forall \phi, t\, \Big(T\tau(\underline{\phi}, \underline{t}, \underline{b}) \vee T\neg\tau(\underline{\phi}, \underline{t}, \underline{b}) \Big)$$

Similarly, let

$$\xi(b) = \exists \phi, t\, \Big(\tau(\phi, t, b) \wedge \neg\tau(\phi, t, b) \Big).$$

Suppose that $T(\exists d < \underline{b}\, \xi(d))$ holds for any nonstandard $b < d_1$. Then by underspill we would have $T\xi(\underline{n})$ for some $n \in \omega$. But we have just shown that this is impossible. So there exists some nonstandard $b < d_1$ such that for any $d \leq b$ and any ϕ, t at most one of $T\tau(\underline{\phi}, \underline{t}, \underline{d}), T\neg\tau(\underline{\phi}, \underline{t}, \underline{d})$ holds. At the same time, we know that at least one of these formulae holds. So $\tau'(\phi, t) = \tau(\phi, t, b)$ is total and consistent. $\qquad \square$

We are very close to showing that we have defined a predicate satisfying full induction. Before we proceed, we have to introduce some new notation. Let η be any formula containing a unary predicate P not in the language of PT^- and let $\xi(v)$ be an arbitrary formula with one free variable. Then by $\eta[\xi/P]$ (or simply $\eta[\xi]$) we mean a formula resulting from substituting $\xi(v_i)$ for any instance of $P(v_i)$ in η. We assume that all the variables in η has been renamed so as to avoid clashes.

Let us give an example. Let $\eta(x, y) = P(x + y) \wedge \exists z\, (z = y \wedge P(z))$. Let $\xi(v) = (v > 0)$. Then

$$\eta[\xi] = x + y > 0 \wedge \exists z\, (z = y \wedge z > 0).$$

Now, basically, we would like to finish the proof in the following way. Let τ' be a total formula defined as in the above lemmata and let η be an arbitrary standard formula from the arithmetical language enlarged with a fresh unary predicate P. Then, applying compositional axioms a couple of times, we see that

$$T(\eta[\tau]) \equiv \eta[T\tau].$$

Let us call this principle the **generalised commutativity**. If this were true, then we could conclude our proof. Namely, by the internal induction principle, we know that

$$\big(\forall x \big(T(\eta[\tau](\underline{x})) \to T(\eta[\tau](\underline{x+1})) \big) \big) \longrightarrow \big(T(\eta[\tau](\underline{0})) \to \forall x T(\eta[\tau](\underline{x})) \big)$$

which, by generalised commutativity, would allow us to conclude that

$$\big(\forall x \big(\eta[T\tau](\underline{x}) \to (\eta[T\tau](\underline{x+1})) \big) \big) \longrightarrow \big(\eta[T\tau](\underline{0}) \to \forall x \eta[T\tau](\underline{x}) \big).$$

Since the choice of η was arbitrary, this precisely means that τ satisfies the full induction scheme.

The generalised commutativity principle in the form stated above does not even quite make sense, since we would have to apply the truth predicate to a formula containing free variables. Therefore, we have to restate it in somewhat more careful manner.

Definition 23. Fix a unary predicate P. Let η be an arbitrary formula from the language containing that predicate. We say that η is in **semirelational** form if the predicate P is applied only to variables rather than to arbitrary terms.

We may always assume that formulae we *use* are semirelational, since we may eliminate any occurrence of $P(t)$ for complex terms t, by replacing it with $\exists x \, (x = t \wedge P(x))$. This is expressed in the following lemma:

Lemma 24. *Any formula is equivalent in first-order-logic to a formula in semirelational form.*

Now we are ready to state generalised commutativity lemma in a proper manner.

Lemma 25. *Let $(M, T) \models \mathrm{PT}^-$. Let $T * \xi(x) = T(\xi(\underline{x}))$ for every x. Suppose that ξ is total and consistent. Let η be an arbitrary standard formula from the arithmetical language extended with a fresh unary predicate P. Then the formula $\eta[\xi]$ is total and consistent, and*

$$(M, T) \models \forall x_1, \ldots, x_n \big(T(\eta[\xi](\underline{x_1}, \ldots, \underline{x_n})) \equiv \eta[T * \xi](x_1, \ldots, x_n) \big).$$

The lemma generalises to the case, where the predicate P is not unary (i.e. ξ may have more than one variable). The proof may be easily adapted to cover this case. We will actually use the lemma for the case with P binary.

Proof. We prove both claims simultanously by induction on complexity of η. Suppose that η is an atomic formula. Then it is either of the shape $s = t$ for some standard arithmetical terms s,t, or of the form $P(x)$.

In the first case, $\eta[\xi] = \eta$, and the following equivalences hold:

$$
\begin{aligned}
T(s(\underline{x_1}, \ldots, \underline{x_n}) = t(\underline{x_1}, \ldots, \underline{x_n})) \quad &\equiv \quad s(\underline{x_1}, \ldots, \underline{x_n})^\circ = t(\underline{x_1}, \ldots, \underline{x_n})^\circ \\
&\equiv \quad s(x_1, \ldots, x_n) = t(x_1, \ldots, x_n) \\
&= \quad \eta[T * \xi](x_1, \ldots, x_n).
\end{aligned}
$$

If $\eta = P(x)$, then $\eta[\xi] = \xi$ and

$$T(\eta[\xi](\underline{x})) = T\xi(\underline{x}) = \eta[T * \xi](x).$$

17

So let prove the induction step. If η is a conjunction or disjunction, then the proof is straightforward (the fact that a conjunction or disjunction of sentences which are either true or false is itself either true or false is an easy application of the compositional axioms of PT$^-$). If $\eta = \neg\rho$, then we know by induction hypothesis that $\rho[\eta]$ is total and consistent. Then by the compositional axiom for double negation for the truth predicate, the formula $\neg\rho[\eta]$ is also total and consistent and the following equivalences hold:

$$
\begin{aligned}
T(\neg\rho[\eta](\underline{x_1}, \ldots, \underline{x_n})) &\equiv \neg T(\rho[\eta](\underline{x_1}, \ldots, \underline{x_n})) \\
&\equiv \neg\rho[T * \eta](x_1, \ldots, x_n).
\end{aligned}
$$

The induction step for quantifier axioms is also simple. Let us prove it for the existential quantifier. Suppose that $\eta = \exists x\, \rho(x, x_1, \ldots, x_n)$. Then

$$
\begin{aligned}
T(\exists x\, \rho[\xi](x, \underline{x_1}, \ldots, \underline{x_n})) &\equiv \exists x\, T(\rho[\xi](\underline{x}, \underline{x_1}, \ldots, \underline{x_n}) \\
&\equiv \exists x\, \rho[T * \xi](x, x_1 \ldots, x_n) \\
&= \eta[T * \xi](x_1, \ldots, x_n).
\end{aligned}
$$

The second equivalence follows by the induction hypothesis and the last equality by definition. So let us check that $\eta[\xi]$ is total and consistent. Suppose that $T(\exists x\rho[\xi](x, \underline{x_1}, \ldots, \underline{x_n}))$ does not hold. Then by compositional axioms for the truth predicate, there is no x such that

$$
T(\rho(\underline{x}, \underline{x_1}, \ldots, \underline{x_n})).
$$

By induction hypothesis, ρ is total and consistent, so for all x we must have

$$
T(\neg\rho[\xi](\underline{x}, \underline{x_1}, \ldots, \underline{x_n})).
$$

This entails, again by compositional clauses

$$
T(\neg\exists x\, \rho[\xi](x, \underline{x_1}, \ldots, \underline{x_n})).
$$

$\square$

Now we are ready to conclude the proof of our theorem.

Lemma 26. *Let (M, T) be any nonstandard model of* PT$^-$ *+INT. Suppose that $\tau'(\phi, t)$ satisfies the claim Lemma 22. Then the predicate $T'(\phi, t)$ defined as $T * \tau'(\phi, t)$ satisfies the full induction scheme.*

Proof. By internal induction principle, the following holds for an arbitrary standard η from the arithmetical language extended with one fresh unary predicate $P(v)$:

$$
\Big(\forall x\big(T(\eta[\tau](\underline{x})) \to T(\eta[\tau](\underline{x+1}))\big)\Big) \longrightarrow \Big(T(\eta[\tau](\underline{0})) \to \forall x T(\eta[\tau](\underline{x}))\Big),
$$

18

Since τ' is total, if we additionally assume that η is semirelational, we can reach the following conclusion by Lemma 25:

$$\Big(\forall x\big(\eta[T*\tau](x) \to (\eta[T*\tau](x+1))\big)\Big) \longrightarrow \Big(\eta[T*\tau](0) \to \forall x\eta[T*\tau](x)\Big).$$

Since η was an arbitrary semirelational formula and any formula is equivalent to a semirelational one, this shows that T' satisfies the full induction scheme. $\square$

The conclusion of the proof of Theorem 20. We have defined a formula $T*\tau'(\phi,t)$ which satisfies full induction scheme and such that for an arbitrary standard $\phi(v_1,\dots,v_n)$ and an arbitrary sequence of terms $(t_1,\dots,t_n)$ the following holds:

$$(M,T) \models T*\tau'(\phi(t_1,\dots,t_n),t) \equiv \phi(t_1,\dots,t_n).$$

Then the formula $T'(\phi)$ defined as

$$\exists t\, T*\tau'(\phi,t)$$

satisfies the uniform disquotation axioms of UTB as well as the full induction scheme. So it defines a predicate satisfying UTB in (M,T). $\square$

This model-theoretic result allows us to make some conclusions concerning relative definability of the introduced theories.

Corollary 27. *The theory* $\mathrm{PT}^- + \mathrm{INT}\restriction_{\mathrm{tot}}$ *does not relatively truth defines* $\mathrm{PT}^- + \mathrm{INT}$.

Proof. We have just checked that every model $(M,T) \models \mathrm{PT}^- + \mathrm{INT}\restriction_{\mathrm{tot}}$ may be expanded to a model of UTB. By Theorem TODO, there exist recursively saturated rather classless models which cannot be expanded to any model of UTB. On the other hand in [4], Theorem 3.3, it has been shown that any recursively saturated model can be expanded to a model of $\mathrm{PT}^- + \mathrm{INT}\restriction_{\mathrm{tot}}$. Thus, there exist models of $\mathrm{PT}^- + \mathrm{INT}\restriction_{\mathrm{tot}}$ which cannot be expanded to a model of $\mathrm{PT}^- + \mathrm{INT}$. This contradicts relative definability. $\square$

4   Weak and Expressive Theories of Truth

In [8], the authors searched for a theory of truth that would simultaneously satisfy two requirements:

1. It could model the use of truth in model theory;

2. It would witness the expressive function of the notion of truth.

The way to satisfy the former is to be model-theoretically conservative over PA. Being such, the theory would not discriminate among possible interpretations of our basis theory. The way to satisfy the latter is to allow for expressing "thoughts" which are not expressible in the basis theory. There are many ways in which a theory of truth can witness the expressive role of the notion of truth. To mention just two (for the rest of the examples the Reader should consult [8]): if a theory of truth is finitely axiomatizable, then it is more expressive (than PA) and if a theory of truth has non-elementary speed-up over PA, then it is more expressive (than PA). There is a canonical construction which produces a theory of truth satisfying both finite-axiomatizability and the speed-up desideratum: the theory has to be (at least partially) classically compositional and it has to prove that all standard instantiations of the induction scheme are true. Without aspiring to any sort of completeness, let us offer the following explication of both properties. We start with a useful definition:

Definition 28. Let $\mathrm{CC}(x)$ denote the disjunction of the following formulae

- $\exists s, t \ \big(x = (s = t) \wedge (T(x) \equiv s^\circ = t^\circ) \big)$

- $\exists \phi, \psi \ \big(x = \phi \vee \psi \wedge (T(x) \equiv T(\phi) \vee T(\psi)) \big)$

- $\exists \phi \ \big(x = \neg \phi \wedge (T(x) \equiv \neg T(\phi)) \big)$

- $\exists \phi \exists v \ \big(x = \exists v \phi \wedge (T(x) \equiv \exists y T(\phi(\underline{y}))) \big)$

- $\mathrm{Form}^{\leq 1}(x) \wedge \forall s, t \Big(s^\circ = t^\circ \rightarrow \big(T(x(s)) \equiv T(x(t)) \big) \Big)$

Informally, $\mathrm{CC}(x)$ says that x is a formula on which T behaves compositionally in the sense of classical first-order logic.

Definition 29. A truth theory Th is *partially classically compositional* if there exists a formula $D(y)$ such that Th proves the following sentences:

1. $\forall y(D(y) \rightarrow \forall x \leq y D(x))$;

2. $D(0) \wedge \forall y(D(y) \rightarrow D(y + 1))$;

3. $\forall y \big(D(y) \rightarrow \forall \phi(\mathrm{dp}(\phi) \leq y \rightarrow \mathrm{CC}(\phi)) \big)$;

where $\mathrm{dp}(\phi) \leq x$ denotes an arithmetical formula representing the (primitive recursive) relation "the depth of the syntactic tree of ϕ is at most x".

If a formula satisfies the first requirement, we say that it is *downward closed*. If a formula satisfies the second one, we say that it is *progressive*. If a formula $D(y)$ is both downward closed and progressive, we will say that it *defines an initial segment*. This is justified, since if $D(y)$ satisfies 1 and 2, then in each model $\mathcal{M} \models$

Th the set $\{a \in M \mid \mathcal{M} \models D(a)\}$ is an initial segment of the model. In fact, being *downward closed* is not a very restrictive condition: if $D(y)$ is progressive, then the formula

$$D'(x) := \forall y \leq x D(y)$$

defines an initial segment. (this corresponds to a model-theoretic fact that each *cut* can be *shortened* to an initial segment). The third condition says that if ϕ is not too complicated (i.e., its complexity belongs to the initial segment defined by D), then T behaves classically on ϕ.

Definition 30. Let $\mathrm{ind}(\phi(x))$ denote the instantiation of the induction scheme with $\phi(x)$, i.e., the universal closure of the following formula:

$$\forall x(\phi(x) \rightarrow \phi(x+1)) \longrightarrow \big(\phi(0) \rightarrow \forall x \phi(x)\big)$$

Following our conventions, we will use $\mathrm{ind}(\cdot)$ to denote an arithmetical formula representing the function which, given a Gödel code of a formula with at most one free variable, returns the Gödel code of the corresponding induction axiom.

Definition 31. A truth theory *proves the truth of induction* if there exists a formula $D(y)$ such that Th proves that $D(y)$ defines an initial segment and

$$\forall \phi(v) \ \Big(D(\phi(v)) \rightarrow T\big(\mathrm{ind}(\phi(v))\big) \Big). \qquad \qquad \text{(T(IND))}$$

We shall say that Th is finitely axiomatisable modulo PA if there is a sentence ϕ such that the logical consequences of Th are precisely the logical consequences of $\mathrm{PA} \cup \{\phi\}$. For example, CT^-, PT^- and WPT^- are finitely axiomatisable modulo PA.

Now we have the following theorem whose unique novelty rests on isolating the features that are usually used to prove the thesis for concrete theories of truth.

Theorem 32. *Assume that*

1. Th *is partially classically compositional and proves the truth of induction and*

2. Th *is finitely axiomatizable modulo* PA,

then Th *is finitely axiomatizable and it has super-exponential speed-up over* PA.

Sketch of the proof. Let $D_1(y)$ define an initial segment on which T is classically compositional. Let $D_2(y)$ define an initial segment on which Th proves the truth on induction. Then $D(y) := D_1(y) \wedge D_2(y)$ defines an initial segment on which

T is classically compositional and proves the truth of induction. Obviously, for every standard natural number n we have

$$\text{Th} \vdash D(\underline{n})$$

In particular, $\text{Th} \vdash \text{UTB}^-$, which for every concrete formula of standard complexity can be proved by external induction on the complexity of its subformulae. Now, for every standard formula $\phi(x_0, \ldots, x_n)$, we can prove $\text{ind}(\phi(\bar{x}))$ in Th in the following way:

1. prove that D defines an initial segment on which T is classically compositional;

2. prove that $D(\ulcorner\text{ind}(\phi(\bar{x}))\urcorner)$ and conclude $D(\ulcorner\phi(\bar{x})\urcorner)$;

3. prove T(IND);

4. using 2. and 3. conclude $T(\ulcorner\text{ind}(\phi(\bar{x}))\urcorner)$;

5. prove $\text{UTB}^-(\text{ind}(\phi(\bar{x})))$;

6. conclude $\text{ind}(\phi(\bar{x}))$.

Observe that, given 1., the proof of $D(\ulcorner\text{ind}(\phi(\bar{x}))\urcorner)$ can be constructed in pure First-Order Logic. Similarly, given 1., all we need to use in proving $\text{UTB}^-(\text{ind}(\phi(\bar{x})))$ are some basic syntactical facts provable in $I\Sigma_1$. Let ϕ be a sentence such that $\text{PA} \cup \{\phi\}$ is a finite-modulo-PA axiomatisation of Th. It follows that for some n, every proof of $\text{ind}(\phi(\bar{x}))$ can be given in $I\Sigma_n + \phi$, hence the theory

$$I\Sigma_n \cup \{\phi\}$$

is a finite axiomatization of Th. To prove that Th has super-exponential speed-up over PA, we show that there is a formula $D'(y)$ which provably in Th defines an initial segment and that

$$\text{Th} \vdash \forall y\big(D'(y) \rightarrow \text{Con}_{\text{PA}}(y)\big)$$

where $\text{Con}_{\text{PA}}(y)$ is a finitary statement of consistency of PA saying that there is no PA proof of $0 = 1$ which can be coded using less than y bits. For the details, see [6], Theorem 9. $\qquad\square$

In [6], it was shown that $\text{PT}^- + \text{INT} \restriction_{\text{tot}}$ satisfies the assumptions of the above theorem. However, as was shown in [4], this theory is not model-theoretically conservative over PA. We shall now show that the right theory to use is $\text{WPT}^- + \text{INT}$.

Proposition 33. WPT$^-$ *is partially classically compositional and* WPT$^-$+INT *proves the truth of induction.*

Proof. Let us define

$$D'(y) := \forall x \leq y \forall \phi \Big(\mathrm{dp}(\phi) \leq x \longrightarrow \big(T(\neg \phi) \equiv \neg T(\phi) \big) \Big).$$

Then it can be easily shown that $D'(y)$ provably in WPT$^-$ defines an initial segment on which T is classically compositional (that $D'(y)$ is progressive is assured by the compositional axioms of WPT$^-$). For convenience, let us define:[12]

$$\mathrm{GC}(x) := \mathrm{Form}(x) \wedge \big(T(\mathrm{ucl}(x)) \equiv \forall \sigma (\mathrm{Asn}(x, \sigma) \rightarrow T(x[\sigma])) \big)$$

where

1. $\mathrm{ucl}(\phi(\bar{x}))$ denotes the universal closure of $\phi(\bar{x})$;

2. $\mathrm{Asn}(\phi, \sigma)$ represents the relation "σ is an assignment for ϕ", i.e., σ is a function defined exactly on the free variables of ϕ;

3. $x[\sigma]$ denotes the result of simultaneous substitution of numerals naming numbers assigned by σ to the free variables of x.

Further define

$$D(y) := D'(y) \wedge \forall \phi(\bar{x}) \Big(|\mathrm{FV}(\phi(\bar{x}))| \leq y \longrightarrow \mathrm{GC}(\phi(\bar{x})) \Big)$$

where, , $|\mathrm{FV}(\phi\bar{x})| \leq x$ represents the relation "$\phi(\bar{x})$ contains at most x free variables". For the sake of definiteness, we assume that $\mathrm{ucl}(\phi)$ starts with a quantifier binding the variable with the least index among the free variables if ϕ. It can be easily seen that $D(y)$ is downward closed. Let us now show that $D(y)$ is progressive. We work in WPT$^-$. Fix arbitrary a and suppose that $D(a)$. Then $D'(a)$ and, as $D'(y)$ is progressive, we have also $D'(a+1)$. Let us fix an arbitrary formula ϕ with less than $a + 1$ free variables and let v be its free variable with the least index. Then the following are equivalent

1. $T(\mathrm{ucl}(\phi))$

2. $\forall x T(\mathrm{ucl}(\phi(\underline{x}/v)))$

3. $\forall x \forall \sigma \big(\mathrm{Asn}(\phi(\underline{x}/v), \sigma) \rightarrow T(\phi(\underline{x}/v)[\sigma]) \big)$

4. $\forall \sigma \big(\mathrm{Asn}(\phi, \sigma) \rightarrow T(\phi[\sigma]) \big).$

---

[12]GC stands for "generalised commutativity." GC(x) expresses that the truth predicate commutes with the whole block of universal quantifiers in the universal closure of x.

The equivalence between 1. and 2. is by the axiom for universal quantifier in WPT^-. The equivalence between 2. and 3. holds because $\phi(\underline{x}/v)$ has $\leq a$ free variables. The last equivalence holds because each assignment for ϕ consists of an assignment to v and an assignment to the free variables of $\phi(\underline{x}/v)$.

We show that $\mathrm{WPT}^- + \mathrm{INT}$ proves the truth of induction on $D(y)$. We work in $\mathrm{WPT}^- + \mathrm{INT}$. Let us observe that for each formula ϕ, we have $\mathrm{dp}(\phi) \leq \phi$ and $|\mathrm{FV}(\phi)| \leq \phi$. [13] Hence if $D(\phi)$, then $D'(\mathrm{dp}(\phi))$ and $D(|\mathrm{FV}(\phi)|)$. In particular, if $D(\phi)$ then T is classically compositional on subformulae of ϕ and T can successfully deal with the universal closure for ϕ. Let us fix an arbitrary formula $\phi(v, \bar{w})$ such that $D(\phi(v, \bar{w}))$. We have to show $T(\mathrm{ind}(\phi(v, \bar{w})))$, i.e.

$$T\left(\mathrm{ucl}\Big(\forall v(\phi(v) \to \phi(v+1)) \longrightarrow \big(\phi(0) \to \forall v \phi(v)\big)\Big)\right) \tag{1}$$

(we skip the reference to $\bar{w}$ and assume that they are bounded by the universal quantifiers from ucl). Since the formula

$$\big(\forall v(\phi(v) \to \phi(v+1)) \longrightarrow \big(\phi(0) \to \forall v \phi(v)\big)\big)$$

contains less free variables then $\phi(v)$, we know that (1) is equivalent to

$$\forall \sigma T\Big(\forall v(\phi(v)[\sigma] \to \phi(v+1)[\sigma]) \longrightarrow \big(\phi(0)[\sigma] \to \forall v \phi(v)[\sigma]\big)\Big). \tag{2}$$

Let us fix an arbitrary σ. Then $\phi(v)[\sigma]$ is a formula with at most free variable v. Let us abbreviate it with $\psi(v)$. Hence it is enough to show:

$$T\Big(\forall v(\psi(v) \to \psi(v+1)) \longrightarrow \big(\psi(0) \to \forall v \psi(v)\big)\Big). \tag{3}$$

Since $\mathrm{dp}(\psi(v)) = \mathrm{dp}(\phi(v))$ and the depth of (3) is equal to $\mathrm{dp}(\psi(v)) + 3$, then T is classically compositional on (3). Hence (3) is equivalent to

$$\forall x\big(T(\psi(\underline{x})) \to T(\psi(\underline{x+1}))\big) \longrightarrow \big(T(\psi(\underline{0})) \to \forall x T \psi(\underline{x})\big)$$

which follows by INT. Hence $\mathrm{WPT}^- + \mathrm{INT}$ proves the truth of induction on D. □

Hence $\mathrm{WPT}^- + \mathrm{INT}$ exemplifies the expressive role of truth. Let us observe that, as it contains no restriction on arithmetical formulae admissible in the axiom of internal induction, it is more natural than $\mathrm{PT}^- + \mathrm{INT} \restriction_{\mathrm{tot}}$.[14] $\mathrm{WPT}^- + \mathrm{INT}$ proves that all arithmetical formulae, and not only total, satisfy induction, which is clearly the idea behind PA. Let us show that despite having such an expressive axiom, it is a model-theoretically conservative theory of truth.

---

[13]Being precise, this is a property of our coding. But most natural codings surely have it.

[14]In [8] authors discuss this restriction in the context of PT^- and admit it is as a possible objection to their theory.

Theorem 34. $\mathrm{WPT}^- + \mathrm{INT}$ *is model-theoretically conservative over* PA.

Proof. Let $\mathcal{M} \models \mathrm{PA}$. For $b \in M$ let $b \in Tr'$ if and only if for some $t_0, \ldots, t_n$ such that

$$\mathcal{M} \models \bigwedge_{i \leq n} \mathrm{Term}(t_i)$$

and some (standard!) $\phi(x_0, \ldots, x_n) \in \mathcal{L}_{\mathrm{PA}}$

$$\mathcal{M} \models \big(b = \ulcorner \phi(t_0, \ldots, t_n) \urcorner \big) \wedge \phi(t_0{}^\circ, \ldots, t_n{}^\circ)$$

Let us observe that with such a definition, we have

$$(\mathcal{M}, Tr') \models \mathrm{UTB}^-$$

To become the appropriate interpretation of WPT^- truth predicate, Tr' requires only one small correction. Let $\sim_\alpha$ denote the arithmetical formula representing in PA the relation of two *sentences* being the same modulo renaming variables (α-conversion). Let us define

$$b \in Tr$$

if and only if for some $\psi \in Tr'$, $\mathcal{M} \models b \sim_\alpha \psi$. Now it can be easily shown that

$$(\mathcal{M}, Tr) \models \mathrm{WPT}^- + \mathrm{INT}.$$

Indeed, compositional axioms are satisfied, since for every $x \in M$ such that $\mathcal{M} \models \mathrm{Form}^{\leq 1}(x)$

$$(\mathcal{M}, Tr) \models \mathrm{tot}(x) \text{ if and only if for some } n \in \omega, \mathcal{M} \models \mathrm{dp}(x) \leq n \qquad (*)$$

and moreover $(\mathcal{M}, Tr) \models \mathrm{UTB}^-$. Hence in verifying compositional axioms we may use the fact that $\models$ is compositional. Let us check the axiom for $\vee$. Suppose $\phi = \psi \vee \theta$ and $(\mathcal{M}, Tr) \models T(\phi)$. Then there exists $\phi' \sim_\alpha \phi$ such that $T(\phi')$ and $\phi' = \ulcorner \phi''(t_0, \ldots, t_n) \urcorner$ for some standard $\mathcal{L}_{\mathrm{PA}}$ formula $\phi''(x_0, \ldots, x_n)$ and $t_0, \ldots, t_n$ terms in the sense of $\mathcal{M}$. If so, then $\phi' = \psi' \vee \theta'$ such that $\psi \sim_\alpha \psi'$ and $\theta \sim_\alpha \theta'$. Also ψ' and θ' are of the form $\psi''(t_0, \ldots, t_n)$ and $\theta''(t_0, \ldots, t_n)$, respectively. By UTB^-, we have

$$\mathcal{M} \models \psi''(t_0{}^\circ, \ldots, t_n{}^\circ) \vee \theta''(t_0{}^\circ, \ldots, t_n{}^\circ)$$

Without loss of generality, assume that $\mathcal{M} \models \psi''(t_0{}^\circ, \ldots, t_n{}^\circ)$. It means that $(\mathcal{M}, Tr) \models T(\psi)$ and consequently $(\mathcal{M}, Tr) \models T(\psi) \vee T(\theta)$. By $(*)$, we have

$$(\mathcal{M}, Tr) \models \mathrm{tot}(\psi) \wedge \mathrm{tot}(\theta) \wedge \big(T(\psi) \vee T(\theta) \big).$$

which completes the proof of one implication. Let us now assume that the above holds. Since we have $\mathrm{tot}(\psi)$ and $\mathrm{tot}(\phi)$, it follows that for some n,k,

$$\mathcal{M} \models \mathrm{dp}(\psi) \leq n \wedge \mathrm{dp}(\theta) \leq k.$$

In particular, $\mathrm{dp}(\phi) \leq \max\{n, k\} + 1$. Let us assume that $(\mathcal{M}, Tr) \models T(\psi)$. Let $\psi \sim_\alpha \psi'$ and $\theta \sim_\alpha \theta'$ be such that $(\mathcal{M}, Tr') \models T(\psi')$. Reasoning as previously, we conclude that $(\mathcal{M}, Tr') \models T(\psi' \vee \phi')$ and hence

$$(\mathcal{M}, Tr) \models T(\psi \vee \phi)$$

which completes the proof of the compositional axiom for $\vee$.

Let us now verify that $(\mathcal{M}, Tr) \models \mathrm{INT}$. Fix an arbitrary formula $\phi(x)$ in the sense of $\mathcal{M}$ and assume that

$$(\mathcal{M}, Tr) \models T(\phi(0)) \wedge \forall x \big(T(\phi(\underline{x})) \to T(\phi(\underline{x+1})) \big)$$

It follows that $\mathcal{M} \models \mathrm{dp}(\phi(x)) \leq n$ for some $n \in \omega$ and for some standard

$$\phi'_0(x, y_0, \ldots, y_k), \phi'_1(x, y_0, \ldots, y_k), \phi'_2(x, y_0, \ldots, y_k),$$

we have

$$\mathcal{M} \models \phi(x) \sim_\alpha \ulcorner \phi'_i(x, t_0, \ldots, t_k) \urcorner$$

for some terms in $t_0, \ldots, t_k$ in the sense of $\mathcal{M}$ and for $i \leq 2$. In particular, by UTB^- we have

$$\mathcal{M} \models \phi'_0(0, t_0{}^\circ, \ldots, t_k{}^\circ) \wedge \forall x \big(\phi'_1(x, t_0{}^\circ, \ldots, t_n{}^\circ) \to \phi'_2(x+1, t_0{}^\circ, \ldots, t_n{}^\circ) \big).$$

But as satisfiability in a model is closed under α-conversion and each two of ϕ'_0, ϕ'_1, ϕ'_2 are α-equivalent, we get that

$$\mathcal{M} \models \phi'_0(0, t_0{}^\circ, \ldots, t_k{}^\circ) \wedge \forall x \big(\phi'_0(x, t_0{}^\circ, \ldots, t_n{}^\circ) \to \phi'_0(x+1, t_0{}^\circ, \ldots, t_n{}^\circ) \big)$$

Hence, by induction in $\mathcal{M}$ we get

$$\mathcal{M} \models \forall x\, \phi'_0(x, t_0{}^\circ, \ldots, t_n{}^\circ)$$

which by UTB^- again gives us $(\mathcal{M}, Tr') \models \forall x T(\phi'_0(\underline{x}, t_0, \ldots, t_k))$. Hence also

$$(\mathcal{M}, Tr) \models \forall x\, T(\phi(\underline{x}, t_0, \ldots, t_k)),$$

which ends the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In order to find a theory satisfying the Fischer-Horsten criterion, we decided to switch the inner logic of the truth theory. It allowed to formulate a very natural theory of truth modelled after Weak Kleene Scheme. Is it possible to realise Fischer-Horsten desiderata using a compositional theory of truth extending PT^-? With the meaning we gave to the term "axiomatic theory of truth", we are not allowed to add more symbols to the language.[15] For the moment, we leave it as an open problem.

---

[15]Without this restriction the answer is trivial: simply take PT^- together with $(\mathrm{WPT}^- + \mathrm{INT})$ but formulated with a different truth predicate symbol.

References

[1] Andrea Cantini. A theory of formal truth as strong as ID_1. *The Journal of Symbolic Logic*, 55(1):244–259, 1990.

[2] Cezary Cieliski. The innocence of truth. *Dialectica*, 69(1):61–85, 2015.

[3] Cezary Cieśliński. Typed and untyped disquotational truth. In Theodora Achourioti, Henri Galinon, José Martínez Fernández, and Kentaro Fujimoto, editors, *Unifying the Philosophy of Truth*, pages 307–320. Springer-Verlag, 2015.

[4] Cezary Cieśliński, Mateusz Łełyk, and Bartosz Wcisło. Models of PT^- with internal induction for total formulae. *The Review of Symbolic Logic*, 10(1):187–202, 2017.

[5] Martin Fischer. Minimal truth and interpretability. *Review of Symbolic Logic*, 2:799–815, 2009.

[6] Martin Fischer. Truth and speed-up. *Review of Symbolic Logic*, 7:319–340, 2014.

[7] Martin Fischer. Deflationism and instrumentalism. In Kentaro Fujimoto, José Martínez Fernández, Henri Galinon, and Theodora Achourioti, editors, *Unifying the Philosophy of Truth*. Springer Netherlands, 2015.

[8] Martin Fischer and Leon Horsten. The expressive power of truth. *Review of Symbolic Logic*, 8:345–369, 2015.

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

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

[11] Volker Halbach and Graham E. Leigh. Axiomatic theories of truth. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2014 edition, 2014.

[12] Leon Horsten. The semantical paradoxes, the neutrality of truth and the neutrality of the minimalist theory of truth. In P. Cartois, editor, *The Many Problems of Realism (Studies in the General Philosophy of Science: Volume 3)*. Tilberg University Press, 1995.

[13] Matt Kaufmann. A rather classless model. *Proceedings of American Mathematical Society*, 62(2):330–333, 1977.

[14] Richard Kaye. *Models of Peano Arithmetic*. Oxford: Clarendon Press, 1991.

[15] Jeffrey Ketland. Deflationism and tarski's paradise. *Mind*, 108:69–94, 1999.

[16] Stewart Shapiro. Proof and truth: Through thick and thin. *Jornal of Philosophy*, 95(10):493–521, 1998.

[17] Saharon Shelah. Models with second order properties II. Trees with no undefined branches. *Annals of Mathematical Logic*, 14:73–87, 1978.

[18] Craig Smoryński. Recursively saturated nonstandard models of arithmetic. *The Journal of Symbolic Logic*, 46(2):259–286, 1981.