Properties of models of weak theories of truth

March 15, 2014

1 Introduction

Our paper concerns models of weak theories of truth. By a "theory of truth" we mean an extension of Peano Arithmetic (henceforth denoted by PA) with an additional unary predicate $T(x)$ with intended reading "x is (a Gödel code of) a true sentence". We call a theory of truth "weak" iff it is a *conservative* extension of PA. Theories of truth are of interest both because of the philosophical context in which they emerged and of insights into the structure of models of PA they provide us with. Let us briefly comment on both these issues.

Truth theories constitute a well established area of research in contemporary epistemology and logic (for a comprehensive introduction and reference see [5]). In particular weak theories of truth has been introduced as a coherent formal framework for an explication of some stances in the debate over the metaphysical status of the notion of truth, specifically so called deflationary theories of truth. Deflationism claims that:

1. the sentences of the form "ϕ is true" do not ascribe any actual property to the sentence ϕ,
2. the meaning of truth predicate is completely analysible in terms of Tarski's disquotation scheme.

The first claim is contemporarily (see [11], [3]) explicated in terms of conservativity of a theory of truth over a theory of syntax (the latter usually modelled as PA). Namely: the claim that the predicate "ϕ is true" does not express any actual property is rearticulated as a thesis that the correct theory of truth should be conservative over PA. This is precisely where the interest in weak theories of truth as defined in the following article comes from.

The second claim is usually explicated as a thesis that the notion of truth is axiomatisable by Tarski's scheme or, more precisely, some syntactic restriction thereof (e.g. Tarski's scheme restricted to arithmetical sentences, for a detailed discussion of this explication of the deflationary theory of truth see [8], for a discussion of both theses see again [5]). Formal theories of truth satisfying both above conditions such as TB an UTB, defined later on in this paper, are subject to investigation as deflationary theories *par excellence*.

Our research in the structure of models of weak theories of truth has been initially motivated by another possible interpretation of the first claim of the deflationary theory of truth, which can be traced back directly to Shapiro's paper [11]. This explication claims that the correct theory of truth should be *model-theoretically conservative* over PA, i.e. every model M of PA should admit an expansion to a model (M, T) of the deflationary theory of truth. Speaking a bit naïvely, if "ϕ is true" really doesn't express any genuine property, then its admissability should not impose any conditions of how the object domain look like, which can in turn be directly explicated as a model-theoretic conservativeness of the theory of truth.

As of this moment, we are highly sceptical towards adequacy of this explication in the debate on deflationism. Even if it is the case that some weak theory of truth is not model-theoretic conservative over PA, this is not a substantial objection to a deflationist, who *precisely* does not agree that Tarskian semantics provides a correct analysis of the relationship between the actual language and its object domain. It might happen however, that in course of the philosophical debate new arguments emerge for this stronger notion of conservativeness as an adequate explication. Model theoretic considerations might be also seen as a tool for fine-grained classification of the weak theories of truth, which strength cannot be measured by merely proof-theoretical considerations, since most important ones, UTB and CT^-, are incomparable. Namely: Th_1 could be deemed not stronger than Th_2 if all models of PA that admit an expansion to Th_2, admit also an expansion to Th_1.

In the following paper we actually prove for three most important weak theories of truth, i.e. TB, UTB and CT^-, the classes of models of PA ($\mathfrak{TB}$, $\mathfrak{UTB}$, $\mathfrak{CT}^-$) that admits an extension to the models of respective truth theory can be linearly ordered by inclusion, namely:

$$\mathfrak{PA} \supset \mathfrak{TB} \supset \mathfrak{RS} \supset \mathfrak{UTB} \supseteq \mathfrak{CT}^-,$$

where $\mathfrak{RS}$ denotes the class of recursively saturated models of PA. Note that $\supset$ means "strict inclusion". Note that we *do not assume* that models of PA we deal with are countable, which would make the right part of the above sequence trivially collapse, due to Barwise-Schlipf theorem.

Seen from the purely model-theoretical perspective weak theories of truth are handy tool for obtaining interesting results about the structure of models of PA. Most striking examples of their implementation include: easy proof of Smoryński-Stavi theorem, proof of the existence of recursively saturated rather classless models of PA (in ZFC; both due to Schmerl see [10]) and the result that countable recursively saturated models of PA have recursively saturated end extensions. Although all these proofs had been established independently, weak theories of truth provided conceptual and uniform way of dealing with those complicated structures.

2 Notation and definitions

In this chapter we would like to introduce key definitions along with some notation. As for the latter we make a number of simplifications which strictly speaking might be ambiguous but on no reasonable reading may invoke any confusion. Considerations of fairly logical nature might be easily obscured by inappropriately heavy coding and putting too much stress on this aspect, which we tried to avoid.

Convention 2.1.

1. PA denotes Peano Arithmetic, and $\mathcal{L}_{PA}$ is the language in which it is formalized (for the sake of definiteness we assume that $\mathcal{L}_{PA} = \{\cdot, +, \leq, 0, 1\}$ where $\cdot$ and $+$ are two argument functions).

2. We use big capital letters M, $N \ldots$ for models of PA even if not stated explicitly.

3. We use $Form(x)$, $Sent(x)$ $Term(x)$, $ClTerm(x)$ to denote formulae defining sets of (Gödel codes of) respectively (arithmetical) formulae, sentences, terms and closed terms.

4. We skip Quine's corners when talking about Gödel codes of formulae, i.e. we write
$$\Phi(\psi)$$
instead of $\Phi(\ulcorner\psi\urcorner)$.

5. We will implicitly assume that variables $s, t, \ldots$ refer to (Gödel codes of) terms and $\phi, \psi, \ldots$ refer to (Gödel codes of) formulae. In particular we write
$$\forall t \quad \phi(t)$$
instead of
$$\forall x(ClTerms(x) \longrightarrow \phi(x)),$$
and we treat
$$\forall \psi \quad \Phi(\psi)$$
in the same fashion. Analogously for the existential quantifier.

6. We write t° to denote the result of formally evaluating the (Gödel code of) term t.

7. We sometimes write the result of syntactical operations with no mention of the operations themselves e.g.
$$\exists t \quad \Psi(\phi(t))$$
stands for
$$\exists t \quad \Psi(Subst(\phi, t))$$

where $Subst(x, y)$ is a formula representing substitution function and

$$\forall\psi \quad \Big(\exists\phi, \theta(\psi = \phi \wedge \theta) \rightarrow \Xi\Big)$$

stands for

$$\forall\psi \quad \Big(\exists\phi, \theta(\psi = Conj(\phi, \theta)) \rightarrow \Xi\Big).$$

where $Conj(x, y)$ stands for the formula representing function which takes two (Gödel codes of) formulae to (the Gödel code of) their conjunction.

8. If $T(x)$ is any predicate then by $Ind(T)$ we mean the set of all instantiations of induction scheme for all formulas in the language $\mathcal{L}_{PA} \cup \{T\}$. We will denote such language by $\mathcal{L}_T$.

We shall now introduce the theories which will be considered in this paper.

Definition 2.2. All the theories are formalized in the language $\mathcal{L}_T$ and are extensions of PA (below we list only additional axioms).

1. TB^- is a theory axiomatized by the scheme (called *Tarski Biconditional* scheme)

$$T\phi \equiv \phi,$$

   where ϕ is a sentence of $\mathcal{L}_{PA}$.

2. UTB^- is a theory axiomatized by the scheme (called *Uniform Tarski Biconditional* scheme)

$$\forall \bar{t}\big(T\phi(\bar{t}) \equiv \phi(\bar{t}^\circ)\big),$$

   where ϕ is a formula of $\mathcal{L}_{PA}$

3. CT^- is finitely axiomatized by the following sentences

   (a) $\forall t, s \quad \Big(T(R(t, s)) \equiv R(t^\circ, s^\circ)\Big)$ where R is $=$ or $\leq$.

   (b) $\forall\phi, \psi \quad \Big(T(\phi \otimes \psi) \equiv T(\phi) \otimes T(\psi)\Big)$, where $\otimes$ is $\wedge$ or $\vee$.

   (c) $\forall\phi \quad \Big(T(\neg\phi) \equiv \neg T(\phi)\Big)$.

   (d) $\forall\phi \quad \Big(T(Qx\phi(x)) \equiv QtT(\phi(t))\Big)$, where Q is $\exists$ or $\forall$.

4. TB, UTB, CT are the extensions of TB^-, UTB^-, CT^- with full induction for enriched language, i.e. $TB = TB^- \cup Ind(T)$, $UTB = UTB^- \cup Ind(T)$, $CT = CT^- \cup Ind(T)$.

Convention 2.3. As suggested by the examples of weak theories of truth in the Introduction, if T is any theory extending PA then by $\mathfrak{T}$ we denote the class of those models of PA which admits an extension to a model of T.

3 Easy or already known results

In this section we show how to prove almost all inclusions mentioned in the introduction. Let us begin with some trivial observations:

Proposition 3.1. $\mathfrak{TB} \supseteq \mathfrak{UTB}$

Proof. This follows immediately from the fact, that TB is a subtheory of UTB. □

Fact 3.2. $\mathfrak{RS} \supseteq \mathfrak{UTB}$

Proof. Fix any model $M \models UTB$ and a recursive type $p(x, \bar{a})$ with parameters $\bar{a} \subseteq M$. Let $\phi(x)$ represent $p(x, \bar{y})$ in PA^-. Since $p(x, \bar{a})$ is a type, for all $n \in \mathbb{N}$

$$M \models \exists c \forall \psi < n \ \Big(\phi(\psi) \longrightarrow T\psi(c, \bar{a}\Big).$$

Hence, by overspill, there is a nonstandard $b \in M$ and a $c \in M$ s.t.

$$M \models \forall \psi < b \ \Big(\phi(\psi) \longrightarrow T\psi(c, \bar{a}\Big).$$

Which proves the thesis. □

In proving the inclusion $\mathfrak{TB} \supset \mathfrak{RS}$, we will need the characterization of $\mathfrak{TB}$, which was observed independently by Fredrik Engström and Cezary Cieśliński.

Proposition 3.3. *$M \in \mathfrak{TB}$ if and only if the set*

$$Th_{\mathcal{L}_{PA}}(M) = \{\phi \in \mathbb{N} \ \mid \ \phi \in Sent_{\mathcal{L}_{PA}} \text{ and } M \models \phi\}$$

is coded in M.

Proof. Fix any model M.
(⇒) Observe that for all $n \in \mathbb{N}$,

$$(M, T) \models \exists x \forall < n \ \ (\phi \in x \equiv T\phi)$$

and use overspill to find a code of the theory of M.
(⇐) Take $T = \{a \in M \ \mid \ M \models a \in c\}$, where c is the code of $Th_{\mathcal{L}_{PA}}(M)$. □

Corollary 3.4. $\mathfrak{TB} \supseteq \mathfrak{RS}$

Proof. Suppose that M is recursively saturated and consider the following recursive type with a free variable x:

$$\{\phi \equiv \phi \in x \ \mid \ \phi \in Sent_{\mathcal{L}_{PA}}\}.$$

Any element of M realizing this type will be a code of the theory of M. Hence by Proposition 3.3 $\mathcal{M} \models TB$. □

Corollary 3.5. $\mathfrak{TB} \subset \mathfrak{PA}$.

Proof. Take any prime model K of a complete extension $Th \neq Th(\mathbb{N})$ of PA. Suppose that K admits an expansion to a model

$$(K, T) \models TB.$$

But then by Proposition 3.3 there would be an element $c \in K$ which codes the theory Th. Since all elements of K are arithmetically definable without parametres, then the formula

$$x \in c$$

would yield an arithmetical definition of truth for Th, contradicting Tarski's theorem. □

Now we proceed to the construction of a model which codes its theory and is not recursively saturated, in this way proving that the inclusion $\mathfrak{TB} \supseteq \mathfrak{RS}$ is strict. It shows up that this is an easy consequence of MacDowell-Specker theorem and the following lemma:

Lemma 3.6. *Suppose that there are $M \prec_{cons} M'$ i.e. M' is a proper, elementary and conservative supmodel of M. Then M' is not recursively saturated.*

Proof. Let M, M' be as in the formulation of the lemma and suppose that M' is recursively saturated. Pick $c \in M' \setminus M$ and let $b \in M'$ realize the following recursive type with free variable y:

$$\{\forall \bar{x} \ \Big(\phi(\bar{x}) < c \longrightarrow (\phi(\bar{x}) \equiv (\phi(\bar{x}) \in y)\Big) \ \mid \ \phi(\bar{z}) \in Form_{\mathcal{L}_{PA}}\}.$$

Consider the set

$$X = \{a \in M | M' \models a \in b\}.$$

Since the extension $M \prec M'$ is conservative, X should be definable with parameters from M. On the other hand note that by definition of b X is exactly the set of codes of formulae from elementary diargam of M, hence undefinable in M by Tarski's Theorem.

□

Theorem 3.7. $\mathfrak{TB} \supset \mathfrak{RS}$. *Moreover, every model M has an elementary extension to $(M', T) \models TB$ with M' not recursively saturated.*

Proof. We prove the "moreover" part which of course suffices. Let us fix any M. Let c be a fresh constant. By compactness the following theory

$$ElDiag(M) \cup \{\phi \in c \ \mid \ \phi \in Sent_{\mathcal{L}_{PA}} \wedge \mathcal{M} \models \phi\}$$

has a model M', which is an elementary extension of M. Note c is a code of the theory of $Th_{\mathcal{L}_{PA}}(M')$. Using MacDowell-Specker theorem we can find

$$M'' \succ_{cons} M'.$$

Since $Th(M') = Th(M'')$, we see that $c \in M''$ is a code of a theory of $\mathcal{M}''$. By Proposition 3.3 M'' can be expanded to a model $(M'', T) \models TB$. But by lemma 3.6 it cannot be recursively saturated. □

Let us now return to the inclusion $\mathfrak{RS} \supseteq \mathfrak{UTB}$. It is easy to show that for a counterexample to equality here we have to search among models with *uncountable cofinality.* By a result of Smorynski-Stavi [?] (restated in [10]) if T is an extension of PA in a language $\mathcal{L} \supseteq \mathcal{L}_{PA}$ such that T contains induction axioms for $Form_{\mathcal{L}}$, then T is preserved in cofinal supmodels, i.e.

$$\mathcal{M} \models T \wedge \mathcal{M} \prec_{cf} \mathcal{N} \Rightarrow \mathcal{N} \models T.$$

Putting it together with the fact that countable and recursively saturated models of PA are resplendent and that UTB is a conservative extension of PA we see that every recursively saturated model of PA with *countable cofinality* can be expanded to a model of UTB. In order to prove the existence of recursively saturated models which do not expand to a model of UTB we will profit from the fact that the interpretation of UTB-truth predicate, if exists, is always a proper class.

Observation 3.8. If $T \subseteq M$ is such that $(\mathcal{M}, T) \in \mathfrak{UTB}$ then T is a proper class on M. Indeed, T is a class, because UTB contains induction axioms for all formulas of enriched language and T is obviously undefinable by Tarski's theorem.

Recall that model $\mathcal{M}$ is *rather classless* if it contains no proper class. The existence of recursively saturated rather classless models of PA was first demonstrated by Matt Kauffman in $ZFC + \diamond$ ([?]). The assumption about existence of $\diamond$-sequence was later eliminated by Shelah (in [?]). We will present another argument, which was given in [10]. The proof of the following theorem can be found in appendix.

Theorem 3.9. *There is a recursively saturated and rather classless model of PA.*

□

4 Main Result

In the following part we will present the most technically involved part of our result i.e.

Theorem 4.1. *Let $(M, T) \models CT^-$. Then there exists T' such that $(M, T') \models UTB$.*

Before we proceed to the proof, let us introduce some notation.

Definition 4.2. Let $\delta(\bar{x})$ be arbitrary formula in the language $\mathcal{L}_{PAP}$ i.e. language of arithmetics with a unary predicate $P(x)$ added . Then for arbitrary formula ϕ with one free variable by $\delta[\phi]$ we mean a result of formally substituting the formula $\phi(x_i)$ for any occurence of $P(x_i)$ in the formula δ (possibly preceded by some fixed renaming of bounded variables in δ, so as to avoid clashes). If $M \models \delta[\phi]$ we will say that ϕ **satisfies** a property δ.

Definition 4.3. By $\phi[\xi \mapsto \delta]$ we mean a result of formally substituting a formula δ for a subformula ξ in a formula ϕ.

Let us quickly give an example, which probably could be more illuminating than a definition.

Example 4.4. Let $\delta(x,y) = \Big(P(x) \equiv P(y)\Big)$. Then

$$\delta[z = z] = \Big((x = x) \equiv (y = y)\Big).$$

Certainly both notions may be formalised in PA. We our now ready to state the main lemma. This is the combinatorial core of our theorem. Essentially it has been proved in [**?**], although in a special case. Basically, the lemma states that the existence of a truth predicate satisfying CT^- allows us to define a predicate satisfying UTB^- and some additional definable properties shared by arithmetical formulae.

Lemma 4.5. *Let δ be arbitrary formula in $\mathcal{L}_{PAP}$. Let $(M,T) \models CT^-$. Suppose that for arbitrary standard arithmetical formula ϕ we have*

$$(M,T) \models \delta[\phi].$$

Then there exists a formula $T'(x)$ in $\mathcal{L}_{PAT}$ with parametres such that

$$(M,T) \models T'\psi(t) \equiv \psi(t^\circ)$$

for arbitrary standard arithmetical formula ψ and arbitrary term $t \in Term(M)$ and moreover

$$(M,T) \models \delta[T'].$$

Proof. We will try to construct an analoge of simplistic arithmetical partial truth predicates

$$\tau(x) = \vee_{i=1}^{n}(x = \phi_i) \wedge \phi_i$$

but in such a way that we can use a form of overspill available in models of CT^-. Let $(\chi_i)_{i<\omega}$ be arbitrary primitive recursive enumeration of arithmetical formulae. Let

$$\begin{aligned}
\xi_{2i}(x) &= \forall \bar{t} \forall j \leq i \;\; x = \chi_j(\bar{t}) \longrightarrow \bigvee_{j \leq i} \Big(x = \chi_j(\bar{t}) \wedge \chi_j(\bar{t}^\circ)\Big) \\
\xi_{2i+1}(x) &= \exists \bar{t} \exists j \leq i \;\; x = \chi_j(\bar{t}).
\end{aligned}$$

Finally, let us define formulae which will play the role of the partial truth predicate τ above.

$$\rho_0 = \xi_0$$

$$\begin{aligned}\rho_{2i+1} &= \rho_{2i}[\xi_{2i} \mapsto \xi_{2i} \wedge \xi_{2i+1}] \\ \rho_{2i+2} &= \rho_{2i+1}[\xi_{2i+1} \mapsto \xi_{2i+1} \vee \xi_{2i+2}].\end{aligned}$$

Obviusly the definitions may be formalised. Here are two simple properties of so defined ρ_i's for arbitrary $i \in \omega$ and $a \in M$.

1. $T\rho_a(x) \rightarrow T\rho_{2i}(x)$.
2. $T\rho_{2i+1}(x) \rightarrow T\rho_a(x)$.

To prove the first one assume that $T\rho_a(x)$. Note that

$$\rho_a(x) = \rho_{2i}[\xi_{2i} \mapsto \xi_{2i} \wedge \gamma]$$

for some nonstandard formula γ. Since ρ_{2i} is a positive formula (i.e. no negation symbol occurs in it) a substitution of $\xi_{2i} \wedge \gamma$ for ξ_{2i} yields a formula stronger (no weaker) than ρ_{2i}. As it is purely a matter of finite boolean calculus, this can be proved in CT^-. Proof of the second one is analogous.

We are now in a position to show that for arbitrary nonstandard a formula $T\rho_a(x)$ satisfies uniform disquotation scheme i.e for arbitrary standard arithmetical ϕ and $t \in Term(M)$

$$T\rho_a\phi(t) \equiv \phi(t^\circ).$$

Let us take any $\phi(t)$. We know that $\phi(t) = \chi_i(t)$ for some $i \in \omega$. Suppose

$$T\chi_i(t^\circ).$$

Then it is easy to see that

1. $T\xi_{2i+1}(\chi_i(t))$
2. $\neg T\xi_{2j+1}(\chi_i(t))$, for $j < i$
3. $T\xi_{2j}(\chi_i(t))$ for arbitrary j.

It is enough to show that $T\rho_{2i+1}(\chi_i(t))$. To this end we will prove by external backwards induction that for any subformula ψ of ρ_{2i+1} righthandside part of ψ is true. By assumption the claim holds for

$$\psi = \xi_{2i}(\chi_i(t)) \wedge \xi_{2i+1}(\chi_i(t)).$$

Now any righthandside of any subformula of ρ_{2i+1} is exactly of one of the two following forms:

1. $\xi_{2j} \wedge \gamma$
2. $\xi_{2j+1} \vee \gamma$.

But by induction hypothesis we can assume that γ is true. Now, since the main connective in the formula ρ_{2i+1} is a conjunction of the form $\xi_0 \wedge \gamma$, we are able to prove in CT^- that

$$T\rho_{2i+1}(\chi_i(t)).$$

By previous observation it follows that

$$T\rho_a(\chi_i(t)).$$

The converse implication is handled in a similar fashion (but now using $T\rho_a(x) \rightarrow T\rho_{2i}(x)$).

Finally, we will prove the lemma. It will be similar to prove of Lachlan's theorem. Define the following sequence of formulas:

$$\begin{array}{rcl} \gamma_0(x) & = & (x = x) \\ \gamma_{i+1}(x) & = & \delta[\gamma_i] \longrightarrow \alpha_{i,i}(x) \\ \alpha_{i,0}(x) & = & \rho_{2i}(x) \\ \alpha_{i,j+1}(x) & = & \big(\forall t \;\; \gamma_i(\chi_{i-(j+1)}(t)) \equiv \chi_{i-(j+1)}(t^\circ)\big) \wedge \alpha_{i,j}(x) \\ & \vee & \neg\big(\forall t \;\; \gamma_i(\chi_{i-(j+1)}(t)) \equiv \chi_{i-(j+1)}(t^\circ)\big) \wedge \rho_{2(i-(j+1))}(x). \end{array}$$

Observe again that this definition may be formalised in PA. We will show that for some $b > \omega$ $T\gamma_b(x)$ satisfies both property δ and uniform disquotation scheme. In understanding the following part it helps to think of $\alpha_{i,j}$ as organised in lower-triangle matrix.

Observe first that if for some c it happens that $T\gamma_c$ doesn't satisfy the property δ, then $T\gamma_{c+1}$ does satisfy it, since it is then equivalent to standard formula.

Let n_c be the least $n < \omega$ such that

$$T\big(\neg\big(\forall t \;\; \gamma_c(\chi_n(t)) \equiv \chi_n(t^\circ)\big),$$

if such n exists and arbitrary nonstandard number otherwise.

Now let us make key observation. If $T\gamma_c$ does satisfy δ but does not satisfy disquotation scheme, i.e. $n_c < \omega$, then neither $T\gamma_{c+1}$ does but $n_{c+1} > n_c$. Indeed, suppose $n_c < \omega$. Then by definition

$$T\alpha_{c,c-n_c}(x) \equiv \rho_{2n_i}(x).$$

But then for $j \geq c - n_c$ we have

$$T\alpha_{c,j}(x) \equiv T\alpha_{c,j+1}(x).$$

In particular

$$T\alpha_{c,c}(x) \equiv T\alpha_{c,c-n_c}(x) \equiv \rho_{2n_c}(x).$$

Since $T\gamma_{c+1}$ is then equivalent to $\alpha_{c,c}$ i.e. to ρ_{2n_c} it follows that

$$n_{c+1} = n_c + 1 > n_c.$$

Now we will finish the proof of the lemma. Suppose that for no nonstandard c does $T\gamma_c$ satisfy both property δ and uniform disquotation. Then it follows that actually no γ_c satisfies uniform disquotation. Indeed: take any $c > \omega$ and observe that if γ_{c-1} doesn't have either of the two properties, then by our previous observation γ_c doesn'y satisfy scheme of uniform disquotation.

So if for all c our γ_c fails to enjoy one of desired properties, then for arbitrary γ_c we have $n_c < \omega$. But then

$$n_c > n_{c-1} > n_{c-2} > \dots$$

form an infinite decreasing sequence of natural numbers. It follows from this contradiction that for some nonstandard c we have both:

$$(M,T) \models \delta[T\gamma_c]$$

and

$$(M,T) \models T\gamma_c\phi(t) \equiv \phi(t^\circ)$$

for arbitrary standard ϕ and $t \in Term(M)$.

□

Now we are ready to prove our theorem. As a matter of fact, we shall obtain slightly stronger result. The predicate T we are going to construct will display additional property that

$$(M,T) \models UTB$$

will be recursively saturated *as a model of UTB*.

Proof. Let $(M,T) \models CT^-$ and let $\tilde{\theta}(y)$ be a formula '$P(x)$ is a compositional predicate for formulae $< y$' i. e. a conjuction of the following formulae:

1. $\forall \phi < y \quad P(\neg\phi) \equiv \neg P(\phi)$.
2. $\forall \phi, \psi < y \quad P(\phi \odot \psi) \equiv P(\phi) \odot P(\psi)$.
3. $\forall \phi < y \quad P(Qx\phi) \equiv Qt\ P(\phi(t^\circ)$,

where $\odot \in \{\wedge, \vee\}, Q \in \{\forall, \exists\}$.

Let θ be a sentence '$\tilde{\theta}$ is inductive' i.e.

$$\Big(\forall x\tilde{\theta}(x) \to \tilde{\theta}(x+1)\Big) \longrightarrow \Big(\tilde{\theta}(0) \to \forall x\tilde{\theta}(x).$$

Let now (ind_k) be some recursive enumeration of instances of the induction scheme in the arithmetical language with an additional predicate $P(x)$. For arbitrary formula ϕ let

$$\begin{aligned} *Ind_0(\phi) &= ind_0[\phi] \\ Ind_{i+1}(\phi) &= Ind_i(\phi)\Big[ind_i[\phi] \mapsto ind_i[\phi] \wedge ind_{i+1}[\phi]\Big]. \end{aligned} \quad (1)$$

*

So $Ind_k(\phi)$ is simply a conjunction of first k instances of the induction scheme for a formula ϕ with parentheses put in reverse order. Be not confused with the fact, that we used a predicate P to define $Ind(\phi)$. It does not occure in our formula anymore.

Let $\tilde{\zeta}(x)$ be defined as

$$P(Ind_x(\rho_x)).$$

So it is a formula saying 'x-th formula ρ_x satisfies first x instances of the induction scheme'. Let *zeta* be defined in an analogous fashion to *theta* i.e.

$$\zeta = \Big(\forall y \;\; \tilde{\zeta}(y) \to \tilde{\zeta}(y+1)\Big) \to \Big(\tilde{\zeta}(0) \to \forall y \;\; \tilde{\zeta}(y)\Big).$$

Let finally

$$\delta = \zeta \wedge \theta.$$

Observe that every standard formula ϕ has the property δ, since $\delta[\phi]$ is simply an instance of the induction scheme. So by our lemma there is a formula $T'(x)$ such that

$$(M, T) \models \delta[T']$$

and T' satisfies uniform disquotation scheme. Since it satisfies the scheme, it is compositional for standard formulae, i.e. for all $k \in \omega$.

$$(M, T) \models \tilde{\theta}(k)$$

So, by overspill we have

$$(M, T) \models \tilde{\theta}(c)$$

for some $c > \omega$.

Now, since ρ_k for $k \in \omega$ are standard formulae, they satisfy full induction scheme. In particular

$$(M, T) \models \tilde{\zeta}(k).$$

So applying overspill once more we get some nonstandard d such that

$$(M, T) \models T'\tilde{\zeta}(d).$$

W.l.o.g. we may assume that $d < c$.

We claim that

$$(M, T'(\rho_d(x))) \models UTB.$$

Since $d < c$ our predicate is compositional, so we may show that it satisfies disquotation scheme exactly as for CT^-. It is enough to show that it satisfies full induction scheme, that is

$$(M, T) \models ind_k(T'(\rho_d(x))$$

for arbitrary $k \in \omega$. By assumption we have

$$(M, T) \models T' Ind_d(\rho_d),$$

but since T' is compositional for formulae $< c$ and ind_k is located in the formula Ind_d on finite syntactic depth, we see that

$$(M, T) \models T' ind_k[\rho_d].$$

Which, by compositionality again, implies that

$$(M, T) \models ind_k[T' \rho_d].$$

So indeed $(M, T' \rho_d(x))$ is a model of UTB.

□

An inspection of the proof shows that the UTB predicate we have defined is of the form

$$T \gamma_a \rho_b(x),$$

for some nonstandard a, b. Thus by Theorem 3.1 of [?] th model we defined is recursively saturated.

5 Appendix

Proof of Schmerl's theorem makes use of some set theoretic notions which we shall now briefly recall.

Definition 5.1. Let α be a regular and uncountable cardinal.

- Subset $A \subseteq \alpha$ is a *club* iff A is closed and unbounded in α, i.e.
  1. for every subset Y of A, if $supY < \alpha$ then $supY \in A$ and
  2. $supA = \alpha$.

  The set of all clubs on α will be denoted by $Club(\alpha)$
- Subset $A \subseteq \alpha$ is *stationary* iff for every $d \in Club(\alpha)$, $A \cap d \neq \emptyset$.
- Function $f : \alpha \longrightarrow \alpha$ is regressive on a subset $A \subseteq \alpha$ iff for all $\beta \in A, \quad f(\beta) < \beta$.

For completeness we cite the following well-known theorem about regressive functions on stationary sets:

Theorem 5.2. *(Fodor's lemma)*
Let κ be an uncountable regular cardinal and $S \subseteq \kappa$ be stationary. Then for every function $f : \kappa \to \kappa$ regressive on S exists $T \subset S$ such that $f \restriction_T$ is constant and T is stationary.

□

We introduce also an enriched notion of satisfaction class.

Definition 5.3. Let $S(\cdot,\cdot)$ be a new binary predicate. We call an interpretation of $S(\cdot,\cdot)$ in a model $\mathcal{M}$ a *counting satisfaction class*[1] iff the following conditions holds for all $k \in \mathbb{N}$ and all ϕ,$\psi \in \Sigma_k$

1. $\forall t, s(S(t = s, k) \iff t^\circ = s^\circ)$;
2. $(S(\phi \otimes \psi, k) \iff S(\phi, x) \otimes S(\psi, k))$;
3. $S(\neg\phi, k) \iff \neg S(\phi, k)$;
4. $S(Qx\phi(x), k+1) \iff QtS(\phi(t), k)$;

where as usual $\otimes \in \{\wedge, \vee\}$ and $Q \in \{\forall, \exists\}$. If moreover for all $k \in \mathbb{N}$

$$S\restriction_k := \{< e, c > \in S \parallel c \leq k\} \neq \emptyset$$

we call such a class *proper*. If

$$(\mathcal{M}, S) \models Ind(S)$$

we call such a class inductive.

The proof of following two observations is standard (the second one can be done exactly in the same way as the proof of Proposition **??**):

Proposition 5.4. *For all $\mathcal{M}$ there is an $\mathcal{N}$ and $S \subset N^2$ such that*

$$\mathcal{M} \prec \mathcal{N}$$

and S is a proper, inductive counting satisfaction class.

Proposition 5.5. *If S is a proper and inductive counting satisfaction class in M, then M is recursively saturated.*

This lemma is one of the main reason for considering proper, nonstandard and counting satisfaction classes:

Lemma 5.6. *If S is a proper and inductive counting satisfaction class on $\mathcal{M}$ and for all nonstandard $b \in M$*

$$X \in Def(\mathcal{M}, S\restriction_b),$$

then $X \in Def(\mathcal{M})$

Proof. Observe that if X is definable in $(\mathcal{M}, S\restriction_b)$ by a Σ_{n+1} formula, then it is definable in $(\mathcal{M}, S\restriction_{b+1})$ by a Σ_n formula. It follows that for all nonstandard b X is definable in $(\mathcal{M}, S\restriction_b)$ by a Σ_1 formula. Hence, by underspill there is a standard n such that X is definable in $(\mathcal{M}, S\restriction_n)$. But for all *standard* k $S\restriction_k$ is arithmetically definable, so X is definable in $\mathcal{M}$. □

[1] We introduce this name for the use in this paper. Schmerl called it just "satisfaction class"

Now we can prove Schmerl theorem:

Theorem 5.7. *If κ is a singular cardinal with uncountable cardinality, then every consistent extension of PA has a κ-like, recursively saturated and rather classless model.*

Lemma 5.8. *If α is a limit ordinal, $cf(\alpha) > \aleph_0$ and $(N_\beta)_{\beta\leq\alpha}$ is a chain of conservative elementary extensions, then N_α is rather classless.*

Proof. Let X be a class on N_α. We show that $X \in Def(N_\alpha)$. Observe that since N_α is an elementary end extension of each of N_β, then for arbitrary β $X \cap N_\beta$ is a class on N_β. Moreover, since for each $\beta < \alpha$, $N_{\beta+1}$ is a *conservative* extension of N_β then $X \cap N_\beta$ is *definable* in N_β. Let us consider the following function $f : \alpha \longrightarrow \alpha$:

$$f(\beta) = \min_\nu[X \cap N_\beta \text{ is definable in } N_\beta \text{ from parameters in } N_\nu].$$

By the observations above this function is well defined. Since in the definition of the chain $(N_\nu)_{\nu\leq\alpha}$ at limit levels we took sums of previously constructed models, f is regressive on the set

$$\mathbb{LIM} \cap \alpha.$$

Hence by Fodor's lemma (5.2) there is a γ such that $f^{-1}(\gamma)$ is unbounded in α. Moreover, since $cf(\alpha) > \aleph_0$, then there is an $n \in \mathbb{N}$ and an unbounded $J \subset f^{-1}(\gamma)$ such that for each $\rho \in J$

$$X \cap N_\rho \text{ is definable in } N_\rho \text{ by a } \Sigma_n \text{ formula.}$$

For each $\nu \in J$ let a_ν be an of N_ν such that

$$N_\nu \models a_\nu \text{ is the least } a \text{ such that } X \cap N_\nu \text{ is definable by } a-\text{th } \Sigma_n \text{ formula.}$$

Let us observe that this is well defined since for each ν, $X \cap N_\nu$ is definable by a *standard* Σ_n formula. From this it also follows that for each $\nu \in J$, $a_\nu \in N_\gamma$. We shall show that for all $\nu \in J$, $a_\nu = a$ some $a \in N_\gamma$. This observation will end the proof, since

1. J is cofinal in α.
2. N_α is a common elementary extension of all N_β for $\beta \in J$.

Let us pick $\nu < \mu \in J$. Since $a_\mu \in N_\nu$ and $N_\nu \prec N_\mu$, then it must be the case that

$$N_\nu \models a_\nu\text{-th and } a_\mu\text{-th subsets are the same.}$$

By elementarity

$$N_\mu \models a_\nu\text{-th and } a_\mu\text{-th subsets are the same.}$$

So in N_μ, $a_\nu = \min\{a_\mu, a_\nu\} = a_\mu$, which ends the proof. □

We now return to the proof of 5.7. Fix arbitrary M and κ as in the statement of the theorem. Let $cf(\kappa) = \lambda > \aleph_0$ and let $(\beta_\alpha)_{\alpha<\lambda}$ be an increasing sequence of ordinals converging to κ. Let M_0 be an elementary extension of M such that

1. $|M_0| = \beta_0$
2. there is a strictly decreasing sequence $(b_\gamma)_{\gamma<\lambda}$ with no nonstandard lower-bound.
3. there is $S_0 \subset M_0^2$ - an inductive, proper and counting satisfaction class for M_0.

Let us put $M_0^* = (M_0, S_0)$. Let $(M_\alpha, S_\alpha)_{\alpha<\beta}$ be a chain of models defined:

1. if $\nu = \mu + 1$ then (M_ν, S_ν) is a conservative, elementary and finitely generated end extension of (M_μ, S_μ)
2. if ν is a limit ordinal and for some $\alpha < \lambda$ $\kappa_\alpha < \nu < \kappa_{\alpha+1}$ then let $(M_\nu, S_\nu) = \bigcup_{\delta<\nu}(M_\delta, S \restriction_{b_\alpha})$.

Finally let $M_\kappa = \bigcup_{\alpha<\kappa} M_\alpha$. We show that M_κ is as required.

First of all, M_κ is recursively saturated, since each of M_α was (by lemma 5.5), and by construction it is a κ-like elementary extension of M. We show that it is rather classless. Aiming at a contradiction suppose X is an undefinable class of M_κ. By the proof of lemma 5.8 there is a $\alpha < \kappa$ such that $X \cap M_\alpha$ is not definable in M_α and by lemma 5.6 there is a $\beta < \lambda$ and $\alpha < \kappa_\beta$ such that $X \cap M_\alpha$ is not definable in $(M_\alpha, S \restriction_{b_\delta})$. But

$$(M_\alpha, S \restriction_{b_\delta}) \prec (M_\beta, S_\beta)$$

so $X \cap M_\beta$ cannot be a class of (M_β, S_β). This contradicts X being a class of M_κ. $\square$

References

[1] C. C. Chang, J. H. Keisler (1990), *Model Theory*, North-Holland, Amsterdam, 3rd ed.

[2] A. Enayat, A. Visser, "Full Satisfaction Classes in a General Setting, Part I".

[3] H. Field, "Deflationist Views of Meaning and Content", *Mind*, New Series 103, ss. 249-285.

[4] Hájek Petr, Pudlák Pavel (1998), *Metamathematics of First Order Arithmetic*, Springer-Verlag, Berlin.

[5] V. Halbach (2011), *Axiomatic Theories of Truth*, Camridge University.

[6] Horsten Leon (2009), „Levity", *Mind*, vol. 118, s. 555-581.

[7] Kaye Richard (1991), *Models of Peano Arithmetic*, Oxford University Press, Nowy Jork.

[8] J. Ketland, „Deflationism and Tarski's Paradise", *Mind* 108, ss. 69-94.

[9] Kossak Roman, Schmerl James H. (2006), *The Structure of Models of Peano Arithmetic*, Clarendon Press, Oxford.

[10] J. Schmerl, „Recursively Saturated, Rather Classless Models of Peano Arithmetic", Lecture Notes in Mathematics 859, 268-282.

[11] S. Shapiro, „Proof and Truth: Through Thick and Thin", *Journal of Philosophy* 95, pp. 493-521.

[12] Smith Stuart T. (1989), „Nonstandard Definability", *Annals of Pure and Applied Logic*, vol. 42, s. 21-43.